Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

On kinematic waves

II. A theory of traffic flow on long crowded roads

By M. J. Lighthill, F.R.S. and G. B. Whitham

(*Department of Mathematics, University of Manchester*)

(*Received* 15 *November* 1954—*Read* 17 *March* 1955)

This paper uses the method of kinematic waves, developed in part I, but may be read independently. A functional relationship between flow and concentration for traffic on crowded arterial roads has been postulated for some time, and has experimental backing (§2). From this a theory of the propagation of changes in traffic distribution along these roads may be deduced (§§2, 3). The theory is applied (§4) to the problem of estimating how a 'hump', or region of increased concentration, will move along a crowded main road. It is suggested that it will move slightly slower than the mean vehicle speed, and that vehicles passing through it will have to reduce speed rather suddenly (at a 'shock wave') on entering it, but can increase speed again only very gradually as they leave it. The hump gradually spreads out along the road, and the time scale of this process is estimated. The behaviour of such a hump on entering a bottleneck, which is too narrow to admit the increased flow, is studied (§5), and methods are obtained for estimating the extent and duration of the resulting hold-up.

The theory is applicable principally to traffic behaviour over a long stretch of road, but the paper concludes (§6) with a discussion of its relevance to problems of flow near junctions, including a discussion of the starting flow at a controlled junction.

In the introductory sections 1 and 2, we have included some elementary material on the quantitative study of traffic flow for the benefit of scientific readers unfamiliar with the subject.

1. Introduction

A new problem, which has arisen in the twentieth century, is how to organize road traffic so that the full benefits of our increased mobility can be enjoyed at the lowest cost in human life and capital. The problem has many sides—constructional, legal, educational, administrative. The early lines of attack were largely intuitive. But, more recently, there has been an increasing tendency to adopt scientific methods, and try to assess the relative merits of different lines of attack by means of controlled experiments. This has been done both by the various authorities responsible for road lay-out, administration and propaganda, and also, more comprehensively, by organizations like the Road Research Laboratory in Great Britain, and the Bureau of Public Roads (formerly the Public Roads Administration) in the U.S.A. (Glanville 1953; Smeed 1952).

An important branch of the subject, with repercussions on all the other branches, is the quantitative study of traffic flow. An account of the experimental methods employed in this field has been given by the head of the traffic-flow section at the Road Research Laboratory (Charlesworth 1950). They include methods for measuring the means and standard deviations of vehicle speed at a point or journey time over a stretch of road, and for measuring the flow (number of vehicles passing a given point per unit of time). Attempts to correlate these variables for roads of particular mean width, mean curvature, etc., are made. Also, traffic performance

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

is studied before and after some change in road conditions, and statistical technique is used to find out whether the change significantly reduces journey times or accidents. Extensive researches on similar lines are carried out in the U.S.A., notably by the Division of Highway Transport Research, and by certain university departments such as the Post-graduate School of Highway Engineering at Yale.

In contrast to the well-developed character of traffic flow as an experimental science, theoretical approaches to the subject are in their infancy. Wardrop (1952) has given a valuable account of such theoretical investigations as have been made. He emphasizes the need for theoretical ideas to be used in conjunction with experimental data and the experience of individuals. It is well known, of course, in all branches of science and technology, that judicious use of theoretical ideas can save a lot of time by suggesting how experimental results obtained under one set of conditions can be extrapolated to another set of conditions. For example, theory may suggest in what form a set of results should be graphed, to give a curve likely to vary as little as possible with change of conditions. It may also suggest what things can most usefully be measured.

The theories which Wardrop (1952) describes are, as might be expected, statistical. First, the kinds of mean values which can be taken are discussed—for example, a 'space mean' over a length of road, or a 'time mean' over an interval of time at a fixed point. The space-mean speed (which we use in this paper) is the length of road divided by the average journey time of vehicles traversing it. It is also the ratio of the flow (vehicles per hour) to the concentration (vehicles per mile). The time-mean speed is somewhat greater because fast vehicles pass a fixed point more frequently (relative to their distribution in space) than slow vehicles.

Wardrop discusses the effect of increase of flow on overtaking. The number of 'desired overtakings' might be expected to increase as the square of the flow, so evidently, beyond a certain value of the flow, the proportion of desired overtakings which are possible must decrease. (For detailed observations on this point, see Norman, 1942.) This would clearly cause a reduction of mean speed with increase of flow, which is observed. He discusses also how traffic with uniform origin and destination may be expected to distribute itself over alternative routes, and he gives useful applications of the 'theory of queues' to the problem of delay at traffic lights (see also Tanner 1953).

In this paper we introduce a quite different method, suggested by theories of the flow about supersonic projectiles and of flood movement in rivers. It is the method of kinematic waves, introduced in part I (Lighthill & Whitham 1955); however, it is not essential to have read part I to understand the account which follows.

Now, a theoretical approach to road-traffic problems using methods from fluid dynamics is limited in advance to a restricted range of problems. Other ranges undoubtedly require statistical treatment of the kind described above, based on the theory of queues or the general theory of 'stochastic processes' (random time series). The 'continuous-flow' approach represents the limiting behaviour of a stochastic process for a large 'population' (total number of vehicles), and is therefore applicable to large-scale problems only—principally to the distribution of traffic along long, crowded roads.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

This 'arterial road' problem is an important one, however, which would be almost impossible to treat by purely statistical methods (though it may later be found desirable to use the present approach only as a first approximation, passing to higher approximations by means of a suitable blend with statistical ideas). To illustrate the theory, we use it to predict (§4) the progress of a traffic 'hump' in a long main road (due to a period of increased inflow at the main feed point), and (§5) the extent of the hold-up which results when such a hump passes through a bottleneck, which is too narrow to admit the increased flow. We also apply the method (§6) to junctions, especially controlled junctions, on long main roads.

The fundamental hypothesis of the theory is that at any point of the road the flow q (vehicles per hour) is a function of the concentration k (vehicles per mile). The evidence for this is discussed at length in §2. The hypothesis implies, as was shown in part I, that slight changes in flow are propagated back through the stream of vehicles along 'kinematic waves', whose velocity relative to the road is the slope of the graph of flow against concentration. A driver experiences such a wave whenever he adjusts his speed in accordance with the behaviour of the car or cars in front of him—for example, on observing a brake light, or an opportunity to overtake. It was seen also in part I that kinematic waves can run together to form 'kinematic shock waves', at which fairly large reductions in velocity occur very quickly. These too are very common on roads, notably at the rear of a traffic 'hump', and behind a bottleneck.

The properties of kinematic shock waves, and of continuous kinematic waves, will be derived again, by purely descriptive arguments, in §2. The more mathematical derivation, which some readers may prefer, will be found in §1 of part I.

The later sections are devoted to examples of the kinds already mentioned. The predictions are found to agree qualitatively with experience, but the extent of quantitative agreement is not yet known. Experiments to determine this are being planned.

It should be mentioned that essentially the same methods and results apply to pedestrian traffic of a congested character. The bottleneck theory (§5) is particularly relevant to the movement of crowds through passages. However, the following exposition is confined to the more serious problem of vehicular traffic flow.

2. The flow-concentration curve

Although the flow q and the concentration k have no significance except as means, the purpose of the theory is to ask how they vary in space and time. However, on a long crowded road this is reasonable, since the means can be taken over relatively short distances or time intervals, and we are interested in variations over much greater distances and times.

The precise definitions of q and k, at a given point x on the road and a given time t, are included in the following instructions for measuring them. Draw two lines across the road, a short distance dx apart, to form a slice of road with the point x in the middle. Take averages over a time interval of moderate length τ,

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

with the time t in the middle. The interval τ must be long enough for many vehicles to pass. Then the flow q is

$$q = n/\tau, \tag{1}$$

where n is the number of vehicles crossing the slice in time τ. The concentration k is

$$k = \frac{\Sigma\,\mathrm{d}t}{\tau\,\mathrm{d}x}, \tag{2}$$

where $\Sigma\,\mathrm{d}t$ means the sum of the times taken by each vehicle to cross the slice. Thus k is the average number of vehicles ($\Sigma\,\mathrm{d}t/\tau$) on the slice of road, divided by the length $\mathrm{d}x$ of the slice; in other words, k is the number of vehicles per unit length of road.

A third important quantity is

$$v = \frac{q}{k} = \frac{\mathrm{d}x}{\frac{1}{n}\Sigma\,\mathrm{d}t}. \tag{3}$$

This is the 'space-mean speed' of Wardrop (1952), being both the ratio of flow to concentration and the ratio of length of slice to average crossing time. Thus it is an average of vehicle speeds weighted according to the time they remain on the slice of road. (If conditions were uniform, on the average, over a much longer stretch of road, v would also be the average speed of all vehicles while they remain on that stretch; the further averaging with respect to time would then be unnecessary, since the fluctuations with time would become small for a long stretch of road. This explains the name 'space-mean speed'.) The time-mean speed, which we shall not use, is the unweighted average speed of vehicles crossing the slice, namely $n^{-1}\Sigma\,(\mathrm{d}x/\mathrm{d}t)$. This exceeds v. If speeds at a point are measured directly (as by a Radar speedmeter), instead of in terms of times, one can still derive the space-mean speed (Wardrop 1952) by taking the 'harmonic' mean of the observed speeds, namely,

$$\left(\frac{1}{n}\Sigma\frac{1}{\mathrm{d}x/\mathrm{d}t}\right)^{-1} = \frac{\mathrm{d}x}{\frac{1}{n}\Sigma\,\mathrm{d}t} = v. \tag{4}$$

Most road-traffic observers have concentrated on measuring q and v, as being the quantities of greatest practical interest. The concentration k must be obtained from such measurements by division. Sometimes, however, k is observed directly by taking photographs of the road from above. Such results are sometimes quoted in terms of mean 'headway' (distance between the fronts of successive vehicles in the same lane of traffic). The mean headway is N/k, where N is the number of lanes travelling in the direction considered.

Vehicle counts are sometimes made by moving observers, especially (Charlesworth 1950; Wardrop & Charlesworth 1954) by observers in cars filtered into the traffic. If an observer moving at uniform speed U records the number of vehicles which pass him, minus the number which he passes, and divides the difference by the total time of observation (say τ), the result is

$$q - kU. \tag{5}$$

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

(A number $q\tau$ of vehicles would pass him if he were stationary, but this is reduced by $k(U\tau)$, namely, the average number of vehicles in the distance U which he travels.) By measuring expression (5) successively for two values of U (in practice, values with opposite signs), q and k may be separately deduced.

This experimental method is closely linked to the basic theoretical idea of this paper. Consider two observers moving with uniform speed U, the second starting, and remaining, a time τ behind the first.* Suppose now that the flow and concentration are changing with time, but that nevertheless the observers adjust their speed U so that the number of vehicles which pass them, minus the number which they pass, is, on the average, the same for each. Then by (5), $q-kU$ is the same for each, and so

$$U = \frac{\Delta q}{\Delta k}, \tag{6}$$

where Δq and Δk are the change in flow and concentration after time τ.

