The Euro and the European Central Bank

*Jeffrey M. Wrase**

In order to form a more perfect economic union, establish a single financial market, provide a high level of employment, promote convergence of economic performance, and secure the benefits of sustainable and noninflationary growth, 11 European countries have established a common currency and a European Central Bank. The formal introduction of the new monetary unit, called the euro, occurred on January 1, 1999. On that date, the old national currencies officially became subunits of the euro, much as the nickel and quarter are subunits of the dollar. (See Table.)

Fifteen countries passed a major milestone on the road to monetary union in 1992, when they signed the Treaty on European Union, commonly called the Maastricht Treaty, which outlined a basic structure for the alliance. Of those 15, only 11 actually joined the European Monetary Union (EMU): two opted out for now, and two others have not yet met the economic criteria established

*Jeff Wrase is a senior economist in the Research Department of the Philadelphia Fed.

TABLE

Conversion rates between the national currencies of the 11 member countries and the new euro were irrevocably fixed at midnight, local time, on December 31, 1998. Between January 1, 1999, and June 30, 2002, one euro will be equivalent to the following amounts of each of the 11 currencies:

Austrian shilling 13.7603
Belgian franc 40.3399
Dutch guilder 2.20371
Finnish markka 5.94573
French franc 6.55957
German mark 1.95583
Irish punt 0.787564
Italian lira 1936.27
Luxembourg franc 40.3399
Portuguese escudo 200.482
Spanish peseta 166.386

Thus, a German mark is a bit over half a euro, and a French franc is a bit more than 15 euro cents.

for membership in the union.[1] The EMU countries decided that the benefits of having one common currency instead of 11 different ones will outweigh the costs, especially given the amount of travel and trade that takes place between these countries.[2]

To facilitate adoption of a single currency, the EMU countries also established, after a great deal of preparation, a European Central Bank (ECB) that sets a single monetary policy for the 11 members. (See *Foundations of and the Legal Framework for the Euro*.) The new setup is similar in some respects to that in the United States, in which the states share a common currency (the dollar) and central bank (the Federal Reserve). The ECB took responsibility for monetary policy on January 1, 1999. We'll have more to say about its functions and operations later.

THE TRANSITION PERIOD

In the transition period — January 1, 1999, to December 31, 2001— a consumer can use the euro for *noncash* transactions, but euro notes and coins will not yet circulate. To buy something with euros during the transition period, a consumer can use a credit card or traveler's check, or she can make an electronic funds transfer or write a check. Euro-denominated bank accounts, credit cards, and traveler's checks have been available since January 1, 1999.

During the transition period, the "optional use principle" applies: no one can be forced to use the euro or be prevented from using it. For example, a bank customer with an account de-

[1]The 11 countries are Austria, Belgium, Finland, France, Germany, Ireland, Italy, Luxembourg, the Netherlands, Portugal, and Spain. Denmark and the United Kingdom opted not to join initially. Greece did not meet the criteria for inflation, long-term interest rates, and ratios of government debt and budget deficits to GDP; all were too high. Movements in the foreign exchange value of Sweden's currency were deemed incompatible with the necessary conditions for adoption of the euro, as was the statute of the Swedish central bank. In any case, the Swedish government, citing lack of popular support, decided that Sweden would not introduce the euro at the beginning of monetary union in 1999. See the European Monetary Institute's 1998 convergence report for details of the economic criteria and evaluation of the individual countries.

[2]For a discussion of the economic and political benefits and costs of European monetary union and policy, see the articles by Barry Eichengreen, Gwen Eudey, Martin Feldstein, Maurice Obstfeld, and Ed Stevens. The book by Peter Kenen provides a broad description of European monetary and economic union.

Foundations of and the Legal Framework for the Euro

There have been many steps toward European economic and monetary integration, dating back at least to 1951 when the treaty that established the European Coal and Steel Community was signed. The history of *monetary* integration in particular began with the Werner Report, published in 1971, which set out a blueprint for the stage-by-stage realization of economic and monetary union. In 1979, the European Monetary System was established: bilateral exchange rates among all currencies in the system were to fluctuate only within narrow preset margins.

In 1989, the Delors Report, which had been commissioned by heads of government at the 1988 meeting of the European Council in Hanover, Germany, laid the foundation for the euro.[a] The report insisted that Europe's economic union, monetary union, and the single market were inextricably linked. It advocated a monetary union characterized by the complete liberalization of capital movements, the full integration of financial markets, the irrevocable fixing of exchange rates via a progressive tightening of the European Monetary System, and the completion of the single market for goods and services. The Delors Report also envisaged a fully independent institution to set the union's monetary policy. The report was endorsed by governments at the European Council summit held in Madrid in 1989.

The Treaty on European Union was agreed to in December 1991 and was signed on February 7, 1992, in Maastricht. This treaty entered into force on November 1, 1993, after it was ratified by all member countries. It forms the basis for economic and monetary union. Annexed to the Treaty on European Union is the statute of the European System of Central Banks and of the European Central Bank.

At a summit held in Madrid on December 15-16, 1995, the heads of government reconfirmed that monetary union would begin on January 1, 1999, and agreed on euro as the name of the single currency. At the same time, they adopted firm dates for the transition period (January 1, 1999 to December 31, 2001) and the final period (January 1, 2002 to July 1, 2002 at the latest).

Early in May 1998, in Brussels, leaders of the European Union formally approved the launch of a single currency on January 1, 1999. After consulting the European Parliament, the European Monetary Institute, and the European Commission, the European Council determined which countries had met the convergence criteria and would therefore be founding members of the European Monetary Union.[b] The method for permanently fixing bilateral exchange rates among the 11 member countries was set, and members of the Executive Board of the European Central Bank were recommended (and subsequently accepted by the European Monetary Institute).

---

[a]The Delors Report was produced by a committee of all European Union central bank governors; the then-president of the European Commission, Jacques Delors; and a number of independent experts.

[b]The European Monetary Institute, the precursor of the European Central Bank, was set up in accordance with the terms of the Maastricht Treaty to prepare for establishing the European Central Bank's functions and monetary policy operations.

nominated in German marks may work for a company that has chosen to pay employees in euros. The customer can choose to switch to a euro account, if the bank offers such accounts, but the bank may not convert the account against the customer's wishes.[3]

Also during this period, many merchants in the 11 member countries of the EMU are marking prices in both the national currency and euros. However, merchants are not legally obligated to show two prices. And although using euro-denominated checking or credit card accounts while still using a national currency may be confusing, it's part of the cost of monetary transition.

There are other costs, aside from shopping inconveniences. For example, Europe's three million plus vending machines will need to be reconfigured at an estimated cost of between $100 and $500 per machine, depending on the machine's age. Every ATM machine will also have to be converted to dispense euros rather than a national currency.

Regulations implementing the Maastricht Treaty forbid contracting parties from altering or terminating contracts because of the introduction of the euro.[4] For example, a contract denominated in French francs will remain in force during and after (if applicable) the transition period, and its terms will be unaltered by the euro's introduction, except that payment may be made in euros.

Overall, legislation governing the euro and transactions made with euros provides a framework to ensure acceptability of the new currency. So far, the transition period has proceeded without major difficulties, a situation that reflects the extensive planning that took place well before the EMU member countries entered into the union.

THE FINAL PERIOD

In the final period, which will begin on January 1, 2002, and will end on July 1, 2002, at the latest, national currency notes and coins will be withdrawn from circulation and euro notes and coins will start to circulate. The old notes and coins will continue to be legal tender during the final period, unless an individual member country decides to remove legal-tender status from its currency before July 1, 2002.[5]

Parallel circulation will pose practical problems for consumers in Europe, who will need to keep two separate sets of notes and coins, and for shopkeepers, who will need two tills. Methods are being sought to shorten the parallel-circulation period, perhaps by stocking cash dispensers with euros only; by giving change in euros only, regardless of the unit of payment; or by removing legal-tender status from the old notes and coins very quickly.

An important determinant of the success of the monetary union is the performance of the institution controlling monetary policy, the European System of Central Banks, including the European Central Bank.

[3]There are exceptions to and restrictions on the principle. For example, if a national law stipulates the use of a national currency unit for certain transactions with the public sector, such as tax payments, citizens and enterprises must respect the use of this denomination. Anyone wishing to purchase new debt issued by member countries of the EMU must use euros, because all such debt is denominated in the new currency. In addition, some issues of private bonds are denominated in euros.

[4]This continuity principle does not apply if a contract contains a clause specifically allowing for renegotiation or termination because of the introduction of the euro.

[5]According to the European Commission, EMU member countries are discussing withdrawal of legal-tender status for national currencies earlier than July 1, 2002, because of the difficulties of maintaining dual circulation for a full six months. See the question-and-answer database "Quest" at the commission's website: http://europa.eu.int.

