Mini-Course on Market Design:
Matching, Carpooling, and Congestion Pricing

Michael Ostrovsky
Stanford GSB

IWGTEA 2026
São Paulo, Brazil

Lecture 1: Two-sided matching with transfers

Lecture 2: Vickrey (1969) model of traffic congestion. Congestion pricing, carpooling, and commuter welfare.

Lecture 3: Bringing it all together: self-driving cars, carpooling, and congestion pricing

Lecture 4: Empirics: Congestion Pricing in New York City

Textbook Solution - Congestion Pricing

- Pigou (1920), Knight (1924), Vickrey (1969)
- Basic logic:
  - “By charging people higher tolls during rush hour, tolls can operate as a congestion tax that channels people with more flexible schedules to drive at different times”
  - “The toll can also encourage people to take different modes of transportation such as the train and public transit, thereby creating cleaner air and benefiting the environment”

  (From Ayres, I., S. Berry, K. Gillingham, and B. Nalebuff (2019). *How we learned to love the toll.* Hartford Courant.)
- “Why You’ll Love Paying for Roads That Used to Be Free” (guest post on the *Freakonomics* blog, reprinted in Greg Mankiw’s Microeconomics textbook).

Problem

- Many people clearly don't seem to love the idea of "paying for roads that used to be free."

- E.g., as just one illustration (out of countless possible ones), consider the ongoing fights over congestion pricing in New York City.

Governor Murphy Announces Filing of Lawsuit to Block New York's Proposed Congestion Tolling Program

07/21/2023

Signs Bipartisan Legislation to Confront the Unjust Taxation of New Jerseyans by Other States

FORT LEE – Governor Phil Murphy today announced the filing of a lawsuit against the U.S. Department of Transportation (USDOT) and the Federal Highway Administration (FHWA) to block the ill-conceived congestion pricing plan put forward by the Metropolitan Transportation Authority (MTA) and New York City and State agencies. In the suit, filed in the U.S. District Court for the District of New Jersey, the State of New Jersey argues that the USDOT and the FHWA violated the National Environmental Protection Act, which requires a full environmental impact review for projects of this projected impact and scope, as well as the Clean Air Act.

Since New York and the MTA first revealed their proposed congestion pricing scheme, Governor Murphy has remained a staunch advocate for the New Jersey commuters, transportation agencies, businesses, and residents who would suffer as a result of this unreasonable and unprecedented proposal.

Kreindler (2024), "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications"

- "I study the peak-hour traffic congestion equilibrium in Bangalore. [...] field experiment with congestion pricing policies"
- "Commuter responses in the experiment reveal moderate schedule inflexibility"
- "My policy simulations with endogenous congestion indicate that optimal congestion charges would lead to a small reduction in travel times, and small commuter welfare gains."
- "Overall, these results suggest limited commuter welfare benefits from peak-spreading traffic policies in cities like Bangalore."

This lecture:

- Commuters are not necessarily wrong in their opposition to congestion pricing: even if the overall impact of congestion pricing on social welfare is positive, its impact on the welfare of *commuters* may not be (Vickrey, 1969).

- Self-driving cars won't save us on their own (Small, 1992).

- The combination of congestion pricing *and* convenient carpooling (broadly understood) can substantially improve all commuters' welfare, *even if the government burns all the revenue from tolls*, and even in the world in which convenient carpooling by itself wouldn't make much of an impact.
  - Note: while in our model "carpooling" involves only two people and no professional driver, the intuition would extent to "intermediate public transportation" that would endogenously bridge the gap between full-scale public transportation and solo driving.

Baseline: the canonical "bottleneck" model of Vickrey (1969)

- $N = 7200$ travelers crossing a bottleneck during their morning work commute. No carpooling (for now).

- Their ideal times for passing the bottleneck are distributed uniformly between $\underline{t} = 8{:}00$ AM and $\bar{t} = 9{:}00$ AM (120 per minute).

- Capacity constraint: only $v = 60$ cars per minute can pass the bottleneck.

- Value of time (cents per minute):

w_h (home)	w_q (traffic)	w_j (job, late)	w_p (job, early)
20¢	0	40¢	10¢

Equilibrium Analysis

Let $q(t)$ denote the amount of time that a commuter will need to spend waiting in a queue if he wants to *leave* the bottleneck at time t. (Thus, he will need to arrive there and start queuing at time $t - q(t)$.)

