Toll have financed highway infrastructure since the Roman Empire. Although toll revenue often may be used for other purposes, many believe that financing highway infrastructure is the main function of tolls. Tolls for congestion pricing are different. They generate revenue, but with the intent of changing travel behavior to make more efficient use of the transportation system, by shifting some drivers to less congested periods, or to other modes, routes, or shared-ride vehicles, so that the traffic flows more freely.

Most products and services supplied in the marketplace rely on pricing to align demand with supply. If demand exceeds supply, prices will rise, and some customers will choose not to buy. Highways, however, are not priced this way, and the prices on the few facilities that are tolled are seldom allowed to vary according to changes in demand.

Motivations for Congestion Pricing
From the beginning of automobile travel, the U.S. system of paying for highways has relied largely on motor fuel taxes, excise taxes, sales taxes, and tolls. Motorists have grown accustomed to these methods of paying for the cost of building, maintaining, and operating highways. A system that sets a price for highway use as a way of reducing congestion would be a major change, potentially affecting where people live and work, locate businesses, and socialize.

Until recently, the congestion pricing of highways mostly had been an academic concept, because the necessary technology did not exist. Recent advances in electronic toll collection, however, have prompted greater interest in congestion pricing. Yet when the topic is discussed in public policy arenas—from state legislatures and governors’ offices to the radio and blogs—the motivations often differ from the efficiency concerns that interest academics.

Reducing congestion during peak periods and improving travel time reliability are important moti-
vations, but so are encouraging transit use, reducing vehicle emissions and energy consumption, and especially providing a funding source for transportation programs and projects. Some may view a system of congestion pricing as a means of changing urban form and promoting regional economic development.

Pricing Signals and Congestion
According to basic microeconomic theory, the demand for a good is directly responsive to the price of the good. If the supply of a good is fixed, the prices can be raised when demand peaks. Examples of time-based pricing include airline tickets on holiday weekends, daytime cell phone use, and midday electricity use. In each case, customers who cause the peak congestion must pay a premium, while users who are willing to purchase at off-peak times—when the resource is less scarce—pay less.

Moreover, the higher prices signal that additional investment in production capacity may be profitable. An airline will raise prices on a popular route to manage demand for the limited number of seats, but at some point the airline may decide that adding another flight to the route would be profitable. Some proponents emphasize that a key advantage of congestion pricing is that it would identify places in the transportation system that warrant investments in more capacity.

But the mostly private users of roads and the mostly public suppliers of road capacity do not receive pricing signals. State and local governments largely have been responsible for the building and operation of roads in the United States, so that the highway system is perceived as a public good. In an economic sense, no one can be excluded from the use of a public good, and one person’s use of the good does not diminish its value to others.

Under conditions of high volume, however; one additional vehicle entering a road system may cause the flow of traffic to slow, creating congestion and delay for others—so that highways are not strictly public goods. When there is no charge for entry, motorists do not consider that they are imposing a cost on others; the resulting market failure is known as congestion.

Pricing for Social Efficiency
Nobel Laureate William Vickrey advanced the idea of congestion pricing during the 1950s and 1960s. Implementation of his ideas was impractical, however, because of the primitive nature of toll collection. By the 1990s, tolls could be collected electronically without stopping vehicles, and several toll roads operated without tollbooths. These technological developments renewed interest in road congestion pricing.

Technology was not the only impediment to congestion pricing, however. Changing the status quo of highway funding and use would create winners and losers. Vickrey had recognized that this would need to be addressed if congestion pricing were to be implemented.

Although congestion pricing can produce an economically efficient solution to road congestion—so that society as a whole gets the most value out of its expenditures—travelers may be made worse off on average if the revenues are not used to increase mobility. Consider these examples:

- Some motorists will choose to pay the congestion charge and continue to use the same road at the same time as before. They will pay more as a consequence. These motorists place a high value on time, and some will be better off because of the travel time savings. Because all motorists pay the charge by choice, all presumably are better off doing this than taking advantage of other options, such as driving at a different time or on a different route, taking transit, or forgoing the trip altogether. On net, however, many still will be worse off than they were before congestion pricing.

- Those who choose an alternative road or mode or who cancel the trip are worse off, because they are not traveling when, where, or how they want.

