A Psychophysical Ordered Response Model of Time Perception and Service Quality: Application to Level of Service Analysis at Toll Plazas

Partha Chakroborty[a,+], Abdul Rawoof Pinjari[b], Jayant Meena[c], Avinash Gandhi[c]

[a] Professor, Department of Civil Engineering, Indian Institute of Technology Kanpur, India
[b] Associate Professor, Department of Civil Engineering, Indian Institute of Science Bangalore, India
[c] Frmr. Graduate Student, Department of Civil Engineering, Indian Institute of Technology Kanpur, India
[+] Corresponding Author

Abstract

This work attempts to bring to the fore the importance of explicitly modelling how time (or other physical quantities like distance) is perceived in the analyses of phenomena or behaviors where time (or other physical quantities) plays an important role. To do so the application area of level of service at toll plazas is chosen. Principles arising out of past work on the psychophysics of time perception are seamlessly incorporated into a novel unified model that accounts for both bias and random errors in perception. Multiplicative error terms arise naturally in this model. Results indicate that it is important to include both bias and random errors while modelling perception as ignoring the bias leads to high variance in the error term. Interestingly the pattern of bias in perception of time seems to remain steady across various waiting line situations from disparate contexts. Finally, the proposed model creates a framework that can be used to determine how time durations are perceived by humans when responses about perceptions are provided through categories like very small, small, etc. whose definitions are also concurrently identified. Such a framework can be used to study a wide range of situations both in transportation (such as route choice, mode choice, gap acceptance, and overtaking) and elsewhere. From a narrower perspective, this methodology can be used, as has been done here, to (i) determine level of service category definitions in terms of observable and "engineerable" variables (actual or measured delay faced by vehicles) even though it is the perceived value (a latent variable) that determines travelers' responses on the quality of service, and (ii) identify parameters that define the systematic bias and randomness in perceptions.

1. Introduction

The duration of time spent in transportation is a concern for users as well as planners and engineers. The quality of service that a transportation facility provides to its users is dependent on how long one spends in that facility to get service. For example, at facilities such as signalized intersections and toll plazas level of service (LOS) is directly related to the delay one experiences. It is not surprising then that many transportation facilities are designed to minimize the time users take while not compromising safety.

Although transportation planners and engineers often design facilities to minimize service time (i.e., the time spent at the facility in order to be serviced) that is objective and measurable (i.e., clock time), users' perception of the quality of service at the facilities depends on their perceived service time that is subjective and not easy to measure. Indeed, user perceptions of time duration and other physical quantities not only influence their ratings of the quality of service of transport facilities but also govern their behavior in transport systems. Examples include driving behavior aspects such as car-following, overtaking, and gap-acceptance and travel behavior aspects such as mode choice and route choice.

1.1 Use of perceived time duration in transportation analysis

Despite the influential role played by human perceptions of physical quantities such as time and distance in deciding user responses, mathematical models of user response and behavior in transportation systems traditionally incorporate such physical quantities without regard to perceived values of those quantities. To overcome this limitation, three different threads of literature in transportation emphasize the need to account for subjective perceptions as opposed to objective measurements of time duration – (a) literature on level of service (LOS) characterization of transportation facilities, (b) literature on human factors in driver behavior, and (c) literature on travel behavior, each of which is briefly discussed here.

Most of the LOS literature, while acknowledging that LOS is dependent on users' perception of the extant conditions, involves the development of empirical models that directly relate user-stated qualitative LOS categories, typically measured on a Likert scale – such as very good, good, bad, and very bad – to objective measurements of some physical quantity, for example, waiting time duration or delay (see, for example, the origin of the LOS concept in Highway Capacity Manual (HCM), 1965 or its evolution over HCM 1985, 2000, 2010; Indo-HCM, 2017; or other commentaries by Choocharukul et al., 2004 or Kikuchi and Chakroborty, 2006). Some studies on LOS at signalized intersections attempt to develop a direct relationship between the user-stated LOS and their subjective perceptions of time duration (Wu et al., 2009; Othayoth et al., 2020), and how various factors other than time duration influence user's perceptions and stated LOS categories. It is worth noting here that since user-stated LOS is typically on an ordinal scale (e.g., Likert scale), ordered response models have often been used to relate perceived LOS to measurable physical quantities. However, most of the LOS literature ignores the psychophysics of how users translate actual waiting time into perceived time, which in turn determines the user's stated LOS of the facility.

