PROBING A TRAFFIC CONGESTION CONTROVERSY: A COMMENT*

Erik T. Verhoef

Department of Spatial Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. E-mail: everhoef@econ.vu.nl.

ABSTRACT. Ohta (2001) claims to have resolved a die-hard controversy on traffic congestion modeling by defining an inverse aggregate demand function that has traffic density as its argument—in Ohta's terminology the 'primitive term.' Using this demand function, Ohta shows that 'hypercongestion' may very well be an optimal stationary state. This contribution argues that at least if what road users demand is completed trips, and if time spent on the road while traveling implies a cost, then Ohta's approach is fundamentally flawed. Also the conclusion that hypercongestion can be optimal is no longer valid.

1. INTRODUCTION

In an article appearing in this issue of this journal, Ohta (2001) attempts to resolve an important controversy in the traffic congestion literature. The controversy concerns the fundamental question of how the demand and supply sides of a market for road use should be represented in order to arrive at a conceptually sound modeling framework for studying the efficiency of congested road usage with and without optimal pricing. Ohta's paper is one in a long tradition of articles, comments and author's replies (see Verhoef, 1999 for a review of the literature). This contribution will not pass without a comment. Specifically, I argue that if what road users demand is completed trips, and if time spent on the road while traveling implies a cost, Ohta's approach is fundamentally flawed. His remarkable conclusion that hypercongestion can be optimal is no longer valid.

2. DIFFICULTIES IN MODELING TRAFFIC CONGESTION IN A DEMAND-SUPPLY FRAMEWORK

The most basic model of traffic congestion considers a homogeneous stationary-state flow of identical vehicles on a single one-lane road of constant capacity that serves travellers from one single origin, the road's entrance, to one single destination, its exit. Congestion on this road results from the fact that if traffic density increases, speed falls. In a stationary state, traffic density D, speed S,

---

and traffic flow F are constant along the road, and constant over time. This is the type of situation considered by Ohta, and also the situation that I consider in my comment. It is certainly not the most realistic representation of traffic congestion. In reality, typically traffic congestion builds up and phases out over some finite (peak) period and occurs on a full network instead of on a single constant-capacity road. But this representation is simply the only type of situation that can be studied with a static model in which only single equilibrium values for S, D, and F prevail.

When the aim is to investigate the equilibrium and welfare properties of this basic model in greater detail, it is essential either to retain explicitly this set of assumptions (e.g., one entrance, one exit, one OD-pair, a single lane with constant capacity, stationary state traffic where S, D, and F do not change over time and do not vary along the road), or to explicitly replace one of these assumptions by another assumption to amend the model structure consistently to reflect this change in assumptions. For instance, as in Verhoef (1999, 2001) when he studies transitional paths towards stationary-state equilibria in a dynamic extension of the original static model. Unfounded—that is, not formally derived—expectations on the properties of the model, based on 'loose reasoning,' may often be inconsistent with the actual properties of the model itself. One example is the intuitively plausible, but in fact incorrect conjecture presumably made by Ohta; namely that a greater density on the road should always eventually somehow lead to a greater number of trips accomplished. In Section 3 I explain why this is not always the case. However, first I discuss the basic static model in some more detail.

In stationary-state traffic conditions—and only then—on a single lane, flow F (vehicles per hour) is identically equal to the product of speed S (kilometers per hour) and density D (vehicles per kilometer).[1] Under plausible assumptions about the technical relationship between D and S, this implies a backward-bending speed-flow function (equilibrium speed as a function of flow) That is, there exists a given maximum flow F^{Max}, and any positive flow level below F^{Max} can be realized as a normally congested stationary-state equilibrium with a relatively high speed and low density, and as a 'hypercongested' stationary-state equilibrium, where the opposite holds. When average generalized travel costs AC are assumed to increase with travel time and hence with the inverse of speed, a backward-bending AC curve can then be derived, showing average costs as a function of traffic flow. This curve is depicted as the AC $= t(F)$ curve in Ohta's Figure 1.