Now, in the circumstances mentioned, the number of vehicles between the observers must remain the same. But the number of vehicles passing any point between the times at which the observers pass it is $q\tau$. Since τ is fixed, it follows that the flow q *remains unchanged along the path of observers travelling with the speed* (6).

In other words, when changes of flow are occurring, the waves which carry such changes through the stream of vehicles travel at a velocity given by equation (6). This velocity, relative to the road, may, as we shall see, be positive or negative. However, it never exceeds $+v$, the space-mean speed; hence the waves are always transmitted backwards relative to the vehicles on the road.

Now, it has been conjectured by many authors that, on any uniform stretch of a road, the flow q is a function of the concentration k. If this is true, equation (6) becomes especially valuable, since it shows that *small* changes of flow are propagated at the speed

$$c = \frac{\mathrm{d}q}{\mathrm{d}k}, \tag{7}$$

which is known if k (or q) is known.

The relationship between flow and concentration has usually been stated in rather different forms. At low values of the concentration, the mean speed $v = q/k$ has been regarded as a function of the flow q (Normann 1942; Normann & Walker 1949; Glanville 1949, 1951). It falls off as q increases, with a slope which is steep for narrow roads but more gradual for wide roads. Wardrop (1952) ascribes the effects of increased flow, in the main, to increased interference with overtaking, which tends to reduce the mean speed to nearer the speed of the slowest vehicles on the road. Doubtless, a general sense of the greater possibility of accidents also contributes to the reduction in mean speed.

At high values of the concentration, however, most writers have regarded the 'mean headway' N/k as a function of the mean speed $v = q/k$. At $v = 0$, the mean

* Imagine them to be cyclists on an adjacent cycle track, so that they can maintain their uniform speed U unimpeded, and in turn will not influence the observed traffic flow (we are not suggesting this as a practical method of observation, but as a convenient way of thinking about the flow).

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

headway takes a value (around 17 ft. in Great Britain) only just greater than the average vehicle length. As v increases, the mean headway increases almost linearly (by about 1·2 ft. for each 1 mile/h increase in speed). Many authors (see, for example, Normann & Taragin 1942) have interpreted such results by saying that a driver allows just enough headway so that no collision will result if the vehicle in front brakes suddenly, and he himself brakes after a certain 'reaction time'. Glanville (1949) points out that the observed rate of increase of headway with speed would correspond to a uniform braking force, equal for both vehicles, and a reaction time of 0·8 s. The reader may easily verify this. Attempts have been made to apply such considerations also at low values of the concentration, but then the greater freedom to overtake alters the situation completely.

Different experimental methods are appropriate for determining these two kinds of relationship. Our contention, however, is that the information obtained from these two sources should be combined into a single curve, and that the curve which sums up all the properties of a stretch of road which are relevant to its ability to handle the flow of congested traffic is a graph of the two fundamental quantities, flow against concentration.

The form of such a curve must be as in figure 1. As the concentration k tends to zero, the flow q must also become zero. Again, in the limiting case of high concentration $k = k_j$ (j for jam) the vehicles travelling in a given direction are packed tight on the part of the road where they are permitted to be; the flow q is then again zero. For some value of the concentration between these two extremes, the flow q must have a maximum q_m, which may be called the capacity of the road.

The deduction in the last paragraph (which a mathematician would call an application of 'Rolle's theorem'!) does not seem to have been clearly made in the traffic-flow literature, except perhaps by Greenshields (1935). Considerable effort has been put into finding a suitable definition of road capacity, but it has not been noticed that the very simple and relevant one 'maximum flow of which the road is capable' is available.*

Experimentally, this was because flow at the particular concentration k_m corresponding to this maximum flow is not often observed, for reasons which will appear later. Flow at smaller concentrations is commonly observed, and described by a speed-flow relation. (A description in such terms is inconvenient for the complete range of speeds, since there are two speeds for a given value of the flow.) Flow at concentrations near to k_j is commonly observed, and described by a headway-speed relation. (This description is unsuitable at low concentrations because headway ceases to have significance when overtakings are prominent.)

To complete the curve satisfactorily, an independent measurement of q_m and k_m (flow and concentration for maximum flow) is desirable, since interpolation between the two measured parts of the curve is very difficult without knowledge

* Normann (1942) introduced a 'theoretical maximum capacity', obtained by assuming that the flow at all concentrations was governed by the theoretical speed-headway curve, but he points out that observed flows are hardly ever more than about half of this 'theoretical maximum'. The maximum here discussed, on the other hand, is the real, experimentally determined, maximum. Again, it should not be confused with a statistical 'extreme value', since the flow-concentration curve represents the average relationship between the quantities.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

of some intermediate point. Fortunately, the theory of this paper provides a special method of measuring these two quantities, as follows.

If a stream of vehicles is stopped, as at a traffic light, and then started again after a considerable delay, as when the lights go green, a system of waves is emitted.* Each carries a particular value of the flow q and concentration k, and hence also a particular value of the wave velocity c, and propagates with this uniform velocity, some forwards and some backwards (see §6 below). One wave alone remains stationary at the original stopping-point. Now this wave has $c = 0$, so by (7) it corresponds to a value of k for which $\mathrm{d}q/\mathrm{d}k = 0$, namely, to $k = k_m$, for which q is a maximum. This shows that the mean flow and concentration measured

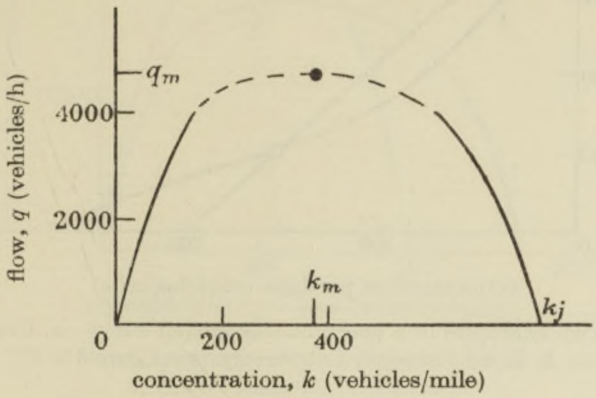

FIGURE 1. A flow-concentration curve.

at the stopping-point itself (after the stream of vehicles has started up, and before all those slowed down by the original stoppage have passed through—the need for these restrictions will become clear in §6) are the required quantities q_m and k_m.

A typical flow-concentration curve constructed in the manner indicated is shown in figure 1. The full line on the left is derived from speed-flow data, that on the right from headway-speed data, and the central point (q_m, k_m) from measurements at the stopping-point after a long line of traffic had been stopped and then allowed to flow forward freely again. The curve refers to a certain one-way three-lane section of the Great West Road, and the speed-flow data were obtained during the period of peak evening traffic between 5 and 7 o'clock. The authors are grateful to the Director of the Traffic and Safety Division, Road Research Laboratory, for permission to use the unpublished results displayed on this curve.†

* A really long lane of vehicles must be stopped if the theory is to be applicable, as will appear later (§6).

† Mr Wardrop has recently indicated to the authors that he would now consider a rather lower value (say 3200) more typical of the flow q_m past a stopping point on this particular stretch of road than the earlier value (round 4700) supplied to the authors and quoted in figure 1. However, R.R.L. measurements for single-lane traffic yield values of q_m of 1500 v/h., so that values of around three times this would be expected for three-lane traffic. If they were not observed, the cars were probably not filling the three available lanes when stopped. The flow q_m will be achieved only if all available lanes are fully used.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

Another method of deriving the curve was used by Greenshields (1935), who plotted $v = q/k$ against k for one-lane traffic, as in figure 2, and drew a straight line through his points. This involved a rather drastic interpolation since there is a large intermediate range where there are no points, and where in fact the true curve probably lies below the straight line. However, the method gives a simple and probably not too inaccurate result, which led to the predicted existence of a maximum flow on any road much earlier than had been inferred elsewhere, as mentioned above. Greenshields introduced a 'kink' at the top of his graph, to

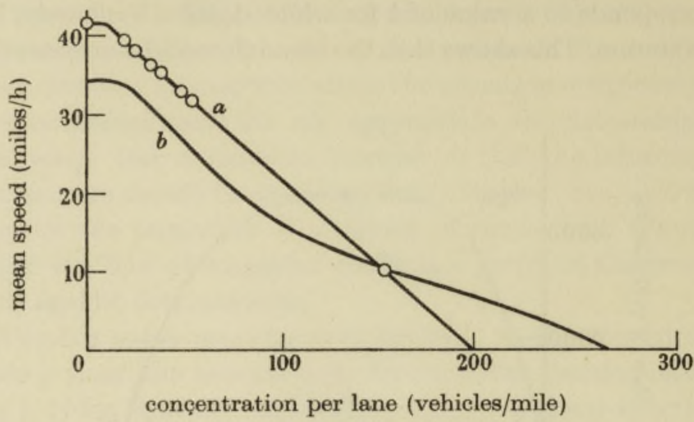

FIGURE 2. Two examples of a speed concentration curve. *a*, Greenshields; *b*, Road Research Laboratory (see figure 1).

make the speed flatten out at the independently determined 'free speed' for the road. A flat portion like this must be expected on any speed-concentration curve, since the mean speed will be unaffected by concentration below a certain limiting value. On a wide road like that of figure 1 this limit may be as much as 50 vehicles per mile.

One may use the word 'crowded' to describe road conditions on which the concentration exceeds this limit. Then a road is crowded if any increase in concentration will lead to a reduction in mean speed. The theory of this paper is applicable only to long, 'crowded' roads.

For comparison with Greenshields's result the curve corresponding to figure 1 is also shown in figure 2, with the densities divided by 3 to allow for the greater number of lanes. In comparing the two curves, one must bear in mind the differences between English and American driving habits and vehicle lengths.

The two curves are shown also in figure 3, as flow-concentration curves per lane of traffic. That of Greenshields is the arc of a parabola with vertex upwards. A portion of the arc near the origin is replaced by a chord through the origin. This corresponds to a range of non-'crowded' conditions, in which the mean speed is constant.

To conclude this section it may be noted that the flow-concentration curve for a particular stretch of road may vary from time to time (especially with the day of the week, but also with the time of day), owing to changes in the proportion of commercial vehicles on the road, or in the quantity of traffic travelling in the

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

opposite direction. Some care is therefore needed in specifying the conditions under which a particular determination of the curve has been made. Again, the variations along a given road, due to differences of width, gradient, curvature, population density, etc., between different stretches of the road, may be very great. The velocity of a wave in any one stretch of road, however, will be given by the slope of the flow-concentration curve for that particular stretch of road, as the argument leading to equation (6) makes clear. The use of the theory in such cases is possible, therefore, and will be fully illustrated in §5.