The driving behavior literature pays significant attention to the human factors involved in the drivers' translation of the microscopic traffic environment around them into perceived stimuli (Wiedemann, 1974; Treiber and Kesting, 2013; Saifuzzaman and Zheng, 2014). While most of this literature does not focus explicitly on the time duration dimension, studies in this area have accumulated substantial evidence that drivers' perceptions of distances and speeds are associated with systematic biases (see, for example, Nilsson, 2000; Castro et al., 2005). Majority of this literature (except van Lint and Calvert, 2018) models perception errors as stochastic processes (Treiber et al., 2006; van Lint et al., 2017) with exogenously assumed distributions as opposed to an explicit model that translates a physical quantity from the traffic environment into a perceived stimulus. Interestingly, these studies do not appear to build on psychophysical laws of human perception of physical quantities, which is an established line of research in the field of psychology.

The travel behavior literature has long recognized that subjective perceptions of travel time have stronger influence than objective travel times on travellers' choices (Clark, 1982; Li, 2003; Grisolia and Ortuzar, 2010; Malokin et al., 2019). More importantly, a stream of studies in this area draws from a widely applied psychophysical law of human perception of physical stimuli, called the Stevens' law (Stevens, 1957). According to this law, as applied to the perception of time duration, perceived duration (d) is a power function of objective duration (D), as below:

$$d = aD^b$$

In the above equation, a and b are the parameters of the perceived duration function. The use of power law in travel behavior research dates to Burnett (1978) and Clark (1982), who used a simple, log-log regression to implement the power law for developing empirical relationships between perceived travel times, surrogated by reported travel times of travellers, and objectively measured travel times (also see Li, 2003

and Tenenboim and Shiftan, 2018 for reviews on the use of Stevens' law for analyzing perceived travel times). Interestingly, log-log regression between perceived and objective time durations were employed by other researchers as well (Fan et al., 2016; Carrion, 2013), albeit without recognizing the power law as an underlying psychophysical mechanism. In other studies (Varotto et al., 2017), perceived travel time is treated as a latent, stochastic variable, without incorporating systematic bias in human perceptions due to non-linear distortion of objective time durations that Stevens' law helps model.

1.2 Gaps in research

Important gaps in research may be synthesized from the above discussion. First, the application of Stevens' power law requires the availability of perceived time durations in addition to objective time durations. Only then can one develop empirical relationships between the perceived time and objective time for subsequent use. However, there may be several situations where objective measurements of time durations are available along with the travelers' discrete choices that are easy to measure – for example, their mode choices and route choices on a nominal scale or their stated LOS responses on an ordinal scale – but subjective measurements of time durations are neither available nor easy to rely on. In such situations, as will be discussed in detail in Section 4.2, it is not possible to estimate the parameters of Stevens' power law to be able to link individual choices to their perceived times. This is one reason why most past LOS studies directly link individual's stated LOS categories to objective measurements of time duration, as opposed to connecting their LOS ratings to an underlying psychophysical process that translates objective measurements to perceived time durations. Therefore, it is important to develop an alternative functional form for the psychophysical model of time perception whose parameters can be recovered using only objective measurements and individuals' choices such as LOS ratings.[1]

Second, Stevens' law implies that (i) there will necessarily be a point at which the perceived time duration is veridical and (ii) perceived time durations must necessarily be systematically biased upward (downward) when the objective times are small (large) in magnitude. In many waiting and travel situations, however, the perceived time durations are generally greater than objective time durations albeit to different extents depending on the magnitude of the objective durations. This is a consistent empirical finding across a variety of studies (Henley et al., 1981; Hornik, 1984; Carrion, 2013; Peer et al., 2014; Fan et al., 2016). As discussed in Tenenboim and Shiftan (2018), waiting time and travel time situations, which the individual is in limited control of, faces uncertainty due to external factors, and would well avoid if given a choice, might generate pessimistic perceptions of the time endured. Therefore, it would be useful to have an