This more or less standard static model of road traffic congestion (discussed in greater detail in Verhoef, 1999) has been subject to fierce debate in the literature. In particular topics discussed include the 'choice' of equilibrium when a downward-sloping inverse aggregate demand function (with flow as the argument) has multiple intersections with the backward-bending AC curve, the

---

[1]In Verhoef (2001), it is shown that this identity does not hold in transitional phases where vehicles accelerate or decelerate.

use of 'flow' as the argument in the inverse demand function, and the stability of hypercongested equilibria (on the upper segment of the backward-bending AC curve). In Verhoef (1999), I attempt to summarize these discussions and to (briefly) review the relevant literature. In that paper, I hypothesize that the confusion surrounding the static model of traffic congestion at least in part stems from the following:

(1) Analysts that disagree over the correct argument to be used in the inverse demand function for road use often have distinctly different 'types of demand' in mind. Because the type of demand considered is typically left implicit, agreement is seldom (perhaps never) reached in the public debates that ensue. Implicitly, the type of demand considered is usually one of the following two archetypes:

   (a) *Continuous demand*, where it is assumed that a stable demand function for the use of the road exists, that does not shift over time, and that implies that any stationary-state equilibrium reached should be understood as a situation on the road that could in principle be everlasting. That is, no fluctuations in S, D, or F would ever occur in a stationary state with continuous demand, so Ohta's interpretation of stationary states (as I understand it from his footnote 2)—an ongoing cycle of alternating static peak equilibria and a static off-peak equilibria— would not be a true stationary state, and indeed cannot be analyzed with a static model in which speeds are constant over time.

   (b) *Peak demand*, where the demand function is assumed to represent the demand for using the road by a finite set of (potential) users, each of whom consider using the road for a single trip during the same (peak) period. So, for a static model, the stationary-state equilibrium considered is then assumed to apply for a finite duration T with the road being completely empty before and after this peak period.[2] Note that, because travel takes time, the peak period at the entrance of the road would start and end L/S time units earlier than at the road's exit in a static model (L denotes the length of the road).[3]

   Clearly, although in reality peak demand is the more relevant type of demand for most cases of traffic congestion a static model is actually much better fitted to describe the more unrealistic case of continuous

---

[2]Or at least, outside the peak period considered the road would be used only by users that do not belong to (nor interfere in any sense with) the finite set of potential users for peak usage. Only then can these other users indeed be ignored completely for welfare analyses for peak-period congestion.

[3]Also note that the static nature of the model implies that the unattractive and unrealistic assumption has to be made that the very first driver departing, who has an empty road in front of him, nevertheless chooses the equilibrium speed S consistent with the traffic density *behind* him. So upstream congestion (behind one's back) must be assumed to work like some sort of 'vacuum', in the sense that upstream congestion cannot be escaped by simply speeding up even if the road in front is completely empty.

demand. The most important reasons are that a static model can neither describe the changes in speeds that would normally be witnessed over an entire peak and that are required for a dynamic equilibrium to hold (in dynamic models normally defined as the constancy of the sum of travel delay costs, schedule delay costs,[4] and—if levied—tolls). Nor can it be used to determine the equilibrium duration T of the peak endogenously. However, the latter is essential also for a static model with peak demand. This is because for any equilibrium number of trips made by the finite set of potential users N it is the duration of the peak T that determines the implied average traffic flow F (assumed constant during a peak in a static model) through $F = N/T$. Hence for any equilibrium level of N, T directly determines the implied equilibrium average cost level $AC(F) = AC(N/T)$. Regardless of these considerations, the most important point here is that it should explicitly state which of these two types of demand is being considered.