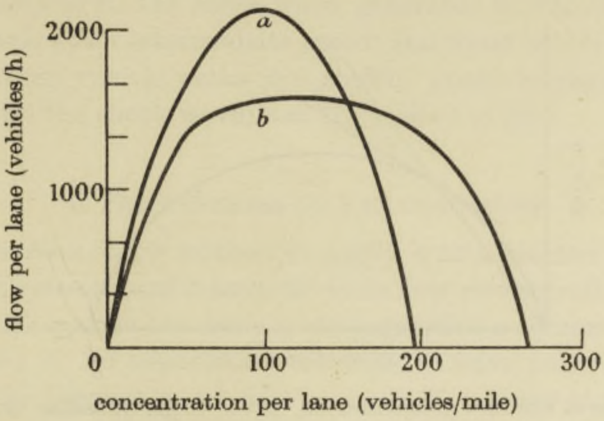

FIGURE 3. Flow-concentration curves per lane of traffic. *a*, Greenshields; *b*, Road Research Laboratory.

3. Use of the flow-concentration curve

To make practical use of the flow-concentration curve for a particular stretch of road, a geometrical expression of the results of §2 is often valuable.

First, note that, corresponding to any point on the curve, the space-mean speed $v = q/k$ (under the conditions represented by that point) is the slope of the radius vector from the origin (figure 4). The speed $c = \mathrm{d}q/\mathrm{d}k$ of waves carrying continuous changes of flow through the stream of vehicles is the slope of the tangent to the curve at the point (figure 4). This slope is the smaller,* provided that the mean speed *decreases* with increase of concentration; in other words, if the road is 'crowded'. For we can write

$$c = \frac{\mathrm{d}}{\mathrm{d}k}(kv) = v + k\frac{\mathrm{d}v}{dk}, \tag{8}$$

which is less than v if $\mathrm{d}v/\mathrm{d}k$ is negative. The velocities c and v are equal only at low concentrations, below the limit (mentioned in §2) at which significant interaction between different vehicles on the road first occurs. At such concentrations, $\mathrm{d}v/\mathrm{d}k = 0$.

To express velocities as slopes in this way is convenient if conditions on a road are to be represented in a space-time diagram. If the road is represented as

* Meaning that waves travel backwards relative to the mean vehicle flow.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

stretching up the paper, with time travelling to the right, then a path on this diagram, representing the motion of a wave or of a vehicle, will have a slope dx/dt equal to the velocity. Since lines of equal slope are parallel, it follows that a mean vehicle path on this diagram must be parallel to the radius vector from the origin to the relevant point on the flow-concentration curve, while a wave must be parallel to the tangent to the curve.

A second use of the flow-concentration curve refers to discontinuous waves. These are likely to occur on any stretch of road when the traffic is denser in front, and less dense behind. For waves on which the flow is less dense travel forward

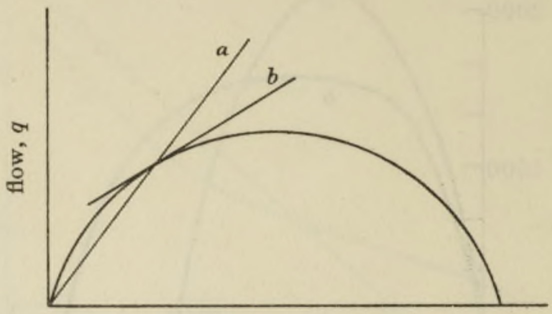

FIGURE 4. Use of the flow-concentration curve. Slope of radius vector (a) gives average velocity of vehicles; slope of tangent (b) gives wave velocity.

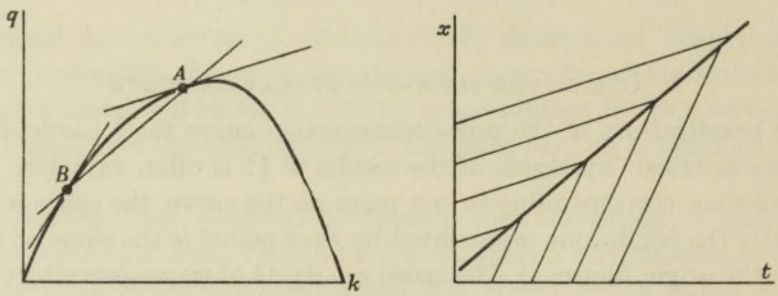

FIGURE 5. Use of flow-concentration curve to predict the local conditions near a shock wave.

faster than, and hence tend to catch up with, those on which the flow is denser. When this happens a bunch of continuous waves can coalesce into a discontinuous wave, or 'shock wave'. When vehicles enter this their mean speed is substantially reduced very quickly. The wave is not totally discontinuous of course, but its duration is not much longer than the braking time that each vehicle needs to make the required reduction of speed.

The speed of a discontinuous wave, or shock wave,* is given by (6) as $\Delta q/\Delta k$, the slope of the *chord* joining the two points of the flow-concentration curve which represent conditions ahead of and behind the shock wave. (Note that the argument

* In future we shall prefer the latter name, suggested by the very strong analogy with shock waves in gases.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

leading to (6) is applicable, provided that the time interval τ between the two observers exceeds the duration of the shock wave. The number of vehicles between two observers with the shock wave between them can remain constant only if they travel at the speed of the shock wave.)

Figure 5 illustrates the use of the flow-concentration curve to predict conditions near a shock wave. The shock wave is shown as a heavy line on the space-time diagram on the right. Ahead of it the flow is denser and the waves (plain lines) are drawn parallel to the tangent to the flow-concentration curve at A. Behind it the concentration is less and the waves travel faster; they are drawn parallel to the tangent to the curve at B. The shock wave, generated by the running together of these waves, travels at an intermediate speed, and must be drawn parallel to the chord AB. The mean vehicle paths (not shown) would be parallel to the radius vectors OB (behind the shock wave) and OA (ahead of it).

4. The progress of a traffic hump

As a first illustration of the method we apply it to a problem where the road is uniform, so that all stretches of it have the same flow-concentration curve. In these circumstances, each continuous wave is propagated at a constant velocity c, since q is constant along it. In a space-time diagram the wave paths are straight lines, each parallel to the tangent to the flow-concentration curve at the corresponding point.

The source of traffic is taken to be at one end of the road, and we consider the case when the inflow rises to a peak and then falls to its original value, producing a traffic hump.* The rise and fall of inflow can be easily measured by an observer at the feed point. A problem of some importance is then: How can the behaviour of the hump as it passes down the road be predicted in advance? For example, when will it reach a given point? Will it spread out, or become more concentrated, and how fast? How will it affect average journey times?

The wave theory gives convenient answers to these questions. Figure 6 shows the wave pattern in a space-time diagram. The wave path starting from the feed point at any time is parallel to the tangent to the left-hand part of the flow-concentration curve at the point which corresponds to the inflow at that time. The waves travel more slowly inside the hump than outside it. Hence the wave paths in figure 6 'fan out' at the front and become concentrated at the rear, where they must ultimately run together.

It must be emphasized that the lines drawn are 'waves' (lines of constant flow, and hence also, for a uniform road, lines of constant mean speed) and *not* vehicle paths. Vehicles go (on the average) faster than the waves, and most vehicles starting at the rear of the hump will in time get through it. On entering the hump a driver has to slow down fairly rapidly (since the lines of constant speed are

* Traffic humps (regions of increased concentration) generated in *this* way have concentrations remaining solely on the left-hand half of the flow-concentration curve. But humps at the much higher concentrations corresponding to the right-hand half have similar properties; the only important difference is that the waves travel backwards relative to the road. Examples of humps of *this* kind occur below, especially in the theory of bottlenecks (§5).

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

bunched together on the right of figure 6), but on leaving it he can increase his mean speed only slowly as he traverses the fan of waves on the left.

Figure 6 gives a clear answer to the question of the speed of the front of the hump, which turns out to be the wave velocity associated with conditions in front of it. Note that this may be considerably less than the space-mean speed (which in turn is less than the time-mean speed) of the vehicles in this region. The other questions noted above can be answered only after the path of the shock wave, which results from the running together of the waves at the rear of the hump, has been determined.

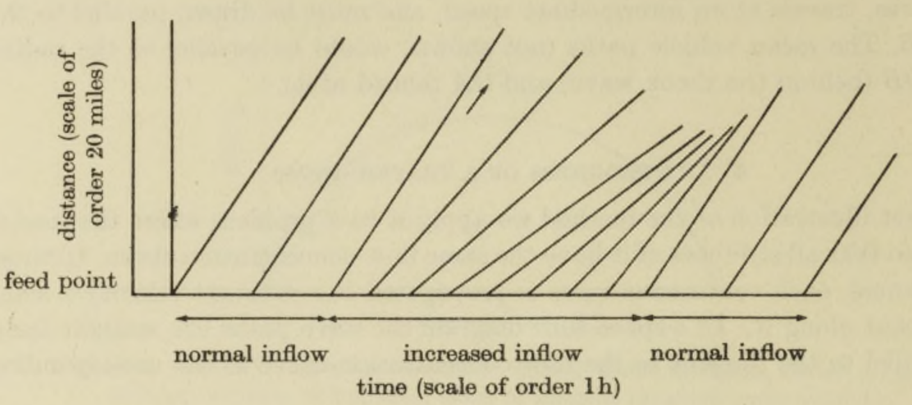

FIGURE 6. Wave forms in traffic hump.

The shock wave starts at the point where two waves first run together, and its progress after that is governed by the simple law stated in §3: at each point of the shock, the two waves which meet there are represented by two points on the flow-concentration curve, and the shock wave path must be drawn parallel to the chord joining those points. This gives a straightforward geometrical step-by-step method for constructing the path of the shock wave.

In practice it is convenient to note that the slope of the chord is approximately the mean of the slopes of the tangents at its end-points, so that the speed of the shock wave is approximately the mean of the speeds of the waves running into it from either side. This approximation is exact for a parabolic arc with vertical axis, such as Greenshields's flow-concentration curve (figure 3). For other smooth curves with nothing approaching a vertical tangent, the approximation is still fairly good, as the known series for the slope of the chord,

$$\frac{q(k_2)-q(k_1)}{k_2-k_1} = \tfrac{1}{2}(q'(k_1)+q'(k_2)) - \frac{(k_2-k_1)^2}{24}(q'''(k_1)+q'''(k_2)) + \dots, \tag{9}$$

shows. In view of the approximate character of the whole theory, the additional approximation is probably worth making wherever it will make an effective simplification.

It certainly makes the shock wave easier to draw in by eye, as no further reference to the flow-concentration curve is then necessary. One has simply to

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

draw a path on the space-time diagram whose slope at any point is the mean of the slopes of the waves running into it from either side. This process is illustrated in figure 7; it can be mastered with only a little practice.