- Those who were using other routes and modes before may be worse off, because new travelers now may be competing with them for the capacity. If congestion pricing increases the throughput of the priced highway, however, congestion on other routes and modes may be reduced. Only a detailed analysis can reveal the traffic impacts.

The distribution of winners and losers, and whether society as a whole is better off with congestion pricing, will depend on several factors, including who pays the tolls, the net effect on travel conditions, how the revenue is spent, and changes in other areas of concern, such as fuel consumption, emissions reduction, and safety.

The current system of financing highways through fuel taxes and vehicle fees creates its own winners and losers, but the social fabric has been
Getting Prices Right

Congestion pricing can be carried out in many ways. One option is to price one or more lanes on a freeway, offering patrons a higher level of service on the tolled lanes. Another option is to price an entire road or collection of roads. A third approach is to establish prices for access to—or travel within—all roads in a specified zone, such as a cordoned central business district. A fourth is to price the entire roadway system.

Congestion pricing on a large scale requires extensive knowledge about congestion levels on all parts of the transportation system simultaneously and an understanding of how each driver’s decision to embark on a trip will affect system congestion. Vickrey pointed out that the decision to travel is made at the beginning of a trip, but the impact of the travel was felt along the entire route and persisted after the trip because of the nature of bottlenecks.

In theory, under congestion pricing, a highway authority must anticipate economically efficient prices; communicate the prices to travelers, who then decide how to respond; and adjust the prices according to the responses—all in real time. Contemporary high-occupancy toll (HOT) lanes demonstrate how this works—prices are set dynamically, based on the traffic level in the priced lanes, and are changed frequently to maintain optimum traffic flow. Extending pricing beyond a single, limited highway corridor to a longer corridor or to a broader system of priced roadways introduces additional technical and political complexity.

Obtaining an economically efficient outcome is only part of the equation. Perceptions of fairness are another, evident in establishing urban transit fares. Transit systems often charge a flat rate, regardless of time of day or of distance traveled. Although some systems—like the Washington, D.C., Metro—charge per distance traveled, with higher prices during peak hours, most systems have constant rates all day. No system charges higher rates on more congested or more popular routes—although the higher rates may be economically efficient. Many travelers would consider this treatment to be unfair by a government-run system.

One option that perhaps is more practical than setting a different price for each minute of the day and each road on the system is a simplified system of user charges based on the time of day, type of road, and general location—for example, central business district, suburb, or rural area. After political compromises, however, the resulting system may not be the most economically efficient but, if done well, still would be more efficient than the status quo.

Effects on the System

Unless all roads are priced, motorists will have opportunities to shift travel to other parts of the system to avoid the charges. These motorists will incur the cost of using a less appealing route or mode. In addition, motorists who previously used the alternate route may experience the negative effects of higher traffic volumes and possibly more congestion. But because freeways carry so much traffic, pricing freeway use may instead have a positive impact on the system—a net win for society. Still, travelers on the priced freeway will be winning at the expense of those who no longer use the facility and of those who now must share the nonpriced roads with the displaced traffic.

The extent to which nonpriced roads will experience an increase, a decrease, or no change in congestion will depend in part on whether some of the revenue can be used to enhance the roads’ capacity and whether improved operations on the priced facility allow higher throughput. The optimum flow on a highway occurs when vehicles travel at about 45 miles per hour. When demand exceeds a certain point, speeds drop precipitously, allowing less throughput despite higher demand. This paradox is borne out daily in stop-and-go conditions on freeways. Therefore if pricing can manage demand to maintain optimal travel speeds, throughput may increase in other parts of the system.
Distributing the Revenue
Distribution of public revenue is an age-old political issue. Concerns about equity and building political support for the congestion pricing concept may create a strong temptation to use revenues to compensate the losers and to spread benefits to favored groups. This may be the only practical way to build support for the concept and still yield a net positive contribution to society.

Examples of potential revenue uses include the following:

- Investing in transit improvements in the affected area;
- Subsidizing improvements to the nonpriced part of the highway system—for example, to parallel arterials;
- Rebating motor fuel taxes;
- Reducing general taxes, such as income or property taxes;
- Awarding grants to affected communities; and
- Allocating toll credits to all drivers, which some may use in full or trade-in any surplus for cash or tax rebates.

