R E V I E W

Vol. 72, No. 6 November/December 1990

THE FEDERAL RESERVE BANK of ST. LOUIS

Federal Reserve Bank of St. Louis
Review
November/December 1990

In This Issue . . .

The brewing industry has evolved over the past 25 years from an industry that concentrated on domestic markets to one that views itself as part of a global market. In the first article of this *Review*, Jeffrey D. Karrenbrock examines "The Internationalization of the Beer Brewing Industry." The author provides a case study of how and why brewers have used merchandise trade, licensing agreements and foreign direct investment to penetrate foreign markets. In addition, Karrenbrock demonstrates how certain economic factors, such as economies of scale and trade barriers, can affect which type of international transaction is used in different markets. A few specific countries' barriers to beer trade are also discussed.

* * *

On October 6, 1979, the Federal Reserve announced its landmark decision to implement monetary policy by targeting the growth rate of M1 rather than some level of the federal funds rate. The Federal Reserve abandoned this practice less than three years later primarily because empirical regularities in the behavior of M1 velocity that had made M1 targeting an attractive strategy seemed to disappear. Although no one has yet offered a definitive explanation for the observed changes in M1 velocity, these changes nonetheless have led to a general discrediting of monetary aggregate targeting as an approach to monetary policy.

In the second article of this issue, "Alternative Measures of Money As Indicators of Inflation: A Survey and Some New Evidence," Michael T. Belongia and James A. Chalfant review the arguments about why M1 velocity may have changed, then discuss more general issues about the construction of a measure of the money stock and the properties one would expect it to exhibit. Noting that the composition and weighting of components in a monetary aggregate each may contribute to measurement error and unexplained shifts in behavioral relationships, the authors then review other recent efforts to construct new monetary aggregates from first principles of economic and statistical theory. They compare the empirical performance of the alternatives against various final use criteria and find that, although some measures offer improvements over the current official aggregates, none can be judged at this time to be the definitive measure of the money stock.

* * *

In the third article in this issue, "The Trade-Related Aspects of Intellectual Property Rights: What Is At Stake?" Alison Butler examines one of the major new issues being negotiated in the current round of multilateral negotiations. With increased trade in high-technology products, intellectual property rights have become an increasingly important and contentious issue in international trade. The protection of in-

tellectual property, which includes such things as patents, copyrights and trademarks, is justified by most economists as being necessary for invention, while others, primarily policymakers, argue that these rights only serve to protect the monopoly power of innovating firms.

The author analyzes the questions surrounding the protection of intellectual property, how innovation is affected by trade, and the benefits of protecting intellectual property internationally. In addition, she examines why the incentive to protect intellectual property differs between industrialized and developing countries and what likely needs to happen before an agreement can be reached.

* * *

Before the crisis in the Middle East this summer, the easing of international tensions had reduced, for many, the urgency for the United States to continue building or even maintain its military strength. As support for allocating the nation's resources to defense weakened, people began to argue about the potential for a significant "peace dividend" available to the U.S. economy.

In the fourth article of this issue, "The Economic Consequences of Reducing Military Spending," Michelle R. Garfinkel examines some issues that have been overlooked by those searching for ways to use the peace dividend. Garfinkel attempts to answer two important economic questions: How can the resources from reduced military spending be reallocated efficiently to nonmilitary uses, and how can we identify what these uses should be? While the recent course of events in the Middle East might appear to detract from the prospects of a large dividend in the near future, it does not detract, the author says, from the relevance of these questions. Assuming that conflict is not a permanent obstacle to realizing a large dividend later, it provides us with more time now to evaluate the various options associated with future defense cuts.

* * *

Jeffrey D. Karrenbrock

Jeffrey D. Karrenbrock is an economist at the Federal Reserve Bank of St. Louis. David H. Kelly provided research assistance.

The Internationalization of the Beer Brewing Industry

THE BEER BREWING INDUSTRY has been undergoing a process of internationalization for the past 25 years. This article examines the roles that three types of international transactions — merchandise trade, licensing agreements and foreign direct investment — have played in this internationalization. As in other industries, a few general economic factors explain much of the increase in international brewing activity. What makes beer brewing a particularly interesting case study is that it provides an opportunity to demonstrate how certain economic factors, such as economies of scale and trade barriers, can affect the internationalization of an industry.

THE INTERNATIONALIZATION OF THE BREWING INDUSTRY

Merchandise Trade

As with most other goods, world merchandise trade in beer has expanded rapidly over the past 25 years (see figure 1). Much of the increase in world beer trade — and in world trade in general — can be attributed to such factors as lower trade barriers, more efficient communication and transportation technology, and growth in real personal incomes. The value of world beer trade increased from $149 million in 1965 to $2.08 billion in 1987, a 14-fold increase; at the same time, world trade in all goods increased to more than 12.5 times its 1965 value. In more recent years, between 1980 and 1987, world trade in beer expanded 83.8 percent, while total world trade grew only 23.4 percent. Despite its rapid growth, trade in beer in 1987 accounted for less than one-tenth of one percent of total world merchandise trade.

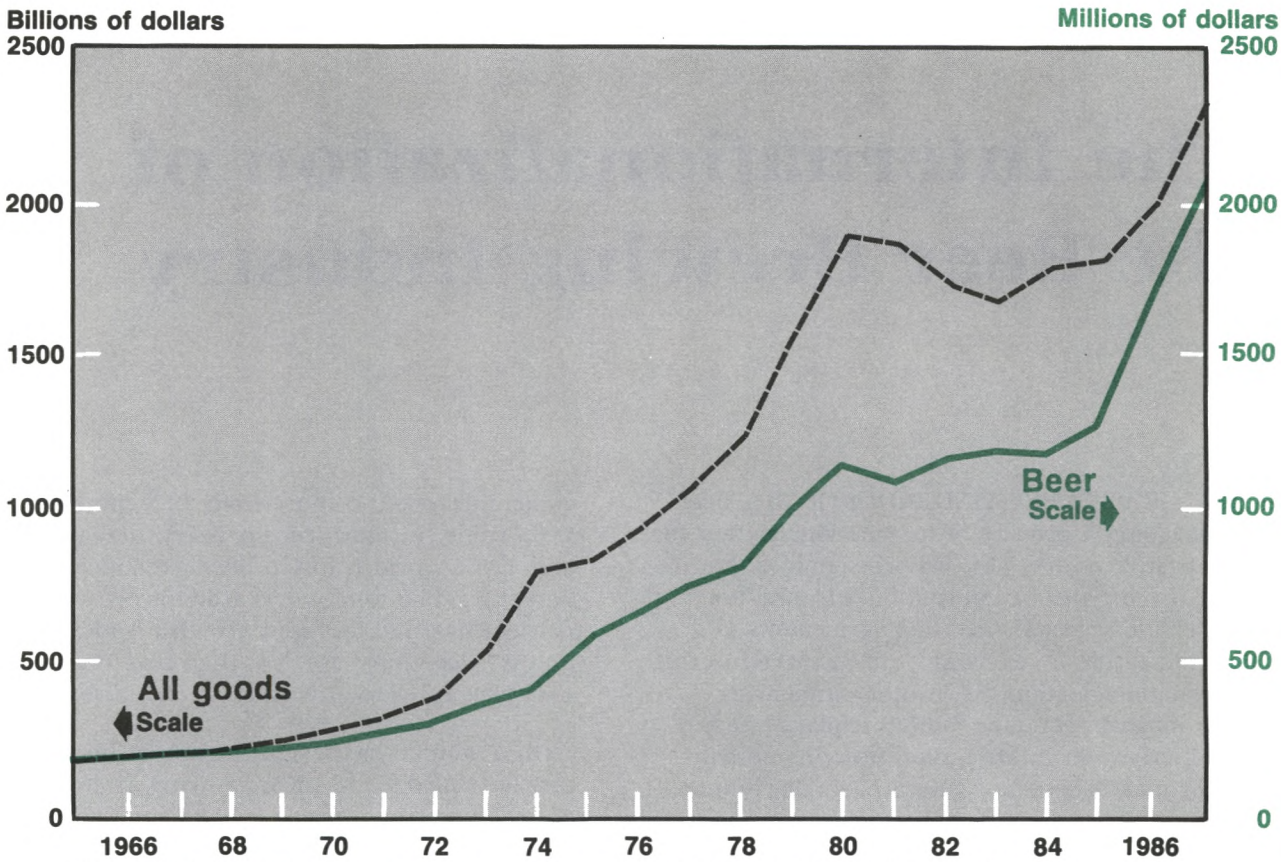
On a volume basis, world trade in beer has nearly tripled since 1965, growing at an average annual rate of 6.5 percent between 1965 and 1987. The largest exporters of beer in this growing market, ranked by volume, are the Netherlands, West Germany, Czechoslovakia, Belgium and Canada (see table 1).[1] The largest importers are the United States, the United Kingdom, France, Italy and West Germany.

Beer imports as a percent of total consumption (IPC) and exports as a percent of production (EPP) are larger for some of the smaller exporters and importers than they are for some of the larger exporting and importing countries. As table 1 shows, among 25 importers, beer IPC ranges from a low of 0.2 percent in Norway to 16.4 percent in Italy. The percent of beer consumption accounted for by imports in the largest beer importing country, the United States, is about 5 percent.

Similarly, among exporters, figures for EPP range from 0.4 percent in the United States to

[1]The *FAO Trade Yearbook* indicates that Mexico was the world's third-largest exporter of beer in 1987.

Figure 1
World Merchandise Trade



SOURCES: *FAO Trade Yearbook* and *International Financial Statistics Yearbook*.

41.6 percent in Ireland. The export numbers as a percent of production for such countries as the Netherlands and Luxembourg, however, are questionable as these countries do a significant amount of re-exporting to other countries (that is, much of the beer reported as exports may simply be imported and then re-exported for consumption elsewhere).[2]

Few of the countries listed in table 1 are strictly importers or exporters of beer. Most of the countries that export beer also import some beer and vice-versa. This pattern of trade is known as intra-industry trade. An examination of the IPC and EPP statistics in table 1 show that intra-industry trade in beer is more important to some countries than others. The largest exporter of beer, the Netherlands, imported only 4.3 percent of the beer it consumed in 1987.[3] Similarly, the two largest importers of beer, the United States and the United Kingdom, exported only 0.4 percent and 1.9 percent of their beer production in 1987. Ireland, on the other hand, exported nearly 42 percent of its production, while importing more than 12 percent of its beer consumption.

The degree of intra-industry trade for a country can be measured using a simple index, calculated for a given product as the absolute value

[2]Ott (1988) notes that the Netherlands has a long history of re-exporting imported goods.

[3]Ott (1988) notes that re-exported goods from the Netherlands are not included in the country's import figures.

Table 1

Selected 1987 Brewing Industry Statistics

	Exports 1000 HL	Imports 1000 HL	Exports As Percent of Production	Imports As Percent of Consumption	Intra-Industry Trade Index 1975	Intra-Industry Trade Index 1987	Average Annual Growth Rate, 1975 to 1987: Exports As Percent of Production	Imports As Percent of Consumption	Per Capita Consumption	Imports Per Capita
Australia	728.3	70.0	3.9%	0.4%	.83	.82	15.0%	17.0%	– 1.7%	14.7%
Austria	361.0	285.0	4.0	3.2	.33	.12	7.4	– 0.7	1.0	0.2
Belgium	2537.0	565.5	18.1	4.7	.23	.64	2.4	– 5.3	– 0.5	– 5.7
Canada	2415.8	448.0	10.5	2.1	.63	.69	13.8	52.0	– 0.4	50.7
Czechoslovakia	2698.0	0.0	12.1	0.0	1.00	1.00	5.9	0.0	– 0.7	0.0
Denmark	1934.0	20.8	22.8	0.3	.98	.98	– 0.3	*	– 0.2	*
W. Germany	5706.0	1301.6	6.2	1.5	.54	.63	8.2	6.3	– 0.2	6.1
Finland	22.8	13.9	0.7	0.4	.60	.24	53.7	171.1	1.9	184.6
France	672.0	2445.9	3.4	11.3	.56	.57	2.4	2.4	– 1.1	1.3
E. Germany	N/A	N/A	N/A	N/A	.25	N/A	N/A	N/A	1.6	N/A
Hungary	0.0	1290.0	0.0	12.6	N/A	1.00	0.0	– 2.86**	2.5	– 0.83**
Ireland	2066.0	415.0	41.6	12.5	.98	.67	2.1	39.0	– 2.0	35.9
Italy	73.1	2162.8	0.7	16.4	.95	.93	20.5	5.7	6.9	12.9
Japan	293.3	224.0	0.5	0.4	.73	.13	1.8	19.6	2.0	22.6
Luxembourg	270.3	42.2	40.9	9.8	.79	.73	– 0.7	0.5	– 0.8	– 0.3
Netherlands	5725.2	534.9	32.6	4.3	.69	.83	6.2	2.4	0.6	2.8
New Zealand	83.2	70.5	2.0	1.7	.68	.08	7.2	36.7	– 0.7	33.9
Norway	19.2	405.0	0.9	0.2	.89	.62	– 12.2	22.6	1.1	26.6
Poland	327.1	106.2	2.8	0.9	.74	.51	2.7	12.3	– 1.5	10.7
Portugal	74.5	34.3	1.5	0.7	.97	.37	10.9	N/A	3.7	N/A
Spain	121.0	735.0	0.5	2.8	.58	.72	14.5	19.3	2.7	22.4
Sweden	44.9	344.4	1.1	7.9	.91	.77	72.0	8.4	– 1.2	7.0
Switzerland	36.7	485.9	0.9	10.5	.75	.86	2.8	6.4	– 0.3	6.1
USSR	N/A	900.0	N/A	1.7	N/A	N/A	0.0	8.8	– 1.4	7.0
United Kingdom	1140.0	4093.1	1.9	6.5	.54	.56	3.1	3.4	– 0.4	2.9
USA	919.7	10991.1	0.4	5.0	.80	.85	22.1	13.7	0.8	14.7

SOURCE: Derived from information in the Brewers Association of Canada's *International Survey: Alcoholic Beverage Taxation and Control Policies.*

1 HL = 100 liters = 26.4 gallons

* Denmark's beer imports were generally small and declining throughout the period. However, a few relatively large increases in imports in later years caused imports as a percent of consumption and imports per capita to grow at average annual rates of 2,218 and 2,146.4, respectively.

** Average annual rate of growth for 1978-1986.

of the difference between exports and imports divided by the sum of exports and imports.[4] If the index is close to zero, the degree of intra-industry trade is substantial. An index value of one indicates that there is no intra-industry trade. This index, labeled the "Intra-Industry Trade Index" and shown in table 1, was calculated for beer trade in 1975 and 1987.

Of the 23 countries for which the index could be calculated for both years, nine countries' indexes rose over the period, indicating less intra-industry beer trade. In 12 countries, the indexes declined, indicating that intra-industry beer trade had increased. Only five countries in table 1 had an intra-industry index value of less than 0.5 in 1987. The majority (70 percent) had an index value of more than 0.5 in 1987, which indicates that intra-industry trade plays a minor role in the brewing industry in general, although it has become more prevalent during the past 15 years.

The growth rates of IPC and EPP provide further evidence of the increasing importance of intra-industry trade to the brewing industry. Of the 20 countries in table 1 reporting increased EPP between 1975 and 1987, 16 also reported increased IPC. Similarly, of the 19 countries reporting increased IPC, 16 reported increased EPP. In sum, merchandise trade in beer has expanded rapidly during the past 25 years, with intra-industry trade playing a small, but growing, role in beer trade.

Licensing Agreements

Brewers also use licensing agreements to make their products available to foreign consumers. A typical license agreement allows a brewer in one country to brew and market the beer of a foreign brewer. One example is Anheuser-Busch's (A-B) licensing agreement with the Canadian brewery John Labatt Ltd. This agreement allows Labatt to brew and market some A-B beers, such as Budweiser and Michelob, in Canada. In return, Labatt pays a royalty fee to A-B. Of course, the licensing brewers insist that the consistent quality of their products be maintained. In essence, the licenser is selling its know-how in brewing a specific beer, the right to use a trademark and the name recognition it has built for that trademark in exchange for a royalty payment from the licensee. There were at least 30 licensing agreements among various brewers around the world in 1987 (see table 2).

Several factors that are not mutually exclusive promote the use of licensing agreements. First, some firms use licensing agreements to circumvent trade barriers. For example, U.S. beers that are brewed in Canada under license agreement are not subject to either the Canadian federal tariff or the discriminatory mark-ups that other imported beers face at the provincial government outlets.[5] Second, the physical qualities of beer promote the use of licensing agreements. Beer is about 90 percent water, so transportation costs can be reduced through local production. In addition, beer has a shelf-life of about three to four months, of which two to three weeks could be taken up by overseas shipment. Also, when companies enter new markets, they often find it more profitable to license existing plants and distribution systems to handle their products rather than build their own plants and establish their own distribution systems.

The import and export figures discussed previously did not include consumption of foreign-held brand names that are brewed domestically under a licensing agreement. Thus, the degree of internationalization is understated when only merchandise trade is analyzed. Information on the amount of beer brewed under licensing agreements is usually closely held by the companies involved, and not much data are publicly available. The Conference Board of Canada, however, has estimated the impact of licensed brewed beer in the Canadian beer market and its findings serve to demonstrate how important licensed production can be.

In the Canadian market, three U.S. brewers, Anheuser-Busch, Coors and Miller, had licensing agreements with the three largest Canadian brewers— John Labatt, Molson and Carling O'Keefe— respectively, in 1986.[6] The Conference

[4]See Gray (1987), pp. 243-49.

[5]The U.S.-Canada Free Trade Agreement eliminates the federal tariffs on beer between these countries, but does not alter the pricing practices of the provincial liquor outlets. See Carter, et al (1989) for a more detailed description of Canadian barriers to beer trade.

[6]Molson and Carling O'Keefe agreed to merge their breweries in 1989. The new company is called Molson Breweries and will continue to brew for Coors and Miller under license.

Table 2

Licensing Accords Between Major Brewers

Anheuser-Busch

Guinness
Guinness brews Bud in Ireland

Labatt
Labatt brews Bud, Michelob & Bud Light in Canada

National
National brews Bud in Israel

Oriental
Oriental brews Bud in Korea

Suntory
Suntory brews Bud in Japan

United
United brews Bud in Denmark

Watney Mann
Watney brews Bud in U.K.

Artois

Whitbread
Whitbread brews Stella in U.K.

Carlton & United*

*Carling O'Keefe**
Carling brews Foster's in Canada

*Courage**
Courage brews Foster's in U.K.

Watney Mann
Watney brews Foster's in U.K.

Castlemaine Parkins

Allied-Lyons
Allied brews XXX in U.K.

Coors

Asahi
Asahi brews Coors in Japan

Molson
Molson brews Coors in Canada

Guinness

Labatt
Labatt brews Guinness in Canada

Heineken

Kirin
Kirin brews Heineken in Japan

Whitbread
Whitbread brews Heineken in U.K.

Kirin

Molson
Molson test brewing Kirin in Canada**

Lowenbrau

Allied-Lyons
Allied-Lyons brews Lowenbrau in U.K.

Asahi
Asahi brews Lowenbrau in Japan

Miller
Miller brews Lowenbrau in U.S.

Molson
Molson brews Lowenbrau in Canada

Miller

*Carling O'Keefe**
Carling brews High Life, Lite in Canada

*Courage**
Courage brews Lite in U.K.

United

Beamish & Crawford
Beamish & Crawford brews Carlsberg in Ireland

*Carling O'Keefe**
Carling brews Carlsberg in Canada[1]

Falken
Falken brews Carlsberg in Sweden

Heileman
Heileman brews Tuborg in U.S.

Suntory
Suntory brews Carlsberg in Japan

Watney Mann
Watney brews Carlsberg in U.K.

*Parent company, Elders IXL.

**No licensing agreement in effect.

SOURCE: *Modern Brewery Age,* July 13, 1987.

[1]John Labatt Ltd. acquired the right to brew Carlsberg in Canada, effective July 1, 1988.

Board estimates that brands produced in Canada under license with U.S. brewers in 1986 may have accounted for as much as 15 percent of beer sales in Canada.[7] This amounts to approximately 2.7 million barrels of U.S. brands produced and sold under license agreements in Canada in 1986. If these estimates are correct, the volume of licensed production of U.S. beers in Canada was more than 17 times the amount of beer exported directly to Canada in 1986 and more than four times the amount of total U.S. beer exports to all countries (exclusive of shipments to U.S. military bases and Puerto Rico).[8] In terms of Canadian consumption, the licensed brewed beer might have accounted for 15 percent of Canadian beer consumption compared with the 2.1 percent of domestic consumption accounted for by imports.

The numerous licensing agreements with breweries in Japan and the United Kingdom might indicate that beer produced under license represents a significant part of the foreign beer consumed in these countries. At least in some countries, beer produced under license clearly accounts for a much larger portion of foreign beer consumption than does imported beer.

Foreign Direct Investment

In addition to merchandise trade and licensing agreements, the internationalization of the brewing industry has been characterized by the increasing production of beer by "foreign-owned" firms. This production reflects the increasing frequency of foreign direct investment (FDI), in which one brewer purchases an existing firm or invests in a new or existing facility in a foreign country. Like licensing agreements, FDI is a substitute for merchandise trade. Firms may be prompted to use FDI for the same reasons they use licensing agreements. In addition, such factors as lower labor and energy costs and less government regulation may also encourage the use of FDI.

Several brewers have invested capital in breweries outside their home countries. Two Australian brewers, Elders IXL and Bond, have used this method of globalization extensively. Elders purchased Courage Ltd. of England in 1986, then purchased Carling O'Keefe breweries of Canada in 1987. Early in 1989, Britain's Department of Trade and Industry blocked the proposed takeover of Scottish & Newcastle Breweries PLC by Elders IXL Ltd.[9] In 1990, Elders IXL announced that its United Kingdom Courage breweries would purchase Grand Metropolitan's U.K. brewing and brands interests and its beer distribution and wholesaling activities. Furthermore, Courage's 4,900 pubs will merge with Grand Met's 3,570 pubs in a joint venture under the name Inntrepreneur Estates.[10] Bond has purchased two U.S. breweries, Pittsburgh Brewing and G. Heileman.

Japanese brewers also have actively invested in foreign brewery operations. In 1989, Tatsuuma-Honke Brewing Co. announced plans to build a sake brewery on the grounds of the Coors brewery in Colorado. In 1990, Asahi Breweries Ltd. announced plans to invest $70 million to open a brewery near Denver, Colorado, where it will produce a dry beer.[11] Finally, the Canadian brewing company John Labatt Ltd. purchased Latrobe Brewing Co. of the United States in 1987.

THE ECONOMICS OF INTERNATIONALIZATION

Underlying the preceding description of the internationalization of the brewing industry are some economic factors. The next section outlines the reasons why demand for foreign beer can exist in a country that already produces some domestic brands and discusses how changing relative prices and rising income can expand the demand for foreign beer. The second section analyzes the basic economic factors that determine the type of international transaction a brewery will use to put its products in the hands of foreign consumers. A more technical presentation of the economics discussed in these two sections is provided in appendixes to this article.

---

[7]Conference Board of Canada (1989), p. 9.

[8]Data for U.S. beer exports to Canada and total U.S. beer exports, exclusive of shipments to military bases, Puerto Rico and the territories, were provided courtesy of R.S. Weinberg & Associates.

[9]Carrington (1989).

[10]See Thornhill and Harris (1990), Harris (1990) and Sherwell (1990).

[11]"Japan's Asahi Plans Brewery in U.S." (1990).

Foreign Demand: The Attributes Approach

Why Demand for Foreign Beer Can Exist— One reason why people consume foreign beer is that they can buy it at a price at which they want more beer than domestic brewers want to produce. That is, the quantity of beer demanded is larger than the quantity of beer supplied domestically at the price of foreign beer, and therefore, some foreign beer is imported to meet the excess demand. Another reason why people consume foreign beer is that at least some consumers prefer the attributes, or characteristics, of the foreign beer over domestic brands. This second possibility is discussed in this section.

In general, consumers purchase beer for the "services" that they feel it can provide.[12] Consumers have a wide variety of beer brands to choose from and, subject to price and income limitations, will choose those brands that have the attributes that most closely match their desired services from drinking beer. Many attributes, such as taste, caloric content, alcohol content and packaging, distinguish one brand from another, and each combination of characteristics offers a distinctly different package of services.

Brewers differentiate their products on the basis of attributes and price. Consumers compare the package of services provided by a particular beer and its price to the services and prices of other brands. If consumers prefer the services of foreign beer over domestic beer, at given market prices, then demand for foreign beer will exist in a country.

Of course consumers do not necessarily consume only one domestic beer or foreign beer. Consumer satisfaction may be maximized by purchasing a combination of domestic and foreign beers. Suppose a consumer prefers the taste of a high-calorie foreign beer over all other domestic brands, but needs to watch his caloric intake and finds the taste of a particular brand of light beer to be acceptable. This consumer might purchase both the foreign and domestic beers, drinking the foreign beer in limited amounts, say, on special occasions, and drinking the low-calorie beer the rest of the time.