As an alternative, or as a check, one has the analytical solution for the shock path (Whitham 1952) which again is based on the approximation noted above. This also can be expressed as a geometrical construction, as follows.* Given the

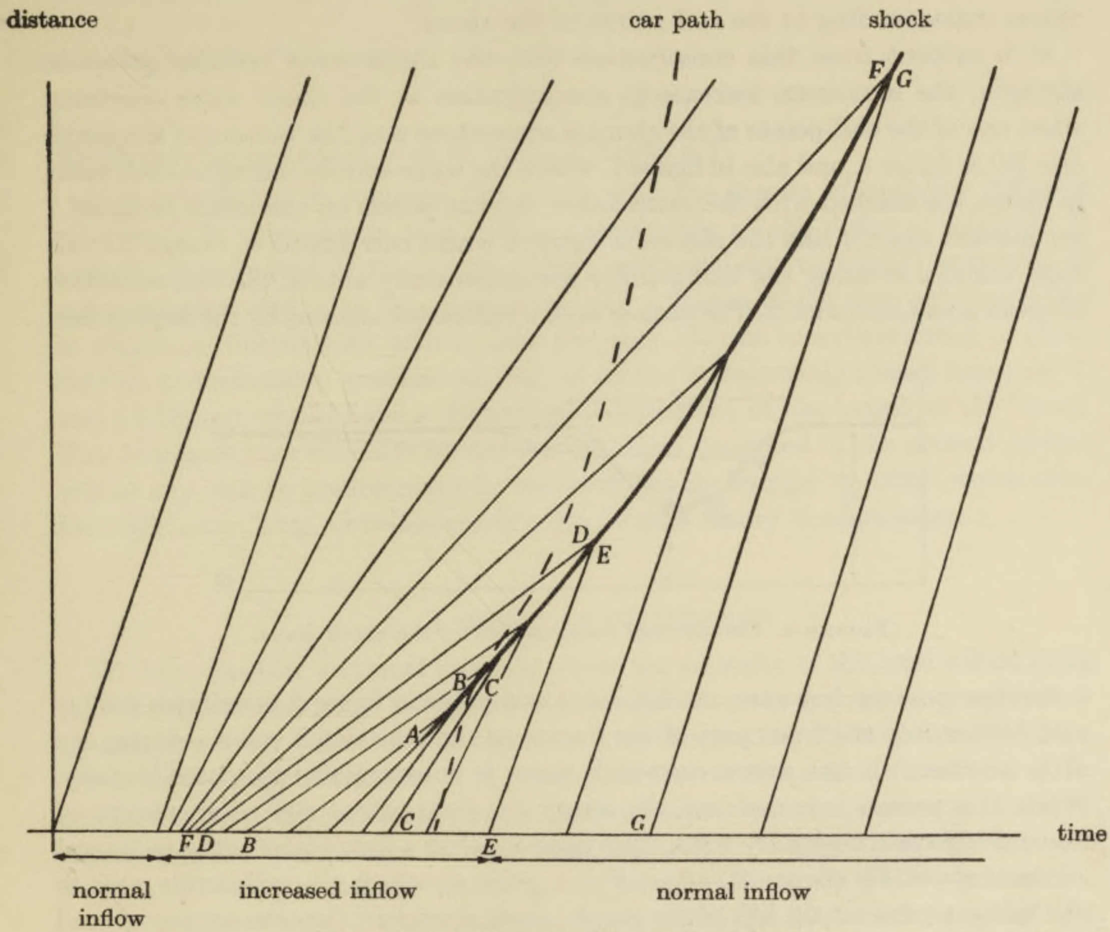

FIGURE 7. Progress of traffic hump with time.

variation with time t of the inflow rate observed at the feed point, plot a graph (figure 8) with the corresponding wave velocity c (the slope of the flow-concentration curve for the observed value of the inflow) as ordinate, and its product with the time, ct, as abscissa. Then the time at which the shock wave first appears is

* The present problem is somewhat simpler than that treated by Whitham (1952), in which our ordinate c, the rate of change of x with respect to t on a wave, is replaced by $F(y)$, the rate of change of $\alpha r - x$ with respect to $kr^{\frac{1}{2}}$; and in which the abscissa is y, the value of x when $r = 0$. The analogous abscissa in our problem is evidently the value of $-x$ when $t = 0$. For the wave which passes $x = 0$ at time t, with velocity c, this is ct. Readers of part I should note that another approach, in which c^{-1} and not c replaced $F(y)$, was found convenient there (§4); however, that approach cannot be used if the flow-concentration curve has a stationary point.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

given by the reciprocal of the slope of the tangent to this graph at its right-hand point of inflexion A; the value of c (or t) at A also determines, through its velocity (or time of origin, respectively), *on which wave* the shock wave first forms. To determine the further progress of the shock wave, draw chords on the graph (e.g. BC, DE, FG) which cut off lobes *of equal area* above and below between them and the curve. Then the slope of any one of these chords is the reciprocal of the time at which the shock wave absorbs the two waves on which c and t have the values corresponding to the end-points of the chord.

It is evident from this construction that the shock wave initially grows in strength, the maximum increase in concentration at the shock wave occurring when one of the end-points of the chord is somewhere near the bottom of the graph (see BC in figure 8, and also in figure 7, where the wave corresponding to each point in figure 8 is marked with the same letter, so that points on the shock in figure 7 are marked exactly like the chords in figure 8 which correspond to them). At this time vehicles entering the hump suffer instantaneously almost the full reduction of speed associated with it. The path of such a vehicle is indicated by the broken line.

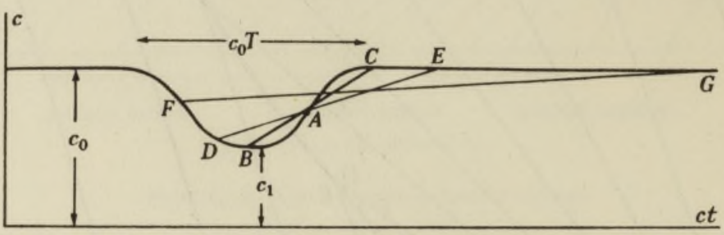

FIGURE 8. Geometrical construction for the shock wave.

As time goes on, however, the left-hand end-point in figure 8 penetrates farther and farther into the front part of the hump, so that the shock wave absorbs, one after another, all the waves on which there is substantially increased density. When this process is completed, the hump has disappeared and what remains of the shock wave is negligibly weak. This happens after a time equal to the reciprocal of the slope of FG (figure 8), where F is a point at which c is sufficiently near to the value it takes on the left of the graph. Note, however, that the section of road satisfying the conditions postulated may in many cases come to an end before the hump is dispersed in this way.

Regarding the hump as a region of increased concentration, it may be asked how the excess of vehicles can effectively vanish in this way. The answer is that the region of increased concentration spreads backwards (relative to the front of the hump, which has a constant mean speed), so that the excess of vehicles is dispersed over a constantly increasing length of road. A quantitative estimate of the process may be obtained if one knows the duration, say T, of the increased inflow at the feed point, the wave velocity c_0 outside the hump and the lowest value, say c_1, of the wave velocity inside the hump. Then the shock wave is at its strongest at a time about

$$\frac{c_0 T}{2(c_0 - c_1)} \tag{10}$$

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

after the time of maximum inflow. At this time (corresponding to BC* in figures 7 and 8) the hump has hardly spread backwards at all; it has simply altered its shape so that the increase of concentration is sudden and the subsequent decrease is spread over the whole length of the hump. Later, the decrease of wave velocity at the shock wave becomes a small quantity δ after a time†

$$t = \frac{c_0(c_0 - c_1)\,T}{\delta^2}, \tag{11}$$

and the length of the hump is then about

$$l = \{c_0(c_0 - c_1)\,Tt\}^{\frac{1}{2}}, \tag{12}$$

which may be compared with its original length $c_0 T$.

It is interesting to compare this result with the results of ordinary diffusion processes. It corresponds to a diffusion coefficient of the order of $c_0(c_0 - c_1)\,T$, namely, the product of the length of the hump and the maximum reduction in wave velocity within it. By comparison, any diffusion which may be present due to statistical fluctuations with a mean free path, or due to a dependence of mean flow on concentration-gradient as well as on the concentration itself (see part I, and §6 below), would have a diffusivity independent of the length of the hump. This indicates that diffusion by the wave process described in the present section will at any rate be predominant for sufficiently long humps—in other words that, for *sufficiently* 'long, crowded roads', the present theory is appropriate.

5. A theory of bottlenecks

We now consider a typical problem where the capacity of the road varies along it. We suppose that some bottleneck is present, where the maximum possible flow q_m falls to a lower value than on the main part of the road. Then, presumably, the whole flow-concentration curve is reduced in its vertical scale. (It may well be reduced in horizontal scale too (that is, k_j may become less), but figures 9, 11 and 13 illustrating the theory have actually been drawn for the case where this does not happen.) The local minimum value of q_m may be called the *capacity* of the bottleneck.

We consider first a stream of vehicles approaching the bottleneck at a flow rate which remains always *less* than its capacity. Then each vehicle suffers simply a temporary reduction in speed as it passes through. The waves are also reduced in speed while in the bottleneck. For the flow q remains constant on any wave, as was shown in §2 independently of whether the flow-concentration curve varies with position. Hence (figure 9) conditions on a wave as it passes through the bottleneck are represented by points of flow-concentration curves all at the same horizontal level. Since the tangent to the lower curves at a given horizontal level has a smaller slope, the wave velocity is reduced inside the bottleneck, and the

* The quantities c_0, c_1 and T are indicated in figure 8, and it is evident that the slope of BC is approximately the reciprocal of (10).

† The area of the hump in figure 8 is about $\frac{1}{2}(c_0 - c_1)\,c_0 T$, and this will be equal to the area above FG, namely, $\frac{1}{2}\delta^2 t$, where t^{-1} is the slope of FG, if (11) holds. Here $c_0 - \delta$ is the value of c at F.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

wave paths behave as in figure 10. Under the conditions illustrated in this figure the delay to each vehicle is relatively small.

Next, we consider the more serious hold-up resulting when, as time goes on, the oncoming flow rate increases above the capacity of the bottleneck. Waves then turn back before reaching the centre of the bottleneck and form a shock wave. This passes back down the main road and forces vehicles to pile up behind the bottleneck at a rate given by the *difference* between the oncoming flow and its capacity. In practice, the oncoming flow would exceed the capacity of the bottle-

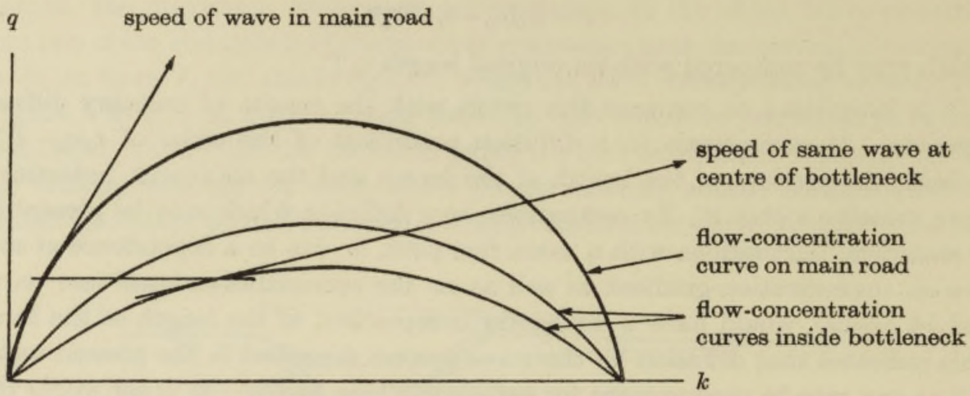

FIGURE 9. Variation of flow-concentration curve in a bottleneck.