Consider a commuter who arrives at work late. If he shifts his arrival time by ε minutes (i.e., arrives even later), his value of time at work is reduced by εw_j, his value of time in the queue is increased by $\varepsilon q'(t) w_q + o(\varepsilon)$, and his value of time at home is increased by $\varepsilon(1 - q'(t))w_h + o(\varepsilon)$.

The total change is therefore

$$-\varepsilon w_j + \varepsilon q'(t) w_q + \varepsilon(1 - q'(t))w_h + o(\varepsilon).$$

First-order condition:

$$-w_j + q'(t) w_q + (1 - q'(t))w_h = 0,$$

or, equivalently,

$$w_h - w_j = q'(t)(w_h - w_q).$$

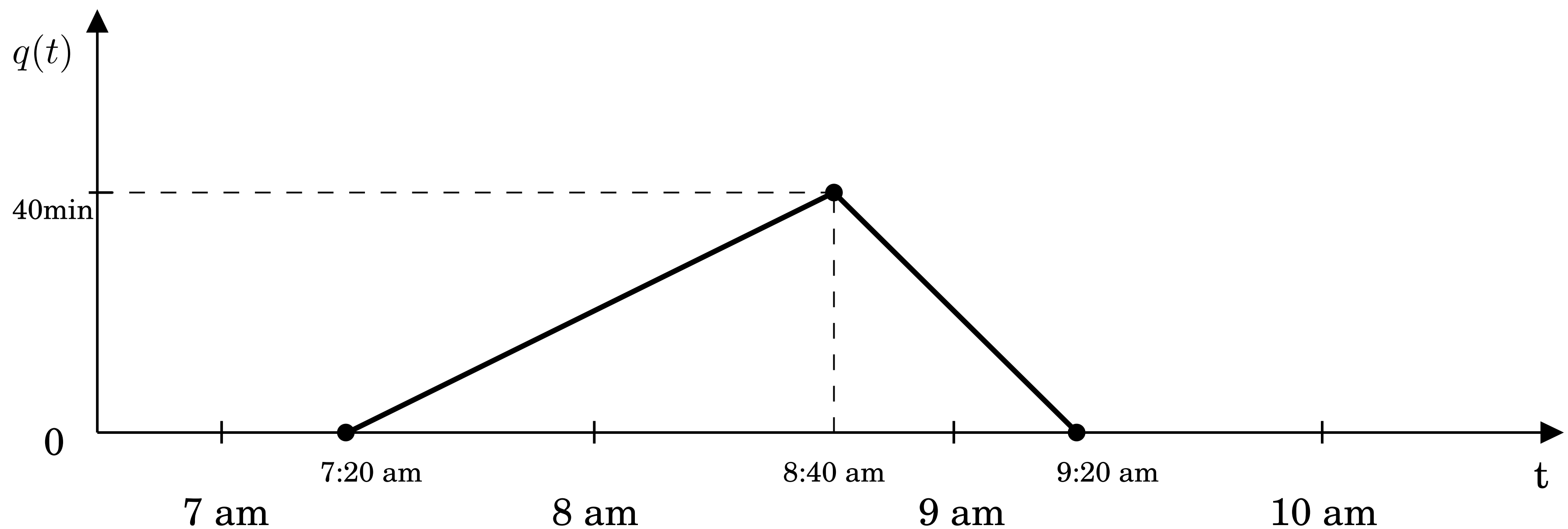
Thus, for a commuter who arrives at work later than his ideal time and passes the bottleneck at time t, we have

$$q'(t) = \frac{w_h - w_j}{w_h - w_q} = -1. \tag{1}$$

Similarly, for a commuter who arrives at work earlier than her optimal time and passes the bottleneck at time t, we have

$$q'(t) = \frac{w_h - w_p}{w_h - w_q} = 0.5. \tag{2}$$

These equations allow us to characterize the overall shape of function $q(t)$ and the corresponding pattern of traffic. Traffic starts building up at 7:20 AM, reaches its peak for commuters leaving the bottleneck at 8:40 AM, and then starts going down, until 9:20 AM, when it ends. Commuters leaving the bottleneck at 8:40 AM get to work at their ideal time, but also spend the longest in traffic: $q(\text{8:40 AM}) = 40$ minutes. A third of commuters (those leaving the bottleneck after 8:40 AM) will be late for work, and two thirds (those leaving the bottleneck before 8:40 AM) will be early.