Growth in the Demand for Foreign Beer— The demand of foreign beer can expand if its price falls relative to the price of domestic beer. If a consumer had been purchasing domestic beer, the relative fall in the price of the foreign beer may be enough to compensate him for any perceived loss in services due to switching from the domestic to the foreign brand. In this case, the quantity demanded of the foreign beer will increase. The decline in the relative price of the foreign beer may also encourage people who already consume it to purchase more. Unfortunately, data on imported beer prices are scarce, and thus the role that changing relative foreign beer prices has played in the globalization of the industry is uncertain.

Increases in consumer incomes can also spur the demand for foreign products. When consumers' incomes increase, they are able to purchase more of all of the products they desire. In general, however, the quantities purchased of some goods, like flour, decline as incomes rise, while quantities of other goods purchased, like furniture, increase as incomes rise. The statistical evidence relating beer consumption to income growth is mixed.[13] Some studies have shown that the quantity of beer consumed increases as income increases, while others have shown the opposite.

Although little work has been done to estimate the relationship between foreign beer consumption and income in general, there is some data to suggest that the demand for foreign beer might be positively influenced by increases in income. All 21 OECD countries in table 1 that provided import data on beer had positive per capita gross domestic product growth between 1975 and 1986; 16 of these reported a positive average annual rate of growth of beer imports per capita and imports as a percent of consumption. In addition, the market for imported beers grew much more rapidly than most domestic beer markets during the late 1970s and early 1980s, a period of income growth for most countries. These figures roughly suggest that per capita income growth has contributed to the internationalization of the brewing industry.

[12]Much of the information in this section on the attributes model is taken from Douglas (1987).

[13]For a review of beer demand estimates, see Ornstein (1980). Also see Heien and Pompelli (1988). The estimated income elasticity of beer in these studies ranged from -0.46 to 0.79.

Where to Produce Beer: Some Economics

Once a firm determines that foreign demand for its products exists, it must determine the lowest-cost method of supplying these products to the foreign market. Should the firm use direct exports, a licensing agreement or direct foreign investment to enter the target market? The answer is relatively simple, in theory, and is based on the principle of profit maximization.

A brewery's total cost of supplying a foreign market is equal to the beer's cost of production plus transportation and distribution costs, marketing costs and overhead. A brewery's cost of producing beer is a function of its production technology and the cost of its inputs, such as labor, agricultural ingredients and packaging materials. Research has shown that the average cost of producing beer declines as production expands.[14] That is, economies of scale exist in the brewing industry. Economies of scale encourage direct exporting when the quantity demanded of foreign beer is relatively small and encourage foreign production either through licensing or foreign direct investment when the quantity demanded is relatively large.

Suppose a U.S. brewer and a Japanese brewer have identical production functions exhibiting economies of scale and that the firms pay the same price for their inputs. That is, their average cost of production curves are equal and are shaped as shown in figure 2. (For simplicity, assume that the U.S.-Japan exchange rate is fixed throughout and, given this exchange rate, Japanese prices are stated in U.S. dollars.) As the brewers expand production, the average cost of producing a unit of the product falls up to a point, after which average costs no longer decline but stabilize. Assume that Japanese demand exists for a beer—called Colony—produced by the U.S. firm. The U.S. firm must determine whether it can supply the Japanese market cheaper by producing Colony domestically and exporting or by producing it in Japan, either under license or by FDI.

Now suppose that supply and demand conditions and price in the U.S. are such that American consumers consume Q_{US} units of Colony, as shown in figure 2. (Note that the supply and demand curves are not shown in figure 2 and the quantity Q_{US} is simply given.) This substantial amount of consumption allows the U.S. firm to achieve significant economies of scale, producing Q_{US} units of Colony at an average cost of C_2 per unit. Also assume that market conditions are such that a relatively small quantity of Colony, Q_{J1}, is demanded in Japan. Since the U.S. brewer is already producing some Colony for domestic consumption, expanding production to meet the extra demand of Colony in Japan would allow the U.S. brewer to move down its average cost-of-production curve from point A to point B, where it could produce Colony for C_1.[15]

The alternative to producing the beer in the United States and exporting it is to produce Colony in Japan. Since Colony is currently not being produced in Japan, the Japanese firm or branch brewery built by the U.S. firm would have to brew the relatively small amount of Colony, Q_{J1}, at a high average cost of production, C_3. In other words, the relatively small quantity of production will not allow the Japanese plants to achieve significant economies of scale. Thus, producing Colony in the United States for export would save the brewer $C_3 - C_1$ per unit of Colony. If the cost of transporting Colony to Japan and distributing Colony in Japan is less than the difference between C_3 and C_1, then the U.S. brewer would maximize profits by exporting Colony to Japan. If the quantity demanded of Colony in Japan were larger, it might be more profitable for the brewer to use a licensing agreement or foreign direct investment.

Suppose that the quantity of Colony sold in Japan grows to Q_{J2} as shown in figure 3, while sales of Colony in the United States remain at Q_{US}. Since the U.S. brewer has exhausted its economies of scale, it cannot produce $Q_{US} + Q_{J2}$ at a lower per unit cost than that for $Q_{US} + Q_{J1}$. The Japanese brewery, however, by increasing production from Q_{J1} to Q_{J2} could now match the U.S. brewer's cost of production because it has also achieved the lowest possible average cost of production. Thus, given equal average

[14]Elzinga (1973), Fuss and Gupta (1981), Keithman (1978) and Scherer (1973) all provide evidence that economies of scale exist in the brewing industry. See Thompson (1985) or any micro-economic text for a discussion of the reasons why economies or diseconomies of scale can exist at the plant level.

[15]For simplicity, we have ignored any quantity response, stemming from lower prices, that might occur in the U.S. market as a result of the expanded output.

Figure 2

Average Cost of Production: U.S. Plant Achieves Economies of Scale

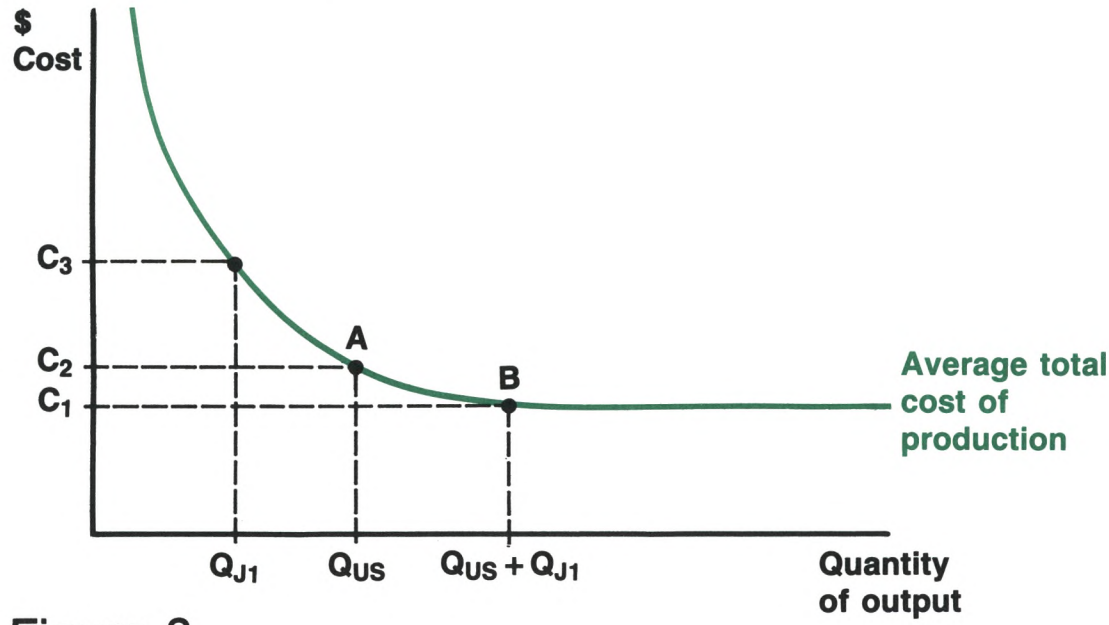
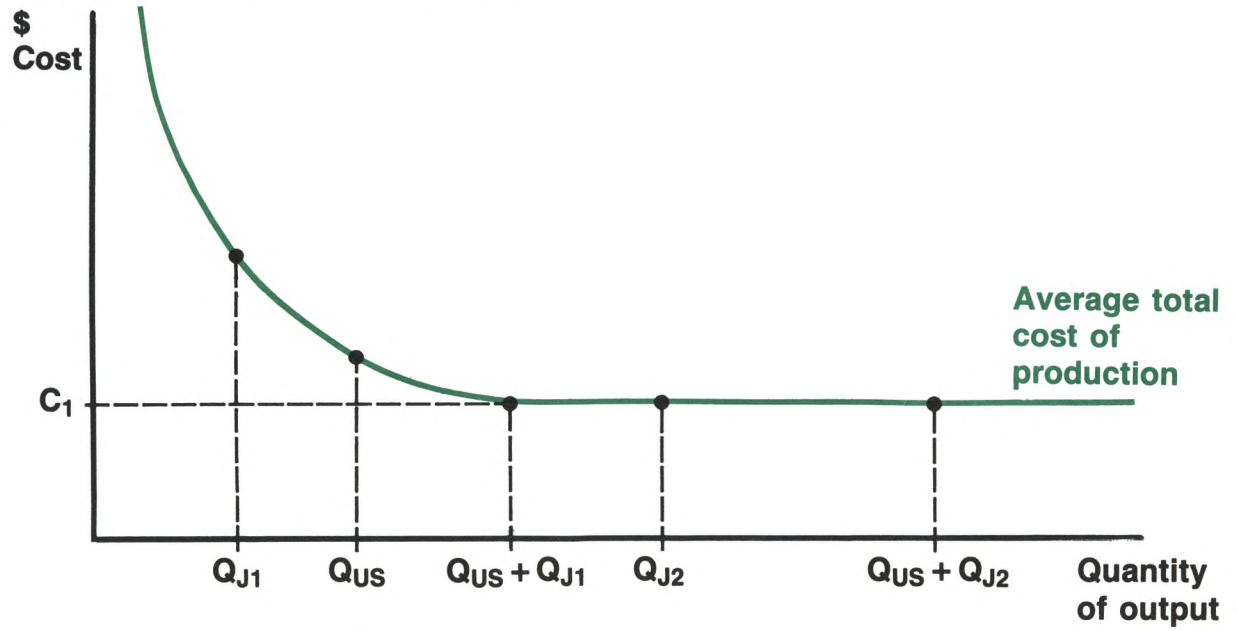


Figure 3

Average Cost of Production: U.S. and Japanese Plants Achieve Economies of Scale



production costs, the U.S. firm will now prefer to either negotiate a licensing agreement with the Japanese brewer or use FDI, thereby saving the additional export-related expenses of shipping Colony overseas and distributing it within Japan.

Like transportation costs, trade barriers also offset production cost advantages. If a target country has high tariffs or distribution systems for imported goods that are relatively costly, production cost advantages in the home country may be offset and licensing and foreign direct investment become the only feasible methods of entering the target market. As shown below, trade barriers have had a significant effect on the choice of licensing agreements and foreign direct investment in the internationalization of the brewing industry.

BARRIERS TO BEER TRADE

Japan

In Japan, two types of barriers inhibit foreign beer from entering the country. The most significant of these is the Japanese distribution system. The Japanese have a complex multi-tiered system, comparable to the U.S. beer distribution system, in which beer moves from producer to wholesaler to consumer.[16] In addition, Japan has little warehouse space, which means shipments are smaller and more frequent than in the United States. Both aspects of the Japanese distribution system raise the cost of distributing beer in Japan, relative to less complex systems. Japan also charges a small customs duty on imported beer. These factors raise the cost of exporting beer to Japan and make licensing agreements or foreign direct investment relatively more attractive methods of selling beer in Japan.

Canada

A GATT panel ruled in 1988 that specific practices of the Canadian provincial governments discriminated against imported beer.[17] Canadian trade barriers include discriminatory mark-ups at provincial liquor outlets and different marketing techniques for foreign beer, such as smaller packages and warm foreign beer sales at the governmental outlets. These non-tariff barriers have prompted U.S. brewers to use licensing agreements in Canada even though several brewers have U.S. plants that are located quite close to the Canadian border. The Australian brewer Elders IXL has chosen to use foreign direct investment to enter the Canadian market. This creates an interesting situation in which a U.S. beer is being made under license in a Canadian brewery that is partially owned by an Australian brewer.

United Kingdom

As in Japan, distribution practices are the main barriers to trade in the United Kingdom. Most beer consumed in the United Kingdom is draft beer, and most of this is sold in pubs. Many pubs are owned outright by breweries, managed by the breweries or leased to individuals who enter into exclusive supply agreements with the breweries. This system was the subject of eight investigations between 1966 and 1986, that focused chiefly on pricing and supply competition.[18] Given the relationship between the pubs and the domestic breweries, foreign label brewers have problems getting local brewers to carry their products in British pubs. Thus, many foreign brewers have chosen to use licensing agreements with domestic firms to penetrate the U.K. beer market. Foreign direct investment has also been used to enter this market.

SUMMARY

The brewing industry has evolved from an industry that concentrated on domestic markets to one that views itself as part of a global market. This internationalization has occurred via the use of merchandise trade, licensing agreements and foreign direct investment. Merchandise trade in beer has developed in an intra-industry pattern, whereas international transactions in licensing agreements and foreign direct investment have not developed, in general, in a bilateral pattern. Licensed production and production at foreign-owned breweries likely account for an unknown, but probably large, part of foreign beer consumption in some countries.

---

[16]See VandeWater and Curley (1990).

[17]G. Heileman Brewing Co. filed a Section 301 trade action against Canada over unfair pricing and distribution practices of provincial governments. See *Daily Report for Executives* (1990).

[18]Brewers Association of Canada (1989), pp. 387-88.

Two conditions must hold for trade in similar goods, such as beer, to occur. First, the foreign product must offer a combination of desired attributes that are not available to domestic consumers from domestic products. Second, it must be profitable to produce the product for the foreign market. Which type of international transaction will be used to supply a foreign market with beer is related to the existence of economies of scale, distribution costs and trade barriers.

REFERENCES

Brewers Association of Canada. *International Survey: Alcoholic Beverage Taxation and Control Policies*, 7th ed. (Ottawa, Canada, June 1989).

Carrington, Tim. "Britain Blocks Elders Takeover of Big Brewer; Government Urges Reforms in Reflecting Firm's Offer for Scottish & Newcastle," *The Wall Street Journal,* March 22, 1989.

Carter, Colin A., Jeffrey Karrenbrock and William W. Wilson. "Freer Trade In the North American Beer and Flour Markets," in Andrew Schmitz, ed., *Free Trade and Agricultural Diversification: Canada and the United States,* (Westview Press, 1989).

The Conference Board of Canada. *The Canadian Brewing Industry: Historical Evolution and Competitive Structure,* (Ottawa, Canada, February 1989).

Daily Report for Executives. (The Bureau of National Affairs, Inc., Washington D.C., May 18, 1990), p. A-3.

Douglas, Evan J. *Managerial Economics: Analysis and Strategy,* 3rd ed. (Prentice-Hall, Inc., 1987), pp. 86-103.

Elzinga, Kenneth G. "The Restructuring of the U.S. Brewing Industry," *Industrial Organization Review* (Vol. 1, 1973), pp. 101-14.

Finnegan, Terri. "International Licensing Pacts on the Rise," *Modern Brewery Age* (July 13, 1987).

Food and Agricultural Organization of the United Nations. *FAO Trade Yearbook,* various issues.

Fuss, Melvyn A., and Vinod K. Gupta. "A Cost Function Approach to the Estimation of Minimum Efficient Scale, Returns to Scale, and Suboptimal Capacity," *European Economic Review* (February 1981), pp. 123-35.

Gray, H. Peter. *International Economic Problems and Policies* (St. Martin's Press, Inc., 1987).

Harris, Clay. "Time for Takeovers, Gentlemen Please," *Financial Times*, March 14, 1990, London, England.

Heien, Dale, and Greg Pompelli. "The Demand for Alcoholic Beverages: Economic and Demographic Effects," *Southern Economic Journal* (January 1989), pp. 759-70.

Horstmann, Ignatius, and James R. Markusen. "Licensing Versus Direct Investment: A Model of Internationalization by the Multinational Enterprise," *Canadian Journal of Economics* (August 1987), pp. 464-81.

International Monetary Fund. *International Financial Statistics Yearbook* (1989), pp. 122-23.

"Japan's Asahi Plans Brewery in U.S." *St. Louis Post-Dispatch*, January 3, 1990.

Keithman, Charles. *The Brewing Industry*, Staff Report of the Bureau of Economics, U.S. Federal Trade Commission (1978).

Lancaster, K. "A New Approach to Consumer Theory," *Journal of Political Economy* (April 1966), pp. 132-57.

_____. *Consumer Demand: A New Approach* (Columbia University Press, 1971).

Ornstein, Stanley I. "Control of Alcohol Consumption Through Price Increases," *Journal of Studies on Alcohol* (September 1980), pp. 807-18.

Ott, Mack. "Have U.S. Exports Been Larger Than Reported?" this *Review* (September/October 1988), pp. 3-23.

Scherer, F.W. "The Determinants of Industrial Plant Sizes in Six Nations," *Review of Economics and Statistics* (May 1973), pp. 135-45.

Sherwell, Chris. "Foster's Brewing Sees Global Sales of A$9bn," *Financial Times,* March 14, 1990, London, England.

Thompson, Arthur A. *Economics of the Firm: Theory and Practice,* 4th ed. (Prentice-Hall, Inc. 1985).

Thornhill, John, and Clay Harris. "GrandMet and Elders Unveil Swap Details," *Financial Times,* March 14, 1990, London, England.

VandeWater, Judith, and John Curley. "Beer Barrels Eastward," *St. Louis Post-Dispatch,* June 11, 1990.

Weinberg, R.S. & Associates. "Tax Free Removals, Destination of Exports and Consumed on Brewery Premises, 1947-1988" (Brewing Industry Research Program, St. Louis, Missouri).

Appendix A
Foreign Demand for Beer: The Attributes Model

The attributes model, introduced by Lancaster (1966, 1971), can be used to show how demand for foreign beer can exist in a country in which domestic brands are already produced.[1] Suppose a consumer chooses between two brands of beer so that, subject to income limitations, his satisfaction from the services provided by the beer is maximized. For simplicity, assume the consumer values only two attributes of beer: taste and low calories. The two types of beer provide these attributes in differing proportions and at different prices.

After sampling both products, the consumer rates each brand on a scale of 1 to 3, 3 being best, for both taste and calorie content, as shown in table A1. Brand F is a foreign beer that tastes great, but is high in calories (thus receiving a low rating on caloric content) giving it a relatively high ratio of taste-to-calorie appeal. Brand D, a domestic beer, does not taste quite as good, but is very low in calories. This beer then has a relatively low taste-to-calorie appeal ratio.

Table A1

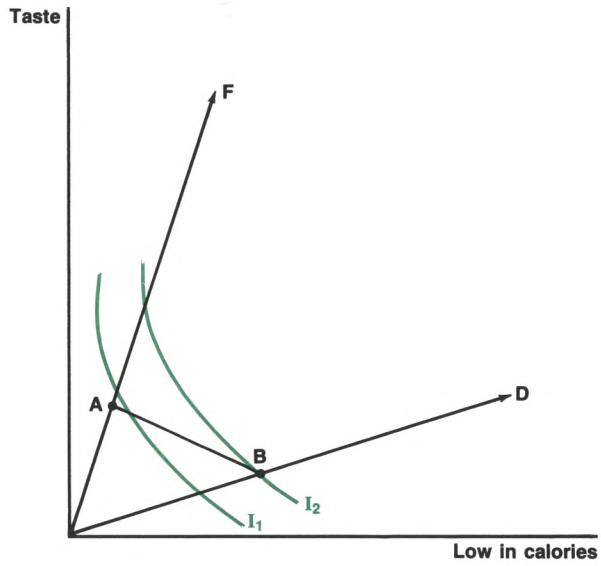
Attribute Ratings and Prices of Three Beers

Brand	Attribute Rating Calories	Taste	Ratio of taste to calories	Bar price per bottle	Bottles per $12
D	3	1	1/3	$2	6
F	1	3	3	$3	4

The amount of each beer the consumer can purchase is determined by his income and the price of the products. Assume that the consumer has decided to spend $12 on beer during a visit to a local bar and the cost of each kind of beer is as shown in table A1. If he spends the entire $12 on only one product, he could buy at most six bottles of D or four bottles of F. Four bottles of F would provide 4 units (4 bottles X 1 unit of caloric appeal per bottle) of the caloric attribute and 12 units (4 X 3) of the taste attribute.

Figure A1

Maximization of Utility by Consuming Only Domestic Beer



The two products are depicted in figure A1 in an attribute space as rays from the origin. The slope of each ray is determined by the ratio of taste to calorie appeal. If the consumer drinks brand F, then he moves out along ray F, absorbing the two attributes in a ratio of 3:1. Points A and B represent the maximum amount of the two attributes that can be obtained by consuming beers F and D, respectively, given the spending constraint of $12.

Joining points A and B provides the consumer's efficiency frontier. The efficiency frontier is the outer boundary of the attainable combination of the two attributes, given the budget constraint of $12. It is called efficient because a utility-maximizing consumer will get more utility by being on the frontier rather than within the frontier, even though these interior points are attainable.

[1]Much of the information in this appendix follows the discussion of the attributes model as presented by Douglas (1987), pp. 86-102.

Figure A2
Maximization of Utility by Consuming Only Foreign Beer

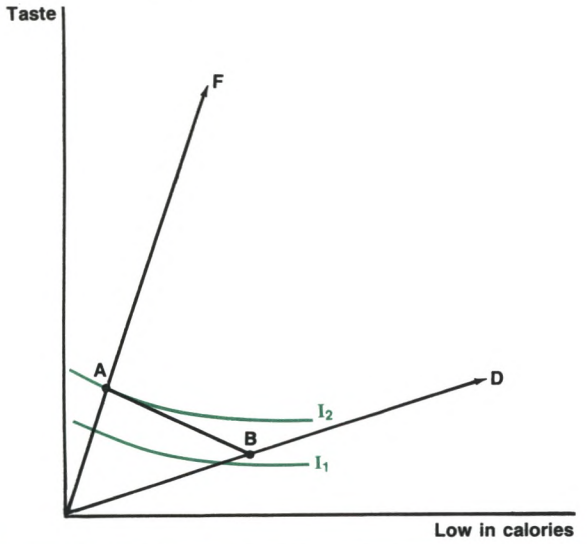
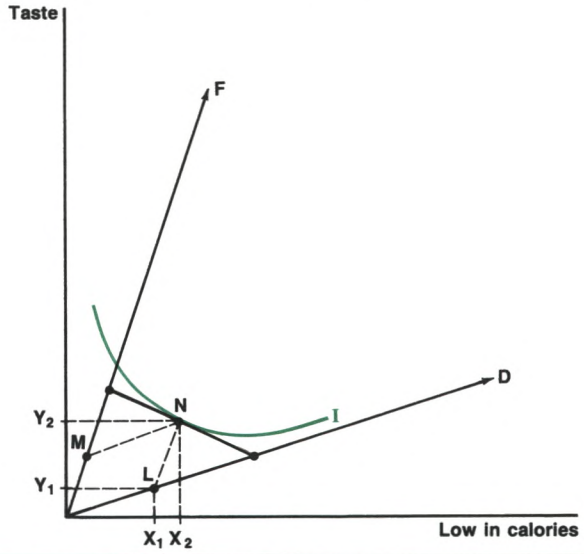


Figure A3
Maximization of Utility by Consuming Both Domestic and Foreign Beer



How do we know which beer, or combination of beers, the consumer will choose? In the attributes model, consumer preferences between attributes can be expressed using indifference curves. Like indifference curves used to express the marginal rate of substitution (MRS) between two products, the attribute indifference curves express the MRS between attributes, and higher indifference curves represent higher levels of utility. The beer consumer's assumed attribute indifference curves I_1 and I_2 have been superimposed on the attribute space in figure A1. Assuming that a consumer wishes to maximize his satisfaction from taste and caloric content, he would choose to be on the highest attainable indifference curve, which occurs at point B in figure A1.

The position and slope of the indifference curves will determine the brand or brands of beer chosen. This particular consumer has an indifference curve that is relatively steep indicating that, compared with a consumer with a flat indifference curve, he is willing to give up a lot of taste to get a few less calories.

Now suppose that a different consumer, who gives the beers the same attribute ratings, is willing to consume a lot more calories to gain a bit better taste. The shape of this consumer's indifference curve would be more flat, and as shown in figure A2, this person would choose the foreign brand F, at point A. Thus, in a society where some consumers prefer the attributes of foreign beers over domestic beers, a demand for foreign beer will exist.