(2) Motivated by the static nature of the basic model, most analysts do not consider *dynamic stability* when discussing the stability of proposed stationary-state equilibria (in particular hypercongested ones have been subject to debate). In fact, the quantity or price perturbations typically considered, assume simultaneous equal changes in the traffic flow at each point along the road and in the arrival rate of users at the road's entrance. Therefore, such stability analyses ignore the intrinsically dynamic causal (lagged) relationship between the arrival rate at the road's entrance and downstream flow levels. A dynamic stability analysis as defined in Verhoef (1999, 2001) instead considers perturbations in the arrival rate of users only, and treats the lagged impact on flow levels at each point along the road as the endogenous result of the combination of the initial stationary-state conditions and the said perturbation in the arrival rate. In Verhoef (1999, 2001), I demonstrate how important it is to study the dynamic stability of stationary-state equilibria. In particular, using the simplest plausible dynamic extensions of the standard static model, I show that hypercongested equilibria on the upper segment of a backward-bending AC curve are dynamically unstable in the sense just defined, and cannot be reached with any possible pattern of arrivals at the road's entrance when starting with an empty road, or even when starting with any stationary-state equilibrium other than the particular hypercongested equilibrium itself (see Section 4 below, for a more detailed discussion). This means that these equilibria cannot be reached as a stationary-state market equilibrium in a dynamic version of the same model, if the initial stationary state is not restricted by assumption to be

---

[4]Schedule delay costs are defined as the costs associated with arriving at the destination at an instant other than the most preferred arrival time.

identical to the equilibrium stationary state considered—the implicit but crucial assumption in purely static approaches.

Clearly, Ohta did not agree with my analysis, and proposes an alternative resolution to the controversy surrounding the static model of traffic congestion. He proposes using density as the argument (in his terminology—primitive) in the inverse aggregate demand and in the average and marginal cost functions. The same 'solution' to the problem was in fact proposed by Evans (1992). Using density, Ohta obtains monotonically upward-sloping average and marginal cost curves that intersect only once with a monotonically downward-sloping inverse demand function, which has traffic density as the argument. This suggests unique and stable equilibria, both in the case with and without (optimal) tolls, and thus appears to avoid some of the difficulties encountered with the standard formulation where flow is the argument in each of these functions. Moreover, according to Ohta's analysis, hypercongested equilibria may very well be optimal (in the sense of economically efficient); see region III in Ohta's Figure 4. I would say that the latter finding is the most remarkable conclusion Ohta draws, with far-reaching policy implications. It means that it may be more efficient to have a given number of users arriving over a given period after having had a trip with a relatively low speed, rather than having exactly the same number of users arriving over exactly that same time interval after having had a trip with a higher speed.

In this comment, I challenge the solution proposed by Ohta on two grounds. First, in Section 3, I argue that if what people demand is trips, the use of density as the argument in the inverse demand function—the single important assumption driving Ohta's results and the single change to the conventional representation he proposes—is fundamentally flawed. Next, in Section 4, I argue that hypercongested equilibria—possibly efficient according to Ohta—are dynamically unstable and cannot even be accomplished in the single road setting used by Ohta.

3. ARGUMENTS IN THE INVERSE DEMAND FUNCTION FOR ROAD USE WHEN PEOPLE DEMAND TRIPS

The crucial assumption in Ohta's analysis concerns the use of 'density' as the argument in the inverse aggregate demand function. It is this assumption that should be appraised critically in order to judge the value of Ohta's contribution. First recall that an inverse aggregate demand function should give the marginal willingness to pay as a function of the quantity consumed of the good considered, and that it is the inverse of the 'normal' demand function that shows the quantity demanded as a function of the price. For the case of road use, the price is usually expressed as the 'generalized costs', for static analyses typically the sum of the toll if levied, plus the monetary valuation of the travel time, plus all other costs of road use (e.g., fuel costs). The latter are often ignored for convenience, and are ignored in the remainder of this comment. The most straightforward way to judge which variable should be used as the argument in

the inverse demand function for stationary-state road use would thus be to identify which is the good that is being demanded.