$q(t)$
40min
0
7:20 am
8:40 am
9:20 am
t
7 am
8 am
9 am
10 am

The total amount of time spent in the queue by all commuters is 144,000 minutes (20 min per commuter). The total delay relative to the ideal arrival time (for those who arrive at work late) is 24,000 minutes (10 min per commuter who is late). Finally, for the commuters who arrive earlier than the ideal arrival time, the average amount of time by which they are early is 20 minutes, for a total of 96,000 minutes.

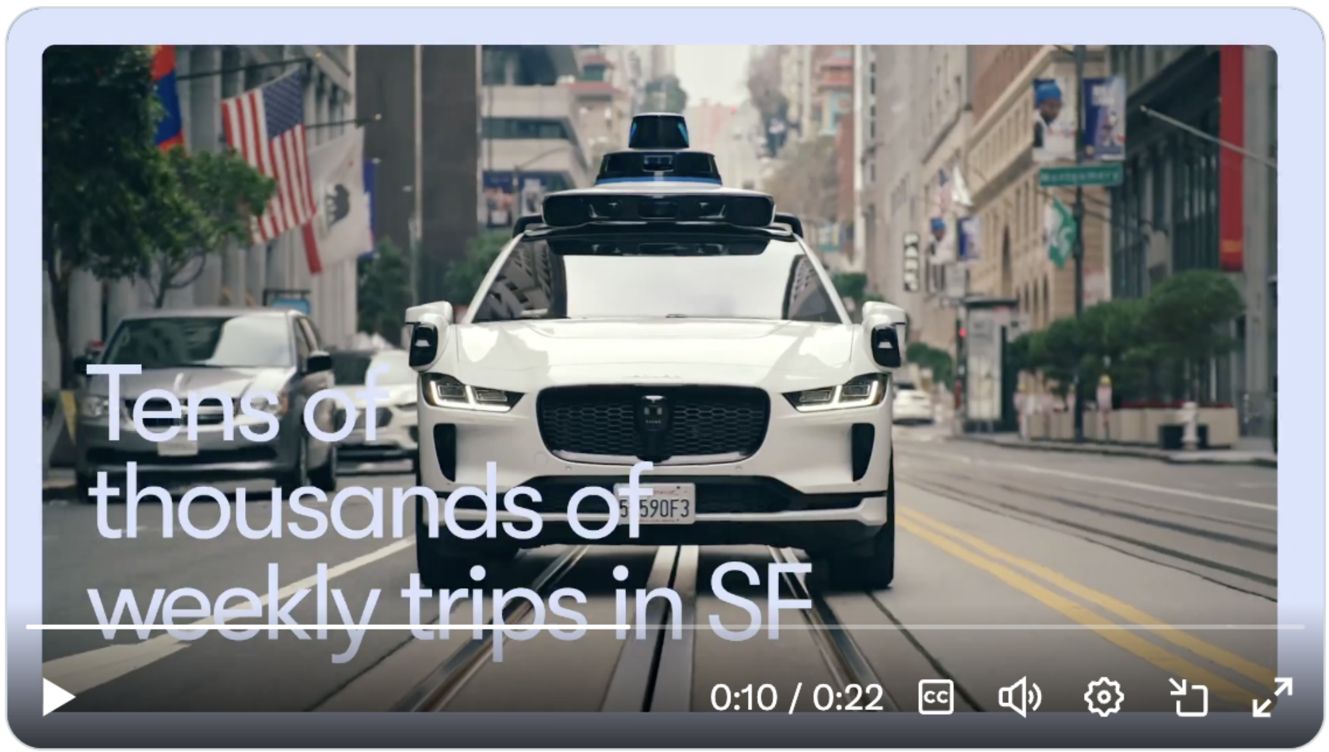
In monetary terms, this corresponds to an aggregate disutility of $\$28{,}800$ from waiting in queue (144,000 min at the disutility of $w_h - w_q = 20¢$) and $\$14{,}400$ from displaced arrival times (24,000 minutes at the disutility of $w_j - w_h = 20¢$ and 96,000 minutes at the disutility of $w_h - w_p = 10¢$).