Figure A3 shows an example of a consumer who would purchase both foreign and some domestic beer. Neither of the beers provides the attributes exactly in the ratio represented by point N. The consumer could reach this point, however, by consuming some of both products. By consuming L units of the domestic brand, the consumer would obtain Y_1 units of taste and X_1 units of caloric appeal. By spending the rest of his budget on brand F, the consumer would travel along the path LN, which has the same slope as ray F, to obtain the $X_2 - X_1$ units of calorie appeal and the $Y_2 - Y_1$ units of taste needed to reach his maximum level of utility at point N. Alternatively, the consumer could have started by consuming M units of brand F and then consumed L units of brand D to reach point N.

Growth in Foreign Demand

The demand for foreign beer can increase if the relative price of the foreign beer falls. As shown in figure A4, when the price of the foreign beer falls, the maximum amount of the foreign brand that can be purchased increases,

Figure A4

The Price Effect Shown by the Attribute Approach

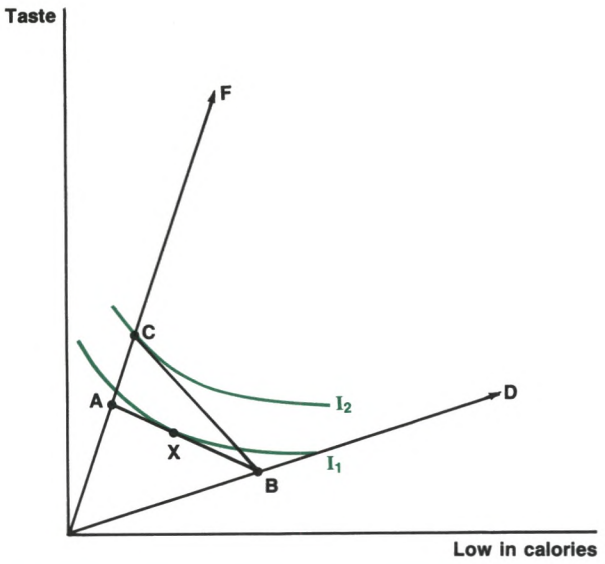
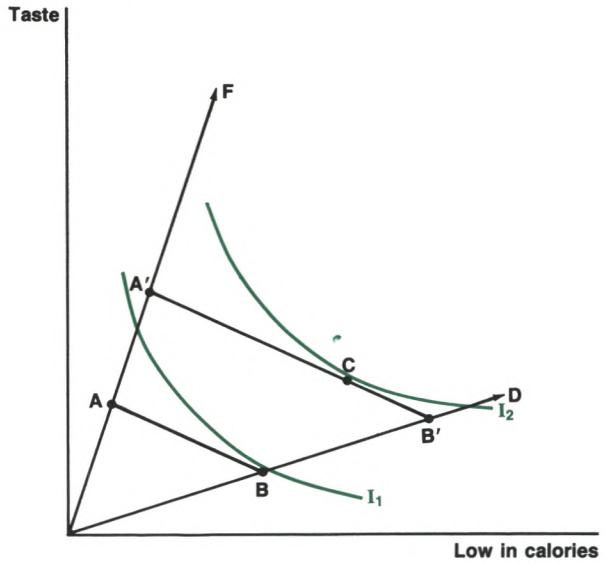


Figure A5

The Income Effect Shown by the Attribute Approach



shifting out the efficiency frontier from AB to BC. In this example, the consumer goes from buying some of both brands at point X to only buying the foreign brand at point C.

Expanding incomes can also explain increased consumer demand for foreign beer. In figure A5, the consumer's increase in income has led to a shift from buying only the domestic beer to buying some of both beers. Initially the consumer's efficiency frontier is AB, the highest indifference curve attainable is I_1 and the consumer purchases only the domestic product D. When the consumer's income increases, the efficiency frontier shifts out in a parallel fashion to A'B', because prices and attributes are fixed and only income is changing. For the given prices and attributes, the increased income allows more consumption of each beer. To maximize utility, the consumer shifts to point C on the higher indifference curve I_2, where he consumes some of both brands of beer. Thus, a higher level of income has induced the consumer to purchase more of the foreign brand of beer.

Of course, the example could have been constructed to show how a shift in the consumer's income could have led to a reduction in the amount of foreign beer purchased. Several researchers have estimated the demand for beer, but no conclusive evidence has been provided as to whether beer consumption expands as a consumer's income expands. Some data, as discussed in the text, however, suggest that the demand for foreign beer might be positively influenced by increases in income.

Appendix B
Whether to Produce Domestically or Abroad: Some Economics

Suppose that a U.S. brewer wants to sell its beer, Colony, in Japan. The U.S. brewer has three methods of supplying Japan with Colony: 1) produce Colony domestically and export to Japan, 2) negotiate a licensing agreement with a Japanese brewer who would brew and distribute Colony in Japan, or 3) produce Colony in Japan either by purchasing an existing Japanese brewery or building a new brewery. Which is the cheapest?

Assume that the long-run average cost (LRAC) curves of any plants where Colony could be brewed are identical. These potential plants include the U.S. parent firm and any branch plant it may establish in Japan, as well as any existing brewery in Japan that could brew Colony under license. In addition, assume that the U.S.-Japan exchange rate is fixed throughout and, given this exchange rate, Japanese prices are stated in U.S. dollars.

Consider the level of demand for Colony in the two countries. In the United States, as shown in figure B1, part of the demand curve is above the LRAC curve, indicating that consumers are willing to pay a price for Colony that is above the U.S. firm's average cost for some levels of production. In Japan, however, if the demand for Colony is relatively small, say D_{J1} as shown in figure B2, then Colony cannot be brewed profitably there. At each quantity along D_{J1}, the price consumers are willing to pay is below the Japanese plant's average cost of producing Colony. Thus, when demand is D_{J1}, direct exports from the U.S. brewery will be the only way the Japanese market might be supplied with Colony. However, if the Japanese demand schedule for Colony were much larger, say, at D_{J2}, then the Japanese brewery could at least cover its production costs for some level of production. Thus, at this larger level of demand, both production in the United States and in Japan are potential routes of supplying Japan with Colony.

Consider the cost of producing different levels of Colony for foreign consumption more closely.

Figure B1
U.S. Production Costs and Demand for Beer

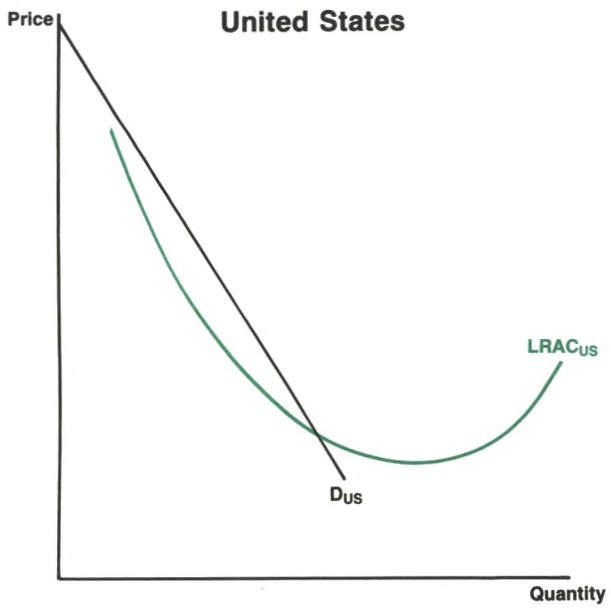


Figure B2
Japanese Production Costs and Two Demand Curves for Beer

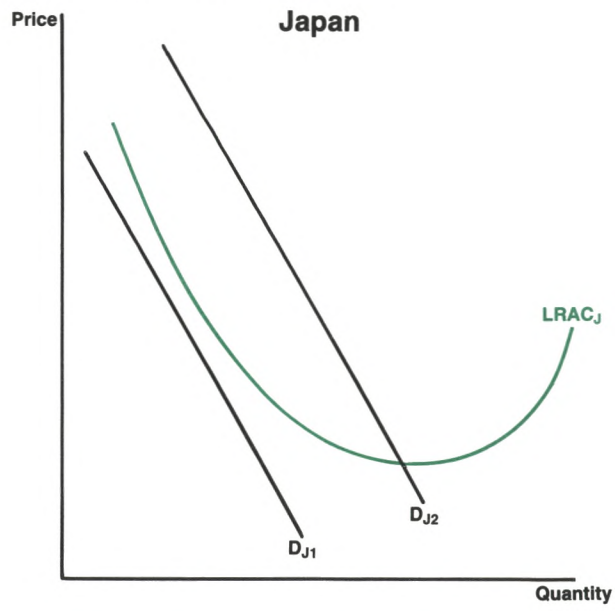


Figure B3
U.S. Plant's Production Costs

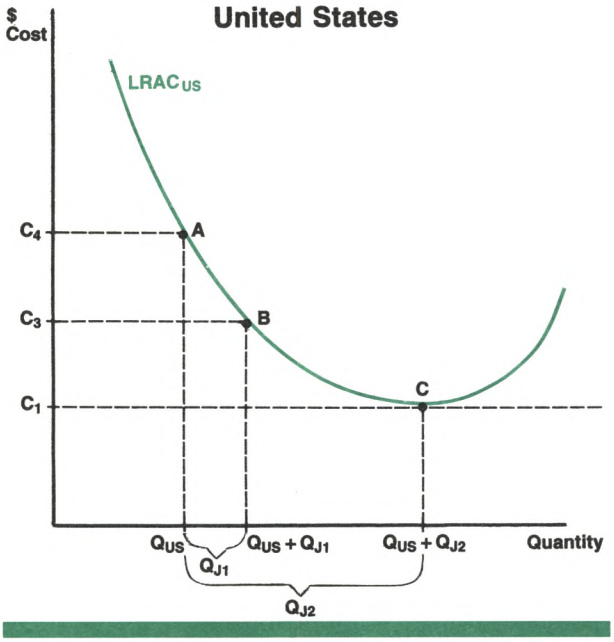
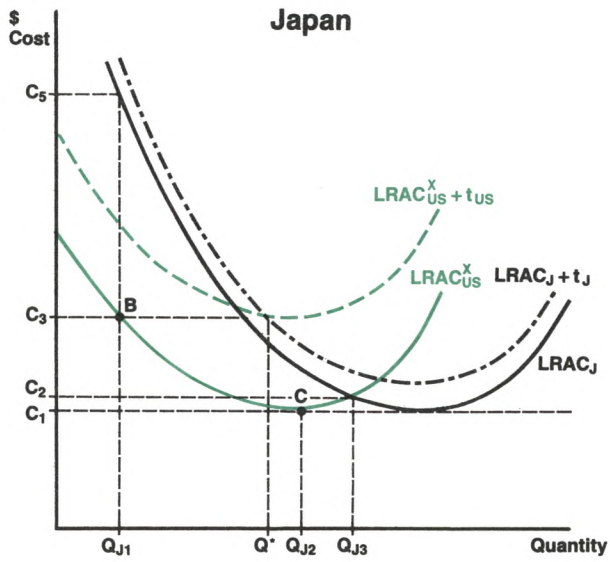


Figure B4
Japanese Plant's Production Costs Compared to U.S. Plant's Cost of Exporting



The goal is to determine which type of international transaction allows the firm to provide Colony to the Japanese market at the lowest cost. First, consider the U.S. brewer's cost of producing and exporting Colony to Japan. Assume for a given market price, the quantity of Colony demanded in the United States is Q_{US}, as shown in figure B3. The U.S. brewer produces this amount at an average cost of C_4 per unit.

Now suppose that for a given price, the quantity demanded of Colony in Japan is Q_{J1}, as shown in Figures B3 and B4. Since the U.S. brewer already produces some Colony for domestic consumption, by expanding production to $Q_{US}+Q_{J1}$ to meet the export demand, the U.S. brewery could move down its LRAC curve from point A to B in figure B3, lowering its average cost of producing Colony from C_4 to C_3.[1] If the quantity demanded of Colony in Japan was larger, at Q_{J2}, the U.S. brewery's average cost of producing it would fall even further to C_1 at point C. By doing a similar analysis for other quantities of U.S. exports, we can develop the U.S. brewer's cost of producing different amounts of Colony for export to Japan, shown by the line $LRAC^{X}_{US}$ in figure B4.

Figure B4 allows a straightforward comparison of the production costs of exports to Japan, $LRAC^{X}_{US}$, with the production of these quantities in Japan, $LRAC_J$. It shows that the average cost faced by the U.S. brewer producing a given amount of Colony for export is lower than the Japanese brewer's average cost, $LRAC_J$, up to the quantity Q_{J3}, but higher for all subsequent levels. This is possible, even when cost curves are identical across countries, because the U.S. plant was already producing Colony for domestic consumption and that by expanding production to meet export demand, the average cost of producing Colony fell. The Japanese plant currently is not producing any Colony; if it were to start brewing Colony for Japanese consumption, it would have to start at a higher cost on its long-run average cost curve. Economies of scale, however, do not continue indefinitely. Consequently, the cost of producing

[1]For simplicity, the analysis ignores any potential sales price decline in the United States that may stem from the brewer achieving greater economies of scale. Thus, the quantity demanded in the United States, after production is increased to meet export demand, is assumed to remain at the same level as before export production occurred.

Colony eventually starts to rise and the Japanese plant can produce Colony cheaper than the U.S. plant after point Q_{J3}.

Thus far, the discussion has focused solely on the cost of producing Colony. Transportation and distribution costs are likely to influence where production is located. Assume that it costs the same per unit to ship and distribute a small amount of Colony as it does a large amount of Colony. Since the U.S. plant has to ship Colony overseas, it is reasonable to assume that its transportation and distribution costs will be significantly higher than a Japanese plant's would be if Colony were produced there. Adding these average per unit transportation and distribution costs to the plant's respective long-run average cost of production curves gives the two dashed lines, $LRAC^{x}_{US}+t_{US}$ and $LRAC_{J}+t_{J}$, shown in figure B4. The U.S. plant can produce, transport and distribute Colony to the Japanese market at a lower cost than the Japanese brewery can up to the quantity Q^{*}. Notice that the additional costs of transportation and distribution have lowered the quantity at which the U.S. brewery can compete from Q_{J3} to Q^{*}. At quantities beyond Q^{*}, the Japanese firm can produce and distribute Colony for less than C_3, giving it a cost advantage over the U.S. brewer.[2]

Because of the cost advantage, for any given quantity of Colony demanded in Japan up to Q^{*}, the U.S. firm would prefer to produce Colony domestically and export the product to Japan. If the quantity of Colony demanded in Japan were greater than Q^{*}, the U.S. firm would either attempt to negotiate a licensing agreement with the Japanese brewer or purchase or build a Japanese brewery for production of Colony.

Whether the U.S. brewer would choose to license production or open a branch brewery in Japan would depend on several factors. Horstmann and Markusen (1987) note that if the licensee and the plant to be built or purchased are equally efficient, then the need to give the licensee the incentive to maintain the reputation of the licenser's product will result in FDI always dominating the use of licensing. They also conclude, however, that if the licensee and branch plant are not equally efficient—that is, if their LRAC curves are not identical—then other factors such as the size of the market, the existence of close substitutes in the target market and the level of interest rates in the two countries will determine whether licensing agreements or FDI will be used.

In addition to production and distribution costs, brewers also face tariff and non-tariff trade barriers, which raise the cost of supplying a country with beer. In terms of figure B4, it is conceivable that the U.S. brewery could have an average cost of production considerably below breweries in Japan, but that trade barriers in Japan are so high that licensing agreements or foreign direct investment become the preferred method of supplying the foreign country at all levels of demand. Here, the $LRAC^{x}_{US}+t_{US}$ curve can be used to incorporate this idea. Let the t_{US} variable now stand for transportation and distribution as well as costs associated with trade barriers, such as tariffs. The existence of trade barriers simply shifts the U.S. brewer's export cost curve upward, pushing Q^{*} closer to the origin.

Other realistic problems associated with international transactions have been ignored in this example. Some of the other factors that would affect how a firm supplies a foreign market include differences in production technology and input costs, government restrictions on foreign investment, costs of negotiating and monitoring licensing agreements, exchange-rate movements and the role other products being produced at the breweries might have on the plant's cost of production.

[2]Of course, the Japanese firm will eventually reach its points of diseconomies of scale and its average cost of production will rise above C_3.

Michael T. Belongia and James A. Chalfant

Michael T. Belongia is an assistant vice president at the Federal Reserve Bank of St. Louis. James A. Chalfant is an associate professor of agricultural and resource economics at the University of California-Berkeley. *Scott Leitz, Lora Holman and Lynn Dietrich provided research assistance.*

Alternative Measures of Money as Indicators of Inflation: A Survey and Some New Evidence

IN OCTOBER 1982, the Federal Open Market Committee (FOMC) officially de-emphasized M1 as an intermediate target variable in conducting monetary policy. The FOMC further reduced the emphasis given to M1 in 1987 by not setting target ranges for its rate of growth. The typical explanations given for this de-emphasis have stressed a breakdown in the 1980s in the relationships between M1 growth and both the rate of inflation and the growth rate of nominal GNP that had characterized the post-war period.[1]

The de-emphasis of M1 as a policy guide has prompted many observers to seek a replacement. Some analysts have argued that financial innovations and deregulation have altered the relationship between interest-bearing and non-interest-bearing assets. One group of individuals, for example, has argued that the interest-bearing components of M2 are good substitutes for the narrower set of assets in M1 and, on this basis, concluded that M2 now is the best monetary aggregate for policy actions. Another group, reasoning that interest-bearing accounts act more as savings balances than transaction accounts, has advocated the narrow measure of M1A (currency plus non-interest-bearing checkable deposits).[2] Still others have advocated the use of weighted monetary aggregates, arguing that the individual assets included in M1, M2 and the

[1]Specifically, in the postwar period through 1981, M1 growth was related closely to nominal GNP growth over a period of one year and to the rate of inflation over a period of three to five years. The breakdown in these relationships also has been discussed as a decline in the growth rate of M1 velocity. See Stone and Thornton (1987), for example, for a survey of explanations that have been offered for this abrupt shift in what had been quite stable relationships. Tatom (1988) and others, in contrast, have disputed whether historical relationships between M1 and nominal variables were fundamentally different in the 1980s.

[2]For the arguments favoring M2, see Hetzel (1988). For the evidence favoring M1A, see Hafer (1984) and Darby, et al. (1989).

broader aggregates possess different degrees of "moneyness" according to the interest they pay. If these assets have different degrees of moneyness, they should be weighted differently when added together.[3]

This article focuses on research that addresses both the composition and weighting issues in deriving a new monetary aggregate from the first principles of demand analysis.[4] It surveys this work and evaluates the performance of two new money stock measures as indicators of future inflation.

WHY IS THE DEFINITION OF "MONEY" IMPORTANT?

What truly constitutes "money" in the real world is an interesting and important issue because, as economic theory predicts, money has powerful effects on economic activity. Unfortunately, theory alone does not offer explicit guidance in choosing a unique and unambiguous real world counterpart to the abstract theoretical concept of "money." Without clear theoretical guidance, many economists have adopted the conclusion that the appropriate definition of money is the one that serves best in the chosen empirical application.[5]

One theoretical framework that attributes an important causal role to money in determining fluctuations in economic activity is the Quantity Theory of Money. This theory is based on the Equation of Exchange, which typically is written as:

$$(1)\quad MV = PT,$$

where M is the quantity of money, V is its velocity of circulation, P is the average price level and T is the volume of transactions. The Quantity Theory derived from this equation predicts that the price level varies directly with the quantity of money. In its strictest form, which has been subjected to empirical testing, the Quantity Theory predicts that the causation runs from changes in money to changes in prices (but not the reverse) and that, accounting for the short-run influences of other variables, changes in money and prices are proportional.[6]

Obviously, the definition of money is important in testing such propositions. The definition of money is also important for monetary policy purposes. Assume, for example, that a central bank believes the implications of the Quantity Theory. It might then direct its policies to achieve a growth rate of money consistent with zero inflation. This raises the issue of how to measure money in the real world. As Laidler (1989) points out, both critics and advocates of the Quantity Theory frequently disagree about the appropriate definition of money to test its implications. Thus, a central bank committed to achieving a policy goal for "P" still faces considerable uncertainty about which "M" should be targeted and how to do so.

FINANCIAL INNOVATIONS AND THE CONDUCT OF MONETARY POLICY

For more than 30 years after World War II, the trend growth rate of M1 velocity was nearly constant at 3 percent. For many economists, this stability made M1 growth a reliable leading indicator of the effects of current monetary actions on future economic activity. As figure 1 shows, however, this seemingly stable trend was disrupted sharply in 1981, perhaps because of the nationwide introduction of interest-bearing checkable deposits in January of that year. Because these deposits not only paid rates of interest that were close to those on savings deposits but also allowed depositors to write checks directly on these deposits, the hypothesis was that at least some portion of NOW account balances were savings. If so, these new accounts combined "idle" savings balances with "active" transaction balances in traditional demand deposits. The larger quantity of balances in M1 accounts, for the same level of GNP, would be reflected in a fall in the growth rate of M1 velocity.[7] More important, uncertainty about the effects of financial innovations on M1,

[3]See, for example, Barnett (1983), Barnett et al. (1984) or Spindt (1985).

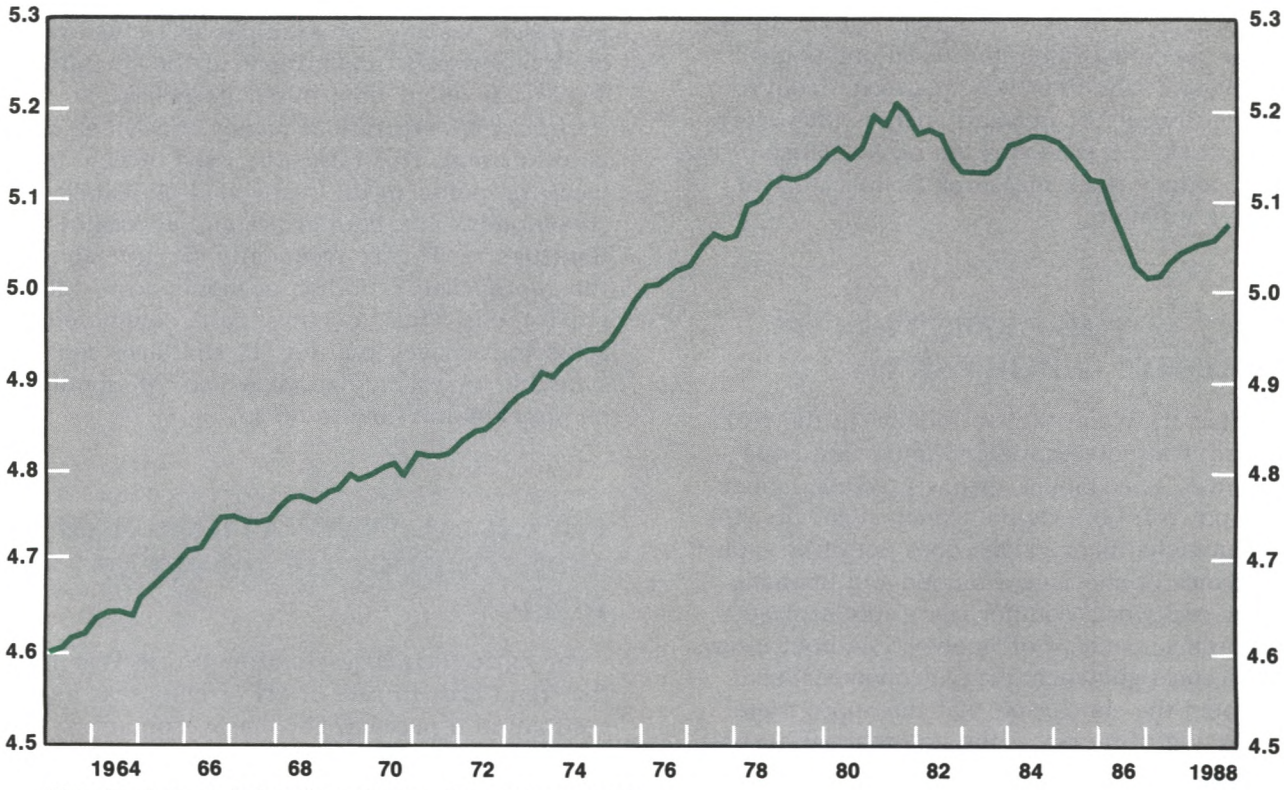
[4]See Belongia and Chalfant (1989) and Swofford and Whitney (1990a,b).

[5]See Friedman and Schwartz (1970), p. 91.

[6]See Laidler (1989) for a thorough review of the Quantity Theory's history, implications and controversies.

[7]See Gordon (1984), for example, for more discussion of this argument.