ORGANIZATION OF THE EUROPEAN SYSTEM OF CENTRAL BANKS

The European System of Central Banks consists of the European Central Bank (ECB), headquartered in Frankfurt, Germany; the 11 national central banks of the EMU member countries; and the four national central banks of the European Union countries not currently EMU members.[6] Except for the inclusion of central banks from outside the monetary union, the structure of the European System of Central Banks looks similar to the setup of the Federal Reserve System. The Eurosystem, composed of the ECB along with the 11 national central banks of countries that have adopted the euro, bears an even closer resemblance to the Board of Governors and 12 regional Federal Reserve Banks. As we'll see, however, there are some important differences.

The most important decision-making body within the ECB is the Governing Council. The Governing Council consists of an Executive Board (six members, including the president and the vice president of the ECB, appointed by common accord of the governments of the 11 EMU countries) and the central bank governors of the 11 EMU countries (appointed by their respective governments).[7] The tasks of formulating and implementing monetary policy are assigned to the Governing Council. In the U.S., these tasks are performed by the Federal Open Market Committee (FOMC), which consists of the Federal Reserve's Board of Governors (seven members, including the chairman and vice chairman, appointed by the President of the United States and confirmed by the Senate) and five Reserve Bank presidents.[8]

The ECB's Governing Council makes key decisions affecting the availability and cost of money and credit in the EMU countries, similar to the task performed by the FOMC. At their respective meetings, the Governing Council and the FOMC make decisions about targets for interest rates and money growth by majority vote. They also vote on the policy to be carried out during the interval between meetings.[9] Each member of the ECB's Governing Council has one vote, so the six-member Executive Board has fewer votes than the governors of the participating countries' central banks. In contrast, on the FOMC, the seven-member Board of Governors has more votes than the five Reserve Bank presidents.[10]

MONETARY POLICY STRATEGY AND IMPLEMENTATION

According to its president, Willem Duisenberg, the ECB will pursue "a stability-

[6]The four countries not in the EMU (Denmark, Sweden, the United Kingdom, and Greece) will be allowed some input into the European System of Central Banks but will not participate in decisions about monetary policy for the 11-country "euro zone."

[7]Another organization in the European System of Central Banks, the General Council, does not have a parallel in the Federal Reserve System. The General Council gives some representation to the four European countries that haven't yet joined the EMU. This council is made up of the president and vice president of the ECB and the governors of *all 15* European national central banks. The General Council's tasks are to provide input concerning monetary and exchange-rate policies for European countries inside and outside the EMU; to collect statistical information; to prepare the ECB's reports and financial statements; and to establish rules for standardizing accounting and reporting of operations undertaken by the national central banks.

[8]The five Reserve Bank presidents are the president of the Federal Reserve Bank of New York and four other Reserve Bank presidents who serve one-year terms on a rotating basis.

[9]The Governing Council of the ECB has agreed to hold its meetings on alternate Thursdays. The FOMC holds eight regularly scheduled meetings per year at intervals of five to eight weeks. Special FOMC meetings or telephone conferences take place if circumstances require discussion or action between regular meetings.

[10]For more details on the institutional structure of the European Central Bank, see Mark Wynne's article, or look at the ECB's website: http://www.ecb.int.

oriented monetary policy." The major goals of the ECB, as set forth in the Maastricht Treaty, are price stability (the primary goal); support of general economic objectives such as high employment; and establishment and maintenance of a stable, credible euro in an open market economy with free competition. These objectives are similar to the goals U.S. law sets for the Federal Reserve: maximum sustainable employment and price stability.

The Governing Council of the ECB has adopted the following definition of price stability: "Price stability shall be defined as a year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below 2%."[11] At the same time, the Maastricht Treaty recognizes that the ECB cannot be held responsible for short-term movements in inflation because there are lags between a change in monetary policy and its effect on prices. In the short term, inflation may also reflect temporary or external shocks over which the ECB has no control.

The Governing Council of the ECB agreed on the main elements of its policy strategy in October 1998. This strategy focuses on the money supply, in particular the growth rate of a broad monetary aggregate labeled M3.[12] Recognizing that too rapid money growth is a primary cause of inflation, the ECB sets a target for average money growth. For 1999, the target is 4.5 percent. On a week-to-week basis, however, the ECB does not attempt to control M3 growth directly. Instead, like the Fed, the ECB conducts monetary policy in the very short run by managing short-term interest rates. By raising or lowering these interest rates, the ECB can indirectly manage money growth.

What will the ECB do if money grows faster or slower than the target? It won't necessarily change short-term interest rates to hit the M3 target; instead, like the Fed, it will try to determine if the difference between actual money growth and the target is due to special factors or is an indication that monetary policy has been too easy or too tight.[13] The results of this analysis and its impact on monetary policy decisions are explained to the public through speeches and published reports.

Before we look at the tools the ECB uses to put monetary policy into effect, let's consider the basics of how the Federal Reserve typically executes monetary policy in the United States.

Federal Reserve's Monetary Policy Actions. The Federal Reserve executes monetary policy in pursuit of its goals mainly through the use of open market operations—the sale or purchase of previously issued U.S. government securities. By purchasing government securities, the Fed increases the supply of reserves in the banking system; by selling them, it reduces the supply. (Reserves consist of cash that banks hold in their vaults along with banks' balances on deposit at the Federal Reserve.) Banks need reserves to settle payments among themselves and to satisfy legal requirements that they hold reserves equal to 10 percent of most balances in checking and other transaction accounts.

[11]See the speech "Monetary Policy in the Euro Area," by ECB President Duisenberg on the Internet at http://www.ecb.int/key/sp990125.htm. The HICP is simply a price index using data constructed in similar fashion by statistical agencies of different member countries.

[12]In the EMU, M3 consists of currency in circulation, overnight deposits, deposits and debt securities with agreed maturity up to two years, deposits redeemable at notice up to three months, repurchase agreements, and money market fund shares. Technically, the ECB focuses on "Harmonised" M3; see footnote 11.

[13]In accordance with the Humphrey-Hawkins Act of 1978, the FOMC sets ranges for annual growth of the U.S. money supply. In recent years, however, rapid financial innovation has made money growth an unreliable indicator of future economic developments in the U.S., so the FOMC has de-emphasized money growth ranges.

Banks that have excess reserves often try to lend them in the federal funds market, generally overnight. Banks that have a shortage of reserves usually try to borrow some in the same market. In the federal funds market, supply and demand interact to determine the quantity of reserves that banks borrow or lend and at what interest rate—the federal funds rate.

The supply side of the federal funds market is influenced by transactions undertaken by the Federal Reserve Bank of New York at the direction of the FOMC—the buying and selling of securities as noted above. The Fed's day-to-day objective is to engineer a supply of reserves that, in conjunction with banks' demand for reserves, achieves a federal funds rate equal or close to a target determined by the FOMC. The target for the federal funds rate depends on the state of the economy relative to the Fed's long-term goals.[14]

ECB's Monetary Policy Actions. The ECB enacts monetary policy through 11 national central banks that buy and sell securities to influence the interbank interest rate. Consequently, the ECB's implementation of monetary policy is slightly more complicated than that of the Federal Reserve System. But its approach is not fundamentally different from the Fed's.

The ECB has three tools for conducting monetary policy: reserve requirements called "minimum reserves," open market operations, and provision of standing facilities.

Minimum reserves. Reserve requirements are applied to a wide range of financial intermediaries in the euro area.[15] Each intermediary's reserve requirement is determined in relation to its balance sheet. Currently, each intermediary must hold reserves in an amount equal to or exceeding 2 percent of its total amounts of these liabilities: overnight deposits, deposits with maturities of up to two years, deposits redeemable at notice of up to two years, debt securities issued with agreed maturities of up to two years, and money market paper.[16] Compliance with the reserve requirement is determined on the basis of an intermediary's average daily reserve holdings over a one-month maintenance period.

The Federal Reserve, by law, also imposes reserve requirements, though only on deposit-taking institutions.[17] Since January 1999, each institution subject to reserve requirements must meet a requirement of 3 percent applied to net transaction accounts totaling between $4.9 and $46.5 million; a 10 percent rate is applied to net transaction accounts above $46.5 million.[18] While these required reserve rates are higher than those imposed by the ECB, they apply to a narrower class of liabilities and a narrower class of financial institutions. We can see the difference if we compare average daily required reserves

[14]Additional details of the Fed's activities in the federal funds market can be found in the book by Ann-Marie Meulendyke.

[15]More specifically, the ECB imposes reserve requirements on *monetary financial institutions,* defined by the ECB and European Community law as resident financial institutions whose business is to receive deposits and close substitutes for deposits, to grant credit, or to make investments in securities. This includes not only banks but also money market mutual funds and some leasing companies. A list of monetary financial institutions subject to reserve requirements is available at the ECB's website at http://www.ecb.int.