[1] One might question why not measure perceived time durations along with objective durations? Doing so requires carefully constructed experimental setup and/or framing of questions. The most widely used approach to measure perceived time durations of any event – whether it is travel, waiting at a signal or a toll plaza – is to ask travelers to report the time duration post the event. Such "reported times" based on a memory recall of the event are associated with their own errors beyond perception errors. As Peer et al. (2014) discuss, reported times need not be the same as perceived times since the individuals are not necessarily aware of their true perception. Further, in situations where people choose from alternatives with different travel times, post-trip reported times may be associated with a justification bias, where the respondents might subconsciously distort their perception of the travel times to justify their choice. Another problem with reported times on a continuous scale is their highly "quantized" nature where humans tend to respond using only a few discrete quantities like 30 seconds, 40 seconds etc. and hardly ever quantities like 33 seconds, 47 seconds., etc. (for example, see Ogden et al., 2020). As regards to the problems associated with memory recall in reported times, one way to reduce them would be to query the travelers immediately after their experiences. However, doing so in real-world situations may not be easy as it may interfere with the tasks that the driver may have at hand. For example, in the manual toll plaza case such querying can only be done during the short interval in which the traveler waits for his payment receipt or change. This limits the number and complexity of the questions that can be posed. Further, since LOS of a traffic facility is a statement of how users of the facility evaluate the service, it becomes imperative to ask the users to classify their evaluation into linguistic categories such as "Very Good", "Good", etc. Therefore, eliciting user responses on their perceived durations on a continuous scale alone does not help with LOS and other such analyses involving users' evaluation of the service they receive. Additionally, asking multiple questions so that responses are available in both linguistic and continuous scales while minimizing the impact of "memory recall" may not always be possible due to constraints on the time available for querying an individual.

alternative functional form for a psychophysical model that has the flexibility to represent perceived durations that are always an overestimate, but to varying degrees, of the objective durations.

Third, while Stevens' law can be used to understand systematic biases in perceptions, this law does not discuss variability due to inherent uncertainty in individuals' perception of time. Yet, it is not inconceivable to expect variability in an individuals' perceived time durations across a series of independent and identical events of equal objective duration (Grondin, 2010). That is, there can be both systematic bias as well as random variability in an individuals' perception of time durations. In this context, another observation called the Weber's law is commonly cited by researchers working in the field of time perception. As will be discussed in Section 3, Weber's law may be interpreted as a representation of a common observation that uncertainty in perception increases with the magnitude of the objective duration being perceived (Allan, 2001). In summary, it would be useful to include such aspects relevant to uncertainty along with the elements relevant to systematic bias in psychophysical models of time perception.

To be sure, the log-log regressions carried out between perceived time and objective time in many transportation studies (Burnett, 1978, Clark, 1982; Fan et al., 2016; Tenenboim and Shiftan, 2018) implicitly include elements of uncertainty, because the log-transformed statistical regression model includes an additive, normally distributed random error term. When such log-transformed regression equations are reverted to the exponential form, the resulting error term becomes multiplicative and log-normal. However, none of these empirical studies motivate any theoretical reasons for the multiplicative nature of the error terms, nor do they connect the error term to psychophysical aspects of time perception. The log-log and log-linear regressions are carried out in many such empirical studies primarily because the dependent variable, perceived time duration, is positive.

Some other studies (Varotto et al., 2017) that treat perceived time duration as a latent, stochastic variable use an additive error specification to recognize the difference between perceived and measured time durations. However, an emerging stream of literature prefers the use of multiplicative error specification over the additive specification for various reasons. For example, Varela et al. (2018) demonstrate that the use of multiplicative terms to represent errors in travel time variables provides a better fit to travel mode choice data than the use of additive error terms. In another line of research, Fosgerau and Bierlaire (2009) derive discrete choice models with multiplicative error terms on utility functions. More recently, Nirmale and Pinjari (2020) demonstrate that variables representing users' perceptions of the choice environment ought to be specified with multiplicative stochasticity for better identification of unordered response discrete choice models with such explanatory variables. However, none of these studies delve into the behavioral reasons why multiplicative errors might be better than additive errors nor connect their models to psychophysics of human perception. In summary, the nature and specification of stochasticity in a model of time perception should be motivated from theoretical/behavioral standpoints as opposed to being a purely statistical exercise based on model fit or identification considerations.