Most analysts—but not Ohta—assume that what is demanded by the individuals using the single road is a trip from the single origin to the single destination. This is consistent with the view that it is 'getting from the origin to the destination' that yields benefits to the road users, and not primarily 'being on the road' per se. If it is indeed trips that are being demanded, a correct demand curve in the case of *peak demand* as defined in Section 2 should give the total number of trips demanded by the finite set of potential users during the peak considered, as a function of the generalized costs during that peak. This means that the inverse aggregate demand function should have the total number of trips accomplished during the peak N as the argument. As implied above, and as I argue in more detail in Verhoef (1999), a static model of traffic congestion typically cannot (or hardly—that is, unless a very peculiar type of schedule delay cost function, identified in Verhoef, 1999, applies) be used to describe such peak congestion accurately, and a dynamic approach would in fact be required. The reason is that schedule delay costs would typically increase with the deviation of a user's actual arrival time from the preferred arrival time, which means that travel times too, would have to vary during the peak to obtain an equilibrium in which the generalized costs (now the sum of schedule delay costs plus travel time costs) are constant during the peak. However, a static model cannot generate nor describe speeds that vary over time. Probably the best-known example of the required dynamic approach is Vickrey's (1969) model of bottleneck congestion that has been used and extended extensively over the last decades by Arnott, De Palma, and Lindsey (1993, 1998) and others. However, if one would be willing to make the assumptions that would make a static model appropriate for analysing peak congestion, then the correct argument in the inverse aggregate demand function used in such a static model would be the total number of trips made during the peak considered N and not density D.

In the case of continuous demand, the demand function should give trips accomplished per unit of time as a function of the generalized cost. This is so because a stable demand function cannot even be identified unless the unit of time is explicitly defined. The stability of the demand function over time namely implies that if x trips are demanded per hour for a given generalized cost level p, $2x$ trips are demanded per 2 hours, $24x$ trips per day, $100 \times 365.25 \times 24 \times x$ trips per 100 years, and $x/3600$ trips per second. Therefore, the total quantity of trips demanded for the generalized cost level p can be given only when the unit of time is defined.[5] The inverse aggregate demand function then has trips accomplished per unit of time as the argument. But when only stationary-state equilibria are considered, because a static model is used, the number of trips completed per unit of time (the outflow rate of vehicles at the road's exit) must

---

[5]Likewise, in a given stationary state, any total number of trips N would be completed with any positive flow F after FT time units.

be equal to the number of trips started per unit of time (the inflow rate at the road's entrance, L/S time units before completion of the trip), and must be equal to the constant flow of vehicles as it can be measured at any point along the road. This is why the analysts mentioned in Ohta's footnote 1 use flow as the argument in the inverse demand function. They consider stationary states of undetermined (possibly infinite) duration, where traffic flow by definition equals the rate at which trips are completed, and use a demand function that is assumed to be stable over time. They do not do this because they think this is the most realistic type of demand to consider for the analysis of traffic congestion, but rather because it is the only type of demand for which a static approach would offer an appropriate modeling framework in the first place.

So with continuous demand, the correct argument in the inverse aggregate demand function when people demand trips would be traffic flow F, again not density D. Thus, using flow as the argument in the relevant functions perfectly reflects that if in a stationary state analysis the unit of time is changed from, say, seconds to hours, the inverse aggregate demand, average, and marginal cost curves all rotate outward by a factor 3600, but that the identified market and optimal equilibria still represent the same stationary state. Marginal benefits, average costs, and marginal costs would keep the same equilibrium value, but the equilibrium number of trips made during one unit of time would increase by a factor 3600, and so would total benefits, total costs, and social surplus realized, as they can be identified as areas in the associated market diagram in a standard way. All this conforms to what should be the case for a conceptually sound stationary-state model.

Instead, Ohta suggests using density D as the argument in the inverse aggregate demand function. This would mean that people primarily demand 'being on the road simultaneously' rather than 'completing a trip'. Ohta's defence to this fundamental criticism of his approach is that this objection would "ignore[s] the fact that these two objects of demand are actually 'inseparable'" (p. 671). However, this is not a convincing argument. First, the representation with, for instance, flow as the argument in a stable inverse aggregate demand function— for continuous demand—does not ignore the said inseparability, because the generalized costs resulting from being on the road during the trip serve as the 'price' for the inverse aggregate demand function with flow as the argument. Second, why would the fact that these two 'objects of demand' (being on the road versus completing a trip) are inseparable mean that one would have to choose density, not flow (with continuous demand) or number of trips (with peak demand), as the argument in the inverse demand function? In other words, Ohta's remark is not a justification for using density.