[16]A lump-sum allowance of 100,000 euros is deducted from an institution's reserve requirement so that banks with 5 million euros or less of reservable liabilities will not have to hold minimum reserves.

[17]In the United States, only *depository financial institutions* are required to hold reserves. According to the Monetary Control Act of 1980, this term covers commercial banks, mutual savings banks, savings and loan associations, credit unions, agencies and branches of foreign banks, and Edge Act corporations.

[18]Compliance with U.S. reserve requirements is generally determined on the basis of an institution's average daily reserve holdings over a two-week maintenance period.

held in March 1999: more than 100 billion euros (almost $109 billion) were held as required reserves in the EMU and around $42 billion were held in the U.S.[19]

Banks and other financial intermediaries in the EMU have to hold a larger amount of reserves than financial institutions in the United States because the ECB assigns reserve requirements a more prominent role in monetary control. The need to hold more required reserves increases the demand to hold reserve deposits at the central bank. By ensuring a large demand for reserve deposits, the ECB can more easily control short-term interest rates by managing the supply of reserves.

In contrast to the Federal Reserve, which cannot legally pay interest on required reserves, the ECB pays interest on minimum reserve holdings. The interest rate reflects short-term money market interest rates prevailing over the reserve maintenance period. Paying interest on required reserves helps to make up for the income banks could otherwise earn by lending the reserves. Without payment of interest on the large quantity of reserves required by the ECB, banks in the euro area might suffer a competitive disadvantage.

Open market operations. Open market operations consist of the purchase and sale of securities initiated by the ECB and executed by the 11 EMU national central banks. In contrast to the Federal Reserve, the ECB accepts a wide range of assets in the conduct of monetary policy operations and does not focus trading on any particular government's securities.[20]

Like the Federal Reserve, the ECB uses open market operations to inject more reserves into or extract reserves from the banking system. By doing so, the ECB keeps the average interbank interest rate, called the EONIA, close to the ECB's target overnight rate, called the main refinancing rate. The ECB's most important open market instrument is a *reversing transaction,* which can be used to make temporary changes in the supply of bank reserves. Suppose, for example, the ECB needs to change the supply of bank reserves for three days. It will instruct national central banks to engage in reversing transactions: the central banks agree to purchase securities from or sell securities to dealers who agree to repurchase or resell them at a specified price three days later. Purchasing securities adds euros to the banking system's reserves; selling them drains euros from the banking system's reserves. When the reversing transactions mature, the initial injection or drain of euros is automatically reversed.[21] Reversing transactions serve as a convenient way for the ECB to deal with short-term pressures on the interbank interest rate, since transaction costs for reversing transactions are low.

The Fed, too, frequently uses reversing transactions called repurchase agreements, or repos, and matched sale-purchase transactions, or MSPs, to deal with short-term pressures on the interbank interest rate. Thus, the ECB and the Fed use the same types of transactions, but give them different names.

The ECB also has other types of open market operations at its disposal: (1) *outright transactions* (operations in which the ECB buys or sells assets in the financial markets to make changes in the supply of euro bank reserves that do not automatically reverse); (2) *issuance of debt certificates* (in which the ECB issues its own debt rather

[19]The average exchange rate in March 1999 was 1.0886 dollars per euro.

[20]Precise definitions of eligible assets for monetary policy operations, as well as eligibility requirements for counterparties in asset transactions, are given in the ECB's September 1998 publication. The list of eligible assets can be found on the ECB's website; it includes private debt and equities as well as government securities.

[21]The interest rate on regular reversing transactions, called the main refinancing rate, is also the rate of interest the ECB pays on required reserves.

than selling some of its assets to absorb euros from the banking system); (3) *foreign exchange swaps* (in which the ECB buys or sells a foreign currency and simultaneously agrees to sell or buy that currency at a specified future date); or (4) *collection of fixed-term deposits* (in which the ECB, to absorb euros from the banking system, invites eligible depositors, such as banks, to make interest-bearing fixed-term deposits at national central banks). These operations would produce long-term changes in the supply of euro bank reserves. Of these operations, the Federal Reserve uses only outright transactions.[22]

Standing facilities. Standing facilities are outlets through which eligible banks can borrow from or lend to the national central banks overnight. Standing facilities provide reserves (when banks borrow) or absorb reserves (when banks lend). The ECB uses standing facilities to signal the general stance of monetary policy and to provide upper and lower bounds for overnight market interest rates.

The ECB provides two types of standing facilities. Any eligible bank can use a *marginal lending facility* to obtain overnight loans from its national central bank. Under normal circumstances, there are no credit limits or other restrictions on banks' access to the facility apart from a requirement to present sufficient assets as collateral.[23] The interest rate on the marginal lending facility is higher than the ECB's target interbank rate; it normally provides a ceiling for the overnight market interest rate, since banks wouldn't borrow from each other overnight at a higher interest rate than that offered at the marginal lending facility.[24]

Any eligible bank can use a *deposit facility* to make overnight deposits with its national central bank. The interest rate on the deposit facility is lower than the ECB's target interbank rate; it normally provides a lower bound for the overnight market interest rate, since banks wouldn't lend to each other overnight at a rate lower than the interest rate at the deposit facility.

A plot of the overnight rate shows its movements over the first half of 1999 (see the Figure). Also shown are the marginal lending rate and the marginal deposit rate, as well as the ECB's target rate (the main refinancing rate). Note that the overnight rate sometimes differs from its target, since the ECB can't control it precisely. The overnight rate always lies between the lower bound provided by the marginal deposit rate and the upper bound provided by the marginal lending rate.

Implementation. On April 8, 1999, the ECB's president provided an example of how the ECB monitors M3 growth relative to its target value

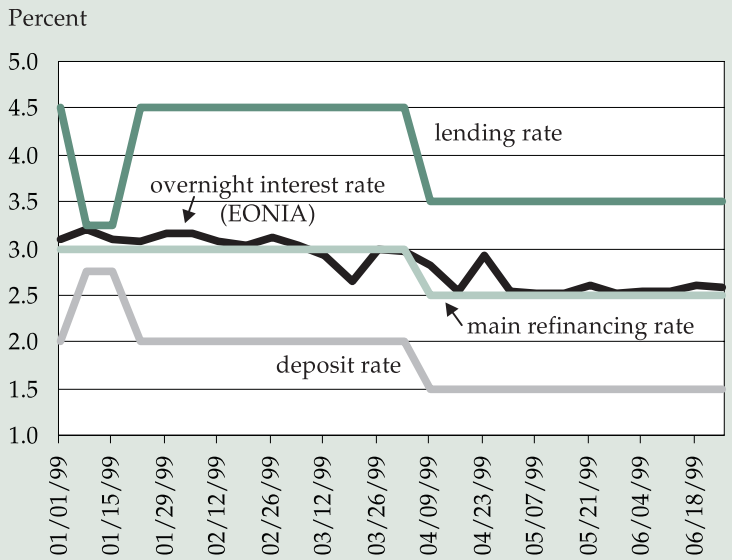
[22]The Fed does have the ability to maintain reciprocal currency arrangements (sometimes called swap facilities) with other central banks, but does not use them to implement monetary policy. Currently, under NAFTA, the Fed maintains swap facilities with Canada and Mexico.

[23]The Federal Reserve also provides a lending facility called the discount window. Eligible depository institutions can borrow from the discount window, typically overnight but sometimes for longer periods, when they face a temporary need for liquidity and cannot readily raise funds from other sources. As is true of the ECB's marginal lending facility, institutions must post collateral to borrow at the discount window. But in contrast to the ECB, the Fed lends at a rate that is usually slightly below policymakers' target for the interbank interest rate, for approved purposes only, and at its own discretion — banks and other depository institutions are not free to borrow as much as they might wish.

[24]In addition to its normal discount window facility, the Fed has chosen to provide a Century Date Change Special Liquidity Facility, from October 1, 1999, to April 7, 2000, as a precaution against unusual funding pressures around the century date change. This special liquidity facility is much like the ECB's marginal lending facility. Eligible depository institutions will be able to borrow as much as they wish from the Fed, provided they have sufficient collateral, at an interest rate 1.5 percentage points above the FOMC's target for the federal funds rate.

FIGURE

ECB Interest Rates and Money Market Rates



EONIA is the euro overnight index average, a weighted average of interest rates on all unsecured overnight lending transactions in the interbank market in the euro zone. The main refinancing rate is the ECB's target interbank rate.

and how it uses tools of monetary policy. At a press conference that day, President Duisenberg discussed a decision to cut the ECB's target for the interbank interest rate from 3.0 percent to 2.5 percent and to cut the interest rate on the marginal lending facility from 4.5 percent to 3.5 percent, as well as that on the deposit facility from 2.0 percent to 1.5 percent. The figure shows the changes in all three interest rates. The decision to cut rates took into account the rate of growth of M3, which remained close to the ECB's target value; inflation, which had been below 1 percent (per year) for several months and thus within the ECB's definition of price stability; and the prospects for overall growth in the euro area, which had weakened.