The disutility from the commute is the worst for commuters whose ideal time for passing the bottleneck is 8:40 AM. Their disutility is $\$8$. The disutility is the lowest for those whose ideal times for passing the bottleneck are 8:00 AM or 9:00 AM. Their disutility is $\$4$.

Detour (but an important one): Won't self-driving cars save us?

Waymo
@Waymo

We're excited to announce that Waymo One is now open to SF riders — no more waitlist or invite codes! This is a key milestone in our mission to be the world's most trusted driver, & we're thrilled to get more of you where you're going safely. Ride today. waymo.com/blog/2024/06/w...



11:00 AM · Jun 25, 2024 · **912.5K** Views

Making being stuck in traffic less painful (Small, 1992)

Suppose the value of time in traffic goes up (say, from 0 to 4¢).

w_h (home)	w_q (traffic)	w_j (job, late)	w_p (job, early)
20¢	4¢	40¢	10¢

The equations characterizing the shape of the "waiting" function $q(t)$ then become:

For a commuter who arrives at work later than his ideal time, we have

$$q'(t) = \frac{w_h - w_j}{w_h - w_q} = -\frac{5}{4}. \tag{3}$$

For a commuter who arrives at work earlier than her optimal time, we have

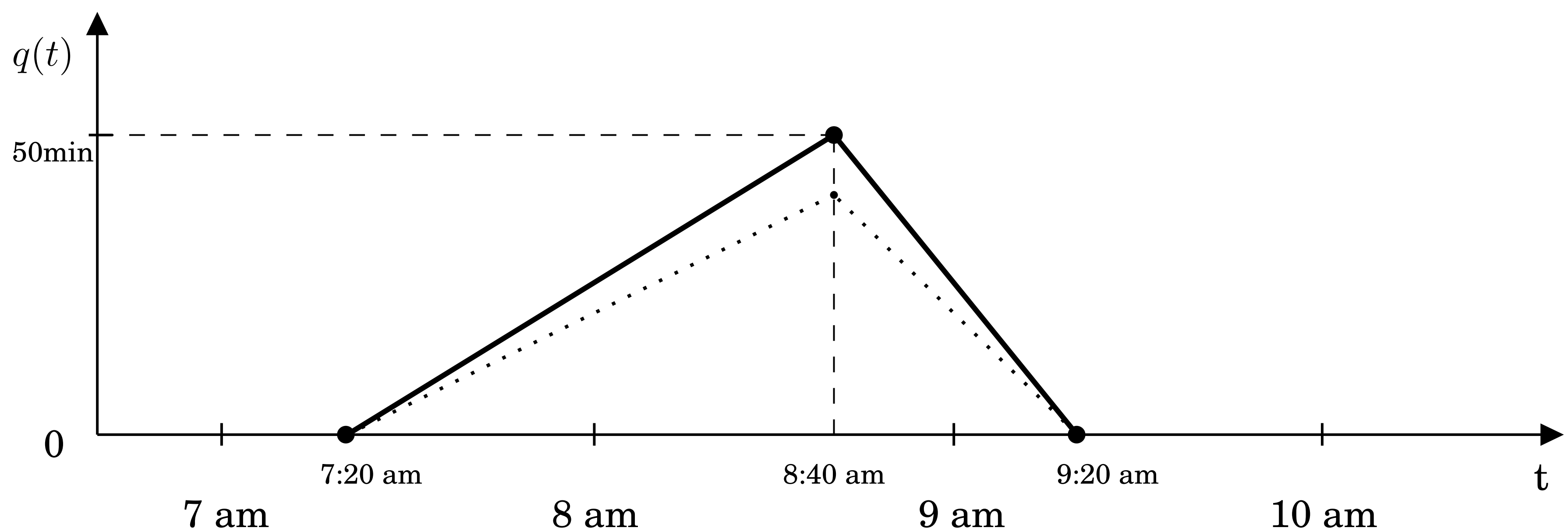
$$q'(t) = \frac{w_h - w_p}{w_h - w_q} = \frac{5}{8}. \tag{4}$$

What does the equilibrium look like now?