Despite the close relation between flow and density, it is crucial to choose the correct argument in the inverse aggregate demand function, particularly for welfare analyses because the assessment of total benefits generated in different equilibria (together with total costs a major element of the overall welfare measure of 'social surplus') crucially depends on this choice. This is so because total benefits realized in a certain equilibrium are found by integrating the

inverse aggregate demand function from zero to the equilibrium quantity. The far-reaching consequences of assuming that people primarily demand to be on the road simultaneously, rather than assuming that they primarily want to complete a trip, is best illustrated by Ohta himself when he derives the result that an equilibrium with hypercongestion can in fact be the unique global optimum (region III in his Figure 4).

For the sake of the argument, ignore the issue of dynamic instability of such hypercongested equilibria here (I postpone this issue to Section 4). The result that hypercongestion can be optimal then correctly follows from the assumptions Ohta makes; the one involving density as the argument in the inverse demand function (p. 663) in particular. But can it in fact be optimal when what people demand is completed trips, and when every extra minute that a given user spends on the road adds to costs only, not to benefits? No, it cannot, and this is easy to demonstrate by taking one such hypercongested 'optimum' proposed by Ohta as a starting point. Denote the flow in this optimum as F, and consider the general case where $F < F^{Max}$. Denote the generalized cost level (exclusive of toll) in Ohta's optimum as $AC^h(F)$, where the superscript h stands for hyper-congested. Similarly, denote the average cost level associated with the same flow in a nonhypercongested equilibrium as $AC^n(F)$. As shown also by the $AC = t(F)$ curve in Ohta's Figure 1, the following inequality applies

$$AC^n(F) < AC^h(F)$$

First, consider a stationary state of undetermined—say infinite—duration that results because a stable demand curve exists (so consider the continuous demand case as defined in Section 2). Every unit of time F trips are completed (the outflow rate at the exit) in Ohta's optimum. The total cost incurred by these F users amounts to $Fc^h(F)$. Now what I would put forward as a superior equilibrium is the one in which the same flow is realized against the lower cost level $c^n(F)$. The same number of trips are completed every unit of time (namely F), but all users incur lower costs. The total benefits realized per unit of time depend on the number of trips completed per unit of time, and are therefore equal between the two equilibria. With lower total costs and equal total benefits for the trips completed per unit of time, the switch would yield a per-unit-of-time net welfare gain of

$$WG = F[AC^h(F) - AC^n(F)]$$

Note that my alternative to Ohta's optimum would require higher tolls (the required difference between Ohta's toll τ^h and my toll τ^n would be equal to $AC^h(F) - AC^n(F)$ to support the same number of trips demanded per unit of time in the two equilibria), but these can be ignored for the welfare comparison because they constitute a transfer only. Also note that for both Ohta's optimum and my alternative, every unit of time the F drivers with the highest willingness to pay (greater than or equal to $AC^h(F) + \tau^h = AC^n(F) + \tau^n$) according to the stable demand function would complete a trip, so that total benefits generated per unit

of time are indeed the same between both equilibria. It is only the difference in total costs that makes my alternative superior to Ohta's global hypercongested optimum.

What if Ohta (as he suggests in footnote 2) does not have continuous demand in mind, but in fact implicitly considers peak demand? Well, a similar reasoning holds. Suppose the hypercongested optimum put forward by Ohta entails departures from the origin over a duration of T time units, say from t_0 to $t_0 + T$. Denoting the hypercongested speed in Ohta's optimum as S^h and the length of the road again as L, this means that arrivals at the destination occur from $t_0 + L/S^h$ to $t_0 + T + L/S^h$, which also lasts T time units. The total number of trips made in Ohta's optimum amount to $N = TF$.