Although the tools available to the ECB allow it to pursue its objectives, what ensures that it can pursue those objectives without facing undue short-term political pressures?

INDEPENDENCE AND ACCOUNTABILITY OF THE ECB

The Maastricht Treaty explicitly set up the ECB as an independent institution free from short-term political pressures. Arguably, the ECB is, by design, one of the most independent central banks in existence. Members of its Governing Council all have relatively long terms of office. Each member of the Executive Board serves a nonrenewable eight-year term, and the central bank governors of the 11 EMU countries serve renewable five-year terms. Moreover, the ECB's mandate makes clear that institutions such as the European Parliament and the governments of EMU member countries may not give instructions to the ECB, nor is the ECB allowed to follow instructions or suggestions from others. However, independence does not imply lack of accountability.

Proceedings of the Governing Council's meetings are kept confidential, to guard against short-term political pressures on individual members. However, just as the Federal Reserve's FOMC announces decisions made at each meeting on

the day of the meeting, the ECB's Governing Council holds a press conference immediately after its first meeting every month. At that time, it releases the "President's Introductory Statement," which is a summary of the council's conclusions from its assessment of economic conditions.[25]

In addition, the ECB publishes other reports to communicate its policy objectives, intentions, and actions. The president of the ECB presents an annual report to the European Parliament, the Council of Ministers, and the European Commission, and the ECB publishes monthly and annual reports, as well.[26] Similarly, the Chairman of the Federal Reserve Board delivers semiannual reports on monetary policy to the U.S. Congress, and the Federal Reserve publishes monthly and annual reports. Also, the president of the Governing Council and the other members of the Executive Board of the ECB, at their own initiative or on request, may be heard by committees of the European Parliament. Like the Federal Reserve, the ECB has made clear its willingness to engage in dialogue concerning its own and other institutions' policies with responsible authorities.

Thus, the ECB is accountable for its policy actions within an institutional structure that provides substantial independence.

CONCLUSION

After decades of planning, the euro was born on January 1, 1999. This new currency is shared by the 11 member countries of the EMU, an economic area whose portion of world output of goods and services, at around 20 percent, is second only to that of the United States. The EMU member countries are now in transition to sole reliance on the euro as the single currency of the union. Guiding the EMU through and beyond its transition phase is the job of the European System of Central Banks, including the new European Central Bank that determines monetary policy for the 11 countries that have adopted the euro. The ECB's policymaking body and tools for setting and conducting monetary policy are similar in several ways to those of the Federal Reserve but also have certain differences.

[25]The "President's Introductory Statement" is similar to the minutes of FOMC meetings. The FOMC publicly releases *minutes* of each meeting about six weeks afterward, and it releases *transcripts* from FOMC meetings after five years.

[26]The Maastricht Treaty requires the ECB to publish quarterly and annual reports covering monetary policy and its other activities. The ECB goes beyond this requirement and, like the Federal Reserve, publishes a monthly bulletin. The ECB also releases weekly financial statements, as does the Fed.

REFERENCES

Eichengreen, Barry. "European Monetary Unification," *Journal of Economic Literature,* 31, September 1993.

Eudey, Gwen. "Why Is Europe Forming a Monetary Union?" Federal Reserve Bank of Philadelphia *Business Review,* November/December 1998.

European Central Bank. "The Single Monetary Policy in Stage Three: General Documentation on ESCB Monetary Policy Instruments and Procedures," September 1998.

European Monetary Institute. "Convergence Report: Report Required by Article 109j of the Treaty Establishing the European Community," March 1998.

Feldstein, Martin. "The Political Economy of the European and Monetary Union: Political Sources of an Economic Liability," *Journal of Economic Perspectives,* vol. 11, no. 4, Fall 1997.

Kenen, Peter B. *Economic and Monetary Union in Europe: Moving Beyond Maastricht.* Cambridge: Cambridge University Press, 1995.

Meulendyke, Ann-Marie. *U.S. Monetary Policy and Financial Markets.* Federal Reserve Bank of New York, 1997.

Obstfeld, Maurice. "EMU: Ready, or Not?" National Bureau of Economic Research Working Paper No. 6682, 1998.

Stevens, Ed. "The Euro," Federal Reserve Bank of Cleveland *Economic Commentary,* January 1, 1999.

Wynne, Mark A. "The European System of Central Banks," Federal Reserve Bank of Dallas *Economic Review,* First Quarter 1999.

Patent Reform: A Mixed Blessing For the U.S. Economy?

Robert Hunt*

At the end of the 1990s, it seems ironic to question the performance of the American patent system. Spending by industries on research and development, measured in inflation-adjusted dollars or as a percent of gross domestic product, has never been higher (Figure 1). Patenting activity in the U.S. has never been higher (Figure 2). The rate of technological advance in sectors such as drugs, computer hardware, and software is simply amazing. Yet there is evidence that devoting even more resources to R&D could further improve our standard of living.[1]

Twenty years ago, the perspective was quite different. Reacting to the most severe recession since World War II, and observing the rapid emergence of Japanese and other foreign competitors in the computer and other high technology sectors, policymakers became increasingly concerned about the technological competitiveness of American companies. There was reason for

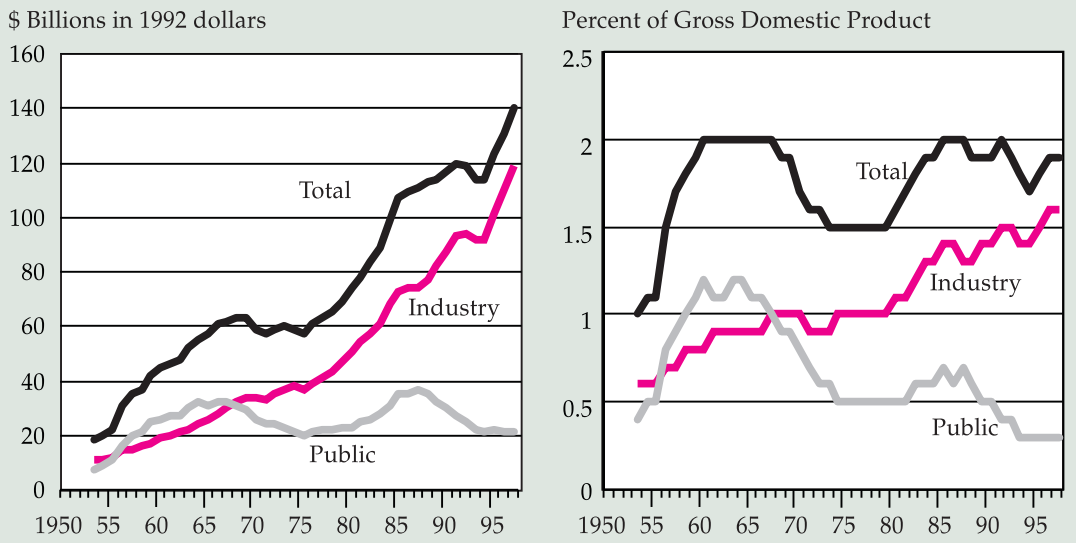
*Bob Hunt is an economist in the Research Department of the Philadelphia Fed.

[1] See, for example, the article by Charles Jones and John Williams.

FIGURE 1

Spending on Research and Development

(in dollars and as percent of GDP)



Source: Bureau of Economic Analysis: National Income and Product Accounts; National Science Foundation; and author's calculations.

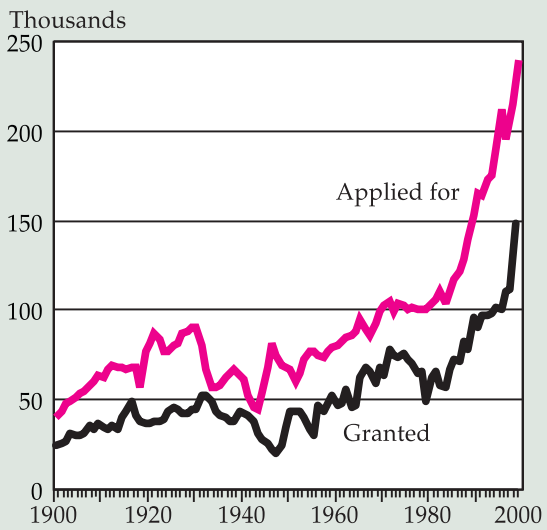
this concern. During the 1970s, private R&D spending and the number of patents issued to U.S. residents stagnated at a time when both were growing rapidly abroad. Productivity growth declined in most developed economies in the early 1970s, but it looked particularly anemic in the United States. From the late 1970s to the mid 1980s, the market share of important industries, such as steel, automobiles, and semiconductors, held by foreign companies increased dramatically.