Figure 1
Log of M1 Velocity[1]



[1]Velocity data are indexed to a base value of 100 in I/1963.

in addition to velocity growth, weakened M1's informational content in predicting future price movements.[8]

This uncertainty has been an important element of monetary policy discussions. The record of the FOMC, for example, shows that, at the meeting of February 8-9, 1983:

> Committee members' views varied considerably on the weight to attach to M1 ... the performance of that aggregate had been subject over the past year or more to substantial uncertainties related to the growing role of NOW accounts and an apparent shift in the behavior of its income velocity ... Only modest allowance was made for the possibility that the new Super NOW accounts would draw funds into M1 from other sources.[9]

Two months later, in a statement to Congress, Federal Reserve Chairman Paul Volcker reported:

> To some extent — but it cannot be measured with any degree of certainty — the decreases in "velocity" may reflect the changing nature of M1; with interest-bearing NOW and Super NOW accounts making up an increasingly large proportion of M1, this aggregate may be influenced by "savings" behavior as well as "transactions" motives. That is a longer-term factor, and the growth in M1 over the shorter-run may have been affected by the reduced level of market rates — particularly relative to interest-bearing NOW accounts — and slowing inflation as well. *The range of uncertainty on these points is substantial* (emphasis added) and has led the Federal Open Market Committee to place less emphasis on M1 in the implementation of policy over the short term.[10]

[8]Monetary control also will be impaired if shifts of savings-type balances into interest-bearing checkables weakens the link between the monetary base and M1.

[9]*Federal Reserve Bulletin*, April 1983, pp. 288-289.

[10]*Federal Reserve Bulletin*, May 1983, p. 338.

Nonetheless, after being "monitored" but not targeted in 1983, M1 was chosen as an intermediate target of monetary policy from 1984 to 1986. In 1987, the FOMC again declined to set target ranges for M1 growth, instead choosing to emphasize M2 and M3 in policy discussions.[11] These official statements point to considerable uncertainty about how interest-bearing checking accounts were being used and how their use might affect the behavior of the monetary aggregates generally, the behavior of M1 specifically, and the relationship of M1 to important economic variables, such as spending and prices.

APPROACHES TO THE DERIVATION OF ALTERNATIVE MONETARY AGGREGATES

Choosing a monetary aggregate requires at least four steps.[12] First, the group of assets to be included in the aggregate must be selected. Second, individual asset categories must be aggregated appropriately. Third, the aggregate must be analyzed in terms of the central bank's ability to control it. Finally, its performance in some ultimate use must somehow be evaluated.

Before the 1980s, those who were skeptical of using the M1 aggregate for policy analysis generally focused on the fourth point, the methodology used to link M1 either to nominal GNP growth or the inflation rate. Many critics, for example, raised issues about the exogeneity of changes in the money stock, reverse causation running from income or prices to money or technical problems associated with estimation of reduced-form equations.[13] While such methodological criticisms continued throughout the 1980s, much more attention in recent years has been focused on the aggregates themselves rather than their final uses. The official money stock measures particularly have been criticized both for the weighting scheme used to aggregate individual asset categories and for the specific assets used to construct an aggregate. These criticisms are discussed in more detail below.

Composition of Monetary Aggregates

Historically, debates about what is money have focused on the things—coins, bank notes, etc.—that should be included in a monetary aggregate. Before the emergence of interest-bearing checkable deposits, the logic behind the assets included in M1 coincided with Irving Fisher's dictum, "Money never bears interest... ."[14] For economists of this school, the monetary aggregate also was viewed as containing only those assets that could be readily exchanged for goods.[15]

Since the birth of money market mutual funds (MMMFs) and, even more so, the nationwide introduction of NOW accounts in January 1981, however, the question of which assets are correctly identified as money has become complicated . Many argue that the historical justification for the group of assets in M1 is no longer clear. What, for example, distinguishes the NOW accounts in M1 from MMDAs or MMMFs, which appear to have characteristics similar to NOWs but are excluded from M1? Conversely, because NOW accounts pay interest and can be used as savings accounts, should they be excluded from a monetary aggregate designed to include those assets that can be readily exchanged for goods?

Tests for Weak Separability and Asset Selection

One approach to deciding which financial assets should be considered as a group for ag-

[11]For more detail on the uncertainty about the behavior of different aggregates, see Hafer and Haslag (1988) or Garfinkel (1990).

[12]Barnett, et al. (forthcoming) argue that this process also involves addressing a fifth step—monitoring how closely the index number tracks the economic aggregator function—that has been overlooked by previous research in this area.

[13]See, for example, de Leeuw and Kalchbrenner (1969), Davis (1969), Gramlich (1971), Goldfeld and Blinder (1972) and B. Friedman (1977), among others.

[14]The full quote is, "Money never bears interest except in the sense of creating convenience in the process of exchange." See Fisher (1963), p. 9.

[15]Ibid., p. 8.

gregation treats financial assets as commodities that are held for the services they provide. In this approach, individuals are assumed to allocate their financial wealth across the spectrum of available assets, according to their preferences for the characteristics of each asset and the relative returns each asset yields. Treating financial assets in this manner leads to inserting them as arguments in the utility function of the representative individual and analyzing the demand for financial assets in much the same way that the demand for other commodities can be studied.[16]

This approach leads to a search for "separable" commodity groups. Separability is related to the notion of multi-stage budgeting in which consumers first allocate their budget across broad expenditure categories such as "shelter," "transportation," "clothing" and "food." Then after allocating, say, 20 percent of their total expenditures to food, further allocations are made within the food group as expenditures are budgeted for "meat," "dairy products," "fruits and vegetables" and so on. The process continues with the gross expenditures for meat allocated across beef, pork and chicken. The result of this process can be pictured as a branching tree diagram in which broad commodity groups are continually divided into smaller, more specific commodity groups.

To decide which financial assets may be grouped under the heading "money," the research task is to find what goods are in the "money" group and what goods are unrelated to utility from the services of monetary assets. This identification can be done formally by testing whether an asset collection is weakly separable from other assets and commodities. Indeed, Barnett (1982, pp. 695-96), has shown that a necessary condition for the existence of a monetary aggregate is that the subset of monetary goods is weakly separable from nonmonetary goods.[17]

Weak separability, a result from demand theory, implies that the consumer's marginal rate of substitution between two assets in a group is unaffected by the quantities of any good not in the group. This ensures that changes in the quantities consumed of goods not in the group will not affect the marginal utilities derived from goods within the group. The practical implication of weak separability is that it permits demand analysis that is concerned only with a specific collection of assets; it does not require that one study all goods in the economy.

To view the problem differently, let some group of assets, A, consist of goods a1, a2 and a3; that is, $A = (a_1, a_2, a_3)$. Assume, however, that this group is not weakly separable from other assets; instead, the separable group also should include asset a4. By arbitrarily cutting off the group to include only a1-a3, changes in the quantity of a4 consumed would affect the marginal rates of substitution within the group of a1-a3 and give rise to seemingly unstable preferences for asset holdings or, from a production function point of view, an unstable transactions technology. In a more concrete sense related to traditional empirical work, the omitted variable a4 would affect an estimated demand function for an aggregate variable constructed from assets a1-a3 by acting as an omitted shift variable. For example, estimating a money demand function based on M1 balances when MMDAs should be part of this group could be the source of instability in such a demand function.[18]

Recent work by Swofford and Whitney (1987, 1988) and by Belongia and Chalfant (1989) has examined the conditions of weak separability for the asset groups that correspond to M1 and M1A (M1 minus interest-bearing checkable deposits), as well as broader collections of assets that include more components of M2. This research checked for the existence of a well-behaved utility function that was consistent with

[16]There has been a steady controversy about whether money can be modeled as an argument in the utility (or production) function; however, Feenstra (1986) has shown that this approach is equivalent to other standard approaches, including the Clower (transactions) constraint; see Clower (1970).

[17]For more detail on weak separability and other points related to finding an admissible monetary aggregate, see Varian (1982, 1983), Barnett (1987), Ishida (1984), Serletis (1987), Fayyad (1986), Hancock (1987) and Belongia and Chalfant (1989).

[18]In fact, studies of money demand functions often include a dummy variable around 1981-82 to capture the effects of financial innovation. Aggregates based on weakly separable groups, however, should be stable and eliminate the need to account for shifts in functional relationships with dummy variables.

the financial balances held in various forms at existing prices and levels of income. Upon finding that the data were consistent with demand theory, the separability tests were then used to identify the specific financial assets, in combination with currency, that were weakly separable from other financial assets in the utility function. These weakly separable asset groups were then candidates for aggregation.

Weighting of Asset Components

Another criticism of M1 as an indicator of monetary actions—specifically when examining why the financial innovations of the early 1980s might be associated with the sharp slowing in M1 velocity growth—focuses on the potential errors associated with the simple-sum weighting scheme used to derive M1 and other official money stock measures. The simple-sum aggregation scheme, which gives the same weight to currency as it does to dollars deposited in accounts with quite different characteristics, violates fundamental aggregation principles, so the argument goes.

For example, a dollar of currency (which is termed pure money because it pays no explicit rate of return) is weighted equally in M1 with a dollar in a NOW account deposit. This weighting occurs even though the NOW account pays a market rate of return that, in recent years, has approximated the return on savings deposits. Thus, even though some NOW account balances may be held as savings, they are included dollar for dollar in M1 with the other assets that traditionally have been considered to be transactions balances.[19] Such equal weights are justified only if the assets are perfect substitutes for each other; both the raw data as well as much empirical work demonstrate that these conditions do not hold.[20]

The fundamental point here is that each dollar in a NOW account should receive less weight than a dollar of currency, which presumably is held purely for the conduct of transactions. Similarly, if the assets collected in M2 are found to be a weakly separable group, each of its components should be weighted differently with the weights declining as own-rates of interest rise. At this point, it should be noted that, while simple-sum weighting is clearly wrong in principle, it may not in the past have mattered empirically; that is, it may not have produced major distortions in the estimated relationships between "money" and other variables. It also should be noted that the interest in weighted monetary aggregates is not new; the deficiencies of simple-sum aggregates were recognized more than 20 years ago.[21]

By formalizing this intuitive argument, Barnett and others constructed a series of Divisia monetary aggregates.[22] A Divisia index, which is merely one of many statistical index number formulas that have desirable aggregation properties, has an advantage over simple-sum aggregation in that it internalizes pure substitution effects. This means that the measured value of the index will not change unless the utility or production functions underlying the index actually produce a different level of utility or output. In contrast, the failure of simple-sum weighting to internalize substitution effects means it can change even when there has been no change in the flow of transaction services from the group of assets.

Unlike the simple-sum weighting scheme currently used, any good index gives different weights to balances held in different deposit categories. In the case of the Divisia index applied in monetary studies, the weights are based essentially on the difference between the rate of return on a pure store of value (typically, a long-term bond) and each asset's own rate of

[19]Some analysts, in fact, have attempted to infer the proportion of NOW balances held for that asset's savings characteristics (its interest rate and deposit insurance) on the basis of turnover data relative to demand deposit turnover rates. See Spindt (1985).

[20]See Barnett (1982) for the technical arguments about the shortcomings of simple-sum weighting. Ewis and Fisher (1984, 1985) and references they cite are sources that find low elasticities of substitution between M1 components and near-monies.

[21]See, for example, Friedman and Schwartz (1970), pp. 151-52.

[22]See Barnett (1983, 1984) and Barnett, Offenbacher and Spindt (1981, 1984).

return.[23] The larger this interest rate differential, the higher the opportunity cost of holding a particular financial asset and the greater its "moneyness." The largest interest differential and, therefore, the greatest "moneyness" occurs for currency, whose rate of return is zero.

The research cited above has put the definition and measurement of monetary aggregates on firmer economic and statistical footing; the specific aggregates it has produced, however, have been subject to criticism. One criticism has been that weighting may be, or at least may have been, unimportant empirically. Specifically, the relative performance of simple-sum vs. Divisia measures against various final uses criteria has been disputed.[24] Perhaps the most important criticism, however, is that this research has failed to analyze whether the re-weighted versions of the official monetary aggregates, M1, M2, M3 and L, represent asset groupings that can or should be aggregated. The shaded insert at right illustrates how these issues can affect the behavior of a monetary aggregate.

DIVISIA M1A AND THE MONEY METRIC INDEX

Previous research that has tested for weakly separable asset groups has suggested two weighted monetary aggregates as possible improvements over traditional simple-sum measures. These new aggregates are Divisia M1A and the Money Metric Index (MMI).[25] The former includes only currency and traditional demand deposit balances, which are weighted by their user costs to construct a Divisia aggregate.[26] The latter includes a wider range of assets: currency, demand deposits, other (interest-bearing) checkable deposits, savings deposits and overnight repurchase agreements. The interested reader is directed to the original sources for detail on the testing strategies that led to the choices of these asset groups.

The velocities for these two measures and M1 are shown in figure 2. Perhaps the most interesting aspect of figure 2 is that, in the early 1980s, when M1 velocity began its heralded

---

[23]A Divisia monetary aggregate is constructed in the following manner: Let q_{it} and p_{it} represent the quantities and user costs of each asset to be included in the aggregate at time t. The expenditure share on the services of monetary asset i in period t is:

$$s_{it} = \frac{p_{it}q_{it}}{\sum_j p_{jt}q_{jt}}.$$

The user cost of each asset is measured as:

$$p_{it} = \frac{P_t^*(R_t - r_{it})\,(1-M)}{1 + R_t(1-M)},$$

where P_t^* is the geometric mean of the CPI and GNP deflator, R_t is the maximum available expected holding period yield, r_{it} is the observed or imputed nominal own rate of return on asset i and M is the average marginal tax rate [see Barnett (1978)]. The growth rate of a Divisia aggregate then can be written as

$$G(Q_t) = \sum_{i=1}^{n} s_{it}^* G(q_{it}),$$

where $s_{it}^* = 1/2(s_{it} + s_{i,t-1})$, and n is the number of assets in the aggregate. For a more detailed explanation of this weighting procedure and its application to the Divisia aggregates, see Barnett (1978) and Barnett et al. (1984).

[24]Barnett (1984), Serletis (1986), Hancock (1987) and Ishida (1984) find the Divisia aggregates to be superior to simple-sum measures in a variety of final uses (e.g., stability of velocity, relationship with nominal spending or prices). Batten and Thornton (1985), against the standard of performance in nominal spending equations, find little difference between the two.

[25]The more narrow measure, Divisia M1A, is derived in Belongia and Chalfant (1989). The broader measure, the Money Metric Index (MMI), is discussed in Swofford and Whitney (1990b). Both groupings were determined by nonparametric tests for weak separability described in Varian (1982). While parametric tests for weak separability have been reported by Serletis (1986) and Hancock (1987), Swofford and Whitney (1986) have noted that such tests are sensitive to the choice of the functional form for the utility or production function.

[26]The user cost formula is derived in Barnett (1978). We use a simplified formula:

$$\text{user cost} = \frac{P_t^*(R_t - r_{it})}{1 + R_t}$$

Although the return on something completely illiquid, such as the return on human capital, would be the preferred choice to represent R_t, practical measurement issues have led to the use of long bond rates in empirical studies.

The approach used by Swofford and Whitney (1990a) actually calculates upper and lower bounds for the index rather than a unique value for each time period. To avoid any problems associated with choosing an incorrect value of their index, we simply used the Divisia weighting scheme to aggregate the asset collection they identified.

Calculating A Divisia Monetary Aggregate

The example below illustrates the consequences of using a monetary aggregate that is composed of the "incorrect" assets or that employs simple-sum weighting of its constituent assets even if they are the proper ones to use. For simplicity, we assume that only four asset categories exist: currency, traditional (non-interest-bearing) demand deposits, NOW accounts (interest-bearing checkable deposits) and savings deposits. We then define M1 as the sum of C + DD + NOWs and M2 = M1 + Savings. From these definitions, we construct four aggregates: simple-sum M1 and M2 and weighted versions of M1 and M2.

For purposes of these examples, we start with the following initial conditions:

Dollar Balances in Each Asset Category

C = 100 DD = 50 NOWs = 50; Savings = 50;
M1 = 200 M2 = 200 + 50 = 250

Interest Rates on Each Asset

Benchmark rate (R) = 0.1; $r_c = 0$; $r_{DD} = 0.03$; $r_{NOW} = 0.05$; $r_{SAV} = 0.05$.

User Costs

A simplified expression for the user cost formula, cited in footnote 26, is:

$p_{it} = \frac{R_t - r_{it}}{1 + R_t}$. Thus, the user cost for each asset is:

C: $\frac{.1 - 0}{1 + .1} = \frac{.1}{1.1} = .0909$

DD: $\frac{.1 - .03}{1.1} = 0.0636$

NOWs: $\frac{.1 - .05}{1.1} = 0.04545$

SAV: $\frac{.1 - .05}{1.1} = 0.04545$

Shares

Expenditure shares are calculated as

$s_{it} = \frac{p_{it}q_{it}}{\sum_i p_{it}q_{it}}$. Thus, for M1 assets:

C: .0909(100) ÷ 14.55 = 62.5

DD: .0636(50) ÷ 14.55 = 21.9

NOWs: .04545(50) ÷ 14.55 = 15.6

Similarly, for M2 assets:

C: .0909(100) ÷ 16.82 = 54.1

DD: .0636(50) ÷ 16.82 = 18.9

NOWs: .04545(50) ÷ 16.82 = 13.5

SAV: .04545(50) ÷ 16.82 = 13.5

The divisors in each column, 14.55 and 16.82, are the sums of expenditures on all assets in that group. With this information, we can illustrate how the two issues of concern—weighting and composition—affect the behavior of each aggregate.

An Illustration: Assume, for quarterly data, $10 is shifted from savings to demand deposits. For the simple-sum aggregates, the dollar changes in levels and corresponding growth rates are:

M1:	**Change**	**Growth rate**
100 + (50+10) + 50 = 210	+10	21.55%
M2:		
100 + (50+10) + 50 + (50−10) = 250	0	0%

Weighted M1: The growth rate of a Divisia index is constructed as the growth rate of each asset category weighted by the average shares of the two periods.[1] New budget shares at quarter t+1 are 59.9 (C), 25.1 (DD) and 15.0 (NOWs). Average shares (S*) for the two periods are 61.2 (C), 23.5 (DD) and 15.3 (NOWs). This yields:

$$0.612*(0) + 0.235*(\text{growth rate of DDs}) + 0.153*(0)$$
$$= 0.235\left[\left(\frac{60}{50}\right)^4 - 1\right]* 100$$
$$= 0.235*(1.0736)* 100 = 25.23\%$$

Weighted M2: For this aggregate, the budget share of DDs rises while that of savings de-

clines. Using the average budget shares across the two periods and recognizing that the changes in C and NOWs both equal zero, we have:

$$0 + .2069 * (107.36) + 0 + .1211 * (-59.04)$$
$$= 22.22 + (-7.15) = 15.07\%$$

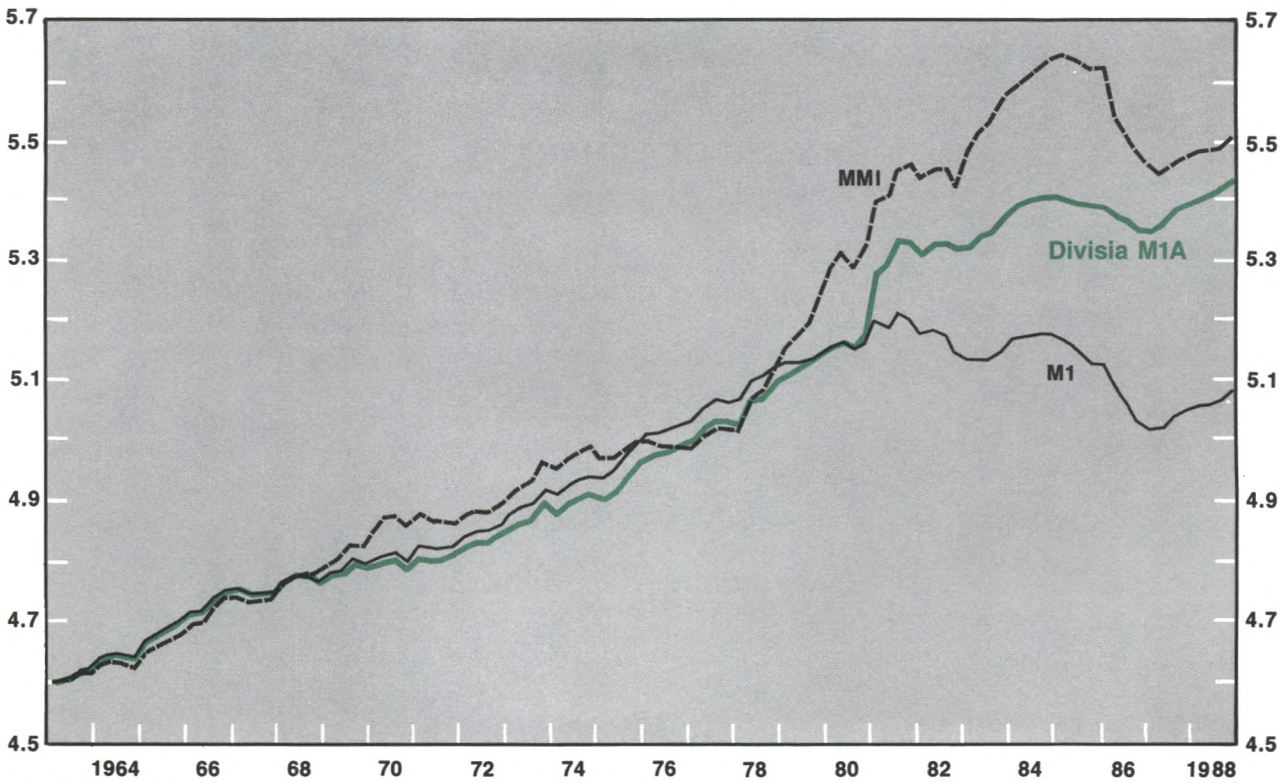
Thus, a simple $10 shift from one category to another produces growth rates ranging from zero (simple-sum M2) to 25.23 percent (Divisia M1).

The intuition behind these results is that $10 has been shifted from one deposit category (savings) to another with greater "moneyness" because demand deposits have a higher user cost, i.e. their own rate of return is lower. Thus, for example, Divisia M2 will show growth while simple-sum M2 will not because $10, while still in M2, has been shifted into a category that receives a larger weight when its growth rate is calculated.[2]

[1]The weighted measures are indexes set equal to 100 in the same base year as opposed to M1 and M2, which are measured in dollars. Thus, comparisons of changes in levels are uninformative.

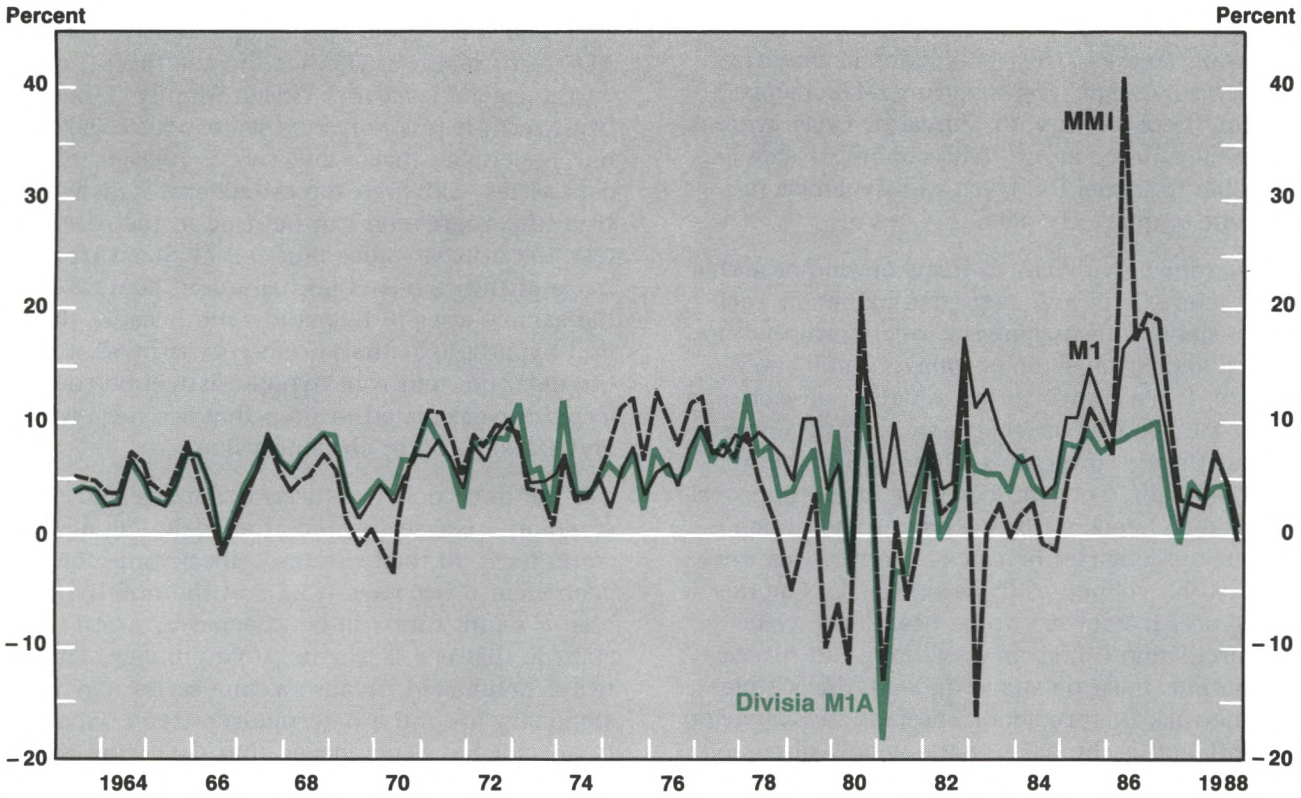
[2]Ideally, we would like to isolate the effect of an interest rate change alone, but this is not done for two reasons. First, economic theory predicts changes in quantities demanded or supplied in response to changes in relative prices. Second, as can be seen from the calculations for the growth rate of a Divisia index, the index will not change unless quantities change.