When switching to the nonhypercongested equilibrium with a speed S^n and a generalized cost level $AC^n(F)$, the same total number of trips N can be realized, arriving during exactly the same time interval but each departing $L/S^h - L/S^n$ time units later from the origin. Similar to the case of continuous demand, one can infer that because it will again be the N users with the highest willingness to pay that will make a trip in both equilibria, total benefits cannot have changed. Total costs however will have reduced by

$$WG = TF[AC^h(F) - AC^n(F)] = N[AC^h(F) - AC^n(F)]$$

Again, Ohta's unique global optimum can easily be improved upon, simply by raising the toll by so much that only the nonhypercongested equilibrium can be the market equilibrium.

So both for the case of continuous demand and for the case of peak demand, it is extremely simple to show that a stationary state with hypercongestion cannot be optimal if what people demand is trips, and if extra time spent on the road by an individual user implies an extra cost. Nevertheless the fact that Ohta proposes such equilibria as possibly globally optimal derives from his assumption that what people demand is being on the road simultaneously. So the fact that just as many users would complete a trip in the nonhypercongested and the hypercongested equilibria compared above, and all would have shorter travel times, is not necessarily a welfare improvement according to Ohta's formulation. This is because according to his demand function what people really want is being on the road simultaneously—rather than leaving it as quickly as possible—and this 'joy' is of course partly diminished when switching to a nonhypercongested equilibrium.

The fact that Ohta derives the possible optimality of hypercongestion correctly given the assumptions he makes (again postponing the discussion of dynamic instability until Section 4) nicely illustrates that his representation is fundamentally flawed if what people in fact demand is a completed trip, and if all time spent on the road in excess of the free-flow travel time contributes only to costs, not to benefits (these are the standard assumptions in models of traffic congestion). However, if Ohta wishes to consider a congested facility where people actually want to be (perhaps a beach, or a touristic road), instead of one

they want to leave as soon as possible (the typical case for a road used by commuters), his analysis would make more sense. However, this is not the context Ohta has in mind, judging by the arguments made in his paper.

4.   THE DYNAMIC INSTABILITY OF HYPERCONGESTED EQUILIBRIA

In the preceding section, I ignored the question of whether hypercongested equilibria can arise as an equilibrium in the first place in the assumed setting with one road with a single entrance and a single exit. That was because I wanted to show that even if hypercongested equilibria could arise in such a setting, they would never be optimal as claimed by Ohta. Now I turn to the second main point in my comment—hypercongestion cannot even arise as an equilibrium in a model with only one single road connecting one single origin to one single destination.

I provide detailed proofs of the dynamic instability of hypercongested equilibria in Verhoef (1999) for a model where traffic flow is treated as a continuum, and in Verhoef (2001) for the corresponding discrete version of it, where vehicles are discrete entities and where drivers behave according to a car-following equation that is consistent with a speed-flow function for stationary states as used in the standard static model of road traffic congestion. The pattern of speed, density, and flow over time for every point along the road is determined fully endogenously as the joint result of (1) the initial conditions on the road and (2) the time pattern of arrivals of new users at the road's entrance. It is not my intention to repeat the proofs of dynamic instability of hypercongestion here, but I do like to briefly summarize some of the discussion in those two articles, given the relevance for Ohta's paper.

What does it mean that hypercongested equilibria are dynamically unstable? It means that in a dynamic extension of the static model, hypercongested stationary state equilibria (on the upper segment of a backward-bending AC curve) can arise if and only if the same equilibrium is assumed to apply as an initial state and if and only if afterwards the arrival rate of new users at the road's entrance is fixed at a level exactly equal to the flow applying in the initial state. However, these same equilibria cannot be reached with any possible pattern of arrivals at the road's entrance following any other initial stationary state on the road, be it hypercongested or nonhypercongested. In particular, when starting with an initial stationary state on the road with an arrival rate of new users at the entrance equal to ρ^0, a flow F^0, a speed S^0, and a density D^0, (with $\rho^0 = F^0 = S^0 D^0$), and next study the consequences of a structural change in the arrival rate of new users to $\rho^1 \neq \rho^0$, the following new stationary or steady states can be shown to be reached:

(1) *The initial stationary-state equilibrium is nonhypercongested, $\rho^1 < \rho^0 < F^{Max}$*:
     the new equilibrium will be the *nonhypercongested* stationary state with a
     flow $F^1 = \rho^1$.