University of California planners King, Manville, and Shoup have suggested that using congestion pricing revenue to compensate groups may make good sense (1). They argue that those who perceive themselves to be losers from congestion pricing are likely to form a strong political resistance to the concept. The targeted distribution of revenue would allow these groups to perceive themselves as winners and give their support to congestion pricing.

Gaining Practical Experience
If the technology necessary for road pricing had been available at the beginning of the motor vehicle era, and if it had been used to capture the full marginal social cost of driving, communities might have developed differently. Urban areas perhaps would be more compact, with greater use of public transportation.

But with little real-world experience of congestion pricing, projecting the outcome is difficult. Most of the pricing experience in the United States has involved minor adjustments in tolls on toll facilities and HOT lanes. Lessons from these might not translate well to other types of pricing, such as the zone-based pricing that has been tried overseas in environments much different from those of the United States.

The complexity of the technical and political aspects of congestion pricing suggests the need to approach these new ideas with caution, by conducting tests and undertaking analyses that are transparent, comprehensive, and methodologically correct. Practitioners also must respect the concerns of the affected constituencies.

The testing of new congestion pricing ideas continues in the United States. The articles in the rest of this issue report on what has been learned so far and on new ideas and insights that are emerging from experiments.

Acknowledgments
The author gratefully acknowledges David Williams of the Oregon Department of Transportation for identifying the need for a plain-language treatment of this topic, and for his review of an earlier version of this work, and the research assistance of Evan Enarson-Hering of Cambridge Systematics and Alexander Heil, formerly of Cambridge Systematics.

Reference

Additional Resources
Lindsey, R. Do Economists Reach a Conclusion on Road Pricing? The Intellectual History of an Idea. 2006.
Only 15 years ago, congestion pricing was in its infancy in the United States. A private pricing project in Orange County, California, was in the planning and design phase, and several other projects were in planning stages with support from the Federal Congestion Pricing Pilot Program. Since then, the Orange County project started operations, and the federal program has supported more than 50 congestion pricing projects and studies in more than a dozen states, with more than 20 projects now operating. The projects implemented or under investigation include the pricing of high-occupancy vehicle (HOV) lanes and new express lanes, the conversion of toll or toll-free facilities to variable tolls, and application of congestion pricing within a region.

**HOT Lane Conversions**

The most common application of congestion pricing in the United States involves the conversion of HOV lanes into high-occupancy toll (HOT) lanes, which allow drivers of vehicles that do not meet the occupancy requirements to buy-in to the lanes by paying a toll that varies by time of day or by the level of congestion or demand. A rationale for converting to HOT lanes is that the HOV lanes are underused, despite increased congestion on the adjacent main lanes.

Electronic tolling ensures high-speed access to the restricted lanes and the setting of rates to maintain the free flow of traffic. In this way, HOT lanes provide travelers facing traffic congestion with new choices. Motorists can choose to continue on the main untolled lanes at the available speed, or pay a toll to gain access to a high-speed alternative, or meet the minimum occupancy requirements and use the high-speed lanes for free. Some major HOT lane conversion projects are summarized in Table 1 on page 9.

The earliest HOT lane conversion was the I-15 FasTrak facility, which opened in 1996 in San Diego, California. The FasTrak tolls vary with the level of demand to maintain free-flowing traffic. Fees can vary as often as every 6 minutes, typically in 25-cent increments. Message signs at the entrance inform motorists of the current fee. Tolls typically vary between $0.50 and $4.00, but can reach $8.00 during peak periods. The average toll rate is approximately $1.25 and seldom exceeds $4.00. Savings in travel time average 20 minutes per journey.

Another early example is I-10 in Houston, Texas. The freeway’s HOV lane, which required a minimum of three occupants (HOV-3), was converted in 1998 to a HOT lane. Drivers of two-occupant vehicles can buy-in to the lanes during the times that three-occupant vehicles have access for free. This QuickRide program increased HOV-2 volume by 40 percent, while the HOV-3 volume decreased by less than 3 percent. The total volume on the HOT lane increased by 21 percent during the morning peak. The average speed on the general-purpose lanes was 25 miles per hour (mph) but exceeded 55 mph on the HOT lane, yielding a 17-minute time savings for the 13-mile trip.

More recent examples of HOT lane conversions include I-25/US-36 in Denver, Colorado, started in 2006; the MnPASS I-394 project in Minneapolis, Minnesota, begun in 2005; and two that opened in