1.3 Current research

This paper proposes a general psychophysical model of time perception that satisfies behavioral patterns expected from two widely cited laws on human perception of time duration – Vierordt's law (1886; from which Stevens' law can be motivated) and the Weber's law. The proposed model accommodates both bias and random error in perceptions and, therefore, allows the analyst to separate systematic bias from random variability in perception. In the context of bias, unlike the Stevens' law, the proposed model is versatile enough to allow situations where the perceived time duration is always overestimated (although to varying degrees) when compared to objective time duration. Depending on the parameter values, the model also allows situations akin to those implied by Stevens' law where perceived time and objective time have a

veridical point. In the context of random variability, the model embeds a multiplicative stochastic term that makes it easy to accommodate an observation based on Weber's law that uncertainty in perception increases with the magnitude of the time duration being perceived.

Importantly, the functional form of the proposed model is such that its parameters are estimable even in the absence of measurements of perceived time durations, as long as data is available on objective time durations and users' rating of their perceived value on an ordinal scale (say, LOS categories). To do so, in the current paper, the proposed psychophysical model of time perception is embedded into an ordered response model that relates objective time durations to users' rating (on a Likert scale) of their perceived values. This not only helps in eliminating the difficulties and uncertainties in measuring perceived time durations, but also provides a psychophysically grounded method to relate user's rating of time perceptions on an ordinal scale to objective measurements of time durations.

The proposed method is applied to model user's LOS at toll plazas based on the delay (or waiting time) they experience. Carefully collected empirical data from three toll plazas in northern India is used in the study. Using this data, the empirical analysis demonstrates how one can use the proposed ordered response model to develop a mapping between observed, actual waiting time and latent, perceived time by converting the threshold parameters of the model from a latent space of perceived time to objective waiting time units. By doing so, this work showcases how the model can be used to develop a relationship between objective and "engineerable" measure of actual waiting time, which can be controlled through engineering and design measures such as the number of toll booths, and subjective ratings of the LOS. From a behavioral standpoint, the empirical results provide evidence of systematic bias in users' perception of waiting time and that the parameters driving human perceptions of time do not vary across individuals as much as the time thresholds that people use to rate the LOS of toll plazas.

The rest of the paper is structured as follows. Section 2 defines and describes the LOS problem, an application for which the proposed model is developed; albeit it is applicable more broadly. Section 3 starts with a discussion of the laws commonly used in the field of psychophysics of time perception and then develops a general model of perception of time duration. The fourth section develops an ordered response model based on perceived waiting time durations that can be used to determine thresholds of various levels of service on a scale of actual or observed delays. The fifth section describes the empirical data, presents and discusses the empirical results, and relates the results to those from other observational studies. The final section summarizes the insights obtained from this work, identifies limitations, and outlines avenues for future research.

2. Problem Statement for Toll Plaza LOS Analysis

As discussed, this work seeks to develop a psychophysical representation that can be effectively used to model individuals' responses (on an ordinal scale) that are based on their perceived physical quantities such as time. It is important that this development of ideas is illustrated using real-world data from a transportation facility. The facility chosen here is the toll plaza where users (drivers) must either stop or slow down to pay toll. A reason for choosing toll plazas in this study is also because, for such facilities, very few studies have been conducted to determine level of service definitions that can then be used to design them. HCM (2000) explicitly acknowledges the need for dedicated research on toll plazas because of its unique nature while HCM (2010) maintains that toll plaza analysis, or the lack of it, remains a limitation of the manual. Unfortunately, despite the similarities, LOS category definitions obtained for signalized intersections cannot be transferred to toll plazas as the expectations of users from these two

different facilities may be different. This is because toll plazas come much less frequently than signalized intersections and that too on roads that otherwise offer higher operating speeds.

Next, the problem of analyzing LOS at toll plazas is described in greater detail so that this work can serve twin purposes. One, develop an estimable model of time perception that can be used in an ordered response model setting. Two, apply the model to determine definitions of various LOS categories at toll plazas.