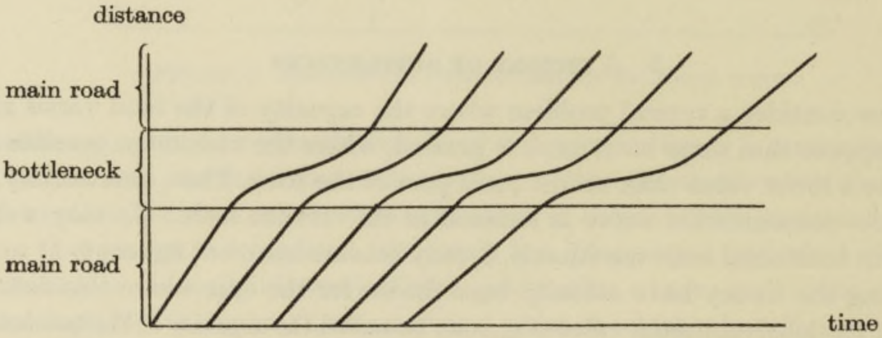

FIGURE 10. Passage of waves through a bottleneck, the capacity of which exceeds the incoming flow rate.

neck only for a finite time, during which the oncoming traffic is in the form of a hump. An important question is the duration of the hold-up resulting from the passage of a given traffic hump through the bottleneck. This will be solved by a detailed study of the shock wave paths.

To understand the formation of the characteristic 'bottleneck shock wave', note that no wave carrying a flow exceeding the capacity of the bottleneck can possibly pass through it, since the flow must remain constant on the wave, and such a large flow is impossible in the centre of the bottleneck. It is not important at which precise point of the bottleneck the wave turns back, but theoretically (if the flow-concentration curve varies continuously through the bottleneck) it should do so at the point where the flow carried by the wave is the maximum possible flow;

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

for here only is the wave velocity (slope of the tangent to the flow-concentration curve) zero. In figure 11, this point is B; the slope of the tangent at C indicates the speed at which the wave will come out of the bottleneck again. Compare the points A, B, C in figure 12, which shows in a space-time diagram the turning back of such a wave. For short bottlenecks, the details of the predicted flow within the bottleneck could not be relied on. However, the qualitative fact that the wave turns back, and its progress beyond C, are predictions on which greater reliance can be placed.

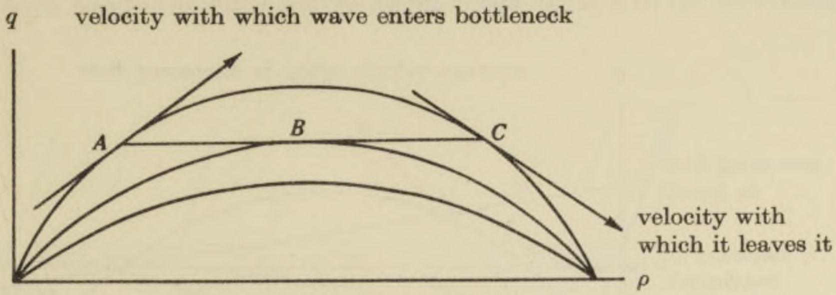

FIGURE 11. Illustrating the 'reflexion' of a wave from a bottleneck.

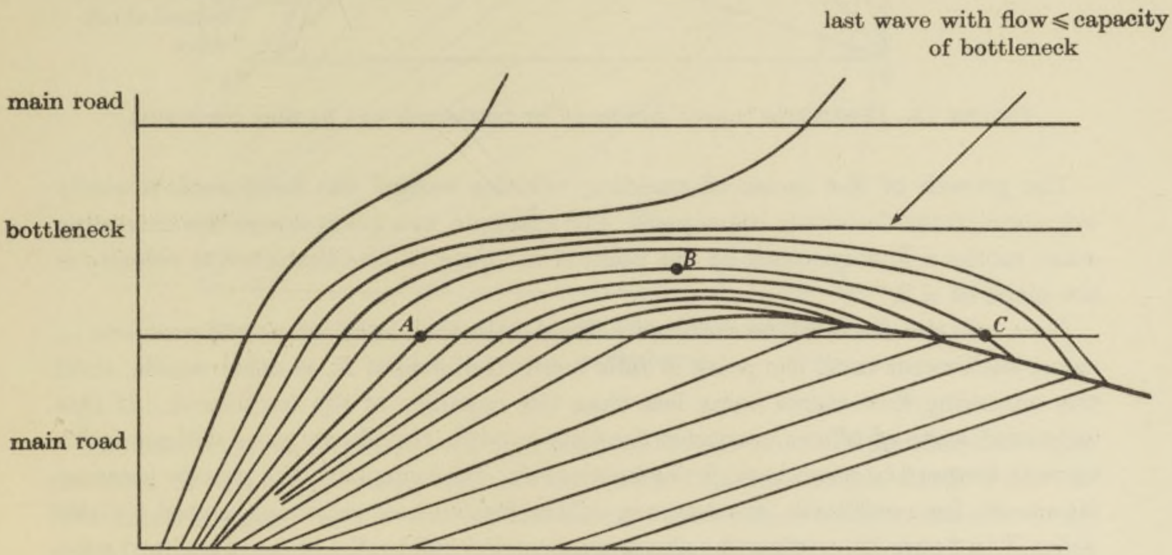

FIGURE 12. Formation of shock wave in the front of a hump as it enters a bottleneck of inadequate capacity.

The need for waves to intersect is at once evident from figure 12, where the beginning of the resulting shock wave is sketched in. This shock wave involves a reduction of flow, so its velocity (the slope of the chord joining points on the flow-concentration curve corresponding to conditions in front and behind) must be backwards relative to the road. As soon as it passes back out of the bottleneck, it must reduce the oncoming flow to almost exactly the capacity of the bottleneck. This is because waves carrying flows less than this have passed through, and waves carrying greater flows have turned back and been absorbed by the shock wave, so

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

that only waves carrying flows approximately equal to the capacity of the bottleneck remain in its neighbourhood. Those just behind it are travelling backwards, corresponding to a point (e.g. B in figure 13) on the *right-hand half* of the flow-concentration curve for the main road, at a flow level corresponding to the capacity of the bottleneck. The speed of vehicles in the slow crawl up to the bottleneck is given by the slope of OB. Conversely, the waves just ahead of the bottleneck are travelling forwards, corresponding to a point (e.g. F in figure 13) on the left-hand half of the curve. Thus, vehicles after passing through the bottleneck are able to accelerate up to a mean speed given by the slope of OF.

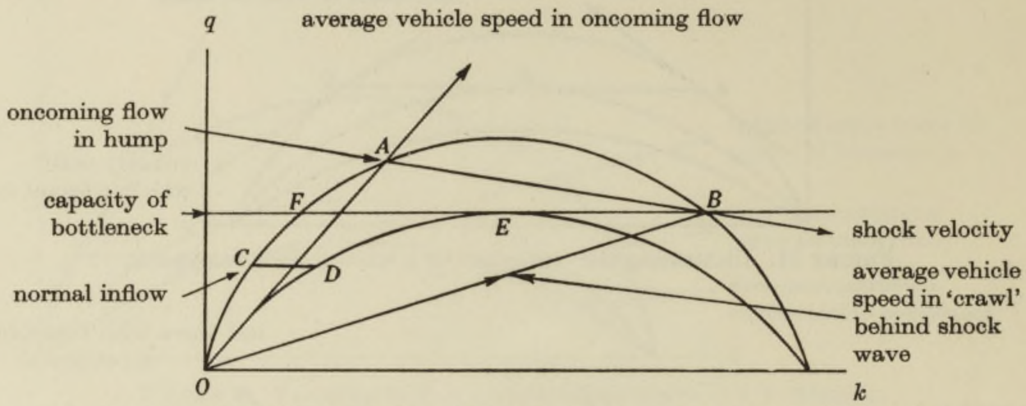

FIGURE 13. Illustrating 'crawl' produced by bottleneck and its final resolution.

The growth of the queue of crawling vehicles behind the bottleneck is easily calculated from the shock-wave path. For example, at a point where the oncoming wave carries a flow specified by the point A in figure 13, the shock-wave velocity is the slope of AB.*

How will the deadlock be resolved? Evidently the shock wave will continue to move backwards until the point A falls below the level of B, in other words, until the oncoming flow starts being less than the capacity of the bottleneck. If this improved state of affairs continues for long enough, the shock wave will move far enough forward to pass through the bottleneck. On doing so it will greatly increase its speed, for conditions downstream of the bottleneck are respresented by the point F in figure 13, so that the shock-wave speed will be the slope of a chord such as CF. Thus, after it has passed back through the bottleneck, the shock wave will be just like the ordinary shock wave in the rear of any traffic hump (§4).

These considerations enable the course of the hold-up, and its approximate duration, to be determined graphically if the approaching hump is known, for example, if the variation of flow with time has been measured at some upstream point. The situation is little changed if there is already a shock wave in the rear of the approaching hump, as is likely in practice to be the case. When this meets the 'bottleneck shock wave', the two shock waves 'unite', a familiar process in

* The fact that increases of concentration from values well below k_m to values well above it are normally made (as here) by means of shock waves, explains why (as noticed in §2) the maximum flow q_m of a road is not often observed.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

gas dynamics. No alternative behaviour is possible, as whatever they become has got to change the flow and concentration from their values behind the hump shock wave to the values associated with the bottleneck crawl. This could not be done by means of two shock waves, for example, because the one behind, which has to make the first increase of concentration, would have a greater speed than the one in front, which is responsible for the final increase to the crawl concentration; this relationship between speeds follows inevitably from the fact that the flow-concentration curve is convex upwards, but, on the other hand, is geometrically impossible since both waves must start at the same time.