These pressures prompted a re-examination of the American system of intellectual property law, which resulted in many significant legislative changes and important changes in the way federal courts decide patent cases. This article considers the effects of an especially important aspect of these changes: many more inventions qualify for patent protection than before. On its face, this would appear to be a good thing, since it might encourage businesses to devote additional resources to developing new products and processes. But economic analysis suggests that the effects of these changes are more complicated than they at first appear. It may well be the case that, in some industries, the rapid technological advances seen in the 1990s have occurred not *because* of these changes in patent law, but *in spite* of them.

THE NATURE OF THE U.S. PATENT SYSTEM

The U.S. Constitution grants Congress the power "to promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries."[2] Thus, the

FIGURE 2
Patent Activity



Source: U.S. Patent and Trademark Office

Constitution permits the government to offer an incentive, in the form of a temporary monopoly, to artists and inventors. Congress quickly took advantage of these powers, passing the first patent act in 1793. The act was drafted by Thomas Jefferson, who was himself a prodigious inventor.

The role of patents envisioned in our Constitution essentially follows economic intuition. It usually costs more, in terms of effort and money, to discover something new than it does to duplicate someone else's discovery. Inventors may work on their discoveries for a variety of reasons. But so long as one of the motivations is the prospect of financial reward, inventors will be concerned about the possibility that others will imitate their discoveries. If an invention can be imitated quickly, the inventor will soon be forced to compete with other suppliers, ones that did not incur the development costs he or she bore. This competition will reduce, possibly even eliminate, the profits an inventor can earn from his or her discovery. In such an environment, then, a discovery not protected by a patent gives the inventor only a fleeting advantage over his or her competitors. Obtaining a patent can reduce this competition because it gives the inventor a temporary monopoly to produce his or her invention. Thus, by helping to ensure a reasonable economic return to inventive activity, patents provide an important incentive to engage in research and development.[3]

But patents also create inefficiencies. Since patent holders have a monopoly over the patented technology, they can charge a higher price than they could charge in a competitive market. In most cases, there will be some consumers willing to buy the product at the competitive price, but unwilling to pay the higher price charged by the patent holder.

Another sort of inefficiency sometimes arises from patents. In many industries, making the best product or using the most advanced pro-

[2]U.S. Constitution, Article I, Section 8. To be precise, a patent grants the right to exclude others from producing a product or using a process covered by the patent's claims.

[3]The significance of patents as an incentive for inventors is sometimes exaggerated. Economic research verifies that patents do provide benefits to inventors, but it has also shown that other factors, such as trade secrets or simply having a head start on the competition, are often just as important. See, for example, the articles by Mark Schankerman; Richard Levin and others; and Edwin Mansfield and others.

cess may require using ideas developed by many different people. Some of those ideas will be patented, so using them requires the consent of the patent owner. While developers and users of technologies have an incentive to reach an acceptable licensing arrangement, the cost of doing so is sometimes quite high. In some cases, an acceptable arrangement is not reached and the parties may resort to litigation.

Two notable examples of this kind of failure include the airplane and the radio in the early years of the 20th century.[4] In both instances, several companies obtained patents covering important aspects of these highly valuable inventions. Unfortunately, they were unable to reach a satisfactory cross-licensing arrangement, and this failure precluded the manufacture of the most advanced aircraft or radios in the U.S. These impasses were broken by the intervention of the U.S. government during the First World War. In the case of aircraft, a successful system of cross-licensing was established, and it continued after the war. In the case of radio, patent rights were essentially suspended for the duration of the war. After the war, the U.S. Navy encouraged the formation of the Radio Corporation of America, which soon held rights to virtually all the important radio patents and a near monopoly position in the emerging industry.

To limit the effects of these kinds of inefficiencies, economists argue that patents should be granted only for novel and valuable discoveries. That is precisely what the American patent system is designed to do. To qualify for protection under U.S. patent law, an invention must be novel, useful, and nonobvious. While the first two criteria are straightforward, the third criterion is less clear. It requires that an invention represent more than a trivial advance over what is already known. This requirement, awkwardly referred to as *nonobviousness,* is typically the most difficult of the three to satisfy.

The idea that only nonobvious inventions should be patentable occurs in some of the earliest patent cases. In a famous 1851 decision, *Hotchkiss v. Greenwood*, the Supreme Court invalidated a patent on doorknobs made of porcelain or clay, arguing that the substitution of these materials for wood or metal was obvious. Thus, the judicial concept of nonobviousness was at least a century old when, in 1952, Congress amended the Patent Act to include a comparable statutory requirement.

In a 1966 case, *Graham v. Deere,* the Supreme Court described how courts should decide whether an invention satisfies the statutory requirement of nonobviousness. First, the court must determine the level of skill of an ordinary practitioner in the field. Next, it must identify the relevant knowledge that existed at the time the invention was made; this is called the prior art. The court must then identify any differences between the claimed invention and the prior art. Finally, the court must determine if those differences would have been obvious to a practitioner of ordinary skill in the relevant field. Other indicators of nonobviousness might also be considered, for example, a long-felt need for the invention, the failure of others to perfect the invention, or commercial success.

How stringent is this requirement of nonobviousness? In *Graham v. Deere,* the Supreme Court invalidated a patent on a combined sprayer and cap used on bottles of household chemicals. The cap, which covers the sprayer, protects the pump and seals off any leaks. The essential elements of the sprayer had been developed by others, but they had never been assembled in this particular way, which made possible the use of automated bottling equipment and reduced handling costs. As a result, the product was highly successful. While the Supreme Court acknowledged that long-felt need and commercial success might suggest the invention was nonobvious, in the end it decided otherwise

[4]For details on the history of these disputes, see the article by Robert Merges and Richard Nelson and the article by Paul Schaafsma.

because the differences between the product's design and that of preexisting ones were minimal.

A more recent example involves semiconductor chips used in computers and other electronic devices. In the early 1980s, courts treated the layout of most semiconductor chips in the same way they treated dress designs: unpatentable variations of a single idea—despite the fact that even minute differences in the layout of a computer chip can significantly improve its performance. In testimony before Congress, Harvard Law Professor Arthur Miller went so far as to say that "as a practical matter, the layout of a chip...will rarely, if ever, satisfy the standard of invention. A chip may be the product of millions of dollars and thousands of hours of effort, but it is the result of hard work, not 'invention.'"[5]

WHAT HAPPENED IN THE 1980s?

During the late 1970s and early 1980s, businessmen and policymakers became increasingly concerned about the apparent deterioration of America's comparative advantage in high technology industries, such as the semiconductor industry. In fact, trends within that industry became a catalyst for dramatic changes in the way the U.S. protects intellectual property.

Semiconductors were invented by American scientists in the late 1940s, and from its beginnings in the 1950s, the semiconductor manufacturing industry was dominated by American companies. The industry's growth was phenomenal. Between 1972 and 1982, the dollar value of semiconductor shipments increased more than 450 percent. If the decline in prices of computer chips during this period is taken into account, shipments in 1982 were 17 times higher than in 1972. Also, employment in the industry increased 71 percent.

So at least until the late 1970s, it would be difficult to argue that the development of the American semiconductor industry was seriously hindered by the lack of patent protection for most semiconductor designs. Indeed, some scholars argue that the industry's rapid technological development could be a consequence of limited patent protection.[6] In industries where technology is advanced by cumulative improvements, the fact that companies are able to copy many of the improvements made by rivals could be beneficial. A healthy amount of *reverse-engineering* allows a firm to incorporate the most advanced technologies, irrespective of their origin, in new designs of its own. Of course, reasonable people may disagree about what they think is a healthy amount of this kind of imitation.

Within the U.S. semiconductor industry, reverse-engineering was a well-established practice. But by the late 1970s, American firms objected to similar behavior by Japanese firms when they began to increase their market share in the more standardized products, such as computer memory chips. The level of competition eventually became so intense that, by the mid 1980s, most American companies abandoned these segments entirely.

When it became clear they could no longer dominate Japanese firms on the basis of production technology alone, American firms attempted to consolidate their comparative advantage in research and development. To do this, they would have to find ways of reducing their competitors' ability to reverse-engineer their products. To that end, American companies began to lobby Congress to increase intellectual property protection for their semiconductor designs. In 1984, Congress created a new form of intellectual property right, called *mask rights*, specially

[5]See the Senate report on S. 1201, one of the versions of the Semiconductor Chip Protection Act considered in the 98th Congress. It should be emphasized that while the layout of computer chips was generally unpatentable, new circuits or new processes for making computer chips could, and often did, qualify for patent protection.

[6]See, for example, the article by Robert Merges and Richard Nelson.

tailored to address the needs articulated by the industry.[7] A critical difference between these mask rights and patents was that the level of originality required to qualify for a mask right was substantially lower than what was implied by patent law's requirement of nonobviousness. Thus, many more semiconductor designs were likely to qualify for protection under a mask right than under a patent.