Different researchers have suggested different variables impact LOS at toll plazas. The list is long and includes obvious variables like delay or waiting time, queue length (see for example, Lin and Su (1994), Klodzinsky and Al-Deek (2002), Aycin (2006), Obelheiro et al. (2011)) to variables such as aesthetics (Navandar et al., 2019). Based on a review of the literature and an understanding of the operations at a toll plaza the following variables are considered as possible determinants of a person's perception of the LOS at a toll plaza: (i) delay at the toll plaza, (ii) the number of vehicles ahead when the person joins a queue (or queue length), (iii) the number of heavy vehicles ahead when he/she joins a queue, (iv) the variance in service times of the vehicles ahead, and (v) the socio-economic characteristics and contextual factors of users of a toll plaza.

Quality of service or a user's response on how good the service is at a toll plaza is based on perceived values of the service measures and not on the actual values. This perceived state is a latent variable as it is not measurable. Hence, the thresholds that distinguish one LOS category from another are defined on the universe of this latent variable. For example, if delay is used as a service measure, then a user's response on the quality of service is assumed to depend on perceived delay. The thresholds that are assumed to exist in the user's mind are therefore defined on the perceived delay scale. However, from an engineer's perspective these thresholds must be defined on the domain of actual (or objective) delay - an observable and measurable quantity. That is, the analyst must not only be able to estimate the thresholds on the perceived delay scale but also be able to uniquely convert them to actual delay units.

Given that toll plazas require users to queue up and face delay like at signalized intersections, LOS at toll plazas is defined using the waiting time or delay that users face at such facilities. Fortunately, in a queuing system such as the one at a toll-plaza the actual waiting time can be predicted and controlled through engineering variables like number of toll booths and time to collect toll. Second, ordered response models are a natural choice for analyzing LOS at any facility since LOS categories are, by definition, ordered, mutually exclusive and collectively exhaustive. Further, the influence of other variables on LOS is incorporated through their assumed effect on the thresholds between different LOS categories. This implies that the threshold value is assumed to depend on variables like queue length, number of heavy vehicles in the queue, service time variation, socio-economic characteristics of the users, etc. Further exposition on variable thresholds is kept for a later section.

In summary, the premise here is that when a user of a toll-plaza is asked how long they had to wait at the plaza (or how good the service was) the user responds by choosing a category from one of the finite number of ordered classes (Very Long, Long, etc.) based on the perceived waiting time (delay), d and not the measured (objective) waiting time, D; of course, D impacts the perceived value d.

Three tasks remain. First, based on the psychophysics of perception of time durations, a model relating actual delay to perceived delay must be developed. The parameters of such a model should be estimable from observations. Second, the latent variable "perceived delay" must be related to the response of toll plaza users. Finally, good quality data on users' perception of the quality of service must be collected and used to estimate and analyze the parameters of the proposed models. For each user who is surveyed the data must include, response on the quality of service (length of delay) immediately after completion of the service, the actual delay faced by the respondent, the number of vehicles ahead of the respondent when

he/she joins a queue, the number of heavy vehicles ahead, etc. Each of the next three sections covers one of the three tasks described here.

3. Understanding and Modelling Perceived Time Duration

Although there does not seem to be any consensus on the nature of the relationship between perceived time and actual time two "laws" seem to be ubiquitous (for example, see Allan 1979, 2001; Fraisse 1984, Grodin 2010, etc.). One is the Weber's law (or Weber fraction) and other is the class of psychophysical laws that propose to relate the actual (objective) time to the perceived time. Although there are few forms in which Weber's law is cast the one most useful for this paper is that presented in Allan (2001).

Allan (2001) states that Weber's law can be represented as:

$$\frac{\sigma_d}{\mu_d} = w \tag{1}$$

where, the numerator and the denominator, respectively, represent the standard deviation and the mean of perceived time (d); and w is a constant often referred to as the Weber fraction. The above equation can be viewed as a formal representation of a common observation that larger (smaller) values of the quantity being perceived have larger (smaller) variability in perception.