Traffic starts building up at 7:20 AM, reaches its peak for commuters leaving the bottleneck at 8:40 AM, and then starts going down, until 9:20 AM, when it ends. Commuters leaving the bottleneck at 8:40 AM get to work at their ideal time, but also spend the longest in traffic: q(8:40 AM) = 50 minutes. A third of commuters (those leaving the bottleneck after 8:40 AM) will be late for work, and two thirds (those leaving the bottleneck before 8:40 AM) will be early.

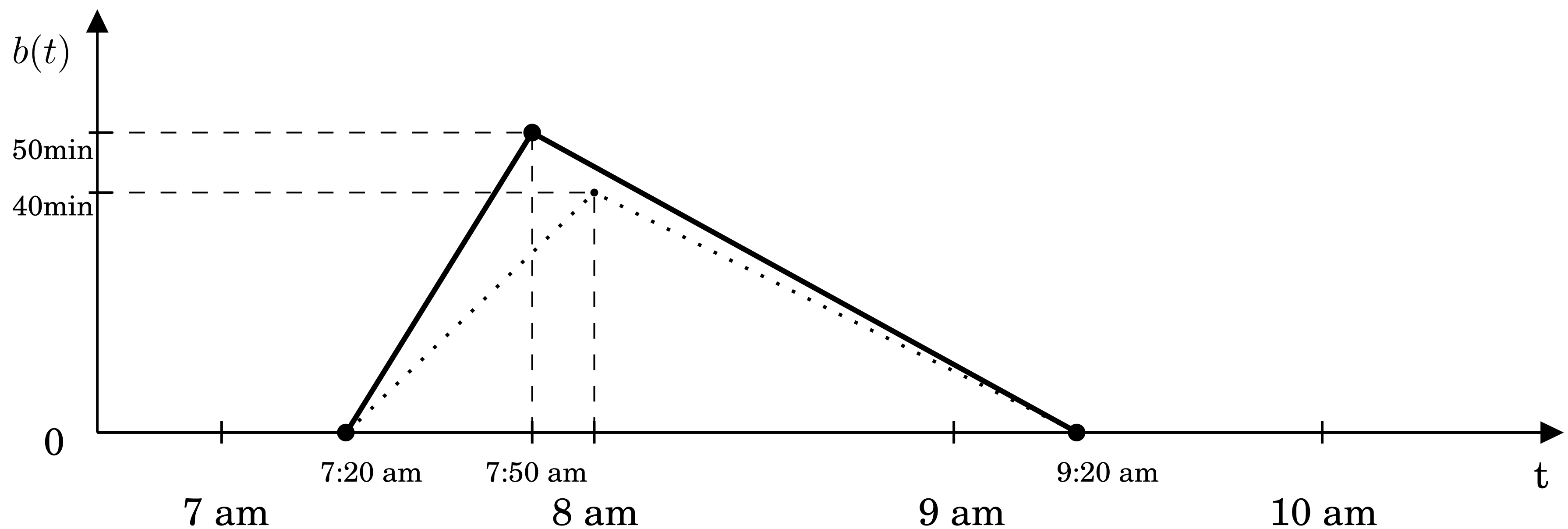
The total amount of time spent in the queue by all commuters is 180,000 minutes (25 min per commuter). The total delay relative to the ideal arrival time (for those who arrive at work late) is 24,000 minutes (10 min per commuter who is late). Finally, for the commuters who arrive earlier than the ideal arrival time, the average amount of time by which they are early is 20 minutes, for a total of 96,000 minutes.

In monetary terms, this corresponds to an aggregate disutility of $\$$28,800 from waiting in queue (180,000 min at the disutility of $w_h - w_q = 16$¢) and $\$$14,400 from displaced arrival times (24,000 minutes at the disutility of $w_j - w_h = 20$¢ and 96,000 minutes at the disutility of $w_h - w_p = 10$¢).