Figure 2
Logs of Velocity of M1, Divisia M1A, and MMI[1]



[1]Velocity data are indexed to a base value of 100 in I/1963.

Figure 3
Growth Rates of M1, Divisia M1A, and MMI



decline, the velocities of both weighted aggregates actually increased. Also interesting is the sharp drop in MMI velocity around 1986, at a time when M1 velocity also fell sharply but Divisia M1A velocity fell only moderately. Since 1986, all three appear to be on similar paths.

These trends are exhibited in different form in figure 3, which depicts the growth rate of each aggregate. Although the three series often track closely, there are notable exceptions. For example, the growth of MMI slows sharply in 1978-79 whereas the other measures continue to grow at rates near 5 percent. In 1980-81, both weighted measures show sharply negative growth rates while M1 growth continues to be positive; this is particularly interesting because the decline in MMI over the same period would refute the idea that this episode merely reflects shifts out of narrow M1A balances into interest-bearing accounts. In general, figure 3 indicates that the behavior of these aggregates was much more divergent throughout the 1980s than in previous decades.

Long-Run Relationships between Monetary Measures and the Price Level

The empirical performance of alternative monetary aggregates and their usefulness as targets for monetary policy might be evaluated in a variety of ways, for example, the use of St. Louis-type spending equations or short-run money demand functions with tests of coefficient stability over time. Either approach is quite sensitive to specification; for example, the results appear to depend crucially on both the sample period chosen and the lag lengths used

for all explanatory variables.[27] Money demand equations also are sensitive to the specification of such things as the scale and interest rate variables and the equations' short-run dynamics. Such short-run empirical controversies, however, are irrelevant for our purposes in this study. As Laidler (1990) has pointed out, many key discussions of money's effect on economic activity specifically emphasize long-run relationships, testing hypotheses with annual or cycle average data, and never argue that economists should be able to model the short-run dynamics inherent in quarterly data.

Therefore, we want to focus on one desirable characteristic of any aggregate monetary variable: stability in its long-run relationship with a variable of primary importance to monetary policy. If we return to the equation of exchange, $MV = PT$, we can see that long-run inferences about future movements in the price level (P) can be made from the behavior of the appropriate money stock (M) if we have some notion about the long-run behavior of output (expressed as the volume of transactions, T). For this analytical framework to be useful, the velocity of circulation (V), given some long-run forecast of output, must be stable. Indeed, this simple framework for predicting the long-run behavior of inflation is the basis of the widely discussed "P*" analysis of Hallman, Porter and Small (1989).

One way to investigate the stability of velocity is to test whether its time series has a unit root. If a time series has a unit root, it is said to be nonstationary; this means that "shocks" to its path may move the series permanently away from its previous pattern. In this case, it may wander from what appeared to be its trend without any force to push it back again. If velocity were a time series that did not exhibit some stable long-term behavior, the Quantity Theory framework would not provide accurate predictions of long-run inflation on the basis of money stock movements.

Velocity, of course, is calculated as $V = (PT)/M$. The velocity of circulation was calculated for Divisia M1A and MMI, and, for purposes of comparison, for previous (M1) and current (M2) monetary targets as well. The Augmented Dickey-Fuller (ADF) test for a unit root then was applied to each velocity series by estimating a regression of the form:

$$(1)\quad \Delta X_t = a + b\Delta X_{t-1} + c\Delta X_{t-2} + dX_{t-1} + \varepsilon_t,$$

where X_t is the natural logarithm of the variable, and Δ is the difference operator.[28] The ADF test examines whether the coefficient, d, on the lagged level of X is significantly different from zero. If it is not, then we cannot reject the null hypothesis that a unit root is present in the data series. Although the estimated t-statistic from this regression can be used in the ADF test, the critical values must be modified from those of the standard t-distribution. Because of the sample sizes in this study and because the null hypothesis is that a unit root is present in the data, the unit root hypothesis will be rejected for calculated t-values that are negative and exceed 2.90 in absolute value.

If the unit root hypothesis is not rejected, equation 1 is re-estimated after including a trend term; in this case the critical value for the coefficient d becomes -3.5 and the null hypothesis of a unit root will be rejected by a test statistic that is a larger negative number. This test is conducted because a time series may be stationary around a deterministic trend. Finally, because it has been alleged that data from the 1960s dominate relationships between money and other variables or that data from the 1980s distort the measurement of M1 and M2, the ADF tests were conducted over three sample periods: I/1963-IV/1980; I/1970-IV/1989; and I/1963-IV/1989.

Table 1A reports the ADF test statistics for the logarithms of each of the four monetary velocities during the three sample periods; the τ_u and τ_T columns refer to the cases without and with trend, respectively. As the results indicate, no velocity measure is stationary in the log levels even after accounting for trend. Thus, we difference the data again and estimate a regression of the form:

$$(2)\quad \Delta^2 X_t = a + b\Delta^2 X_{t-1} + c\Delta^2 X_{t-2} + d\Delta X_{t-1} + \varepsilon_t,$$

where the test again is whether the coefficient, d, is significantly different from zero. As before,

[27] See Friedman and Kuttner (1990) for evidence on the effects of a changing sample period and Thornton and Batten (1985) for a discussion of the sensitivity of such results to lag length selection.

[28] For more detail on the test, see, for example, Dickey, et al. (1986), especially pages 16-17. An FPE criterion was used to determine the number of lagged values for ΔX_{t-i}.

Table 1A

Augmented Dickey-Fuller Tests on Monetary Velocities

Monetary variable	I/1963 - IV/1980		I/1970 - IV/1989		I/1963 - IV/1989	
	τ_u	τ_T	τ_u	τ_T	τ_u	τ_T
M1	0.49	−1.19	−1.92	−1.26	−1.88	−1.05
M2	−1.25	−1.99	−2.33	−2.33	−2.72	−2.69
Divisia M1A	1.44	−0.32	−0.64	−1.66	−0.34	−1.79
MMI	1.53	−0.23	−0.91	−1.46	−0.73	−1.72

Table 1B

Augmented Dickey-Fuller Tests on Growth Rates of Monetary Velocities

Monetary variable	I/1963 - IV/1980		I/1970 - IV/1989		I/1963 - IV/1989	
	τ_u	τ_T	τ_u	τ_T	τ_u	τ_T
M1	−5.84*	−5.87*	−3.10*	−3.29*	−3.86*	−4.08*
M2	−4.50*	−4.51*	−4.34*	−4.31*	−5.31*	−5.29*
Divisia M1A	−5.59*	−5.96*	−4.55*	−4.52*	−5.38*	−5.36*
MMI	−3.54*	−3.86*	−3.01*	−3.00*	−3.62*	−3.61*

NOTE: An asterisk denotes statistical significance at the 5 percent level.

if the null hypothesis is not rejected, equation 2 is re-estimated with a trend term added.

The results of this estimation are reported in table 1B. In this case, the results again find nothing to distinguish any of the monetary indexes as indicators of long-run trends in inflation as each measure apparently exhibits a stable growth rate of velocity in each sample period. If one raises the significance level of the ADF test statistic to the 0.01 level, however, critical values for the τ_u and τ_T tests of about −3.6 and −4.15 would indicate that only the velocities of M2 and Divisia M1A show the existence of a single unit root over time. In contrast, doubts would be raised about the stability of the growth rate of MMI velocity in each of the three sample periods and about M1 velocity at least in the 1970-89 interval; the stability of M1 velocity over the entire 1963-89 sample would be a borderline call.

The upshot of these cursory results is that, for the specific final use criterion chosen (stationary velocity), a badly designed monetary aggregate (M2) may perform better than one that is measured in a manner consistent with both economic and statistical theory (MMI). Conversely, some aggregates that have considered the question of asset collection and weighting (Divisia M1A) perform better than some of the official aggregates (M1).[29]

[29]As was noted in the paper's introduction, a monetary aggregate also must be controllable by the central bank to be useful as an intermediate target variable. Applying a test used by Belongia and Chalfant (1989) to each of the four aggregates used in this article produces the following results:

(1) $\dot{Div\,M1A}_t = -2.25 + 1.05 * \dot{AMB}_t$; $\bar{R}^2 = .34$; DW = 1.53
(2.16) (7.49)

(2) $\dot{M1}_t = -3.21 + 1.37 * \dot{AMB}_t$; $\bar{R}^2 = .56$; DW = 1.39
(3.69) (11.72)

(3) $\dot{M2}_t = 2.98 + 0.78 * \dot{AMB}_t$; $\bar{R}^2 = 0.23$; DW = 0.86
(2.98) (5.76)

(4) $\dot{MMI}_t = -4.79 + 1.38 * \dot{AMB}_t$; $\bar{R}^2 = 0.18$; DW = 1.19
(2.32) (4.96)

The intuition of this test is to look for a close, contemporaneous relationship between the instrument of monetary control (the monetary base-AMB) and the target. By this standard, Divisia M1A and simple-sum M1 both appear to be controllable. Although the coefficient for AMB is statistically different from zero for M2 and MMI, the fits ($\bar{R}^2$ = 0.23 and 0.18) are poor and the low DW statistics suggest that other variables may have important influences on this relationship. Overall, these results do not make a strong case for the controllability of either M2 or MMI.

SUMMARY

The breakdown in the 1980s of historical relationships between the growth rates of M1 and both the price level and nominal income has led researchers to look for explanations of these changes. Some have argued that the historical relationships were spurious and that the current experience demonstrates the true relationship. Others have argued that the traditional relationships still hold after minor revisions to the estimating equations. Still others have used indirect evidence to make adjustments to the official measures of various monetary aggregates. None of these efforts has been completely successful in resolving what has been termed "the velocity problem".

This article has reviewed still another approach to this problem: constructing monetary aggregates from first principles, using economic and index number theory to determine which assets should be included in an aggregate as well as how they should be weighted in aggregation. The two measures discussed, Divisia M1A and the Money Metric Index, exhibited quite different velocity behaviors; Divisia M1A velocity is consistent with a stable long-run relationship with the aggregate price level while MMI velocity appears not to be. Similarly, M2 velocity appears to be stable over time whereas M1 velocity does not. If price stability is an important objective of monetary policy, monitoring the growth rates of certain monetary aggregates can be a viable approach to policymaking. At the same time, there is still no definitive answer in terms of all its final uses to the question: What is money?

REFERENCES

Afriat, S. N. "The Construction of Utility Functions from Expenditure Data," *International Economic Review* (1967), pp. 67-77.

Barnett, William A. "The User Cost of Money," *Economics Letters*, Vol. 1 (1978), pp. 145-49.

_____ . "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory," *Journal of Econometrics* (September 1980), pp. 11-48.

_____ . *Consumer Demand and Labor Supply: Goods, Monetary Assets and Time* (Amsterdam: North Holland, 1981).

_____ . "The Optimal Level of Monetary Aggregation," *Journal of Money, Credit and Banking* (November 1982, part 2), pp. 687-710.

_____ . "New Indices of Money Supply and the Flexible Laurent Demand System," *Journal of Business and Economic Statistics* (January 1983), pp. 7-23.

_____ . "Recent Monetary Policy and the Divisia Monetary Aggregates," *The American Statistician* (August 1984), pp. 165-72.

_____ . "The Microeconomic Theory of Monetary Aggregation," in William A. Barnett and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics* (London: Cambridge University Press, 1987).

_____ . "Developments in Monetary Aggregation Theory," *Journal of Policy Modeling* (Summer 1990), pp. 205-58.

Barnett, William A., and Seungmook Choi. "A Monte Carlo Study of Tests of Blockwise Weak Separability," Mimeo, Department of Economics, University of Texas at Austin, December 1986.

Barnett, William A., Melvin Hinich and Piyu Yue. "Monetary Policy with the Exact Rational Expectations Monetary Aggregates," in, Michael T. Belongia, ed. *Monetary Policy on the 75th Anniversary of the Federal Reserve System* (Kluwer Academic Publishers, forthcoming).

Barnett, William A., Edward Offenbacher and Paul Spindt. "New Concepts of Aggregated Money," *Journal of Finance* (May 1981), pp. 497-505.

_____ . "The New Divisia Monetary Aggregates," *Journal of Political Economy* (December 1984), pp. 1049-85.

Barnett, William A., and Paul A. Spindt. *Divisia Monetary Aggregates: Compilation, Data and Historical Behavior*, Staff Study #116, Board of Governors of the Federal Reserve System, May 1982.

Batten, Dallas S., and Daniel L. Thornton. "Are Weighted Monetary Aggregates Better than Simple-Sum M1?" this *Review* (June/July 1985), pp. 29-40.

Belongia, Michael T., and James A. Chalfant. "The Changing Empirical Definition of Money: Some Estimates from a Model of the Demand for Money Substitutes," *Journal of Political Economy* (April 1989), pp. 387-97.

Clower, Robert W. "Is There an Optimal Money Supply?" *Journal of Finance* (May 1970), pp. 425-33.

Darby, Michael R., Angelo R. Mascaro and Michael L. Marlow. "The Empirical Reliability of Monetary Aggregates as Indicators: 1983-87," *Economic Inquiry* (October 1989), pp. 555-85.

Davis, Richard G. "How Much Does Money Matter? A Look at Some Recent Evidence," Federal Reserve Bank of New York *Monthly Review* (June 1969), pp. 119-31.

Deaton, Angus, and John Muellbauer. *Economics and Consumer Behavior* (London: Cambridge University Press, 1980).

de Leeuw, Frank, and John Kalchbrenner. "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization—Comment," this *Review* (April 1969), pp. 6-11.

Dickey, David A., William R. Bell and Robert B. Miller. "Unit Roots in Time Series Models: Tests and Implications," *The American Statistician* (February 1986), pp. 12-26.

Ewis, Nabil A., and Douglas Fisher. "The Translog Utility Function and the Demand for Money in the United States," *Journal of Money, Credit, and Banking* (February 1984), pp. 34-52.

_____ . "Toward a Consistent Estimate of the Substitutability Between Money and Near Monies," *Journal of Macroeconomics* (Spring 1985), pp. 151-74.

Fayyad, Salam K. "A Microeconomic System-Wide Approach to the Estimation of the Demand for Money," this *Review*, (August/September 1986), pp. 22-33.

Feenstra, Robert C. "Functional Equivalence Between Liquidity Costs and the Utility of Money," *Journal of Monetary Economics* (March 1986), pp. 271-91.

Fisher, Irving. *The Purchasing Power of Money,* Reprints of Economic Classics (Augustus M. Kelley, 1963).

Friedman, Benjamin M. "Even the St. Louis Model Now Believes in Fiscal Policy," *Journal of Money, Credit, and Banking* (May 1977), pp. 365-67.

Friedman, Benjamin M., and Kenneth N. Kuttner. "Money, Income, Prices and Interest Rates After the 1980s," Working Paper, WP-90-11 Federal Reserve Bank of Chicago (July, 1990).

Friedman, Milton, and Anna J. Schwartz. "Money and Business Cycles," *The Review of Economics and Statistics* (February 1963, Part 2), pp. 32-64.

_____ . *Monetary Statistics of the United States* (Columbia University Press, 1970).

Garfinkel, Michelle R. "The FOMC in 1989: Walking a Tightrope," this *Review* (March/April 1990), pp. 17-35.

Goldfeld, Stephen M., and Alan S. Blinder. "Some Implications of Endogenous Stabilization Policy," *Brookings Papers on Economic Activity* (3:1972), pp. 585-640.

Gordon, Robert J. "The Short-Run Demand for Money: A Reconsideration," *Journal of Money, Credit and Banking* (Part 1, November 1984), pp. 403-34.

Gramlich, Edward M. "The Usefulness of Monetary and Fiscal Policy as Discretionary Stabilization Tools," *Journal of Money, Credit, and Banking* (May 1971), pp. 506-32.

Hafer, R. W. "The Money-GNP Link: Assessing Alternative Transactions Measures," this *Review* (March 1984), pp. 19-27.

Hafer, R. W., and Joseph H. Haslag. "The FOMC in 1987: The Effects of a Falling Dollar and the Stock Market Collapse," this *Review* (March/April 1988), pp. 3-16.

Hallman, Jeffrey J., Richard D. Porter and David H. Small. "M2 Per Unit of Potential GNP as an Anchor for the Price Level," Staff Studies, No. 157, Board of Governors of the Federal Reserve System, April 1989.

Hancock, Diana. "Aggregation of Monetary Goods: A Production Model," in William A. Barnett and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics* (London: Cambridge University Press, 1987).

Hetzel, Robert L. "M2 and Monetary Policy," Federal Reserve Bank of Richmond *Economic Review* (September/October 1988), pp. 14-29.

Ishida, Kazuhiko. "Divisia Monetary Aggregates and Demand for Money: A Japanese Case," Bank of Japan *Monetary and Economic Studies* (June 1984), pp. 49-80.

Klein, Benjamin. "Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money," *American Economic Review* (December 1974), pp. 931-49.

Laidler, David. "The Legacy of the Monetarist Controversy," this *Review* (March/April 1990), pp. 49-64.

_____ . "The Quantity Theory is Always and Everywhere Controversial—Why?," Seminar Paper 89-03, Centre for International Economic Studies, University of Adelaide (July 1989).

Offenbacher, Edward K. *The Substitutability of Monetary Assets*, unpublished Ph.D. thesis, University of Chicago, 1979.

_____ . "Disaggregated Monetary Asset Demand Systems: Estimation and an Application of the Preference Independence Transformation," Special Studies Paper No. 144, Board of Governors of the Federal Reserve System, July 1980.

Serletis, Apostolos. "Monetary Asset Separability Tests," in William A. Barnett and Kenneth J. Singleton, eds., *New Approaches to Monetary Economics*, (London: Cambridge University Press, 1987).

Serletis, Apostolos, and A. Leslie Robb. "Divisia Aggregation and Substitutability Among Monetary Assets," *Journal of Money, Credit, and Banking* (November 1986), pp. 430-46.

Spindt, Paul A. "Money is What Money Does: Monetary Aggregation and The Equation of Exchange," *Journal of Political Economy* (February 1985), pp. 175-204.

Stone, Courtenay C., and Daniel L. Thornton. "Solving the 1980s' Velocity Puzzle: A Progress Report," this *Review* (August/September 1987), pp. 5-23.

Swofford, James L., and Gerald A. Whitney. "Flexible Functional Forms and the Utility Approach to the Demand for Money: A Nonparametric Analysis," *Journal of Money, Credit and Banking* (August 1986), pp. 383-89.

_____ . "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure, and Money," *Review of Economics and Statistics*, (August 1987), pp. 458-64.

_____ . "A Comparison of Nonparametric Tests of Weak Separability for Annual and Quarterly Data on Consumption, Leisure and Money," *Journal of Business and Economic Statistics* (April 1988), pp. 241-46.

_____ . "Bounding an Economic Monetary Aggregate Under Nonhomothetic Preferences," *Journal of Business and Economic Statistics* (January 1990a), pp. 137-42.

_____ . "A Money Metric Index of an Economic Monetary Aggregate," *Economic Inquiry* (1990b), forthcoming.

Tatom, John A. "Are the Macroeconomic Effects of Oil-Price Changes Symmetric?, " in Karl Brunner and Allan H. Meltzer, eds., *Stabilization Policies and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, (Spring 1988), pp. 325-68.

Thornton, Daniel L., and Dallas S. Batten. "Lag-Length Selection and Tests of Granger Causality Between Money and Income," *Journal of Money, Credit, and Banking* (May 1985), pp. 164-78.

Varian, Hal R. "The Nonparametric Approach to Demand Analysis," *Econometrica* (July 1982), pp. 945-74.

_____ . "Non-parametric Tests of Consumer Behaviour," *Review of Economic Studies* (January 1983), pp. 99-110.

Wenninger, John. "The M1-GNP Relationship: A Component Approach," Federal Reserve Bank of New York *Quarterly Review* (Autumn 1984), pp. 6-15.

Alison Butler

Alison Butler is an economist at the Federal Reserve Bank of St. Louis. Lora Holman provided research assistance.

The Trade-Related Aspects of Intellectual Property Rights: What Is At Stake?

The pirating of U.S.-financed research and development discourages innovation, denies markets to American exports, and threatens technological progress. Protection of intellectual property rights preserves America's technological edge, which is a key to our continued international competitiveness."

—Clayton Yuetter,
U.S. trade representative, 1986

THE SENTIMENTS expressed by the former U.S. trade representative above could have easily been made by trade representatives from any industrialized country about their own country. In the last few years, many industrial countries have become increasingly concerned over the lack of international protection of intellectual property rights (IPR).

In a 1988 study, the U.S. International Trade Commission (USITC) attempted to estimate the economic effects of intellectual property rights infringement by foreign countries on U.S. firms.[1] In its survey, firms estimated their losses at $6.2 billion in exports sales in 1986. These losses accounted for approximately 1.4 percent of the total export sales of products which are covered by some form of intellectual property rights that year. Firms estimated additional losses of $1.8 billion in sales in 1986 due to imports that infringed on their domestic IPR. In a related study, Feinberg and Rousslang (1990) estimate that the ratio of lost profits to worldwide sales of firms selected from the USITC study ranged from 0.05 percent for the extractive sector to 3.6 percent for the entertainment sector.[2] Since firms incur research and development (R&D) expenses to develop intellectual property, these lost sales and profits defer firms from undertaking the risky process of developing intellectual property in the future.

Frustrations over the lack of effective international agreements regarding IPR have led the United States to unilaterally create measures to deal with what it perceives as unfair trading practices. In particular, the so-called "Special 301" clause of the Omnibus Trade and Competitiveness Act of 1988 requires the U.S. Trade Representative to identify countries that do not

[1]USITC (1988). For limitations of the survey and data, see the study.

[2]See also Feinberg (1988).

adequately protect or enforce IPR. The trade representative then has the authority to bring an unfair trade practice case against that country. Although several countries have been put on a "priority watch list," no cases have yet been initiated by the U.S. Trade Representative under the "Special 301" designation.

Countries throughout the world, particularly the more industrialized ones, are interested in increasing the amount of protection and enforcement of IPR internationally. Their concerns are seen in the current negotiations of the General Agreement on Tariffs and Trade (GATT). GATT is the principal rule-making body for international trade, whose goal is to eliminate trade barriers that reduce the free flow of goods. In the current round of negotiations, the so-called Uruguay Round scheduled to end in December 1990, one of the 15 negotiating groups is devoted to developing an agreement regarding "trade-related aspects of intellectual property rights, including trade in counterfeit goods."[3] Currently, intellectual property rights are explicitly excluded from GATT's auspices.

There are many economic and legal issues related to the protection of intellectual property.[4] The analysis in this paper focuses only on the trade-related aspects of IPR, that is, the effect of differential (and uncertain) IPR across countries on trade and the benefits and costs of creating international standards for protecting intellectual property.

INTELLECTUAL PROPERTY

Intellectual property is an invention, idea, product or process that has been registered with the government and that awards the inventor (or applicant) exclusive rights to use the invention for a given period of time. It confers "...the right to exclude others from making, using, or selling the invention within the national territory."[5] (For a discussion of why intellectual property is awarded protection, see shaded insert on pages 39 and 40.) Industrial property protection is generally awarded only to new and useful products and production processes. Intellectual property is protected by the government in a variety of ways. Copyrights are awarded to protect original works of authorship, such as literary, artistic and musical works; trademarks allow a manufacturer exclusive rights to a distinguished name, symbol or mark.[6] As noted above, these rights are only valid in the countries in which a patent application has been awarded. This lack of an international IPR system could have a significant impact on the amount of innovative activity.[7]

CURRENT PROTECTION OF IPR

The protection awarded different types of intellectual property varies substantially across countries. As shown in table 1, most countries have patent lengths of 15 years or more.[8] Historically, an innovation was awarded patent protection for the number of years approximately equal to the amount of time it took to train two apprentices (14 years).[9] Although the apprentice system has been obsolete for many years, patent lengths have remained essentially unchanged.