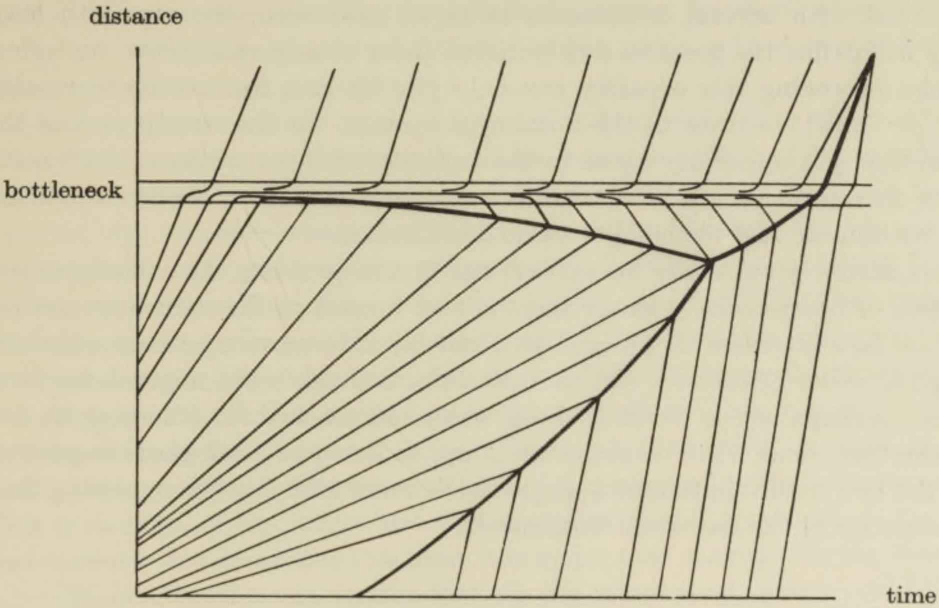

FIGURE 14. Resolution of 'crawl' by arrival of the shock wave in the rear of oncoming traffic hump.

The case when a bottleneck crawl is resolved by the union of the shock wave in the rear of the approaching traffic hump with the 'bottleneck shock wave' is illustrated in figure 14. The path of the shock wave formed by this union is easily traced, since it is still governed by the condition that the flow in front of it is equal to the capacity of the bottleneck—the concentration taking the *greater* of the two values compatible with this flow rate *upstream* of the bottleneck, and the lesser one downstream of it. It is important to notice that the only data required for estimating the course of a bottleneck hold-up in this manner are the flow-concentration curve for the main road, the capacity of the bottleneck, and the variation of inflow with time measured at some upstream point.

As a final theoretical point, it may be noted that the flow near the bottleneck during the crawl is steady. It has often been remarked that the increase of speed on the passage of vehicles (or crowds) through a bottleneck under steady conditions is similar to the effect of a Laval nozzle on the flow of a gas. The above analysis shows how close the similarity is. Upstream of the bottleneck the waves

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

are propagated upstream (as sound waves can be in subsonic flow); downstream of it they are propagated downstream (as sound waves must be in supersonic flow). As the centre of the bottleneck is approached, the mean speed v is increased, and the wave velocity relative to the mean vehicle speed (namely, $v-c$) is decreased, so that both are equal, just as the fluid velocity equals the velocity of sound in the throat of a Laval nozzle. The only essential difference* between the two situations is that the gas is able to transmit disturbances forwards as well as backwards relative to the mean flow. It is this that made the above analysis of the *transients* in the traffic flow problem so much easier than it is in the problem of the Laval nozzle.†

On a road with several bottlenecks in rapid succession, the one with least capacity will define the greatest flow possible under steady conditions. An inflow of vehicles exceeding this capacity can only pile up in a continually increasing 'queue' or 'crawl' in front of the bottleneck system. In the steady part of the flow, the flow q is uniformly equal to the capacity of this narrowest bottleneck, while the concentration k takes the larger value appropriate to this flow upstream of that bottleneck, and the smaller value downstream.

The transients could easily be worked out in this problem. As a hump enters the system of bottlenecks, a shock wave is first formed at the narrowest one (at least if the flow increases slowly enough), and begin to move upstream. If there is a slightly wider bottleneck farther upstream, a shock wave might later form there too, perhaps before the first shock wave had reached it. However, in due course the first shock wave would catch it up, as its speed backwards is greater, and so the two would unite into a single shock wave reducing the oncoming flow to the capacity of the narrowest bottleneck.

6. Some notes on traffic flow at junctions

In this section we attempt a preliminary study of how the method of this paper might be used to predict traffic behaviour at road junctions of various kinds. First, we consider junctions which are not 'controlled' (either by police or by traffic lights).

The simplest junctions are those where minor roads introduce new traffic on to, or abstract traffic from, such a long arterial road as has been considered in the preceding sections. This is normally achieved without significant impedance to the traffic on the major road. Vehicles wishing to enter it have to wait until they can do so without causing obstruction. Vehicles leaving the major road have often to

* A less essential, though more spectacular, difference is that in the traffic problem the typical 'unchoked' flow is totally supersonic, instead of totally subsonic. But in both problems both possibilities exist.

† Students of gas dynamics may wonder, on reading this, whether a rough approximation to the calculation of transients in a Laval nozzle might not be made by regarding them as kinematic waves, on the approximation (*accurate* only for steady flow) that the stagnation enthalpy is everywhere constant. This is found to give the wave velocity as $v-a^2/v$ instead of $v-a$ (where a is the local speed of sound), so that its quantitative value would be small, but it might indicate qualitative behaviour reasonably correctly.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

slow down, or even stop for a time, before they can leave it, but they usually signal their intention in time to enable vehicles behind to pass them on the appropriate side with little loss of speed.

The effect of such a junction on a wave moving past it along the major road, is then to change the flow carried by the wave by an amount equal to the 'mean net inflow rate' from the minor road. This rate is defined as the difference between inflow and outflow, smoothed (as a function of time) by averaging it over such a time τ as was considered in §2. If the road is 'crowded', in the sense defined in §2, the change in flow will change also the speed c of the wave, as well as the mean vehicle speed v. In a space-time diagram, therefore, the waves bend slightly at junctions (backwards where the net inflow is positive, forwards where negative). These rules enable the arterial road theory of §§4 and 5 to be corrected for minor inflows and outflows at junctions.

However, there is a limit to the amount of inflow (especially) which is possible under those conditions. Further, this limit becomes more and more reduced as the flow on the major road increases. These are truisms. It might be thought, however, that the limit was just that increase of flow which would be required to raise the flow on the major road to the maximum possible. The real limit, however, is always much less than this. For inflow under most conditions can occur only when gaps in the traffic pass the junction. As the flow increases, such gaps become rarer and rarer, and for large enough flows, but still well below the capacity of the road, the gaps may be too rare to permit any significant inflow at all.

At cross-roads, where some traffic on the minor road seeks to cross the major road, a closely similar limit exists on the *total flow* originating from the minor road. (This is the sum of the inflow and the cross-over flow.) Evidently, if the minor road carries a flow exceeding this limit, the major road may act for the time being as an effective bottleneck, for the flow on the minor road, which could then be treated by the theory of §5.

It will now be clear why stoppages occurring at junctions under heavy flow conditions can often be resolved by sending a policeman to control the junction. If he stops successively the traffic on the major, and then on the minor, road, the flow originating from each will be approximately the maximum for the road during nearly all the period when the other road is stopped (see §2 above, and also the discussion which follows). The total flow can therefore be made fairly near to this maximum (or, if the capacities of the roads are different, to a weighted mean of them), and this will be greater, as just explained, than what can be achieved under uncontrolled conditions. To achieve best results, the policeman gives each road a time allocation proportional to the flow originating from it. Where traffic lights are installed, one can allocate times on the basis of a mean ratio of flows over an extended period, or else use a vehicle-actuated system of a type calculated to give a better approximation to the optimum at any instant.

Where major roads meet at the same level, a roundabout is preferable to a simple controlled crossing. For this to remain effective under the heaviest traffic conditions, the circular arcs of road which compose the roundabout should each have a capacity equal to one-quarter of the sum of the capacities of the roads radiating

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

from it. For on the average each vehicle uses half the total number of arcs, so that the average flow in an arc will be half the total inflow, or one-quarter of the total flow (inflow and outflow) on all the radial roads. When there are four of these, the argument indicates a width for each arc equal to that of one of the radial roads. Since excessive width for the arcs reduces safety, it may be that these limits should be closely followed.

To conclude the paper, we describe an attempt to discover whether the theory can be successfully applied to flows on a small scale, by using it to predict the effect on the oncoming flow of the compulsory stops and starts at a controlled junction.

First, consider the effect of a sudden stoppage (as when traffic lights turn red) on a *uniform* oncoming flow. It sends a shock wave back into the oncoming stream, at which the flow is reduced to zero and the concentration increased to approximately k_j, the maximum concentration of which the road is capable. (As when the union of two shock waves was discussed in §2, there is no possibility but a single shock wave in this situation, since if there were more than one wave involved the velocity forwards of the wave making the *first* increase in concentration would have to be *greater* than that of the others, and this is impossible because all originate at the same place and time, and the first wave must be at the rear.) The speed of the shock wave is the slope of the chord on the flow-concentration curve which joins the point representing the oncoming flow to the point $(k_j, 0)$.

A more difficult question, where the limitations of the theory become apparent, is what happens when the traffic is permitted to flow forward again (as when the lights turn green; we ignore, to start with, complications due to some vehicles seeking to turn right or left at the junction). The solution, when the assumptions of the theory are retained without change, will first be given in detail (it was already indicated in §2) and afterwards criticized.

The front vehicle can accelerate unhindered to a speed characteristic of an unimpeded road, but the theory ignores the time taken for adjustments of speed (consequent on changes of concentration) to be made. Hence, it represents the front of the stream as moving off instantly at a mean velocity equal to the 'free' mean speed v_F. The wave velocity is also v_F, both being the slope of the flow-concentration curve at the origin. At the same time a wave starts backwards through the stream of waiting vehicles, giving the signal to start. This has a (negative) velocity equal to the slope c_j of the flow-concentration curve at the right-hand limit (corresponding to 'jam' conditions). In between these two extremes there is room for waves of all intermediate velocities, each carrying a corresponding mean vehicle velocity. Since in conditions when the wave velocity is greatest in front there is no tendency for waves to run together and form shock waves, we may suppose that only continuous waves will be present and so that the increase in speed will be achieved through a fan of waves of all possible velocities.

Figure 15 shows the shock wave produced when the lights turn red, and the postulated fan of continuous waves appearing when they turn green, in a space-time diagram. A typical vehicle path is shown as a broken line. The stationary wave which remains at the stopping point is that referred to in §2; the flow across

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

this point is q_m. Figure 15 shows also the 'weakening' of the shock wave when it is caught up by the fan; evidently, its speed must be rapidly reduced when the flow behind it begins to climb up the flow-concentration curve.

If the period of stoppage ('red period') is T_r, and the period of permitted flow T_g, then on the average the total number of vehicles $q_i(T_r+T_g)$ coming up (at the inflow rate q_i) during the complete cycle will pass across the stopping point during the time T_g only if

$$q_i(T_r+T_g) < q_m T_g. \tag{13}$$

This sets an upper limit

$$\frac{T_g}{T_r+T_g} q_m \tag{14}$$

to the inflow (from the road in question) which can be handled by the controlled crossing, without leading to a queue of increasing length. This limit (14) is the 'capacity' of the controlled crossing, when regarded as a bottleneck.