$q(t)$
50min
0
7:20 am
8:40 am
9:20 am
t
7 am
8 am
9 am
10 am

Traffic patterns



"Old" equilibrium (dotted line): $40 \times 60 = 2400$ cars in the queue at 8am, arrival rate 120 cars/minute before 8am, 30 cars/minute after 8am (recall that the outflow rate is constant at 60 cars/minute)

"New" equilibrium (solid line): $50 \times 60 = 3000$ cars in the queue at 7:50am, arrival rate 160 cars/minute before 7:50am, 26.67 cars/minute after 7:50am

Equilibrium with road pricing (still no carpooling; Vickrey, 1969)

Socially efficient outcome is to have all drivers pass through the bottleneck starting at 7:20 AM and ending at 9:20 AM, at a uniform rate of 60 cars per minute, with no queuing.

Question: What price schedule supports this outcome?

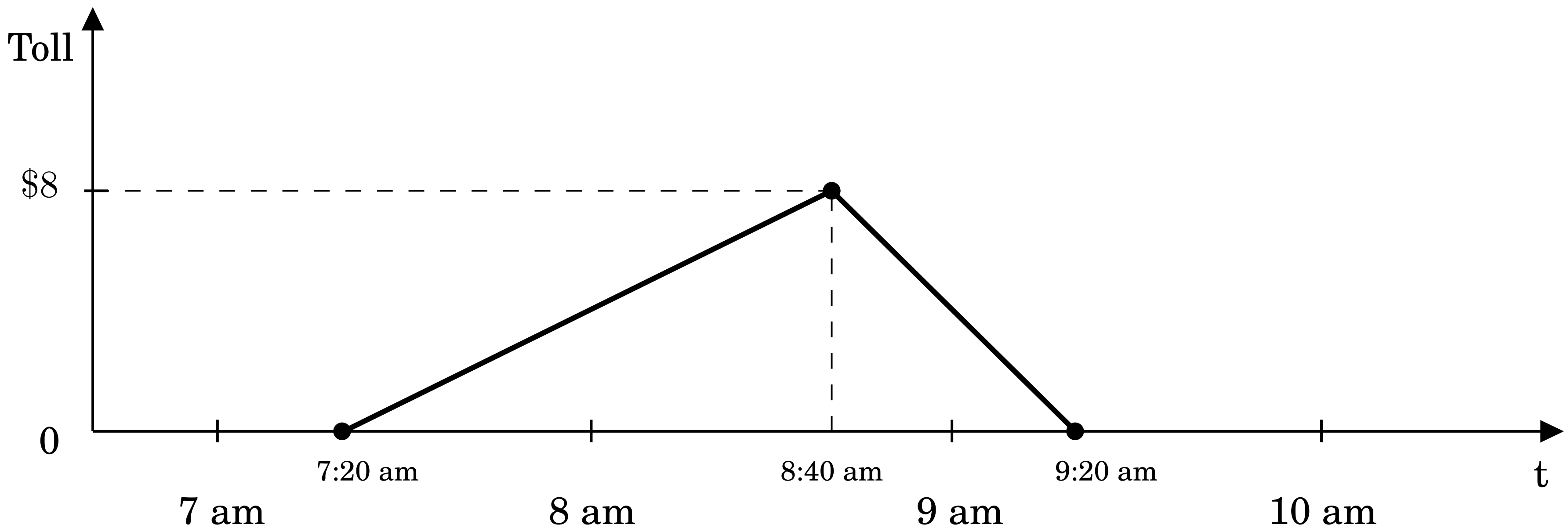
If $p(t)$ denotes the toll at time t, and a (late) commuter considers shifting his bottleneck passing time from t to $t+\varepsilon$, the change in his payoff will be $-w_j\varepsilon + w_h\varepsilon - p'(t)\varepsilon + o(\varepsilon)$. Thus, in the time interval in which commuters are late for work, the F.O.C. implies

$$p'(t) = w_h - w_j = -20¢.$$

Similarly, in the time interval in which commuters are early, we have

$$p'(t) = w_h - w_p = 10¢.$$

Answer: The price for passing the bottleneck starts at zero at 7:20 AM, rises linearly to \$8 at 8:40 AM, and then drops linearly back to zero at 9:20 AM.