Often, however, the actual period in which a firm can sell its product under patent protection is shortened considerably; for example, tests for

---

[3]GATT (1990), p. 11.

[4]For a discussion of some of the legal issues regarding IPR in international trade, see Meessen (1987). To narrow the focus of the paper, two issues have been ignored: which items should be protected by intellectual property regulations and what type of protection—patents or copyrights—is appropriate for certain types of goods, such as software.

[5]USITC (1988).

[6]Other types of intellectual property, such as trade secrets and mask works, constitute a small percentage of all intellectual property and are not discussed here. In addition, because there are no internationally agreed-upon definitions for intellectual or industrial property, these definitions should be considered chiefly as guidelines.

[7]Intellectual property protection is important insofar as innovation is desirable. Some analysts argue (see, for example, Nordhaus, 1969; Grossman and Helpman, 1989) that innovation increases growth worldwide, improving the quality of life through such things as better medicines, improved living conditions and safer production processes. Of course all innovations may not be beneficial, nor do all people necessarily benefit from all innovations (see, for example, Kamien and Schwartz, 1982 and Maskus, August 1990). In this paper, it is assumed that innovation has a net positive effect on a country's growth and on growth worldwide.

[8]All countries referred to are members of the World Intellectual Property Organization (WIPO). All statistics in this section, unless otherwise cited, are from WIPO (1988).

[9]Benko (1987).

Table 1
Domestic Protection

Intellectual property	Number of countries	General length
Patents[1]	18	5-10 years
	28	10-15 years
	57	15-20 years
Copyrights[2]	8	25 years
	75	50 years
	9	50 + years

[1]From date filed.

[2]From author's death.

SOURCE: World Intellectual Property Organization (1988)

product safety (which, for pharmaceuticals, can take up to 10 years) are included in the life of a patent. Recently, both the United States and the European Community proposed legislation that would increase the effective length of patent protection—that is, the amount of time a firm can actually market or use a product before its protection expires.[10]

Several additional factors affect the amount of protection intellectual property is awarded in any given country. As tables 2 and 3 show, many countries, including some industrial ones, exclude certain industries from patent protection. Those excluded are primarily high-technology industries with very high R&D intensities that tend to produce expensive products. Although currently a substantial amount of innovation oc-

[10]See *Congress and the Nation* (1985) and "Patents for Pharmaceuticals" (1990).

Table 2
Patent Exclusions

Product	Number of countries that exclude	Industrial countries[1]	Percent industrialized
Pharmaceutical products (PHARM)	47	8	17.0%
Animal varieties (ANIM)	59	18	30.5
Methods for treatment of humans or animals (TREAT)	59	18	30.5
Plant varieties (PLANT)	57	18	31.6
Biological processes producing plant or animal varieties (BIO)	56	18	32.1
Food products (FOOD)	34	8	23.5
Computer programs (COMP)	48	20	41.7
Chemical products (CHEM)	21	2	9.5
Nuclear inventions (NUC)	13	2	15.4
Pharmaceutical processes	10	2	20.0
Food processes	9	3	33.3
Microorganisms	8	1	12.5
Substances from microbiological processes	6	1	16.7
Cosmetics	2	0	0.0
Fertilizers	2	0	0.0
Mixture of metals and alloys	2	0	0.0
Agricultural machines	1	0	0.0
Anticontaminants	1	0	0.0
Methods of agriculture or horticulture	1	0	0.0

[1]These are the Western European countries, Japan, Canada, Australia, United States, New Zealand and Iceland.

SOURCE: World Intellectual Property Organization (1988)

Table 3
Selected Patent Exclusions by Country

Industrialized Countries	PHARM	ANIM	TREAT	PLANT	BIO	FOOD	COMP	CHEM	NUC
Australia	x					x	x		
Austria-Belgium		x	x	x	x		x		
Canada		x	x	x	x	x	x		
Denmark		x	x	x	x	x	x		
Finland-Norway	x	x	x	x	x	x	x		
France		x	x	x	x		x		
Germany		x	x	x	x		x		
Greece	x	x[1]	x[1]	x[1]	x[1]		x[1]		
Iceland	x					x			
Ireland									
Italy		x	x	x[1]	x		x		
Japan			x				x		x
Luxembourg		x	x[1]	x	x		x[1]		
Netherlands		x	x	x	x		x[1]		
New Zealand	x					x			
Portugal	x	x		x	x	x	x	x	
Spain	x	x	x	x	x		x	x	
Sweden		x	x	x	x		x		
Switzerland[2]		x	x	x	x		x[3]		
United Kingdom		x	x	x	x		x		
United States									x
Total	8	18	18	18	18	8	20	2	2
Developing Countries									
Asian	13	6	8	6	5	7	2	6	3
India	x		x	x		x		x	x
Indonesia									
Pakistan	x								
Rep. of Korea	x					x		x	x
Thailand	x	x		x	x	x	x		
African	8	20	18	20	20	5	18	1	0
Western Hemisphere	10	8	7	8	8	8	2	6	3
Brazil	x	x	x	x	x	x	x	x	x
Mexico	x	x	x	x	x	x	x	x	x
Venezuela	x					x		x	
European	8	7	8	5	5	6	6	6	5
Eastern	7	4	5	4	2	6	3	6	5
Total	39	41	41	39	38	26	28	19	11
TOTAL	47	59	59	57	56	34	48	21	13

NOTE: See table 2 for definitions.

[1]If filed through the European Patent Convention (EPC).

[2]Includes Liechtenstein

[3]Liechtenstein excludes if filed through the EPC.

SOURCE: World Intellectual Property Organization (1988)

curs in some excluded industries—for example, genetic engineering—it generally takes place in countries that do not exclude them from protection. For example, although European firms account for 82 percent of world investment in industrial plant and other biotech assets, only 2 percent of it was spent in Europe. Similarly, 77 percent of all patents in biotechnology were issued in the United States and Japan, which offer the most extensive patent protection.[11]

The length of copyright protection, on the other hand, is a more standardized measure across countries, with the majority of countries issuing copyright protection for the remainder of the author's life plus 50 years.

Other factors that influence the extent of protection for intellectual property are the enforcement of existing laws and the amount of copyright and patent protection in other countries. These factors are particularly important when examining the trade-related aspects of IPR.

TRADE-RELATED ASPECTS OF INTELLECTUAL PROPERTY: SOME THEORETICAL ISSUES

Why IPR Is an International Issue

The problem with the lack of an international system of rules regarding IPR occurs when the cost of copying an innovation (including the cost of penalties if caught) is less than that of either purchasing or leasing the technology itself.

There are primarily two ways that a product patent can be infringed upon: First, a copy of the product can be produced, with essentially the same characteristics (although not necessarily the same quality) as the original, but with no pretense of being the original product. Typically, such copies are sold at a lower price. An example of this occurred in 1976 when Kodak introduced a line of instant cameras that were similar to those already patented by Polaroid. In 1986, U.S. courts ruled that Kodak had infringed on Polaroid's patents and awarded the Polaroid Company $5.7 billion; this amount was reduced to $909.5 million in 1990 on appeal.

The second type of patent infringement would occur if a company produced a camera similar to Polaroid's and labeled and sold it as a Polaroid. This is an example of a "counterfeit" product. In this instance, Polaroid's trademark as well as its patents would be infringed upon.

International trade can affect even non-traded products that are protected by IPR. For example, suppose a pharmaceutical firm, call it SAW, recently developed a new product called NOCOLD, that cured the common cold and had no side effects. If the firm intended only to sell NOCOLD in its own country (call it the North), it would file for a patent only in the North.

Even if SAW does not export NOCOLD, however, the rate of return earned could be affected by international trade. For example, a firm in another country could create either a counterfeit version or a cheaper imitation of the product and export it to the North.[12] Similarly, a country that is producing NOCOLD legally could export it back to the North.[13] If sales of the foreign-produced product in the North reduce SAW's sales of NOCOLD and its profits on the product, the firm would expect a lower return on NOCOLD as well as all prospective products. Because the expected rate of return affects the decision to develop new products, the firm undertakes fewer projects in the future. Thus, protecting intellectual property is not just a domestic issue.

If the company that infringed upon a patent in either of the two ways was not a U.S. company, the patent laws in both the United States and the country in which it was located would have to be considered. Unless an agreement states otherwise, a patent or copyright is valid only in the country in which it was issued.[14] Thus, if Kodak had produced and sold its cameras in countries that did not honor Polaroid's patent, Polaroid would have been unable to sue Kodak. Because many patent violations occur across national borders, differences in pa-

[11]"Bugs that Divide" (1990).

[12]Not all countries have made it illegal to import products that infringe on domestic patents.

[13]Parallel trade, which occurs when a product produced under a patent in one country is exported to another country and competes with the patented product in that country, is legal in many countries and for intra-EC trade.

[14]There are surprisingly few agreements among countries to honor each other's patents. This is discussed in greater detail later.

The Economics of Innovation

There are essentially two types of technological innovations: *process* innovations, which are new production processes or improvements on existing technology, and *product* innovations, which are the creation of new products or improvements on existing products.[1] Both types of innovation are patentable. Because the economics of these two are essentially the same, the discussion focuses on production innovation for simplicity.

Intellectual property has the unusual (although not unique) property that the knowledge it contains is not depleted with use. For example, no matter how many times the formula for aspirin is used, the formula itself (that is, the knowledge contained in the patent) remains unchanged. As a result, the marginal cost of using this knowledge (e.g., the formula for aspirin) is zero. For economic efficiency, this knowledge should be made available to anyone interested, because doing so does not diminish the stock of knowledge (or reduce the number of times aspirin can be made). Over time, however, such a policy would have some unfortunate consequences.

Generally, innovation is the result of investment expenditures on research and development (R&D). Because expenditures on R&D occur before a new product is created, the firm's decision to incur these costs involves considerable uncertainty. The expected rate of return on R&D, which is the present discounted value of the stream of net operating profits divided by the present value of the R&D costs, has to be at least as great as the opportunity cost of resources devoted to R&D—that is, the expected rate of return that would have been earned if the same resources allocated to R&D were invested elsewhere.

While the opportunity cost of capital is easy to determine (it is simply the interest rate), the rate of return on R&D is more difficult to ascertain. It depends on how much R&D must be spent before a new product is discovered and developed, how much demand there will be for the new product, and how much production costs will be. The return on R&D also depends on the time the firm can produce the product exclusively and therefore earn economic profits. While the first three factors are outside the scope of this paper, the last factor demonstrates how IPR can effect the incentive to innovate.

In the absence of government intervention, maintaining exclusive rights to an innovation for any period of time is often difficult. Given that the marginal cost of using the knowledge created by the innovation is zero, one could conclude that governments have no reason to award these rights. Without assigning exclusive rights to produce the innovation, however, the amount of time the innovating firm can produce the product is both less certain and likely shorter; any other firm that can figure out how to make the product could also produce it without changing the knowledge associated with the innovation. For example, a firm that did not discover the formula for aspirin but, instead, was able to produce it would reduce the return earned by the innovating firm. This is true even though entry by the noninnovating firm in this market does not diminish the innovating firm's ability to produce aspirin. This reduced return on the investment in R&D appears to increase efficiency by promoting competition; however, it also reduces the number of R&D projects that will be undertaken. If, however, the government assigns property rights to

[1]The economic reasons for copyright protection, essentially the same as those for patent protection, are not discussed separately. The reasons for protecting trademarks, however, differ from those for patent and copyrights, as trademarks are thought to provide information for consumers about the quality of a product. Protection of trademarks is intended to ensure that they have some informative value to consumers, rather than to protect an idea itself.

innovations (and enforces them), then the amount of time the product can be produced exclusively will increase, raising the rate of return on R&D, which in turn has a positive effect on the amount of innovation.[2]

Of course, the world is not certain. There is no way of knowing in advance whether the R&D expenditures will produce an economically viable product. Intellectual property rights are a way of rewarding firms for incurring the risk associated with R&D by increasing the expected rate of return on R&D, thereby making more projects possible.

As long as innovation is considered desirable, assigning property rights to intellectual property is one way to encourage firms to innovate.[3]

[2]There is some disagreement as to how much IPR increases innovation (see, for example, Maskus, 1990). The question is a difficult one to measure empirically, because it requires determining what a firm would do in a situation that has not occurred. One attempt to do so is Mansfield (1986), who finds some role for patents in encouraging innovation.

[3]For a general discussion of the role of R&D on innovation, see Kamien and Schwartz (1982). In addition, the issue of determining the optimal length and breadth of a patent (within a country), has recently been discussed in the *RAND Journal of Economics* (Spring 1990) and *Economic Inquiry* (October 1984).

tent laws across countries and the lack of an international enforcement system affects the incentives associated with innovation.

IPR and International Trade: An Example

Suppose that SAW wanted to sell NOCOLD internationally as well. A firm can do this one of three ways, all of which are affected by the state of international property rights. For simplicity, the example below compares the case in which only one of the two countries protects IPR.

Direct Exports—First, SAW can export NOCOLD directly to another country, which we will call the South. If the product is not protected by a patent in the South, either because SAW has not filed for protection there or because the South does not protect IPR, cheaper copies that use the same formula as NOCOLD could be legally sold. Similarly, if the trademark is not protected, counterfeit products that are indistinguishable from NOCOLD can also be legally sold. These copies decrease SAW's total sales and profits associated with producing and selling NOCOLD. In addition, if the product has a trademark that is strongly associated with SAW, the counterfeit product, if of significantly poorer quality, could adversely affect the reputation of the firm and further harm the sales of both NOCOLD and future products of SAW.

Licensing—SAW can also license the technology to produce NOCOLD to a firm in the South. The Southern firm, in turn, pays SAW a royalty fee. Without protection of IPR, however, a firm has no incentive to pay licensing fees if it is cheaper to copy the product than pay those fees. For industries where licensing represents a significant proportion of its revenues, the lack of IPR can cause a substantial loss of revenue for a firm. For example, motion picture firms receive considerable revenues from licensing fees paid by foreign distributors. Every "pirated" copy of a movie that is sold or shown represents a loss of revenue for the movie company.[15]

Foreign Direct Investment—The third possibility, called foreign direct investment, occurs if SAW locates production facilities in the South to produce NOCOLD. Again, however, the lack of IPR in the South has a similar effect on innovation as discussed previously. Because SAW has no guarantee that it can control the production of NOCOLD in that country, the expected rate of return on the innovation is smaller, reducing the profitability of foreign investment in the South; as a result, fewer firms will engage in such foreign direct investment. A country's lack of adequate protection of IPR could be particularly costly in this case because the country is foregoing some of the benefits of foreign direct investment, such as new resources, training and

[15]For a discussion of the effects of international piracy in motion pictures in the United States, see USITC (1988) and Plock (1989).

employment. Obviously, counterfeiting firms might provide some of these same benefits; since they develop no new products themselves, however, they are dependent on others for innovations to copy.[16]

As long as the direct cost of counterfeiting (or copying), including the likelihood and penalties associated with being caught, is less than the profits earned by the firms doing the copying, firms will continue to pirate technology. Copying, however, lowers the rate of return earned by innovators and therefore the overall amount of innovation. In the long run, the resulting decline in the amount of innovation reduces the counterfeiting that can be done by firms in the South. On an aggregate level, the citizens of both countries end up worse off. There is less innovation, fewer products overall, and worldwide growth is therefore lower.

While the vast majority of countries have laws protecting intellectual property, the arguments presented above describe what occurs in countries that either have weak IPR or simply fail to enforce whatever IPR they do have. Given the apparent advantages of protecting intellectual property, why are there so many problems with trade in goods affected by intellectual property rights? Part of the answer to this question can be found by examining the current international system of regulating intellectual property rights.

CURRENT SYSTEM OF INTERNATIONAL IPR

The main organization responsible for international agreements on IPR is the World Intellectual Property Organization (WIPO), which is administered under the auspices of the United Nations. WIPO's objectives include administering international treaties and agreements on intellectual property rights and encouraging the protection of IPR worldwide.[17] Currently, 125 countries are members of WIPO.[18] Table 4 provides a partial list of these countries and of signatories of several treaties administered by WIPO. The two major international agreements on IPR are the Paris and Berne Unions, which deal with industrial property (including patents) and copyrights, respectively. (For a description of the two agreements, see shaded insert on page 43.) In addition to these unions, other international property right agreements administered by WIPO cover such things as industrial design, satellite transmissions and registration of trade and service marks.[19] Not all members of WIPO, however, are signatories to these unions and treaties, nor are the number of signatories for each agreement identical (see table 4).[20]

Differing Patent Regulations

One important issue regarding the trade-related aspects of intellectual property is the problem of differential patent regulations across countries. For example, while most countries determine patent eligibility based on a first-to-file basis, the United States employs a first-to-invent rule. As a result, patent protection for the same invention could be awarded patent protection to different applicants depending on whether the actual inventor was the first to file.

WIPO has had some success in trying to standardize the process of obtaining patents internationally through the Patent Cooperation Treaty (PCT). This treaty allows applicants to file for a patent in a central office and specify in which of the signatory countries (shown in table 4) it wishes the application to have effect. This process reduces the costs of filing for patents by centralizing the search and examination work associated with determining patent eligibility. The PCT is also designed to increase the time an applicant has to decide whether to withdraw the application for foreign patents. A firm might choose to withdraw its patent application and avoid the expense associated with translating the patent application into the local language and finding a local patent agent if the demand for the product to be patented is likely to be too small.

---

[16]Whether or not foreign direct investment is desirable for developing countries is a separate issue and is discussed elsewhere. (See, for example, Hood and Young, 1979.) For the purposes of this paper, we assume that the lack of IPR does not mean a country doesn't want foreign direct investment—there are more straight-forward ways to reduce or eliminate foreign investment.

[17]For a list of organizations that are solely administered by WIPO, see WIPO (1988), p. 8.

[18]These statistics exclude the German Democratic Republic because of the unification with the Federal Democratic Republic that occurred on October 3, 1990.

[19]See WIPO (1990a).

[20]The United States, for example, did not sign the Berne Union until 1988.

Table 4

Signatures of Selected International Intellectual Property Agreements

Industrialized Countries	Members of Treaties			
	WIPO	Berne	Paris[1]	PCT
Australia	x	x	x	x
Austria-Belgium	x	x	x	x
Canada	x	x	x	x
Denmark	x	x	x	x
Finland-Norway	x	x	x	x
France	x	x	x	x
Germany	x	x	x	x
Greece	x	x	x	
Iceland	x	x	x	
Ireland	x	x	x	
Italy	x	x	x	x
Japan	x	x	x	x
Luxembourg	x	x	x	x
Netherlands	x	x	x	x
New Zealand	x	x	x	
Portugal	x	x	x	
Spain	x	x	x	x
Sweden	x	x	x	x
Switzerland[2]	x	x	x	x
United Kingdom	x	x	x	x
United States	x	x	x	x
Total	24	24	24	19
Developing Countries				
Asian[3]	27	9	17	4
India	x	x		
Indonesia	x		x	
Pakistan	x	x		
Rep. of Korea	x		x	
Thailand	x	x		
African	40	25	35	14
Western Hemisphere	23	14	11	2
Brazil	x	x	x	x
Mexico	x	x	x	
Venezuela	x	x		
European	11	11	12	4
Eastern	7	7	7	3
Total	101	59	75	21
TOTAL	125	83	99	43

[1]Dominican Republic, Iran, Nigeria, San Marino and Syria are not members of WIPO.

[2]Includes Liechtenstein.

[3]Hong Kong, Singapore and Taiwan are not members of WIPO.

SOURCE: World Intellectual Property Organization (1990a)

The Paris and Berne Unions: Some Background Information

"The Paris Union, the official name of which is the *International Union for the Protection of Industrial Property*, was founded by a Convention signed in Paris in 1883 and last revised in 1967.

Under the Convention, each member State must accord the same protection to the inventions, trademarks and other subject matter of industrial property of the nationals of the other member States as it accords to those of its own nationals. The Convention also provides certain facilities to foreigners; for example, it allows them, without losing their claim to novelty, to file their applications for patents up to a year after first filing in the country of origin. It contains provisions concerning the conditions under which a State may license the use of a patented invention in its own territory if, for example, the owner of the patented invention does not exploit it in such territory.

The Berne Union, the official name of which is the *International Union for the Protection of Literary and Artistic Works*, was founded by a Convention signed in Berne in 1886 and last revised in 1971.

Under the Convention, each member State must accord the same protection to the copyright of the nationals of the other member States as it accords to that of its own nationals. The Convention also prescribes some minimum standards of protection; for example, that copyright protection generally continues throughout the author's life and for 50 years thereafter. It includes special provisions for the benefit of developing countries."

WIPO, *General Information*, 1990

Ninety percent of all patent applications are filed in the 43 countries that make up the membership of the PCT.[21] The usefulness of the PCT is demonstrated by the surge in the number of applications received by the PCT, from 2,625 in 1979 to 14,874 in 1989.[22] In addition, the number of countries designated for patent protection by applicants rose from 6.6 percent to 15.8 percent of member countries during the same time period.

Enforcement of IPR

As mentioned previously, the enforcement of existing IPR is another big problem. Although WIPO administers several international accords regarding IPR, few laws currently enforce these treaties. The Paris Union only requires signatories to give foreigners national treatment, that is, award them the same rights as they give their own citizens.

The Berne Convention includes two enforcement provisions. First, it established the presumption of authorship so that the author does not have to prove it; rather, challengers of the copyright would have to provide evidence to the contrary. Second, copies that infringe on a copyright are subject to seizure in any country in which the work has a copyright. This also applies to imports of reproduced materials from countries where the work is not protected.

WIPO does not have an international dispute settlement mechanism whereby an applicant (or country) can file a complaint against another country's implementation of the treaties. The only recourse under the agreement is to bring a case before the International Court of Justice.[23] That court, however, can only arbitrate cases

[21]The statistics in this section are from WIPO (1990a, 1990b).

[22]The potential of this agreement has just begun to be realized. For example, the United States, which is a signatory of the PCT, had 66,850 applications for patents by foreigners. (U.S. Department of Commerce, Patent and Trademark Office, 1989).

[23]The court's ruling is not binding for all members, however. See WIPO (1988).

that relate to the interpretation or application of the Paris and Berne Unions. There appears to be no penalty for ignoring the ruling of the Court.

WIPO does require countries to give applicants access to the same legal remedies for patent infringement as they do their own nationals. This requirement, however, is subject to existing laws for patent violations. In fact, many countries do not have explicit penalties associated with violations of IPR, and few impose civil penalties. According to WIPO, as of 1988, only 14 countries had laws requiring the seizure of infringing patent articles, 20 countries granted compensation of damages and 11 destroyed infringing goods. There is somewhat more protection for products with trademarks. WIPO is now preparing a study that will examine the possibility of establishing a new treaty which creates a dispute mechanism to arbitrate possible violations of international IPR agreements.

Other Issues

The problems described above explain why there is considerable concern about the current international system of IPR. Having an international set of rules regarding the protection of intellectual property reduces the uncertainty associated with innovation and increases the expected return earned on those innovations. This does not mean that all countries must have the same degree of protection—GATT, for example, has different rules regarding the acceptable amount of tariffs for industrial and developing countries. Rather, an explicit set of agreements, along with an effective mechanism to mediate disputes, could significantly decrease the loss of earnings associated with copying and counterfeiting innovated products. In addition, such agreements could reduce the information costs of determining the amount of protection of intellectual property. These costs can be substantial for firms that intend to sell their products internationally, and, as described above, must file for a patent in each country where it wants to sell the product. Similarly, given the current system in which countries can choose to exclude specific products from patent protection, firms are faced with the choice of selling their products in some markets with no protection or avoiding certain markets altogether. Developing countries that protect IPR will increase its access to new technology because innovating firms will have stronger incentives to produce and sell their products in countries that protect intellectual property.

PROBLEMS IN REACHING INTERNATIONAL AGREEMENTS ON IPR

Why are there so few (and such weak) international agreements on IPR? One primary difficulty in protecting and enforcing IPR is that the incentives to do so differ across countries, particularly between countries that are technology exporters and those that are technology importers.[24] Generally, less innovation occurs in developing countries; instead, in the absence of licensing or foreign direct investment, firms in these countries tend to produce goods whose production technology has become standardized or whose patent protection has expired.

Firms that successfully pirate technology in many developing countries are often able to produce essentially the same product at substantially lower production costs. Because there is less innovation in developing countries, the cost of not protecting IPR (reduced future innovation) is often less, at least in a static sense, than the gain associated with selling these products.