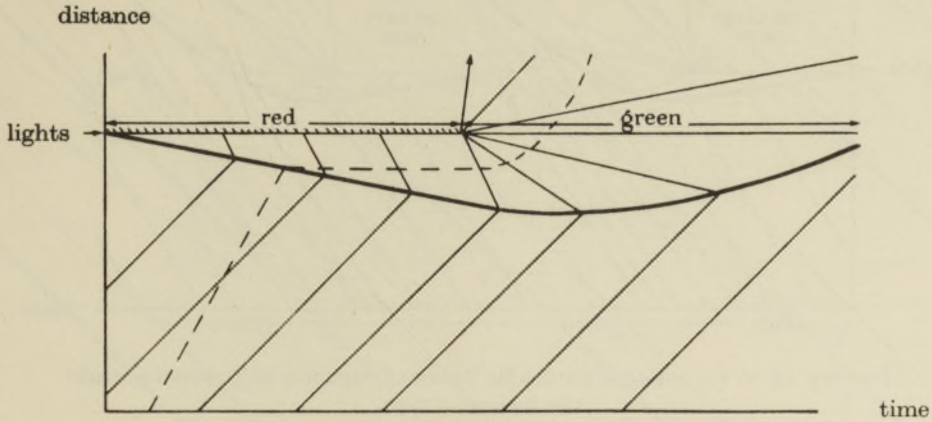

FIGURE 15. Uniform incoming flow stopped for a time and then started again.

If condition (13) is satisfied, that is if the inflow is less than the capacity, then the maximum flow q_m at the stopping point cannot be maintained during the whole period T_g, but only for a reduced period of length T_f (during which the crossing is running 'full') given by the equations

$$q_i(T_r+T_f) = q_m T_f, \quad T_f = \frac{q_i T_r}{q_m - q_i}. \tag{15}$$

After a time T_f, then, the flow ceases to be that carried by the wave which remains at the stopping point, and this must be because the shock wave in figure 15 has moved forward again and passed through the stopping point. This is illustrated in figure 16. Behind the shock wave the flow is the undisturbed inflow q_i. After passing through the stopping point it is just the ordinary shock wave in the rear of any traffic hump (§4).

A simple construction for the path of the shock wave is obtained as follows. The number of vehicles crossing a wave such as OA in figure 16 (on which the flow is q and the concentration k, and whose speed is c) is $(q-kc)\,t$, by §2, if t is the time difference between O and A. This number of vehicles must equal the number

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

$q_i(t+T_r)$ going up to the stopping point in the time $t+T_r$ since the stream was first stopped, minus the number $k_i ct$ left at time t in the distance ct between the stopping point and A.

Thus
$$(q-kc)\,t = q_i(t+T_r) - k_i ct, \tag{16}$$

or
$$t = \frac{q_i T_r}{q - q_i - c(k-k_i)}. \tag{17}$$

This equation can be used, with $x = ct$, to trace the shock-wave path on the (x, t) diagram, if in both k is varied from 0 to k_j, the corresponding values of q and of $c = \mathrm{d}q/\mathrm{d}k$ being deduced from the flow-concentration curve. (Note that equation (15) is a special case of (17), with $t = T_f$, $c = 0$, $q = q_m$.)

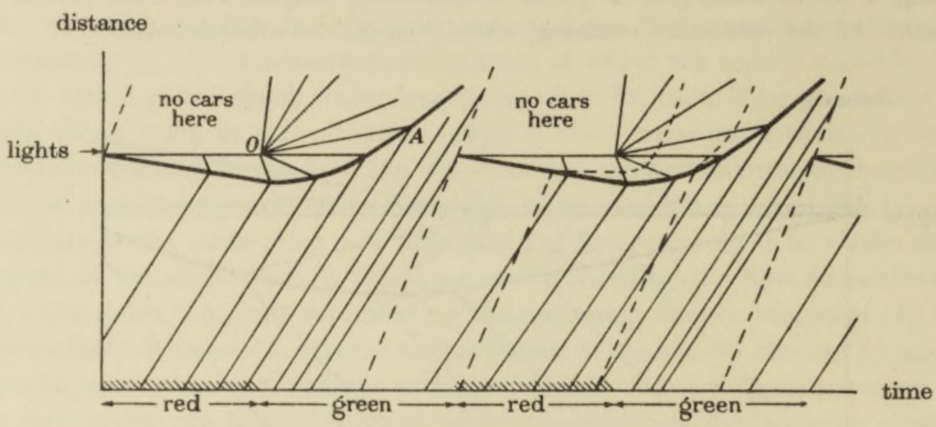

FIGURE 16. Wave pattern for traffic lights of capacity sufficient to admit the incoming flow.

If the incoming flow begins to exceed the capacity of the controlled junction, the shock waves of figure 16 do not get clear of it in time; each then collides with the shock wave sent out at the beginning of the next stopped period. They unite, and in turn collide with the next shock wave, and so on. If the excess incoming flow is maintained, these collisions (of shock waves, not cars!) must occur farther and farther back, and thus become a less and less significant feature of the situation. When this has happened the residual behaviour indicated by the theory is quite simple. Each shock wave (figure 17), on being formed at the stopping point, reduces the full flow q_m to rest. As it moves backwards (e.g. at G) the traffic it stops is travelling more slowly. From D onwards (figure 17) however, it does not reduce the flow completely to rest. Finally, at a very large distance behind the stopping point (e.g. at A) little reduction in vehicle speed occurs at shock waves and their effect has almost been ironed out into a typical bottleneck crawl.

The quantitative details of this familiar oscillating-speed crawl behind a choked controlled junction can be obtained by a device similar to that used above in the unchoked case. At a point in figure 17 such as A, a distance x behind the stopping point, let the values of k, q and c carried by the waves which run into the shock wave at A be distinguished by suffixes 1 (for the first) and 2 (for the second). Then

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

the difference in the number of vehicles crossing the whole of each of these waves is equal to the flow across the stopping point during the time T_g; in symbols,

$$(q_1 - k_1 c_1)\frac{x}{(-c_1)} - (q_2 - k_2 c_2)\frac{x}{(-c_2)} = q_m T_g. \tag{18}$$

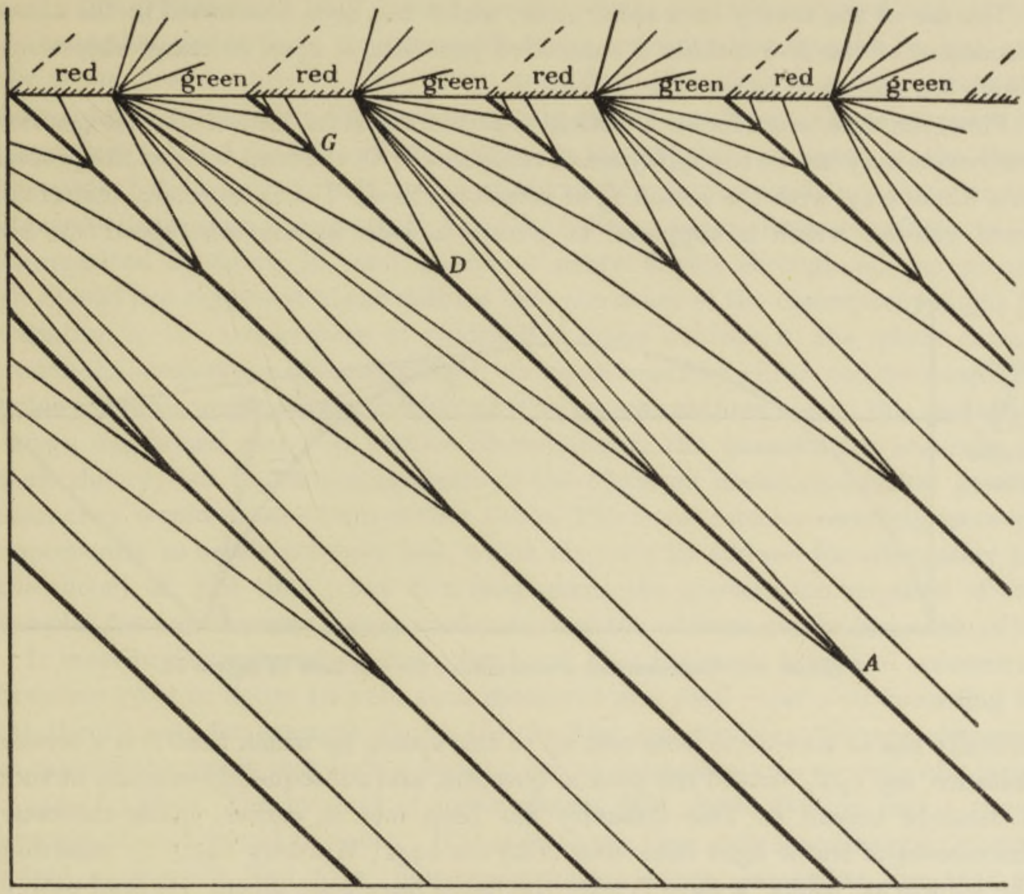

FIGURE 17. 'Crawl' generated by traffic lights when inflow exceeds capacity.

Note that the wave velocities c_1 and c_2 are negative, so that the times taken by the waves to reach A are $x/(-c_1)$ and $x/(-c_2)$ respectively. Also, since these differ by an amount $T_r + T_g$, we have

$$\frac{x}{(-c_1)} - \frac{x}{(-c_2)} = T_r + T_g. \tag{19}$$

Eliminating x from (18) and (19), we can write the result as

$$k_1 - \frac{1}{c_1}\left(q_1 - \frac{q_m T_g}{T_r + T_g}\right) = k_2 - \frac{1}{c_2}\left(q_2 - \frac{q_m T_g}{T_r + T_g}\right). \tag{20}$$

Geometrically, this means that a line drawn across the flow-concentration curve at a level corresponding to the capacity of the controlled junction (figure 18) will have the property that tangents drawn to the curve (e.g. AB and AC, or DE and DF) from points on the line have slopes equal to the slopes of waves meeting

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

at points on the shock wave (e.g. A or D in figure 17) where there is a transition between the states represented by the two points of contact of the tangents (e.g. B and C, or E and F). When one of the tangents cannot be drawn, the point F (where $q = 0$ and $k = k_j$) must be used as an end-point instead (see e.g. GH and GF corresponding to the point G in figure 17).

The use of the theory on a small scale, which has been illustrated in the above discussion of the flow behind a controlled junction, is open to many objections, which are discussed below, one by one.