Psychophysical laws of time perception started with Vierordt's observation (1868; also see Lejeune and Wearden, 2009) that in retrospective estimations (where the subject, in the present case a toll payer, does not know in advance that he/she will be expected to estimate the time he/she spent waiting after the waiting is over) short durations are overestimated and long durations are underestimated. In a more generic sense, these laws are based on various observations that perceived (or estimated) times have systematic bias when compared to actual time. One of the popular forms used to capture and model this systematic bias is referred to as the Stevens' power law (Stevens, 1975; Antonides et al., 2002, Grondin and Plourde, 2007). One can recast this law (in the notation used here) as follows:

$$\mu_d = aD^b \tag{2}$$

where D is the actual (objective) duration, and a and b are positive constants. Note, μ_d is a monotonically increasing function of D. Also, a value of b smaller than unity implies that shorter durations, $D < a^{1/(1-b)}$, are on an average overestimated and longer ones, $D > a^{1/(1-b)}$, underestimated, and a value of b greater than unity implies the opposite[2]. Most studies, especially the ones that deal with retrospective estimation of durations in the range that drivers often face at manual toll plazas (up to around 6 to 7 minutes) report a value of b that is significantly less than unity; for example, Brown and Stubbs (1988) report values of 0.32 and 0.38 while Grondin and Plourde (2007) report an average value of around 0.5.

In a more generic sense (as opposed to the specific form given in Equation (2)) one can assume that

$$\mu_d = g(D) \tag{3}$$

where $g(D)$ is a monotonically increasing function of D.

Motivated by the principles expressed through Weber's law and the psychophysical laws the following model relating d to D is proposed in this paper:

$$d = g(D).\varepsilon \tag{4}$$

[2] At $D = a^{1/(1-b)}$, $\mu_d = D$; the point at which on an average perception is veridical. This is the indifference point.

where, ε is random error term to allow variability in perception and specified with a mean of unity and a variance of v (note $\varepsilon \in (0, \infty)$). Later in this work, while estimating the parameters, the assumption of constant variance was also tested in order to examine Getty's (1975, 1976) idea of a generalized Weber's law that implies heterogeneous variance of ε. This test has been done despite observations that support the constancy of Weber fraction. For example, in a relatively recent study Wu et al. (2009) present some data on the mean of perceived waiting time (in the range 0 to 120 seconds) and the standard deviation of the perceived waiting time at signalized intersections. A simple analysis (although Wu et al. (2009) did not do this) with their reported data indicates that the ratio of standard deviation to mean remains more or less constant around 0.5.

As can be seen, the proposed model (Equation (4)) satisfies Weber's law stated in Equation (1) with the Weber fraction being equal to $\sqrt{v}$. Also, $E(d) = \mu_d = g(D)$. Note, the use of a multiplicative error term in Equation (4) allows the model to naturally incorporate the idea of Weber's law (i.e., larger (smaller) values of the time being perceived have larger (smaller) variability in perception). Doing so with an additive error term would require the less clean option of using error terms with quantity dependent variance or error terms with distributions conditioned on the quantity. Finally, the fact that the perceived delay must lie on the positive semi-infinite interval (i.e., d is always greater than or equal to zero) can be naturally accounted for using multiplicative errors whose distributions are defined on a positive support. If, instead, additive errors are used then externally defined truncation machinery will be required to ensure that the perceived values always remain positive.

Another aspect of the proposed model is the choice of the expression to be used to capture the systematic bias in perceptions. That is, since $E(d) = \mu_d = g(D)$, the function $g(D)$ can be used to embed systematic bias in perception. One possibility is to use Steven's power law, where $g(D) = aD^b$. However, it is shown in the next section that neither of the parameters of such a power function is estimable in the ordered response model set-up being proposed for the present study. Hence, in this work the following form is proposed:

$$g(D) = \left(\alpha e^{-D/k} + \beta\right)D \tag{5}$$

Here, k is a user specified positive constant chosen to scale D, while α and β are parameters that take positive values and, respectively, affect the shape and location of the systematic bias term. The function in Equation (5) within a finite domain of D can be made to behave much the same way as Stevens' power function over the same domain for $b < 1$. (In order to obtain behavior similar to the power function for $b > 1$ over a finite domain, the sign of the exponent in Equation (5) must be made positive or k can be assumed to be a negative constant).

The term in the parentheses on the right side of Equation (5) is a factor, say $B(D)$, that represents the systematic bias in the perception of time durations. If $B(D) < 1$ then the actual delay, on an average, is underestimated and vice versa when $B(D) > 1$. The negative sign in the exponent indicates that it is expected (based on previous research in the field of psychology) that lower values of D will be on an average overestimated and higher values underestimated. The ratio $\frac{\alpha}{\beta} (= \frac{B(0)}{B(\infty)} - 1)$ captures how differently one set of values of D are perceived than another set. Appendix A presents a more detailed description (with figures) of the variations of both $B(D)$ and $g(D)$ with D.