An argument often made by developing countries is that there is "excessive" protection of IPR in industrial countries. The optimal amount of protection of intellectual property is difficult to determine within a given country. This issue becomes more complicated in an international context, because what is optimal from a domestic perspective may not be optimal from an international standpoint. For example, even from a long-run perspective, the optimal amount of IPR protection can differ across innovating and non-innovating countries.[25]

---

[24]This terminology comes from Maskus (1990). This analysis can also be used within a country, where some regions have industries with a lot of innovation (such as the Silicon Valley in California), and others have very few innovating industries. For an example of a model that looks at innovation and technology transfer both within and across countries, see Butler (1990).

[25]Chin and Grossman (1988), for example, find that the desired amount of protection between innovating and non-innovating countries depends on the specification of social welfare used and that strong IPR may not improve global efficiency. Diwan and Rodrik (1989), in a different theoretical framework, also find that the optimal amount of protection between the innovating and non-innovating countries coincide only if welfare in the two regions is weighted equally.

Many poorer developing countries find it difficult to pay the higher price innovated goods would cost if they are given patent protection in their country. Developing countries therefore have little incentive, at least in the short run, to protect commodities they need (or want) but could not afford to buy if protection is awarded. An example of this is seen in the pharmaceutical sector, which is awarded patent protection in essentially all industrial countries, but not in many poorer developing countries, which might otherwise have difficulty purchasing medical supplies.

Another argument for not protecting IPR in technology-importing countries is that reverse engineering, the process by which firms take products apart to learn how to produce them, enables firms to learn how to develop new products themselves and therefore aids in a country's development. As the technological know-how improves in a country, these firms begin to innovate themselves. When this stage is reached, protection of IPR generally begins to increase.[26]

Given the current level of protection of IPR, creating increased standards of protection, at least in the short run and for the least-developed countries probably in the long run, will redistribute income from technology-importing, developing countries (whose residents will now have to pay more for these types of products) to countries which innovate or already protect IPR. As a result, success in negotiating increased international protection for IPR will likely require some concessions to the developing countries in other areas of trade.[27] These issues are currently being discussed by WIPO and GATT.

CONCLUSION

The incentive to protect intellectual property rights, as well as the actual amount of protection awarded, differs widely across industrial and developing countries. Nevertheless, recent negotiations under WIPO and GATT (which, as of this writing, is yet to be concluded) regarding the trade-related aspects of intellectual property rights indicate increased support for international agreements on intellectual property rights and a realization that such agreements benefit both industrial and developing countries. In addition, as the amount of inventive activity and the number of countries engaged in innovation increases, the trend toward more cooperation and protection of intellectual property is likely to continue.

REFERENCES

Benko, Robert P. "Intellectual Property Rights and the Uruguay Round," *World Economy* (June 1988), pp. 217-31.

_____. *Protecting Intellectual Property Rights: Issues and Controversies* (American Enterprise Institute for Public Policy Research, 1987).

"Bugs that Divide." *Economist*, July 28, 1990, pp. 57-58.

Butler, Alison. "Endogenous Innovation in a North-North Model of the Product Cycle," Federal Reserve Bank of St. Louis Working Paper 90-007 (1990).

Chin, Judith C., and Gene Grossman. "Intellectual Property Rights and North-South Trade," *Princeton University Discussion Papers in Economics*, No. 143 (November 1988).

Congress and the Nation. Vol. 6 (Congressional Quarterly Inc., 1985), p. 548.

Diwan, Ishac, and Dani Rodrik. "Patents, Appropriate Technology, and North-South Trade," National Bureau of Economic Research Working Paper No. 2974 (May 1989).

Feinberg, Robert M. "Intellectual Property, Injury, and International Trade," *Journal of World Trade* (April 1988), pp. 45-56.

Feinberg, Robert M., and Donald J. Rousslang. "The Economic Effects of Intellectual Property Right Infringements," *Journal of Business* (January 1990), pp. 79-90.

Finger, Michael J., and Andrzej Olechowski, eds. *The Uruguay Round: A Handbook on the Multilateral Trade Negotiations* (The World Bank, 1987).

General Agreement on Tariffs and Trade. *GATT: What It Is, What It Does* (1990).

Grossman, Gene M., and Elhanan Helpman. "Product Development and International Trade," *Journal of Political Economy* (December 1989), pp. 1261-83.

Hood, Neil, and Stephen Young. *The Economics of Multinational Enterprise* (London: Longman Group United, 1979).

[26]This has recently occurred in Korea, which, according to a 1987 study by the International Trade Commission, was the third-largest source of U.S. losses resulting from violations of IPR. Recently, however, Korea has taken steps to improve its enforcement of IPR, and the number of patents issued in the United States to Korean nationals has risen substantially, although it is still small in absolute value (U.S. Department of Commerce, Patent and Trademark Office, 1989). Of course, U.S. pressure has provided additional impetus for Korea to increase IPR protection.

[27]For a review of some of the problems in negotiating IPR in the Uruguay Round, see Maskus (1990), Benko (1988) and Finger and Olechowski (1987).

Kamien, Morton I., and Nancy L. Schwartz. *Market Structure and Innovation* (London: Cambridge University Press, 1982).

Mansfield, Edwin. "Patents and Innovation: An Empirical Study," *Management Science* (February 1986), pp. 173-81.

Maskus, Keith. "Normative Concerns in the International Protection of Intellectual Property Rights," paper prepared for meetings of the International Seminar in International Trade at the National Bureau of Economic Research (August 2-3, 1990).

_____ . "Intellectual Property," in Jeffrey J. Schott, ed., *Completing the Uruguay Round: A Results-Oriented Approach to the GATT Trade Negotiations* (Institute for International Economics, 1990), pp. 164-79.

Meessen, Karl M. "Intellectual Property Rights in International Trade," *Journal of World Trade Law* (February 1987), pp. 67-74.

Nordhaus, William D. *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (MIT Press, 1969).

"Patents for Pharmaceuticals." *Financial Times,* May 15, 1990

Plock, Ernest. "International Piracy of Motion Pictures," in Ann Main, "Pursuing U.S. Goals Bilaterally: Intellectual Property and 'Special 301'," *Business America* (September 25, 1989), p. 7.

U.S. Department of Commerce, Patent & Trademark Office. *Commissioner of Patents and Trademarks* Annual Report (January 1989).

United States International Trade Commission. "Foreign Protection of Intellectual Property Rights and the Effect on U.S. Industry and Trade," Report to United States Trade Representative, Investigation No. 332-245, Publication 2065 (February 1988).

World Intellectual Property Organization. *General Information* pamphlet (Geneva, January 1990a).

_____ . *WIPO Newsletter* (February 1990b).

_____ . *Existence, Scope and Form of Generally Internationally Accepted and Applied Standards/Norms for the Protection of Intellectual Property* (Geneva, September 1988).

Yuetter, Clayton. News Release. "Statement by Ambassador Clayton Yuetter Concerning Intellectual Property Rights Protection," *Office of United States Trade Representative* (April 7, 1986).

Michelle R. Garfinkel

Michelle R. Garfinkel is a senior economist at the Federal Reserve Bank of St. Louis. Scott Leitz provided research assistance.

The Economic Consequences of Reducing Military Spending

BEFORE THE CRISIS in the Middle East this summer, the easing of international tensions had reduced, for many, the urgency for the United States to continue building or even maintain its military strength. As support for allocating the nation's resources to defense weakened, people began to argue about the potential for a significant "peace dividend" available to the U.S. economy.[1]

How should the savings from reduced defense spending be put to use to enhance our nation's welfare? Many analysts, concerned primarily about the effects of large public deficits, have argued that the savings should be applied to reduce the government's need to borrow.[2] Others have voiced concern that a reduction in defense expenditures will generate unemployment, at least temporarily, while resources are reallocated to productive activities in the civilian sector.[3] Consequently, they have argued the initial savings should be used to ease this adjustment—perhaps by increasing expenditures on training programs.[4] Many other policy recommendations have been made.

Although the current situation in the Middle East raises doubt that there will be any significant dividend in the near term, it does not detract from the relevance of such recommendations. Instead, it provides us with more time to evaluate the various options associated with future defense cuts.

In reviewing the economic implications of reduced military spending, this article examines some issues that have been overlooked by those in search of ways to use the "peace dividend." The article begins with a brief analysis of recent trends and the prospects for future cuts in military spending to see how large a dividend might be. Some simple economic principles are then employed to assess how the peace dividend, regardless of its actual magnitude, might be used to achieve diverse economic goals.

---

[1]For example, see Pennar and Mandel (1989).

[2]See, for example, Schultze (1990).

[3]For example, see "Peace Dividend or Recession?" (1990). Also see Pennar and Mandel (1989), who report the results of a study of the short- and long-term effects of reducing the defense budget by 5 percent a year in real terms from 1991 to 1994 and keeping it constant thereafter. Although this study predicts enhanced economic growth in the long run, it also predicts some short-term losses. Also see Ellis and Schine (1990), who report the results of a study indicating that as many as 1 million defense-related jobs (or 20 percent of all jobs in defense-related activities) could be lost by 1995. Although other analysts have argued that the employment losses could be insignificant (see, for example, Uchitelle (1990)), some defense contractors have already begun to cut production and employment.

[4]See, for example, the bill proposed recently by Senator Pell (S.2097).

Table 1

Trends in U.S. Defense Spending in the Past 50 Years

Decade	Total defense spending[1] Nominal	Real	Share of GNP
1940s	$ 35.8	$285.5	18.3%
1950s	40.2	187.0	10.1
1960s	59.2	216.4	8.4
1970s	90.3	200.0	5.8
1980s	238.3	274.4	6.1

[1]Billions of dollars averaged over the decade. The real figures are in 1989 dollars.

HOW BIG MIGHT THE DIVIDEND BE?

To place the discussion of the economic implications of reduced defense needs into perspective, it is helpful to examine recent patterns of military spending. Table 1 shows three very different perspectives of U.S. military spending over the past 50 years. We can see that nominal military spending in the United States has grown considerably since the 1940s, rising from an average of $35.8 billion per year during the decade that included World War II to an average of nearly $240 billion per year during the 1980s. After adjusting for inflation, however, we see a somewhat different picture. Real military spending declined sharply immediately after the WWII decade; and, its pattern since the 1950s has been more erratic, with its net rise by the 1980s being considerably less dramatic than suggested by the nominal figures. Finally, as the table shows, military spending as a fraction of gross national product (GNP) has fallen markedly over the past five decades. Its share in the 1980s was only one-third of that in the 1940s.

While military spending might now represent a fairly small proportion of GNP, it is still an important component of economic activity. Defense spending was $301.1 billion in 1989—more than $1200 per person in the United States.[5] This number suggests that there could be a considerable dividend from a *large-scale* disarmament.

Even without the Middle East crisis, however, the military spending cuts likely to have occurred in the near future would have been quite small relative to the whole economy. In a recent study, for example, the U.S. Congressional Budget Office (1990) estimated the effects of a proposal by the North Atlantic Treaty Organization (NATO) for limiting conventional forces in Europe.[6] It found that the treaty would generate an *annual* savings of about $3 billion in 1990 prices. Although $3 billion seems large, it constitutes less than 1 percent of the total Defense Department's budget authority for fiscal year 1990 and less than .06 percent of 1989 GNP. This amounts to less than $13 per year per U.S. citizen.

This estimated savings from the prospective reduction in military spending pales in comparison to earlier U.S. disarmament efforts following wartime periods.[7] After World War II, for example, defense spending fell by about $57.3 billion from 1945 to 1946, almost 27 percent of GNP in 1945. From 1953 to 1954, after the Korean War, defense spending fell by $7.4 billion, almost 2 percent of GNP in 1953.

Forecasts of the actual size of future defense cuts, of course, are subject to much uncertainty. The budget proposed by President Bush in January 1990 for fiscal year 1991 called for reducing the defense budget by 2 percent after

[5]During the Reagan Administration's military build up, defense spending grew at a 5.0 percent annual rate after adjusting for inflation; in comparison, the annual growth rate in real GNP over this same period was only 3 percent. Earlier this century, before WWI, military spending's share of GNP was less than 1.5 percent. (Because data on military spending before the 1940s are not entirely consistent with the data presented in table 1, they are not shown here.)

[6]NATO's proposal then called for a reduction of its own and the Warsaw Pact's ground capabilities in Europe and for a reduction in their tactical aircraft capabilities that would leave an advantage for NATO. (Ground capability is measured by army units intended to fight on ground. Tactical aircraft capabilities are measured by fighters and bombers intended to fight in the air and on the ground using conventional weapons.) It called for a 27 percent reduction in a selection of NATO weapons. Assuming an equally proportionate reduction by all NATO members, this requirement implies that the U.S. would have had to remove and destroy 600 tanks, 122 armored personnel carriers, 112 pieces of artillery, 189 helicopters and 105 aircrafts from Europe. Furthermore, about 30,000 U.S. troops would have had to be withdrawn from Europe and demobilized.

[7]As noted below, however, the economic implications of a given reduction in military spending depend quantitatively on whether the reduction follows a war or occurs during a relatively peaceful period.

adjusting for inflation over the fiscal years 1991 to 1995.[8] Others argued for defense spending cuts as much as 5 percent in real terms per year from fiscal years 1992 to 1994, achieving an annual savings of about $60 billion (in 1989 prices) starting in 1994.[9] If cuts of this magnitude were implemented, the implied savings would constitute about 1.2 percent of GNP for the year 1989, nearly $250 per year on a per capita basis.

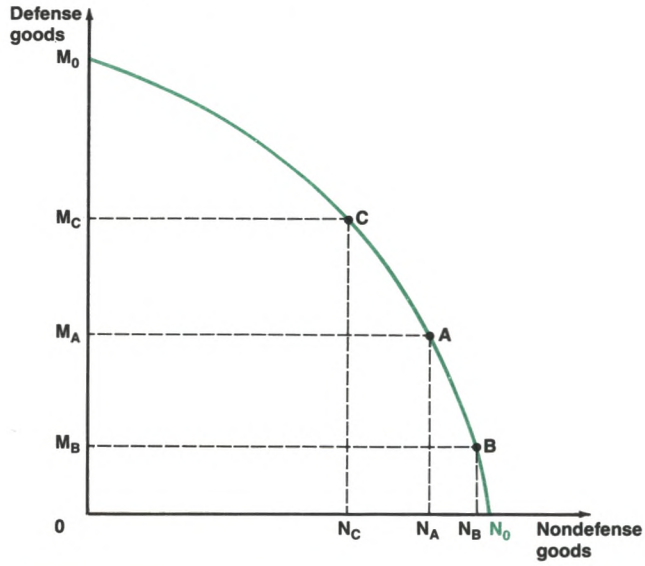
And the savings could be even larger. Indeed, on the day of the Iraqi invasion of Kuwait, President Bush announced that, although the invasion indicates a need to maintain a strong military force, U.S. armed forces could be reduced by 25 percent over five years given the recent changes in Soviet-U.S. relations.[10]

EFFECTS OF THE DIVIDEND: SOME BASIC ECONOMIC PRINCIPLES

Generally speaking, the trade-off between competing uses of resources implies that a reduction in real military spending would provide a "dividend" in the form of increased real private and public consumption and investment opportunities—that is, increased resources available for the production of consumption and investment goods. To quantify these increased opportunities over time, the annual peace dividend is defined here simply as the annual reduction in real military spending.[11]

Figure 1 illustrates this trade-off given the nation's resource and technology constraints by means of a hypothetical production possibility curve (PPC) for defense goods (national security) and nondefense goods (public and private investment and consumption). This curve depicts the maximum quantities of defense goods (M) and nondefense goods (N) that can be produced simultaneously for given amounts of capital and labor inputs.

Figure 1
PPC: Defense and Nondefense Goods



Assuming that resources are fully utilized, the economy is always operating on the frontier regardless of the level of military spending. If no resources were allocated to the production of defense goods, for example, the total output of nondefense goods would be shown as N_0. Producing defense goods thus requires sacrificing the production of some goods for investment and consumption. The opportunity cost of providing a specific level of national security, for example, M_A, is the value of the lost production of nondefense goods, $N_A - N_0$. Conversely, if the economy were originally operating at point A, cutting military spending out entirely would imply increased production of nondefense goods, $N_A - N_0$. Thus, with a reduction in military spending (say $M_B - M_A$), the annual dividend would imply increased opportunities for the production of nondefense goods ($N_B - N_A$). The shaded insert on pages 50 and 51 contains a discussion of the welfare implications of increased opportunities for investment and consumption afforded by a reduction in military spending.

[8] "Peace Dividend or Peace Recession?" (1990).

[9] "The Peace Economy" (1989).

[10] Dowd (1990). Specifically, the Pentagon plan then called for cutting the armed forces by 500,000 troops from the current level of 2.1 million. But, without a clear resolution of the ongoing conflict in the Middle East, any reduction in military spending might seem optimistic.

[11] This definition envisions the dividend as a flow—i.e., as the term is normally understood. Thus, a permanent cut of X dollars in real terms per year implies an X dollar dividend each year indefinitely into the future. Using a present discounted value concept, these flows over time can be translated into a stock concept: X/r, where r is the constant real interest rate.

Will a Dividend Necessarily Enhance "Social Welfare?"

Some interesting social welfare implications of lower military spending can be illustrated within the simple PPC framework. Figure 2 depicts the same PPC shown in figure 1; added to the figure are two indifference curves that show specific combinations of defense and nondefense goods which yield constant levels of national "welfare" or "utility." The curve labeled u indicates a higher level of utility than the one labeled u′. The shape of the indifference curves reflects the notion of diminishing social marginal utility. For example, the nation is "less willing" (must be given considerably more nondefense goods) to decrease consumption of defense goods and consume more nondefense good at point Y compared with point Z.

At the point of tangency (labeled X) between the indifference curve associated with utility u and the PPC, the nation's utility is maximized, given the available resource and technology constraints. At this optimum level of consumption, the marginal utility trade-off between defense and nondefense goods (the slope of the indifference curve, u) equals the marginal rate of transformation between them (the slope of the PPC). The optimal quantities of defense and nondefense goods are M_x and N_x, respectively. If preferences remain unchanged, a move to any other production combination on the PPC (or inside) would result in a lower level of national welfare. Thus, although a move from point X to point Y, for example, produces a positive dividend in the form of increased resources available to produce nondefense goods, it would actually reduce the nation's welfare.

There are two scenarios, however, under which a decline in military spending could enhance welfare by creating increased consumption and investment opportunities. First, the original quantities of M and N could have been suboptimal. If, for example, the economy were operating at a point to the left of X on the PPC (for example, point Z), military spending would be considered too high from

Figure 2
Optimal Levels of Defense and Nondefense Goods

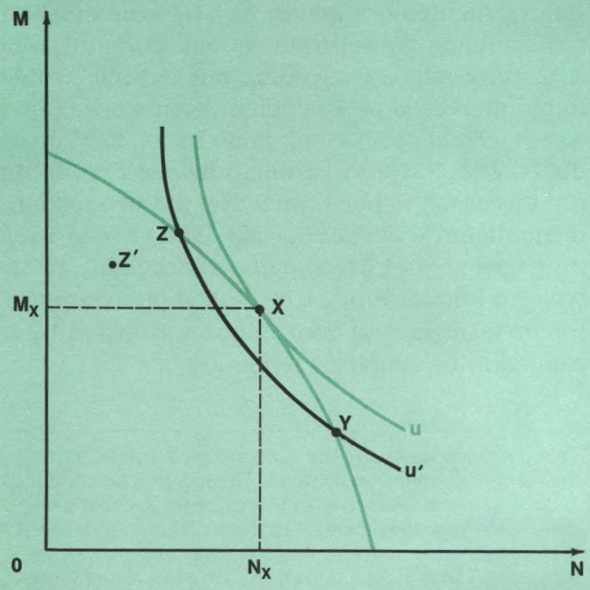
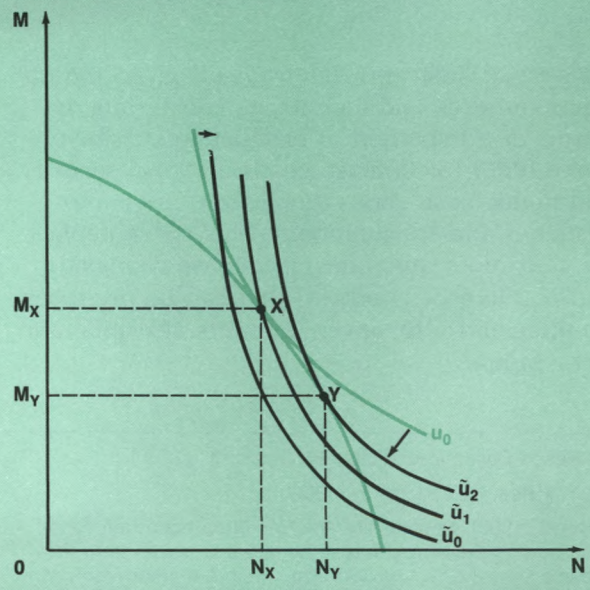


Figure 3
A Preference-Induced Change in the Optimal Levels of Defense and Nondefense Goods



a social welfare point of view. Alternatively, the original point could have been inside the PPC above point X, such as Z′. In this case, military spending is also too high and, more important, the economy is operating below its potential output levels.

Second, a cut in military spending could generate welfare gains if it were associated with an unanticipated decline in the severity of external threats. This seems to be the example most relevant to the current situation involving U.S.-Soviet relations. The reduced value of maintaining existing U.S. military strength in the face of detente can be captured best by presuming there has been a general clockwise movement in the indifference curves. This movement is illustrated in figure 3, where the new indifference curves are labeled $\tilde{u}$. If national security can be maintained with less military spending, a smaller quantity of M with the same quantity of N (N_X) produces the same utility level $\tilde{u}_0 = u_0$. Further, the marginal value of increasing M relative to the cost of foregone consumption of N falls. The new welfare maximizing combination of M and N along the PPC is denoted by Y, where N increases to N_Y and M decreases to M_Y. The net "welfare effect" of the dividend, $M_Y - M_X$, would be reflected in the increase from $\tilde{u}_1$ to $\tilde{u}_2$.

The concave shape of the PPC reflects diminishing marginal returns in productive transformation. That is, the amount of nondefense goods that must be sacrificed to produce one more defense good increases as M increases. For example, in the figure, the move from C to A and the move from A to B involve identical reductions in military spending. Starting at the point with a higher level of military spending (M_C), however, that reduction in military spending implies greater additional production of nondefense goods than when starting at point B. (That is, $N_A - N_C > N_B - N_A$.) Hence, assuming resources are fully employed, a given reduction in the production of defense goods when the current level is low—for example, during peacetime—would imply smaller additions to consumption and investment opportunities than when the current level is high—for example, during wartime.

THE CROWDING-OUT EFFECT OF MILITARY SPENDING AND THE LONGER-RUN EFFECT OF THE PEACE DIVIDEND

These allocative effects of the peace dividend can also have implications for the amount of resources available for production in general. Specifically, any additional investment adds to the future resource base, thereby enhancing future output growth and investment and consumption opportunities. Hence, even if the decline in military spending were temporary, for example lasting only one year, its benefits in terms of increased productive capacity could be realized over many years. This longer-term effect can be modeled in the framework presented above by an outward shift of the PPC curve over time.[12]

Some analysts believe that, in recent decades, military spending has been excessive, reducing the residual supply of productive resources (that is, capital and labor) available for private and nondefense public investment and thereby weakening the economy.[13] According to this "depletion" theory, the effects of higher levels of military spending are reflected in lower rates of investment and, consequently, lower rates of economic growth. Thus, the principal result of reduced military spending would be greater investment and economic growth.

Evidence on the "crowding-out" effect of military spending is based primarily on empirical analyses relating changes in military spending's share of GNP to the GNP shares of other broad categories of expenditures; typically, the studies focus on the effect on private investment's

[12]Of course, one might argue that the level of military spending could positively influence the position of the PPC. Because military spending enhances a nation's ability to protect its resources, it might serve to increase the nation's future resource base by encouraging more investment.

[13]See, for example, Dumas (1987), Melman (1988) and Du Boff (1989). In contrast, Weidenbaum (1990) argues that, as defense spending has fallen relative to GNP, the effects of such spending on the U.S. economy have become less significant.