First, the time taken for each vehicle to accelerate to its desired speed is ignored, whereas it may not be negligibly small compared with the time scale of the process as a whole (say, with the period T_g of permitted flow). This is especially true of the front vehicle, which is supposed to pursue a path at constant speed v_F, but

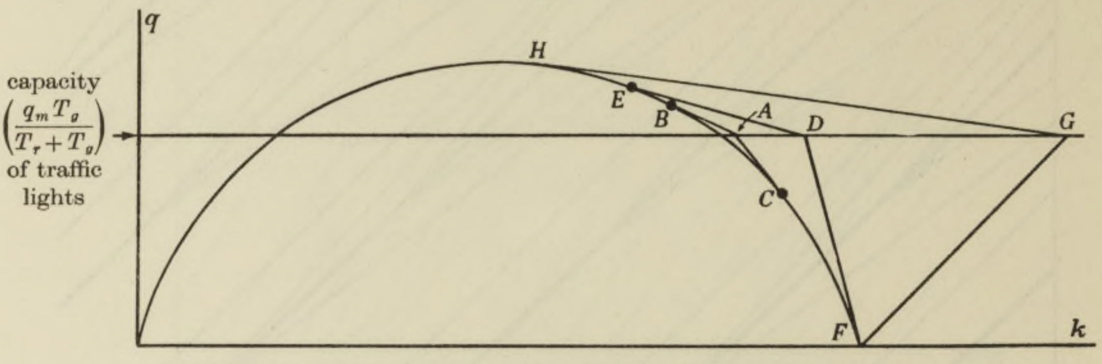

FIGURE 18. Geometrical construction for the flow of figure 13.

actually has to accelerate from rest up to this speed, by which time it is a certain distance, say $v_F T_l$, behind the path in question, and subsequently remains at such a distance behind it. This difficulty has been met in earlier, queue-theoretic, discussions of traffic light behaviour (Clayton 1941; Wardrop 1952) by regarding the vehicle as 'losing' a time T_l after the lights have gone green. Its final path is that which it would have if it accelerated to speed v_F *instantaneously* after an initial delay T_l. It is possible, therefore, that the present theory may be reasonably correct provided that the period of stoppage T_r is taken to include this 'lost time' T_l, which must in turn be deducted from the period of permitted flow T_g. The 'lost time' is of the order of 5 to 10s. The existence of this lost time is an important argument for keeping the periods of stoppage and permitted flow fairly long, so as to achieve a total flow at the junction as near to q_m as possible. Conversely too great a period increases average vehicle delay, and Wardrop (1952) has shown that there is in any given case a cycle length which renders this average delay a minimum.

A second objection is that the theory ignores the fluctuations in inflow over times comparable with T_g or T_r. It is just these fluctuations which lead to the phenomena (alternating quiet and 'busy' periods) studied in the theory of queues. Another way of phrasing the objection is to say that the times T_g and T_r are not large compared with the time τ needed (§2) to obtain smooth mean values of flow

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

and concentration. This objection is certainly valid under relatively easy traffic conditions. It seems likely, however, that when the road is 'crowded' in the sense used in this paper the general picture of the starting flow given above may be relevant, the variations of inflow serving only to alter the positions of the shock waves at any instant.

A third objection is that certain measurements of traffic stopped and started indicate that under these conditions the mean concentration may be far less, and the mean speed far greater, for a given mean flow, than the values taken from the flow-concentration curve under more nearly steady conditions. Measurements on these lines known to the author include an unpublished set made at the Road Research Laboratory in 1954 and a study of flow in the road tunnel under the Meuse at Rotterdam (Aangenendt, Van Gils & Boost 1951). Both sets of results are summarized in figure 19, and a smooth curve drawn through all the points. Dr Smeed has suggested to the authors that the cause of the discrepancy might be variation in the acceleration of vehicles: if many vehicles in the queue cannot match the acceleration of those in front, the mean headway will exceed the minimum value tolerable under steady conditions. The present authors regard this and other causes mentioned above as serious limitations on the quantitative accuracy of their theory, but find the magnitude of the observed departures rather greater than they would expect from such a cause. The front vehicles certainly have an opportunity to accelerate very fast, which may not be allowed for adequately by the theory of 'lost time'; but in a long queue the acceleration required of the vehicles towards the rear is very moderate, and few of them can be incapable of it.

It must be remembered, on the other hand, that the mean flows and concentrations recorded in figure 19 were each measured *at a fixed point*, and according to the theory (see, for example, figure 16) the flow and concentration are changing so rapidly at a point that such a method can at most obtain an average of a large range of values. The method of measurement (§2) shows that in fact a time-average of each would be taken. It is easy to see that such an averaging process would in practice lead to a 'flow-concentration curve' somewhat like that of figure 19. Far behind the stopping point the mean would be taken over a period of very high values of concentration (and low values of flow) and (after the shock wave has passed forward) a longer period of very low values of concentration (and only moderate values of flow). The means would then be well in the left-hand half of the area under the true flow-concentration curve, and near the bottom. But, near the stopping point, there would be a longer period before the shock wave moves forward, and for much of this period the flow would be near its maximum. When averaged with a shorter period of low concentration this would give points in the middle of the area under the curve, somewhat below the top.

Future experiments will perhaps show whether or not this is the major cause of the discrepancy revealed in figure 19.* In the meantime, it is perhaps worth

* Very recent work by Wardrop has already gone some way towards confirming this. By taking means over very short distances (only twice the headway) and replotting the flow-concentration curve, he obtains points lying on a curve which is at least parallel (instead of perpendicular!) to the 'headway' curve of figure 19, although somewhat below it.

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

noting one or two directions in which the present theory could be improved in its application to small-scale flows. First, there is the 'blend with statistical ideas' suggested in §1, but this is too difficult to be treated briefly, and the compounding of this blend is postponed to a later paper.

A second extension is to exclude a 'diffusion' effect due to the fact that each driver's gaze is concentrated on the road in front of him, so that he adjusts his speed to the concentration slightly ahead. This gives a dependence of flow on concentration gradient, which leads to an effective diffusion exactly as noted in part I, §6. Such diffusion 'spreads out' the shock waves; in fact, drivers do **not** reduce speed instantaneously at shock waves, because they see them coming.

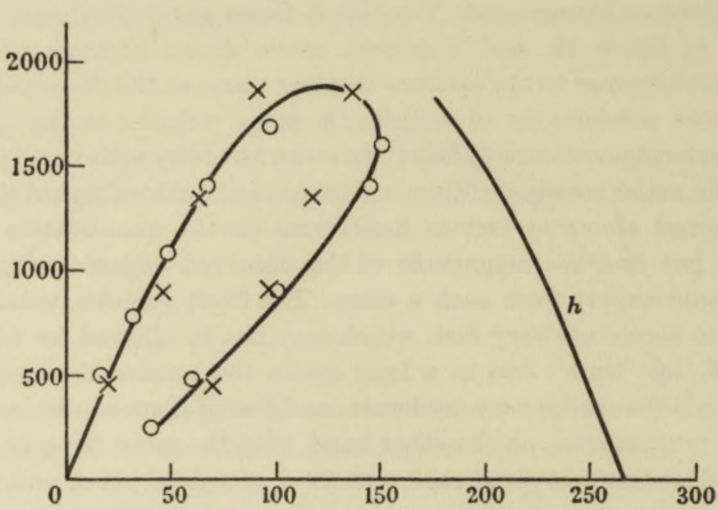

FIGURE 19. Observed mean flows and mean concentrations in traffic stopped and started. ○, Rotterdam tunnel; ×, Road Research Laboratory; h, 'headway' curve.

A third extension is to include an 'inertia' effect due to the fact that a driver must apply accelerator or brake to reach his desired speed and neither is instantaneously effective.

When both the last-named extensions have been applied, one reaches an equation of motion of a general form

$$\frac{\partial q}{\partial t}+c\frac{\partial q}{\partial x}+T\frac{\partial^2 q}{\partial t^2}-D\frac{\partial^2 q}{\partial x^2}=0, \tag{21}$$

where T is the inertial time constant for adjustments of speed, and D is the diffusion coefficient, or decrement of flow for unit concentration-gradient. This is very similar to the equation governing waves in rivers (part I) when higher-order effects are taken into account. The new terms may be expected to introduce similar additional effects in traffic flow on a small scale to those found in certain river flows. In particular, something analogous to 'roll waves' might sometimes arise, in which a uniform flow is unstable and tends to degenerate into a succession of rapid accelerations and even more rapid retardations. This sometimes happens to a long

Downloaded from http://rspa.royalsocietypublishing.org/ on October 5, 2018

convoy of vehicles which are expected to keep in line, even when the front vehicle maintains a uniform speed.

The behaviour of the flow behind a controlled junction could in principle be evaluated on the basis of an equation such as (21), but until the experimental information is clearer such extensions of the theory would seem to be premature.

References

Aangenendt, J. J. M., Van Gils, J. F. L. & Boost, A. G. M. 1951 Studies of traffic and of the conditions which influence it. *Ninth Congress of the Permanent International Association of Road Congresses*, Sec. 2, Qun. III, Rep. No. 42 (Paris).

Charlesworth, G. 1950 Methods of making traffic surveys especially 'before and after' studies. *J. Instn Highw. Engrs*, **1**, 42.

Clayton, A. J. H. 1941 Road traffic calculations. *J. Instn Civ. Engrs*, **16**, 247.

Glanville, W. H. 1949 The study of road traffic. London: Instn Civ. Engrs. (Joint summer meeting with Inst. Civ. Engrs, Ireland, pp. 41–68.)

Glanville, W. H. 1951 Road safety and road research. *J. Roy. Soc. Arts*, **99**, 114.

Glanville, W. H. 1953 Road research and its bearing on road transport. *J. Inst. Transp.* **25**, 186.

Lighthill, M. J. & Whitham, G. B. 1955 On kinematic waves. I. Flood movement in long rivers. *Proc. Roy. Soc.* A, **229**, 281.

Greenshields, B. D. 1935 A study of traffic capacity. *Proc. Highw. Res. Bd, Wash.*, **14**, 448.

Normann, O. K. 1942 Results of highway capacity studies. *Publ. Rds, Wash.*, **23**, 57.

Normann, O. K. & Taragin, A. 1942 *Résumé of previously published material on highway capacity*. Washington: U.S. Public Roads Administration.

Normann, O. K. & Walker, W. P. 1949 Highway capacity, practical application of research. *Publ. Rds Wash.*, **25**, 201.

Smeed, R. J. 1952 Traffic and traffic studies in the U.S.A. *Roads and road construction*, **30**, 138–42 and 166–69.

Tanner, J. C. 1953 A problem of interference between two queues. *Biometrika*, **40**, 58.

Wardrop, J. G. 1952 Some theoretical aspects of road traffic research. *Proc. Instn Civ. Engrs*, pt. II, **1**, 325.

Wardrop, J. G. & Charlesworth, G. 1954 A method of estimating speed and flow of traffic from a moving vehicle. *Proc. Instn. Civ. Engrs*, pt. II, **3**, 158.

Whitham, G. B. 1952 The flow pattern of a supersonic projectile. *Commun. Pure and Appl. Math.* **5**, 301.