What was occurring in the semiconductor industry was also being felt in many other industries. By the late 1970s, there was considerable dissatisfaction with how federal courts were deciding patent cases, especially the frequency with which the courts were invalidating patents. In addition, there is some evidence that patents were being treated differently by federal courts in different parts of the country. This impression contributed to forum shopping by litigants, increasing the cost and delay associated with patent cases. In 1982, Congress created a new federal appeals court, the Court of Appeals for the Federal Circuit, to hear all appeals of patent cases and certain other cases.[8] It was hoped that a single court of appeals would contribute to more uniform decisions by federal district courts across the country. But the decisions of this new court also changed the way federal courts apply the test for nonobviousness.

The early decisions of this new court accomplished many things. In particular, these decisions increased the attention that courts pay to secondary factors, such as long-felt need or commercial success, when evaluating the obviousness of an invention. While these factors had long been considered by the courts, the traditional view was that secondary factors would rarely, if ever, overcome the conclusion of the multipart inquiry described earlier. For example, in one case, a federal district court considered the validity of a patent for a fastener used to attach shelves to the inside walls of refrigerators. The court concluded the invention was an obvious combination of features contained in existing fastener designs and invalidated the patent. The court refused to consider secondary considerations, in particular the product's commercial success, arguing those factors could not overcome the conclusion reached in a review of the prior art. On appeal, the new court reversed this decision, arguing that secondary factors must be considered and that, in this case, they outweighed the conclusion reached in the traditional three-step analysis. The new appeals court reached similar conclusions in a number of other decisions.[9]

It wasn't long before it was clear that the new court was deciding patent cases differently from the appeals courts that preceded it. In the first 25 years after the passage of the 1952 Patent Act, patents issued by the U.S. Patent and Trademark Office were subsequently invalidated in 60 percent of the cases decided by federal courts of appeal.[10] A 1985 study found that in a majority of patent cases reviewed by the newly created appeals court, the court determined the patent in question was nonobvious. That rate stands in contrast with the lower courts, where 30 percent of the patents reviewed were found to be

[7]Mask rights were created by the Semiconductor Chip Protection Act of 1984, 17 U.S.C. 901-914.

[8] The Federal Courts Improvement Act of 1982, 28 U.S.C 1295.

[9]For example, in a 1983 decision, *Stratoflex, Inc. v. Aeroquip Corporation,* the court stated: "Indeed, evidence of secondary considerations may often be the most probative and cogent evidence in the record. It may often establish that an invention appearing to have been obvious in light of the prior art was not."

[10]This statistic is based on an unpublished study cited in Steven Szczepanski's article. One should be careful about reading too much into statistics of this sort. Only a small fraction of all patents are involved in some form of litigation, and only a small fraction of those cases are appealed.

nonobvious. So it is hardly surprising to learn that the new court was twice as likely to reverse a lower court's finding that an invention was obvious than to reverse a finding of nonobviousness by a lower court (31 percent vs. 14 percent, respectively).[11]

What is the significance of all these decisions? About a decade after its creation, one practitioner wrote, "Many patent attorneys believe that the obviousness defense is dead and that the cause of death lies in the decisions of the Court of Appeals for the Federal Circuit."[12] Another expert argued that "as a result of these changes, patents today are more likely to be held valid than, perhaps, at any time in our history."[13]

ARE MORE PATENTS NECESSARILY BETTER?

Do these changes explain the recent surge in R&D activity and the improvement in U.S. competitiveness? Many believe that the federal circuit's decisions reduced uncertainty about the enforceability of patents, a belief that, in itself, would make them more valuable. A number of decisions increased the presumption of patent validity—that is, courts now require more evidence before concluding a patent is invalid. Other decisions made it easier for a patent holder to obtain preliminary injunctions, court orders banning a potential infringing activity before the question of infringement is definitively decided. And it does appear that centralizing the appeals process for patent cases has succeeded in reducing disparities in the treatment of patents across federal district courts.

Probably the greatest single impact of the federal circuit's decisions during the 1980s was to make patents easier to obtain by relaxing the nonobviousness requirement. Wouldn't this also encourage additional private investment in R&D? Somewhat surprisingly, the answer is unclear. In fact, it is possible that making patents easier to obtain might actually reduce R&D activity, especially in high technology industries. What explains this paradoxical result?

R&D Investments Are Related to Their Expected Return. Companies, and at least some individual inventors, make decisions about their R&D activities in the same way they make other investment decisions. In other words, they calculate how much they can expect to earn from different R&D projects and allocate their resources to the ones with the highest expected returns. The higher these expected returns are, the more a firm will be willing to invest in that project.

The expected return from an R&D project is determined by a variety of factors: the cost of the R&D, the chances of making a significant discovery, the likelihood it can be patented, and the flow of profits earned over the life of a patent. The timing of those profits matters because the sooner they are earned, the sooner they can be used to reward investors or reinvested in new projects. This means that profits earned in the near term are more valuable than profits expected to be earned far into the future.[14]

Changing Patent Law Affects the Return to R&D. Patent law matters because it affects the expected return to an R&D project in two ways: it determines the probability that a given discov-

[11]See Donald Dunner's 1985 article. These findings are re-confirmed for the 1982-94 period in Dunner's 1995 article.

[12]See Ronald Coolley's 1994 article. The title of Robert Desmond's 1993 article—"Nothing Seems Obvious to the Court of Appeals for the Federal Circuit"—is also suggestive.

[13]See Lawrence Kastriner's 1991 article.

[14]In other words, inventors discount the value of future profits to take into account the time value of money. Discounting allows inventors to compare the revenues earned in the future to dollars being spent today. The longer they must wait to earn profits on an invention, the more heavily they will discount those profits.

ery can be patented, and it also influences the flow of profits earned over the life of a patent. Relaxing patentability criteria, in particular adopting a weaker standard of nonobviousness, will have two opposing effects on the return to R&D. As patentability criteria are relaxed, a larger share of future innovations will qualify for patent protection. Firms enjoy the benefit of being able to protect more of their inventions from imitation. But firms also lose, because their ability to imitate their rivals' inventions is reduced. Each firm must now compete with rivals that, over time, will receive more patents of their own. As a result, the profits earned from a given patent tend to be smaller and may not last as long. Hence, the value of a patent declines.[15]

The question for policymakers is: which of these two effects is more important? If the probability of obtaining patent protection rises more than the value of patents declines, the expected return to R&D will increase. This should stimulate additional R&D investments and more innovation. On the other hand, if the probability of obtaining patent protection rises less than the value of patents falls, the expected return to R&D will fall. This would discourage firms from engaging in R&D, which, in turn, would reduce the rate of innovation.

Weaker Patentability Criteria and High Technology Industries. It turns out that the effect of relaxing patentability criteria on R&D activity in a given industry depends on the initial rate of innovation in that industry, which, in turn, depends on the opportunities for technological improvement and the resources devoted to perfecting those improvements. Some industries innovate more rapidly than others. In the semiconductor industry, for example, entirely new generations of computer processors and the technology to make them are developed every few years. In other industries, such as steel, it may take several decades to develop a new technology and replace an existing one.

Consider an industry that, prior to a change in the patent system, innovates slowly. In this environment, competition from new technologies takes a long time to develop, so a patentable invention is likely to be highly valuable. In such an industry, relaxing the standard of nonobviousness increases the chances that a firm will obtain a patent that is likely to generate profits for a long time. In addition, it will take a relatively long time before other firms make even the marginal discoveries that would now qualify for patent protection. So the loss of profits to this increased competition won't occur until far into the future. The effect of this increased competition on the value of patents is likely to be small, then, because profits earned far into the future are worth a lot less to the firm than profits earned today. So in the case of an industry that initially innovates slowly, the effect of an increase in the probability of obtaining a patent is probably more important than the decline in the value of patents. So a weakening of patentability criteria is likely to increase the expected return to R&D, and therefore R&D activity and the rate of innovation, in industries that initially innovate slowly.

Now consider the case of an industry that, prior to a change in the patent system, innovates more rapidly. In this environment, new technologies are invented more frequently and, if protected from imitation, very soon compete with the existing technologies. An invention in this industry generates less profits, over less time, than an invention of comparable significance in an industry that innovates more slowly. Consequently, other things equal, individual patents in this industry are less valuable. As a result,

[15]Formal models that illustrate this point include my 1999 working paper and the paper by Ted O'Donoghue and the one by Olivier Cadot and Steven Lippman. It should be noted that this research is not saying that inventions have become intrinsically less valuable. They continue to make possible better or less costly products, or both. But an increase in competition reduces the profits that can be earned on them.

firms do not stand to gain as much from an increase in the likelihood of obtaining patent protection. But once patentability criteria are relaxed, a firm's rivals are able to patent their inventions more easily, which increases their ability to become a market leader. This further reduces the value of the firm's own patents. In the case of an industry that initially innovates rapidly, the decline in patent values is likely to be more important than the increase in the probability of obtaining a patent. So a weakening of patentability criteria is more likely to *reduce* the expected return to R&D, and therefore R&D activity and the rate of innovation, in industries that initially innovate rapidly.