(2) *The initial stationary-state equilibrium is nonhypercongested,* $\rho^0 < \rho^1 < F^{Max}$: the new equilibrium will be the *nonhypercongested* stationary state with a flow $F^1 = \rho^1$.

(3) *The initial stationary-state equilibrium is hypercongested,* $\rho^1 < \rho^0 < F^{Max}$: the new equilibrium will be the *nonhypercongested* stationary state with a flow $F^1 = \rho^1$. So hypercongestion will disappear after a reduction in the arrival rate of new users at the entrance. Note that for a move to the hypercongested stationary state with a flow $F^1 = \rho^1$, the speed on the road would be required to decrease after a reduction in the arrival rate, which would require the density to increase above D^0. However, the reduction in the arrival rate means that instead the density will *decrease* below D^0 (for the first drivers especially near the entrance, but finally along the entire road), so that speeds will increase.

(4) *The initial stationary-state equilibrium is hypercongested,* $\rho^0 < \rho^1 < F^{Max}$: the new equilibrium will be a steady state in which the initial hypercongested stationary state with a flow $F^0 = \rho^0$ will sustain on the road, and in addition a queue will be growing before the road's entrance at a rate equal to $\rho^1 - \rho^0$. Note that, for a move to the hypercongested stationary state with a flow $F^1 = \rho^1$, the speed on the road would be required to increase after an increase in the arrival rate, which would require the density to decrease below D^0. However, the increase in the arrival rate means that the density will instead increase above D^0, particularly near the entrance. Therefore, speeds near the entrance will decrease (but because every individual driver accelerates, the original flow level F^0 will sustain; compare footnote 2) and a queue will form before the road's entrance.

(5) *The initial stationary-state equilibrium is nonhypercongested,* $\rho^0 < F^{Max} < \rho^1$: the new equilibrium will be a steady state with a flow $F^1 = F^{Max}$ on the road, and in addition a queue will be growing before the road's entrance at a rate equal to $\rho^1 - F^{Max}$.

Proofs are provided in Verhoef (1999, 2001). Together, these five points imply that there is no transitional path that would lead us to any of the hypercongested equilibria suggested by the standard analysis. Therefore, the candidate equilibria on the upper segment of the AC curve can be ignored for market analyses because these equilibria are dynamically unstable. Instead, these five points suggest that the true supply curve (with flow as the argument) for the use of a single road would coincide with the lower branch of the AC curve for nonhypercongested speeds, and at F^{Max} would not bend backwards as suggested in standard expositions, but instead rise vertically to represent queuing costs (see Figure 3 in Verhoef, 1999). A queue of constant length in a stationary-state equilibrium with $\rho = F = F^{Max}$ could be the result—with elastic continuous demand—after a transitional phase as described under point (5) has taken place. This would occur if the inverse aggregate demand function has no intersection with the upward sloping part of the AC curve (see Verhoef, 1999, for further

details). Now, one might object that hypercongestion is often observed in reality, implying that the above reasoning must be wrong. However, I argue that

> "the conclusion that the hypercongested equilibria suggested by the standard static model of road traffic congestion are in fact dynamically unstable, should not be mistaken to imply that every observation of hypercongested speeds in reality actually falsifies the model presented here. In contrast, the present model would suggest that whenever such speeds are observed, it is unlikely that the cause is to be found in flow congestion on that road itself, but that the true reason for such speeds may often be a downstream bottleneck. Therefore, optimal pricing rules should then not primarily be based on the road's characteristics, but rather on the bottleneck's capacity, which will in general imply different optimal road prices." (Verhoef, 2001, p. 532).

The reason I bring this all up here is twofold. First, I want to point out that alternative views on the controversy studied by Ohta have been proposed recently. Second, and more important, I find it worth mentioning that the hypercongested equilibria that may be globally optimal according to Ohta, are in fact dynamically unstable for the single road setting he uses, and therefore, cannot even be reached as a market equilibrium (with or without tolling) in any plausible dynamic generalization of the static model used. The implicit but highly restrictive assumption in static models that the initial conditions are always equal to the equilibrium generated is the single crucial assumption that is fully responsible for hypercongestion appearing candidate market equilibria in these static models. When a static model is used only because it is a convenient way of describing and analyzing stationary states that in are reached in reality only after some transitional phase—the way I would suggest that static models should be used—hypercongested equilibria as suggested by the standard static model can thus be ignored as candidate market equilibria.