Table 2

Broad Categories of Real Expenditures as a Share of Real GNP[1]

Decade	C	I	X	Gnm	Gm
1940s	54.7%	12.1%	0.5%	10.1%	23.0%
1950s	58.4	16.5	0.1	11.8	13.2
1960s	59.5	16.7	−0.4	14.3	10.3
1970s	62.2	17.2	−0.7	14.9	6.2
1980s	64.5	16.9	−1.2	13.6	6.2

[1]Each category is converted into real terms using its implicit price deflator. Separate price deflators were used for federal government spending and state and local government spending, but both defense and non-defense federal expenditures were deflated by the same number.

share of GNP.[14] Table 2 shows the trends in personal consumption expenditures (C), private domestic investment (I), nondefense public expenditures (G_{nm}) on goods and services, defense expenditures (G_m) and exports net of imports (X) as shares of GNP over the past 50 years. To focus on the *real* effects of military spending, each broad category of spending is converted into real terms by dividing it by its own price deflator and by real GNP.[15]

As the table suggests, real military spending has crowded out *all* categories of real expenditures, not just real private domestic investment. As real military spending's share fell from the 1940s to the 1980s by 16.8 percentage points, real consumption's share rose by nearly 10 percentage points. Real private investment's share and real nondefense public spending's share also rose over this period, though less dramatically.[16] Of course, in a broad sense, the substitution observed between nondefense and defense public expenditures is consistent with the crowding-out notion; in this case, public investment on the nation's infrastructure—that is, highways, airports, mass transit, water systems—was crowded out.[17]

To be sure, the size of the peace dividend, as defined here, is independent of its allocation among the production of private and public consumption and investment goods. But its longer-term implications depend on that allocation. The best or "socially optimal" reallocation of resources among investment and current consumption depends on the nation's willingness to forego current consumption in order to invest and thereby enhance future consumption possibilities. The greater this willingness, the more likely the resources from a reduction in military spending will be devoted to additional investment rather than additional current consumption. The lower the nation's willingness to forego current consumption to enhance future consumption, the lower will be the proportion

[14]See Gold (1990) for an extensive survey of this evidence. Also see Adams and Gold (1987) and U.S. Congressional Budget Office (1983). It should be noted that identifying the degree to which military spending has resulted in lower expenditures on other goods and services and lower economic growth is difficult. The problem lies in determining how investment and other expenditures would have behaved if military spending had been different—in an extreme case, if it had been zero. Because reduced-form parameters relating defense expenditures to other expenditures would not be independent of the policy regime, estimates of these parameters might provide little information on how a permanent change in military spending (i.e., a policy regime change) would influence other expenditures. In addition, it is important to note that, if higher military expenditures result in higher levels of GNP, lower shares of investment, for example, need not imply a crowding-out effect of military spending.

[15]A comparison of tables 1 and 2 reveals that real military spending as a fraction of real GNP was higher than nominal military spending as a fraction of nominal GNP from the 1940s to the 1970s, but slightly lower during the 1980s. This divergence reflects the difference between the general price level of defense goods and that of all goods and services. The price level for defense goods, on average, was lower than the general price level between the 1940s and the 1970s, but higher during the 1980s. The focus on real military spending here is intended to emphasize the importance of its real allocative effects. Failure to account for relative price movements masks these effects.

[16]Smith (1977) finds that the crowding-out effect of military's share of income on investment's share of income is nearly one to one for 14 OECD nations during the 1960s. However, Boulding (1973), Edelstein (1989) and Aschauer (1989a) argue that this effect is not empirically relevant for the United States. Also see Browne (1989), who questions the validity of the argument that military research and development has crowded out civilian spending on research and development by "depleting" our nation's scientists and engineers. As is well-known, military R&D has produced important innovations that have been applied successfully to production activities in the civilian sector; one commonly cited example is the computer. In addition, Browne argues that this crowding-out effect on R&D presumes that the supply of scientists and engineers is fixed. A greater demand for highly skilled labor, however, has influenced its supply, although with the usual lag.

[17]As an annual average of GNP, net public infrastructure investment fell from approximately 2.3 percent in 1960-65 to about 0.8 percent in 1980-85 (Du Boff (1989), p. 7, table 2). Aschauer (1989b) argues that the recent reduction in public capital, including infrastructure, might be responsible for the recent decline in productivity.

of the savings that is allocated to investment. Thus, in contrast to the suggestion of the depletion theory briefly described above, the dividend from reduced military spending need not result in significantly greater rates of investment and economic growth. The extent to which the dividend will affect economic growth depends on how it is allocated among the production of investment and consumption goods.

ALTERNATIVE USES OF THE DIVIDEND

In recent decades, decreases in nominal military spending typically have been associated with increases in other public expenditures in nominal terms. Indeed, after falling to 19.5 percent in the 1950s from 25.1 percent during the 1940s, total public spending (federal plus state and local) on goods and services has remained roughly constant as a fraction of nominal GNP, around 20 percent; only the composition of those expenditures changed. Although there could be reasons why this pattern might persist in the upcoming decades, many analysts have questioned whether a continuation of this pattern is either likely or even desirable. Nevertheless, the basic question to be addressed should be couched in real terms: What should be done with the peace dividend?

Increase Nondefense Public Expenditures

Some analysts have argued that leaving nondefense public expenditures alone and using the reduction in military spending to either decrease the deficit or lower taxes is not the best use of the peace dividend. Instead, many of them believe that at least part of the savings from arms reduction might best be used to increase nonmilitary government spending—specifically, to rebuild the nation's infrastructure.[18] In terms of the economic framework above, this policy would shift out the PPC, increasing future production and consumption opportunities.

Others have argued that, unless the fall in military spending is somehow offset, resource utilization and, hence, economic activity will fall as well.[19] In this view, which lies within the standard "Keynesian" paradigm, the government should use part (or all) of the savings to finance additional public expenditures, including noninfrastructure expenditures, such as welfare programs, to offset the negative effect of reduced military spending on aggregate demand.

This argument assumes that military spending in particular or public expenditures in general enhance social welfare not only by providing additional public goods, but by increasing employment and thereby stimulating the economy—that is, by inducing the use of idle resources. It implies that, without the increases in military spending or, more generally, public expenditures during the post-WWII period, the economy would have operated below its potential output capability (that is, inside its PPC).

Although there is evidence that a permanent decrease in military spending can produce a permanent decline in aggregate output,[20] the decline in output could be generated, in part, by a voluntary reduction in the supply of labor. In other words, this evidence does not necessarily imply that a permanent decline in military demand, without an increase in other public spending, would cause these productive resources to become involuntarily idle on a permanent basis.

Reduce Taxes

Some analysts would like us to consider an alternative policy that leaves other government expenditures unchanged and uses the dividend to reduce taxes. By increasing individuals' after-tax income, this policy might induce individuals to decrease their supply of labor, which would in turn decrease output without leaving labor resources involuntarily idle.[21]

---

[18]For example, see Du Boff (1989) and Melman (1988).

[19]For example, see Bolton (1966), especially pp. 37-41.

[20]See Barro (1981). In studying the output effects of government expenditures in the United States, he distinguishes permanent from temporary components of military spending. He finds that the effect of increases in temporary military spending (essentially wartime expenditures) on output was nearly one-for-one; increases in permanent military spending also increased output, but by less than the change in military spending.

[21]See Barro (1981) for a theoretic discussion of the effects of government expenditures on output. In support of this line of reasoning, Dunne and Smith (1990) find that for the United States, military spending does not "cause" unemployment. Riddell (1988) argues, in a more Marxian spirit, that the government's apparent bias for military over non-military expenditures is driven by its objective to maintain international order so as to maximize the profitability of U.S. capital. This possible endogeneity of military spending calls many empirical analyses that treat military spending as exogenous into question. Also see Garfinkel (1990a), who uses a game-theoretic model to show how military spending can be driven by aggregate economic activity through the government's motive to prevent other nations from extracting its citizens' resources.

It should be noted that a permanent decline in measured output, triggered by the impact of reduced military spending (and reduced taxes) on leisure, does not necessarily reflect a deterioration in social welfare. Because leisure has value, welfare could increase even if consumption did not. Further, the theory described above suggests that, although individuals would work less, they might actually consume more nondefense private goods because their disposable income has increased (their tax liabilities have declined). On net, their welfare would have increased as long as this new outcome were chosen voluntarily.[22]

Reduce the Public Deficit

Some analysts, who view large public deficits as harmful to the economy, have argued that the government should use the peace savings to reduce the public deficit.[23] In particular, the large deficits (public dissavings) of the past decade are thought to have caused a decline in total national savings—that is, the sum of private and public savings. Since a decline in total savings decreases the residual supply of credit available to private borrowers, large public deficits are considered by many to have pushed up expected real interest rates (interest rates adjusted for expected inflation).[24] Thus, using the dividend to reduce the public deficit would decrease expected real interest rates and thereby stimulate both investment activity and the production of goods, such as exports and new homes, whose sales are sensitive to movements in interest rates.

Although the U.S. savings rate appears to have declined in recent years, how much of this decline can be blamed on large public deficits is unclear.[25] The argument that public deficits influence the national savings rate is based on a number of potentially questionable assumptions. One is that individuals do not view tax cuts that increase public borrowing (holding the level of government spending constant) as increasing their future tax liabilities. Instead, individuals feel wealthier and increase their consumption in response to such tax cuts. Although they also might respond by increasing their savings, the increase in private savings is assumed to be insufficient to keep total savings from falling. In this view, for a given level of government expenditures, public deficits decrease total savings and increase aggregate demand.[26]

Does the Timing of Taxes Matter?

Other analysts argue that individuals believe reductions in current taxes associated with additions to public debt must be financed eventually with additional future taxes. In light of the increase in their future tax liabilities, individuals increase their savings. Conversely, they respond to a decrease in the public deficit, for a given level of government expenditures, by decreasing their savings. In either case, consumption is unaffected.

In this view, often referred to as the "Ricardian" view, individuals behave as if a decrease in the public deficit results in an equal decrease in the present discounted value of their future tax liabilities. This argument builds on the assumption that public debt will be retired eventually out of future taxes. If this view is valid, private and public savings for a given level of government expenditures should be perfectly negatively correlated, while total savings should be unrelated to public savings.[27]

---

[22]As discussed below, however, labor and capital resources could be left involuntarily idle temporarily as the economy adjusts to the reduced military demand.

[23]See, for example, Schultze (1990). Also, see Chrystal and Thornton (1988) for a related discussion of the effects of deficit spending.

[24]Indeed, this effect on interest rates is thought to be the mechanism through which military spending has crowded-out investment. By enhancing the productivity of private capital, however, military spending could have had a "crowding-in" effect that would have offset its crowding-out effect. But Aschauer (1989a,b) presents evidence that does not support the notion that additions to the stock of military capital add to the productivity of private capital. Moreover, the references cited in footnotes 14 and 16 provide evidence that military spending does not crowd out private investment.

[25]Schultze (1990) estimates that national savings as a percentage of national income (i.e., net national product) has fallen from an average of 8 percent during the three decades before 1980 to 3.3 percent during the first three quarters of 1989. Some analysts, however, question the notion that savings is too low in the United States; their skepticism is based on problems with the conventional measurements of savings. Cullison (1990) provides a useful survey of this literature.

[26]See Thornton (1990) for a theoretical discussion of the link between total national savings and public deficits.

[27]Against this Ricardian view, one might argue that deficits could be financed either through increased seigniorage or income taxes in future generations. In either case, the "burden" of the current deficit could be shifted and budget deficits, holding government expenditures fixed, could affect economic activity. See Barro (1989) for a brief discussion of the empirical evidence on the effects of budget deficits. While recognizing the problems associated with testing the Ricardian proposition, Barro argues that the existing evidence lends more support to the Ricardian view than to the alternative view (p. 52).

To be sure, using part of the annual dividend to reduce the public deficit could increase total investment and total consumption. According to the Ricardian view, however, the amounts of these increases do not depend on whether taxes are cut or the deficit is reduced. A cut in the deficit reduces future tax liabilities, but the timing of the tax cuts does not matter.

Because of the distorting nature of the income tax system, however, the equivalence between taxes and debt creation implied by the Ricardian view would be, at best, a rough approximation. Economic theory predicts that proportional income taxation distorts individuals' decisions about consumption and labor supply. These distortions are costly and, other things being equal, the severity of the distortion increases as the tax rate increases. Consequently, if the federal government wants to minimize the costs associated with these distortions, given the path of future government expenditures, it should smooth income taxes over time.[28] This modified version of the Ricardian view suggests that tax reductions, rather than deficit reductions, would be a preferable use for the peace dividend.[29]

THE TRANSITIONAL COSTS OF ECONOMIC ADJUSTMENT

In thinking about how the savings from reduced military spending could be used, it is important to consider how the economy adjusts to unanticipated changes in resource uses. Reduced military spending will produce a negative short-run effect on production, as labor and capital resources are shifted from military to civilian uses. During the transition period, some resources will be unemployed or underemployed.

A Historical Perspective

Previous disarmaments have been associated with sizable reductions in economic activity.[30] From the first quarter of 1945 to the first quarter of 1946 (peak to trough), for example, nominal GNP fell at a seasonally adjusted annualized rate of $22.8 billion or 10.3 percent. These numbers understate the magnitude of the decline in output as, during this period, there was a considerable acceleration in inflation.[31] In real terms, GNP fell 18 percent over this period, and it was not until the third quarter of 1952 that the level of real GNP had fully recovered. Although this decline in real GNP is large, it is an overstatement of the drop in national welfare. While the level of employment fell substantially, the unemployment rate rose very little. Instead, a substantial number of workers, particularly women, voluntarily withdrew from the labor force.

The transition to peacetime after WWII was facilitated, in part, by government policies. Tax reductions and transfer payments (unemployment and veteran benefits) left disposable (net of taxes) income nearly unaffected by the massive reduction in military spending. Thus, demand for consumption goods rose to offset partially the decline in military demand.

The sharp decline in military spending that followed the end of the Korean War was associated with a mild recession. From the second quarter of 1953 to the second quarter of 1954 (peak to trough), real GNP fell 3.2 percent. Declines in defense spending and other federal government expenditures, combined with inventory decumulations, were the driving forces here. By the first quarter of 1955, however, real GNP had climbed well above its previous peak.

Factors Influencing The Adjustment Costs

The aggregate adjustment cost for any reduction in military spending can be measured by the real value of resources (labor and capital) left idle involuntarily during the transition. For a given cut in defense spending, the magnitude

[28]See Barro (1979). Also see Garfinkel (1990b) for an extension of this theory to include conscription as an additional tool for financing public expenditures to avoid the distortions of income taxes, particularly during periods of severe military needs.

[29]In this view, deficits are necessary to smooth out the distortionary effects associated with taxes. Hence, a temporary increase in government spending should be financed with debt and a temporary decrease in government spending should result in a budget surplus; this tax-smoothing view of debt creation predicts that deficits are temporary phenomena. Although historical evidence supports this positive theory of debt creation—see Barro (1979), for example—it is unclear whether the current deficit is only a temporary phenomenon. Indeed, the magnitude and persistence of the peacetime deficits during the 1980s are unprecedented in U.S. history.

[30]See Bolton (1966) and references cited therein for a more detailed examination of these periods of disarmament. Much of the discussion here draws from this work. Data are taken from Balke and Gordon (1986).

[31]This acceleration was driven, in part, by the removal of price controls in 1946.

of these costs depends on the speed with which labor and capital resources can be transformed to meet new demands.

The speed of resource transformation, in turn, depends on the degree of specialization of resources used in the military sector. This specialization has two dimensions. First, certain industries, occupations and firms are highly dependent on military demand. Second, military production is highly concentrated in several regions of the United States. Such specialization will slow the adjustment process.

A given reduction in military demand now might generate relatively greater adjustment costs than those associated with large-scale disarmaments following wartime periods. During wartime periods, resources normally used to produce nondefense goods are mobilized quickly and, presumably, on a temporary basis; after the war, resources are rechanneled easily into their original civilian productive activities. In contrast, during peacetime defense firms and their employees expect that demand for their product is essentially permanent. To the extent that these firms and employees have a comparative advantage in the production of defense goods, they are less likely either to diversify their operations into civilian markets or be able to do so in the event of an unanticipated permanent reduction in military spending.[32]

EVALUATING ALTERNATIVE PROGRAMS

To evaluate society's options for using the savings from reduced military spending, we must address two related economic issues. The first issue, already discussed, concerns what are the best uses of the dividend or the new "highest-valued" uses of the resources previously used in the military (presumably their previous highest-valued uses). The second issue concerns how to rechannel resources efficiently from their military uses to their new highest-valued uses.

Some people have advocated establishing public programs—for example, training programs—to lessen the costs of adjustment borne solely by those closely linked to the military sector. A bill introduced recently by Senator Pell (S.2097), for example, seeks to establish a program through which grants would be made to assist state and local governments in developing economic adjustment plans—for example, job retraining and finding alternative uses for defense facilities. These grants would be funded, in part, from the savings from reduced military spending.[33] Another bill introduced by Senator Coats (S.2682) is intended to aid defense contractors in diversifying their operations into nondefense markets. Through tax incentives, this bill would encourage defense contractors and their employees to adopt employee stock ownership plans (ESOPs) to finance the corporate restructuring necessary to adjust to the reduction in military demand.

These programs are aimed at distributing the adjustment costs and the benefits of reduced military spending equitably; however, they are unlikely to effect an efficient reallocation of resources.[34] Consider, for example, a program that increases public nondefense expenditures on goods that are most easily produced by the capital and labor resources originally employed for the production of defense goods. While this program might well limit the adverse impact of reduced military spending otherwise borne by those firms and individuals highly dependent on military demand, it is clearly inefficient from the nation's point of view. Unless increased spending on these other goods were deemed desirable on a permanent basis, this policy would not provide firms and individuals with the incentives to channel their resources to new higher-valued uses. Instead, it would merely prolong the process of adjustment and delay the realization of the full benefits from a permanent reduction in military spending.

Nevertheless, the redistributive effects of reduced military spending should not be

[32]In contrast, one might believe that defense contractors, having learned from past experience with sharp declines in military demand, would have diversified their operations to exploit commercial opportunities. While such diversification would provide insurance against large losses to these firms and their employees in the event of an unexpected decline in military demand, past efforts in this direction have not been particularly successful. See Weidenbaum (1973) and Ellis and Schine (1990).

[33]Representative Boxer's proposal (HR5327) is similar; however, it would penalize defense contractors who close their firms unless a contract has been canceled. The states of Washington and California already have initiated adjustment plans. See Ellis and Schine (1990).

[34]Although the evidence on the effectiveness of manpower programs (for example, the Manpower Development and Training Act) is mixed, some studies find that manpower policies have been successful in raising the earnings of training program participants. See, for example, Ashenfelter (1978).

dismissed as unimportant or irrelevant in choosing how the peace dividend ultimately will be used. The question of who reaps the gains and who bears the costs of an unanticipated reduction in military spending is an important aspect of the problem and will play an important role in the solution.

CONCLUSIONS

This article has examined some possible effects of a permanent reduction in military spending. In principle, the present discounted value of the implied dividends from such a reduction, in terms of increased consumption and investment opportunities, could be substantial. Through increased private and public investment, reduced military spending implies greater economic growth and, hence, greater consumption and investment opportunities in the future. The important economic questions are, How can these resources be reallocated efficiently to nonmilitary uses? and, How can we identify what these uses should be? This article has introduced and discussed the economic issues that must be addressed in answering these questions; much further analysis and discussion will clearly be needed before these questions can be answered adequately.

Of course, given the recent course of events in the Middle East, the reduction in military spending in the near future might not be large enough to generate any sizable dividend. Thus, debate over whether the savings from reduced military spending should be used to reduce the public deficit, redistributed to taxpayers through tax cuts or be used to rebuild the infrastructure might seem premature. Because many communities, firms and individuals are affected by even small reductions in military spending, however, the "micro" costs of these adjustments will not be ignored in the political decision-making process.

But temporary transitional costs do not justify abandoning the effort to reduce the amount of resources allocated to military spending. Moreover, the Middle East situation is not a permanent obstacle to realize a large dividend in the future. While cuts in defense spending are not expected to be particularly large now, a resolution of the Middle East conflict will permit a much larger dividend in the future.

REFERENCES

Adams, Gordon, and David Gold. *Defense Spending and the Economy: Does the Defense Dollar Make a Difference?* (Washington, D.C.: Defense Budget Project at the Center on Budget and Policy Priorities, July 1987).

Aschauer, David. "Does Public Capital Crowd Out Private Capital?" *Journal of Monetary Economics* (September 1989a), pp. 171-88.

_____. "Is Public Expenditure Productive?" *Journal of Monetary Economics* (March 1989b), pp. 177-200.

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings,"*Review of Economics and Statistics* (February 1978), pp. 47-57.

Balke, Nathan S. and Robert J. Gordon. "Historical Data," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change* (Chicago: The University of Chicago Press, 1986), pp. 781-850.

Barro, Robert J. "The Ricardian Approach to Budget Deficits," *Journal of Economic Perspectives* (Spring 1989), pp. 37-54.

_____. "Output Effects of Government Purchases," *Journal of Political Economy* (December 1981), pp. 1086-121.

_____. "On the Determination of the Public Debt," *Journal of Political Economy* (October 1979), pp. 940-71.

Bolton, Roger E. "Defense Spending: Burden or Prop?" in R.E. Bolton, ed., *Defense and Disarmament* (Prentice-Hall, Inc., 1966), pp. 1-53.

Boulding, Kenneth E. "The Impact of the Defense Industry on the Structure of the American Economy," in Bernard Udis, ed., *The Economic Consequences of Reduced Military Spending* (Toronto: D.C. Heath and Company, 1973), pp. 225-52.

Browne, Lynn E. "Defense Spending and High Technology Development: National and State Issues," *New England Economic Review*, Federal Reserve Bank of Boston (September/October 1988), pp. 3-22.

Chrystal, K. Alec, and Daniel L. Thornton. "The Macroeconomic Effects of Deficit Spending: A Review," this *Review* (November/December 1988), pp. 48-60.

Cullison, William E. "Is Saving Too Low in the United States?" Federal Reserve Bank of Richmond *Economic Review* (May/June 1990), pp. 20-35.

Dowd, Maureen. "Bush Says Military Can Be Cut 25% in 5 Years," *New York Times*, August 3, 1990.

Du Boff, Richard B. "What Military Spending Really Costs," *Challenge* (September/October 1989), pp. 4-10.

Dumas, Lloyd J. "National Security and Economic Delusion," *Challenge* (March/April 1987), pp. 28-33.

Dunne, Paul, and Ron Smith. "Military Expenditure and Unemployment in the OECD," *Defence Economics: The Political Economy of Defence, Disarmament and Peace,* Vol. 1, No. 1 (1990), pp. 57-74.

Edelstein, Michael. "What Price Cold War: Military Expenditures and Private Investment in the U.S., 1890-1980," mimeo (January 1989).

Ellis, James E., and Eric Schine. "Who Pays For Peace?" *Business Week* (July 2, 1990), pp. 64-70.

Garfinkel, Michelle R. "Arming as a Strategic Investment in a Cooperative Equilibrium," *American Economic Review* (March 1990a), pp. 50-68.

_____. "The Role of the Military Draft in Optimal Fiscal Policy," *Southern Economic Journal* (January 1990b), pp. 718-31.

Gold, David. *The Impact of Defense Spending on Investment, Productivity and Economic Growth* (Washington, D.C.: Defense Budget Project at the Center on Budget and Policy Priorities, February 1990).

Melman, Seymour. "Economic Consequences of the Arms Race: The Second-Rate Economy," *American Economic Review* (May 1988), pp. 55-59.

"Peace Dividend or Peace Recession?" *The Economist* (April 7, 1990), pp. 31-32.

Pennar, Karen, and Michael J. Mandel. "The Peace Economy: How Defense Cuts Will Fuel America's Long-Term Prosperity," *Business Week* (December 11, 1989), pp. 50-55.

Riddell, Tom. "U.S. Military Power, the Terms of Trade, and the Profit Rate," *American Economic Review* (May 1988), pp. 60-65.

Schultze, Charles L. "Using the Peace Dividend to Increase Saving," *Challenge* (March/April 1990), pp. 11-17.

Smith, Ron P. "Military Expenditure and Capitalism," *Cambridge Journal of Economics* (March 1977), pp. 61-76.

_____. "Military Expenditure and Capitalism: A Reply," *Cambridge Journal of Economics* (September 1978), pp. 299-304.

Stevenson, Richard W. "Suppliers Brace for Arms Cuts," *New York Times,* June 18, 1990.

Thornton, Daniel L. "Do Government Deficits Matter?" this *Review* (September/October 1990), pp. 25-39.

Uchitelle, Louis. "Economy Expected to Absorb Effects of Military Cuts," *New York Times* (April 15, 1990).

U.S. Congressional Budget Office. *Defense Spending and the Economy* (Government Printing Office, February 1983).

_____. *Budgetary and Military Effects of a Treaty Limiting Conventional Forces in Europe*, Special Study (January 1990).

Weidenbaum, Murray. "Defense Outlays Have Declining Impact on Economy," *Christian Science Monitor,* February 7, 1990.

_____. "Industrial Adjustments to Military Expenditure Shifts and Cutbacks," in Bernard Uris, ed., *The Economic Consequences of Reduced Military Spending* (Toronto: D.C. Heath and Company, 1973), pp. 253-87.

Federal Reserve Bank of St. Louis
Post Office Box 442
St. Louis, Missouri 63166

The* Review *is published six times per year by the Research and Public Information Department of the Federal Reserve Bank of St. Louis. Single-copy subscriptions are available to the public free of charge. Mail requests for subscriptions, back issues, or address changes to: Research and Public Information Department, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, Missouri 63166.