In sum, any positive effect on the expected return to R&D and the rate of innovation resulting from weaker patentability criteria is most likely to occur in industries that originally innovated more slowly. Any negative effect on the expected return to R&D and the rate of innovation is most likely to occur in industries that originally innovated more rapidly. Thus, relaxing the nonobviousness requirement of patent law may not be a very effective way to encourage more rapid advancement in high technology industries.

BUT WHAT ABOUT THE SURGE IN R&D ACTIVITY?

The timing of the changes in patent law, the subsequent surge in R&D activity, and the apparent improvement in American technological competitiveness convinced many attorneys and policymakers that these changes worked as intended.[16] But determining whether weaker patentability criteria really explain those improvements is a difficult empirical question. One problem with associating cause and effect is that patent law changed in so many ways during the 1980s. Outside of a few technology fields, it would be difficult to identify exactly which change in the patent system was the cause of some desirable or undesirable outcome. In addition, during the 1980s, the U.S. economy experienced very large swings in macroeconomic conditions and a dramatic restructuring of its manufacturing sector. Finally, during this same period, a large and sophisticated venture capital market emerged, significantly increasing access to capital for start-up companies in certain industries. Separating all of these influences is no easy feat.

For example, what do we make of the surge in patenting activity in the U.S. during the 1980s and 1990s? Are inventors patenting a higher share of their discoveries? Or are they making more discoveries and patenting many of those? Or is it both? One way to sort out these explanations is to look at the trend in patenting across countries. Evidence that patenting surged in the U.S., but not elsewhere, might be explained by the relaxation of patentability criteria in the 1980s. A surge of patenting in the U.S. by foreign inventors might reinforce this conclusion, especially if there was no comparable increase in patenting abroad. Conversely, a surge of patenting activity in many countries might be better explained by an increase in technological opportunities worldwide. And evidence that U.S. inventors increased their patenting abroad as much as they increased their patenting at home might be better explained by an increase in technological opportunities in the U.S.

In a recent article, economists Samuel Kortum and Josh Lerner examined trends in patenting in the U.S., Europe, and Japan. They found that European inventors increased their patenting in the U.S. in the late 1980s, but that trend was not sustained in the 1990s. Japanese inventors significantly increased their patenting activity, both at home and abroad, during the 1980s. But this is a continuation of a trend evident from the 1960s. Meanwhile, American inventors signifi-

[16]The articles by Lawrence Kastriner and Gerald Sobel are good examples of the optimistic perspective in the legal community.

cantly increased their patenting activity in the U.S. and abroad. The authors concluded these changes in aggregate patenting activity were better explained by an increase in technological opportunities in the U.S. than by a change in the treatment of patents by U.S. courts.

Kortum and Lerner also looked for evidence of a change in the value of patents during the 1980s. Unfortunately, there is not a great deal of information about the initial value of patented inventions. But there are ways to infer something about the value of patents as they get older. In the U.S. and in a number of European countries, patent owners must pay "renewal fees" to keep their patents in force the first few years after they are issued. In the U.S., patents issued after 1980 are subject to renewal fees in the fourth, eighth, and 12th years of the patent. Paying these fees is not mandatory, but if they are not paid, the patent expires at the renewal date rather than at the end of the patent's full term (20 years in the U.S.). If an owner chooses to pay a renewal fee, it is probably because he or she believes the patent remains sufficiently valuable to justify bearing the cost of the fee.[17]

Kortum and Lerner cited recent evidence that patent renewal rates fell during the first half of the 1990s, which suggests a decline in the residual value of patents.[18] This drop-off in renewal rates is consistent with the argument that making patents easier to obtain in the U.S. caused the profits earned on patents to erode more quickly. But that is only one of many possible explanations for an apparent decline in the value of patents. Since patent renewal rates have declined in other countries, perhaps other explanations may be more important.

Changes in patentability criteria could affect the rate of innovation by changing the expected return to R&D. So it may be helpful to look at more direct evidence of changes in the expected return to firms' R&D programs, for example, the stock market's valuation of R&D investments made by publicly traded companies. To derive estimates of this sort, economists use what is called the hedonic approach, which attempts to allocate a firm's stock market value to various characteristics, including its tangible and intangible assets.[19] An important component of a firm's intangible assets is its investments in R&D.[20] Investors presumably value a firm's R&D investments based on their assessment of the potential output: new technologies that contribute to the growth and profitability of the firm. The stock market's valuation of R&D investments should respond to changes in patent law that affect the profitability of developing new products and processes.

Economist Bronwyn Hall has reported that the market value of R&D investments made by

[17] The articles by Ariel Pakes and Mark Schankerman explore the relationship between patent renewal decisions and the remaining value of patents. Patent owners will always choose to renew their most valuable patents, but they may choose not to renew their less valuable ones. The profitability of different patents varies dramatically. The vast majority of patents are of little value, while a small proportion of patents are extremely valuable.

[18] Patent renewal fees increased significantly in 1990, so the authors reported changes only for years after the fees increased.

[19] This approach is described more thoroughly in Bronwyn Hall's 1999 working paper and Zvi Griliches' 1990 article and his 1984 book.

[20] In the U.S., firms' spending on research and development is expensed, i.e., deducted from revenues when calculating current profits. Some economists argue that it is more appropriate to think of R&D expenditures as an investment that contributes to a firm's stock of intangible capital. Standard accounting does not report a stock of intangible capital so this measure must be constructed using a firm's R&D expenditures and an assumption about the rate of depreciation of these investments. For details, see my 1996 working paper and Leonard Nakamura's 1999 article in the *Business Review.*

about 1000 publicly traded companies increased throughout most of the 1970s, then began to decline after 1983. The decline was especially pronounced in the electrical and computing industries, a fact Hall attributed to more rapid technological obsolescence and the competitive effects of entry by new firms.[21] Around 1990, the market's valuation of R&D investments began to rise again. Hall's findings are the result of many factors, but they provide scant support for the idea that changes in patent law increased the market value of R&D investments during the 1980s.

In previous research, I examined the market valuation of R&D investments made by a dozen American semiconductor companies from 1976 to 1994. I found that if only a firm's own R&D investments were taken into account, there was a significant increase in the market value of those investments, but that it occurred after 1989—more than five years after the significance of the changes in the patent system were widely known. While it is possible that those changes explain this increase, the long delay between the alleged cause and its effect suggests that alternative explanations cannot be ruled out.

In the semiconductor industry, the R&D activity of a company's rivals is very important. The widespread practice of reverse-engineering suggests that firms learn a great deal from each other's products, which are themselves the result of considerable research and development. That suggests the possibility of a spillover — the value of a company's own research might be affected by the research conducted by its rivals. Of course, the firm's rivals are doing the same thing, and that means they could soon be producing a similar chip that competes directly with the firm in the product market. That competition is likely to depress prices and, therefore, profits, which could reduce the market value of the firm.

In a 1996 study I analyzed three types of effects that R&D investments might have on a firm's market value: a direct effect, measured by the firm's own R&D investments; a competitive effect, measured by the R&D investments of its rivals; and a spillover effect, measured by the interaction of the firm's own R&D investments with those of its rivals. Using statistical techniques and data on a dozen American semiconductor companies, I was able to confirm that a change in the relationship between these variables and the firms' market value did occur at some point in the 1980s.[22]

In the early part of the decade, the R&D activities of its rivals tended to reduce a firm's market value (the competitive effect). During this period, the contribution to a firm's market value made by its own R&D investments (the direct effect) was quite small, but this contribution was higher the more the firm's rivals spent on R&D (a positive spillover). These results can be explained in a number of ways, but they are cer-

[21]It might at first appear that Hall's finding conflicts with the fact that private spending on R&D increased significantly during the 1980s. But a decline in the value of R&D investments amid rising R&D spending can be explained by an increase in the supply of funds available for investment in R&D projects. If the cost of funding R&D investments declined, firms would be able to invest in R&D projects that would otherwise be unprofitable.

[22]To be precise, I regressed the ratio of each firm's market to book value on a constant, the ratio of the firm's R&D capital to its physical capital, a comparable ratio for its rivals, and an interaction of these two ratios. The direct effect is captured by the coefficient on the firm's ratio of R&D capital to physical capital. That should tell us something about the value of its R&D investments relative to its investments in physical capital. The competitive effect is captured by the coefficient on the comparable ratio for the firm's rivals. That should tell us something about the extent of any loss in market value attributable to reverse-engineering by its competitors. The spillover is captured by the coefficient on the interaction of the two ratios. That should tell us something about the contribution of the firm's own reverse-engineering efforts to its market value.

tainly consistent with an environment in which firms were able to reverse-engineer improvements embodied in each other's designs and incorporate them in new designs of their own.