Toll
$8
0
7:20 am
8:40 am
9:20 am
t
7 am
8 am
9 am
10 am

This outcome eliminates traffic congestion, but still has serious shortcomings:

- None of the commuters are better off after the introduction of congestion pricing: each of them simply pays in dollars what he or she used to pay in time!
  - Redistribution is a hard, politically charged issue.

- Since all commuters arrive at work at the same times as in the original "congested" base case, the costs of suboptimal arrival times are not eliminated.

We now introduce the possibility of carpooling

- Ostrovsky and Schwarz (2025), “Congestion Pricing, Carpooling, and Commuter Welfare.”

Equilibrium with carpooling (and no road pricing)

Two people per car, no distinction between the passenger and the driver.

Carpooling incurs on each of the carpooling partners a small disutility $\Delta > 0$, which enters their payoff functions additively.

No road pricing.

Equilibrium: identical to the one we analyzed first. Nobody carpools. Regardless of what everyone else is doing, an individual commuter's best response is to drive solo.

Carpooling + Congestion Pricing

Simple policy

Recall our assumptions about carpooling: two people per car, no distinction between the passenger and the driver, each of the carpooling partners incurs a small disutility $\Delta > 0$ vs. driving solo. What happens if we now introduce road pricing?

Simple policy: Charge a constant toll of 2Δ at all times (or a slightly higher toll to make incentives strict).

Equilibrium outcome: all commuters will carpool, all will arrive at work at their ideal time, and there will be no traffic congestion. In effect, while the capacity of the road measured in "cars" remains constant at 60 per minute, its capacity measured in "commuters" doubles to 120 per minute—without adding more lanes.

Question: can we do better?

Start with the outcome of the simple policy, and consider a small number $\varphi > 0$ of commuters passing the bottleneck right before 9:00 AM. Each one of them incurs a disutility of Δ from carpooling, and so the total disutility they incur is $\varphi\Delta$.

Suppose now that instead of having them carpool over the time interval $[9{:}00 \text{ AM} - \varphi/120, 9{:}00 \text{ AM}]$, we have them drive solo, over the time interval $[9{:}00 \text{ AM} - \varphi/120, 9{:}00 \text{ AM} + \varphi/120]$.

On average, each of these commuters is delayed by $\varphi/240$ minutes, at the cost of $w_j - w_h = 20$ cents per minute. Thus, their total cost of delay is $\varphi^2/12$ cents, which is less than or equal to $\varphi\Delta$ for all $\varphi \leq 12\Delta$ (where Δ is measured in cents).

The optimal φ maximizes $\varphi\Delta - \varphi^2/12$, and is thus equal to 6Δ.

Analogous calculation for commuters passing the bottleneck right after 8:00 AM; it is optimal to have $\psi = 12\Delta$ commuters drive solo instead of carpooling.

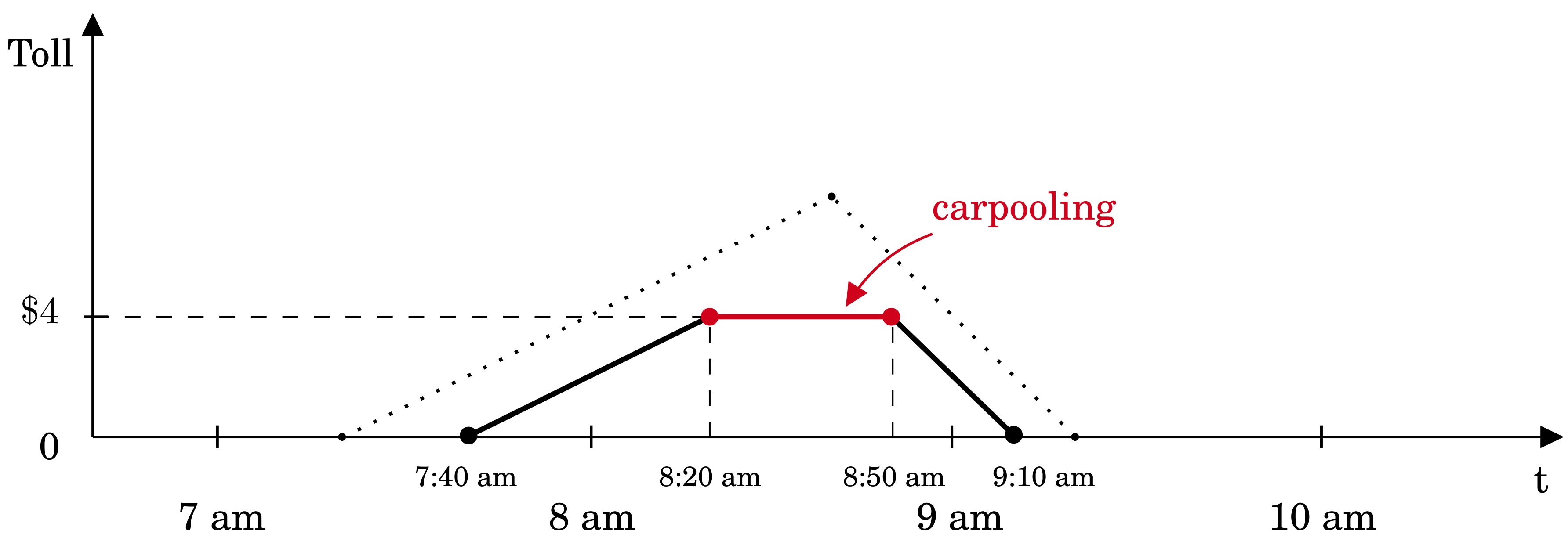
Equilibrium with road pricing and carpooling