5. CONCLUSION

In my comment, I challenge Ohta's contribution on two grounds. First, I argue that if what road users demand is completed trips, and if extra time spent on the road while traveling implies an extra cost for a given user, then Ohta's analysis is fundamentally flawed. His conclusion that hypercongestion can be optimal is invalid. Second, I argue that these hypercongested equilibria are dynamically unstable, and therefore, cannot even be reached as a stationary-state equilibrium in a dynamic extension of the static model studied by Ohta for any initial conditions other than that equilibrium itself.

I anticipate that Ohta will have the opportunity to write a rejoinder, replying to my comment. I also anticipate that he will not agree with the views I have expressed in the foregoing. In attempting to reject my objections, I challenge Ohta to do at least the following:

(1) To explicitly state whether he considers 'peak demand' or 'continuous demand'

(2) To make explicit whether in his representation, people primarily demand to be on the road or primarily want to make a trip. If the latter, then why is density the correct argument in the inverse aggregate demand function, rather than the total number of trips accomplished (if he considers peak demand) or the number of trips accomplished per unit of time which equals traffic flow in a stationary state (if he considers continuous demand)

(3) To demonstrate rigorously why for a static model with single values of S, D, and F applying throughout, the switch from a hypercongested to a nonhypercongested equilibrium with the same flow, as discussed in Section 3 above, would either:

   (a) mean that fewer users complete a trip per unit of time (if he assumes continuous demand); or

   (b) mean that fewer users complete a trip over the T time units that a peak period would last at the road's exit (and indeed at each point along the road, albeit that the peak starts and ends at a later clock time the closer one gets to the road's exit) in the hypercongested optimum envisaged by Ohta (if he assumes peak demand); or

   (c) not reduce the travel costs for each of the F users arriving every unit of time (if he assumes continuous demand); or

   (d) not reduce travel costs for each of the $N = FT$ users completing a trip during the T time units that this peak lasts at the road's exit (if he assumes peak demand)

(4) If none of the four points, (a)–(d), above can be demonstrated rigorously, explain why social surplus is higher and not lower in the hypercongested equilibrium with a flow F than in the nonhypercongested equilibrium with the same flow—although total benefits cannot be higher because the same number of trips is being made (during the peak for peak demand, per unit of time for continuous demand), and total costs will be higher because each user will have spent more time on the road before completing the trip

(5) To demonstrate rigorously how on a single road of constant capacity with a single entrance and a single exit, hypercongestion can build up in a dynamic setting at all if the initial condition entails free-flow speeds—for instance when the road was first opened (for continuous demand), or at say 3:00 a.m. (for peak demand in a morning rush hour), and users are homogeneous.

REFERENCES

Arnott, Richard, André de Palma and Robin Lindsey. (1993). "A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand," *American Economic Review*, 83, 161–179.
———. 1998. "Recent Developments in the Bottleneck Model," in: K. J. Button and E.T. Verhoef (eds.). *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility*. Cheltenham: Edward Elgar, pp. 79–110.
Evans, Alan W. 1992. "Road Congestion: The Diagrammatic Analysis," *Journal of Political Economy*, 100, 211–217.

Verhoef, Erik T. 1999. "Time, Speeds, Flows and Densities in Static Models of Road Traffic Congestion and Congestion Pricing," *Regional Science and Urban Economics*, 29, 341–369.

———. 2001. "An Integrated Dynamic Model of Road Traffic Congestion Based on Simple Car-following Theory: Exploring Hypercongestion," *Journal of Urban Economics*, 49, 505–542.

Vickrey, William S. 1969. "Congestion Theory and Transport Investment" *American Economic Review*, 59, 251–260.