At some point in the late 1980s or early 1990s, circumstances began to change. R&D investments made by a firm's rivals no longer reduced its own market value (the competitive effect), and in some cases actually increased it. At the same time, a firm's own R&D investments contributed significantly more to its market value than before. In other words, the direct effect had increased. But now there was a negative spillover: R&D investments made by its rivals reduced, rather than increased, the market value of a firm's own R&D investments. These changes are consistent with a shift from an environment of significant reverse-engineering to one relying more heavily on patent protection. One interpretation of the reversal of the competitive effect is that firms shifted away from competing directly in product markets and, more often than before, were supplying state-of-the-art components for their rivals' products. One interpretation of the reversal of the spillover effect is that firms were now able to use patents to preclude rivals from developing certain technologies.

The overall effect of these changes was that, once the spillover effect is taken into account, the market value of R&D investments for this group of semiconductor companies during the late 1980s and early 1990s was either the same as or lower than it was in the early 1980s. These results do not support the idea that granting mask rights or otherwise making patents easier to obtain raised the expected return to R&D among established firms in the U.S. semiconductor industry.

Nevertheless, both private R&D spending and patent activity in the industry increased significantly during the 1980s and early 1990s. It may be that other factors were more important than changes in the treatment of patents by U.S. courts. For example, a number of scholars point out that most American manufacturers retreated from certain industry segments and concentrated on products that were less susceptible to reverse-engineering.[23] Others argue that companies adapted to changes in the patent system in ways not anticipated by supporters of those changes. For example, Bronwyn Hall and Rose Marie Ham describe what they see as a trend toward strategic patenting, in which firms try to assemble large patent portfolios in the hopes of gaining leverage in cross-licensing negotiations with their competitors.[24]

CONCLUSION

Economic intuition in itself cannot tell us whether the weaker nonobviousness requirements adopted in the 1980s resulted in less R&D activity than would have occurred without those changes. But it does show that such an outcome is possible and that it is more likely to occur in rapidly innovating industries. Consequently, these changes tend to favor traditional industries over high technology ones. If policymakers remain concerned about encouraging innovation in high technology industries, they should also be concerned about whether the changes adopted in the 1980s advanced or retarded progress toward that goal.

The relatively small amount of empirical research that has been done so far is not favorable to the view that the recent, and impressive, increases in private R&D spending and patenting can be explained by the changes in patent law that occurred in the 1980s. A great deal more research needs to be done to reach a definitive conclusion about the effects of adopting weaker patentablity criteria. But the theoretical and empirical work we have available today suggests there is good reason to exercise caution before adopting similar changes in the future.

[23]See, for example, the articles by Steven Kasch, John Rauch, and Robert Risberg.

[24]See their 1999 working paper for an examination of patenting activity within the semiconductor industry.

REFERENCES

Cadot, Olivier, and Steven A. Lippman. "Fighting Imitation with Fast-Paced Innovation," Working Paper 97/73, INSEAD (1997).

Coolley, Ronald B. "The Status of Obviousness and How to Assert It as a Defense," *Journal of the Patent and Trademark Office Society,* 76 (1994), pp. 625-44.

Desmond, Robert. "Nothing Seems Obvious to the Court of Appeals for the Federal Circuit: The Federal Circuit, Unchecked by the Supreme Court, Transforms the Standard of Obviousness Under the Patent Law," *Loyola of Los Angeles Law Review,* 26 (1993), pp. 455-90.

Dunner, Donald R. Introduction to "The United States Court of Appeals for the Federal Circuit—The First Three Years," *AIPLA Q. J.*, 13 (1985), pp. 185-94.

Dunner, Donald R., J. Michael Jakes, and Jeffrey D. Karceski. "A Statistical Look at the Federal Circuit's Patent Decisions, 1982-1994," *Federal Circuit Bar Journal,* 5 (1995), pp. 151-80.

Griliches, Zvi, ed., *R&D, Patents and Productivity.* Chicago: University of Chicago Press, 1984.

Griliches, Zvi. "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature,* XXVIII (1990), pp. 1661-1707.

Hall, Bronwyn. "Industrial Research During the 1980's: Did the Rate of Return Fall?" *Brookings Papers: Microeconomics,* 2 (1993), pp. 289-343.

Hall, Bronwyn. "Innovation and Market Value," Working Paper No. 6984, NBER (1999).

Hall, Bronwyn, and Rose Marie Ham. "The Patent Paradox Revisited: Determinants of Patenting in the U.S. Semiconductor Industry, 1980-94," Working Paper 7062, NBER (1999).

Hunt, Robert M. "Nonobviousness and the Incentive to Innovate: An Economic Analysis of Intellectual Property Reform," Working Paper 99-3, Federal Reserve Bank of Philadelphia (March 1999).

Hunt, Robert M. "The Value of R&D in the U.S. Semiconductor Industry: What Happened in the 1980s?" mimeo, 1996.

Jones, Charles I., and John C. Williams. "Measuring the Social Return to R&D," *Quarterly Journal of Economics,* 113 (1998), pp. 1119-35.

Kasch, Steven P. "The Semiconductor Chip Protection Act: Past, Present, and Future," *High Technology Law Journal*, 7 (1992), pp. 71-105.

Kastriner, Lawrence G. "The Revival of Confidence in the Patent System," *Journal of the Patent and Trademark Office Society,* 73 (1991), pp. 5-23.

Kortum, Samuel, and Josh Lerner. "What Is Behind the Recent Surge in Patenting?" *Research Policy,* 28 (1999), pp. 1-22.

REFERENCES (continued)

Levin, Richard C., Alvin Klevorick, Richard R. Nelson, and Sidney G. Winter. "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3 (1987), pp. 783-820.

Mansfield, Edwin, Mark Schwartz, and Samuel Wagner. "Imitation Costs and Patents: An Empirical Study," *Economic Journal*, 91 (1981), pp. 907-18.

Merges, Robert P., and Richard R. Nelson. "Market Structure and Technical Advance: The Role of Patent Scope Decisions," in Thomas M. Jorde and David J. Teece, eds., *Antitrust, Innovation, and Competitiveness.* New York: Oxford University Press, 1992.

Miller, Arthur R. Statement before the Subcommittee on Patents, Copyrights, and Trademarks of the Senate Committee on the Judiciary, quoted in Semiconductor Chip Protection Act of 1984, S. Rpt. 98-425 on S. 1201 (May 2, 1984).

Nakamura, Leonard. "Intangibles: What Put the *New* in the New Economy?" Federal Reserve Bank of Philadelphia *Business Review,* July/August 1999.

O'Donoghue, Ted. "A Patentability Requirement for Sequential Innovation," *RAND Journal of Economics,* 29 (1998), pp. 654-79.

Pakes, Ariel, "On Patents, R&D, and the Stock Market Rate of Return," *Journal of Political Economy*, 93 (1985), pp. 390-409.

Rauch, John G. "The Realities of Our Times: The Semiconductor Chip Protection Act of 1984 and the Evolution of the Semiconductor Industry," *Journal of the Patent and Trademark Office Society,* 75 (1993), pp. 93-124.

Risberg, Robert L., Jr. "Five Years Without Infringement Litigation Under the Semiconductor Chip Protection Act: Unmasking the Spectre of Chip Piracy in an Era of Diverse and Incompatible Process Technologies," *Wisconsin Law Review,* 24 (1990), pp. 241-77.

Schaafsma, Paul. E. "Patent Mapping: A Graphical Model of Patent Rights," *Journal of the Patent and Trademark Office Society,* 77 (1995), pp. 953-71.

Schankerman, Mark. "How Valuable Is Patent Protection? Estimates by Technology Field," *RAND Journal of Economics,* 29 (1998), pp. 77-107.

Schankerman, Mark, and Ariel Pakes. "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period," *Economic Journal* 96 (1988), pp. 1052-76.

Sobel, Gerald. "The Court of Appeals for the Federal Circuit: A Fifth Anniversary Look at Its Impact on Patent Law and Litigation," *American University Law Review,* 37 (1988), pp. 1087-1139.

Szczepanski, Steven Z. "Licensing or Settlement: Deferring the Fight to Another Day," *AIPLA Q.J.*, 15 (1987), pp. 299-323.

CASES CITED

Hotchkiss v. Greenwood, 11 How. 248, 13 L.Ed. 683 (1851).

Graham v. Deere, 383 U.S. 101 (1966).

Simmons Fastener Corporation v. Illinois Tool Works, 739 F.2d 1573 (Fed. Cir. 1984).

Stratoflex, Inc. v. Aeroquip Corp., 713 F.2d 1530 (Fed. Cir. 1983).