Socially optimal commuting pattern:

- solo driving from (8:00 AM - $\Delta/10$) until (8:00 AM + $\Delta/10$),
- carpooling from (8:00 AM + $\Delta/10$) until (9:00 AM - $\Delta/20$), and
- solo driving from (9:00 AM - $\Delta/20$) until (9:00 AM + $\Delta/20$).

The price schedule supporting this outcome starts at zero at 8:00 AM - $\Delta/10$, rises linearly to 2Δ at 8:00 AM + $\Delta/10$, stays at 2Δ until 9:00 AM - $\Delta/20$, and then drops linearly back to zero at 9:00 AM + $\Delta/20$.

Optimal toll schedule for $\Delta = \$2$



Welfare comparison under different regimes

	Cost ($/morning)						
	Displaced Arrival	Waiting in Queue	Disutility from Carpooling	Total Social Cost	Total Commuter Cost	Social Cost per Commuter	Average Commuter Cost
No tolls, no carpooling	14,400	28,800	-	43,200	43,200	6.00	6.00
No tolls, improved comfort	14,400	28,800	-	43,200	43,200	6.00	6.00
No tolls, carpooling	14,400	28,800	-	43,200	43,200	6.00	6.00
Tolls, no carpooling	14,400	-	-	14,400	43,200	2.00	6.00
Simple tolls, carpooling, $\triangle$ = $3	-	-	21,600	21,600	43,200	3.00	6.00
Simple tolls, carpooling, $\triangle$ = $2	-	-	14,400	14,400	28,800	2.00	4.00
Simple tolls, carpooling, $\triangle$ = $1	-	-	7,200	7,200	14,400	1.00	2.00
Optimal tolls, carpooling, $\triangle$ = $3	8,100	-	5,400	13,500	35,100	1.875	4.875
Optimal tolls, carpooling, $\triangle$ = $2	3,600	-	7,200	10,800	25,200	1.50	3.50
Optimal tolls, carpooling, $\triangle$ = $1	900	-	5,400	6,300	13,500	0.875	1.875

Conclusion

- Autonomous transportation, road pricing, or carpooling are unlikely to solve our traffic woes *individually*.

- The combination of intelligent road pricing and convenient carpooling can substantially increase commuter welfare *even if the government burns all the revenue from tolls*.

- Key driving force: Without carpooling, introducing tolls simply replaces the time cost of traffic with the equivalent dollar cost, which does not benefit commuters. With carpooling, the monetary cost is fundamentally different from the cost in the form of time wasted in traffic: the former can be split between the carpooling partners while the latter cannot. As a result, monetary tolls create incentives to carpool that are absent under time costs, which in turn leads to an increase in road capacity and throughput (as measured in commuters, not cars), which subsequently leads to a decrease in the market-clearing price.