As will be shown later, in the ordered response set-up, where the purpose is to find the thresholds distinguishing the various ordered classes, the parameter that matters is the ratio $\frac{\alpha}{\beta}$ and not the individual values of α and β. Given its importance and repeated occurrences in the rest of the text this ratio is referred

to as γ. Finally, it must be ensured that $g(D)$ is a monotonically increasing function. It can be shown that for $\gamma \in (-1, e^2)$, $g(D)$ is a monotonically increasing function (see Appendix A). Since, in this analysis $\gamma \geq 0$, so for all $\gamma \leq e^2$ or $\gamma < 7.39$ the monotonicity property is satisfied.

Thus, the perceived delay based on which individual *i* responds to queries on the length of his/her waiting time is modeled as follows (see Equations (4) and (5)):

$$d_{i,m} = \left(\alpha e^{-D_{i,m}/k} + \beta\right) D_{i,m} \cdot \varepsilon_{i,m} \tag{6}$$

where, $d_{i,m}$ is the delay as perceived by individual *i* at the mth toll plaza and $D_{i,m}$ is the actual delay faced by the same individual (objective delay); $\varepsilon_{i,m}$ is the multiplicative perception error for the same individual. Note, for the present, the parameters, α and β are assumed to be same for all individuals in the population and the errors $\varepsilon_{i,m}$ are assumed to be identically and independently distributed with a mean of unity and variance of v.

The proposed model in Equation (6), to best of the authors' knowledge, is the first attempt at combining Weber's Law and Vierordt's Law (or alternatively, the basis on which Stevens' power function is proposed), the two pillars of the psychophysics of time perception, into a single model of perceived time durations. In fact, the structure of the proposed model (see Equation (4)) can be used as a basis for modelling perceived quantities in general; it may happen that the specific form of $g(D)$ may vary from one quantity to another (for example, distance perception may require a different form than time duration perception). Further, as mentioned earlier and shown later, the proposed form of $g(D)$ provides at least two distinct advantages over the Stevens' power function. Recall, (i) parameters of the proposed function, unlike those of Stevens' power function, can be estimated in the ordered response modelling setting and (ii) the proposed function is more general than the power function, for it does not impose the restriction that there must necessarily be a value of *D* where perception has to switch from overestimation to underestimation (or vice versa).

4. Modelling Response

In this study, users were asked how long they had to wait to pay toll just as they handed the toll to the collector at a manual toll booth. Each user was expected to give their answer in terms of one of the five linguistic classes – "VERY SHORT (VS)", "SHORT (S)" "ACCEPTABLE (A)", "LONG (L)", and "VERY LONG (VL)". In a way, the users were asked to rate their waiting time at the toll plaza and as with most rating schemes these classes are ordered in that the threshold demarcating say VS from S is necessarily lesser than the threshold demarcating S from A, and so on.

To analyze the response of users and its relation to the actual delay, an ordered response modelling setup is assumed. One can refer to Greene and Hensher (2010) for more on such models. It is assumed that respondents (toll-payers) have thresholds on a continuous, latent variable (often referred to as propensity) that they use to answer. Here, it is hypothesized that toll-payers use perceived delay, $d_{i,m}$ (a variable latent to the analyst) to categorize their waiting time into one of the five classes. To simplify the notation the linguistic classes are given numerical values as follows: VS is 1, S is 2, A is 3, L is 4 and VL is 5. Mathematically, the response of the ith individual at the mth toll plaza, $R_{i,m}$, can be written as:

$$R_{i,m} = \begin{cases} 1 & 0 \le d_{i,m} < \tau_m^1 \\ 2 & \tau_m^1 \le d_{i,m} < \tau_m^2 \\ 3 & \tau_m^2 \le d_{i,m} < \tau_m^3 \\ 4 & \tau_m^3 \le d_{i,m} < \tau_m^4 \\ 5 & \tau_m^4 \le d_{i,m} < \infty \end{cases} \tag{7}$$

where, τ_m^r is the threshold value of perceived delay that demarcates the boundary between the rth and $(r+1)^{th}$ class or category at the mth toll plaza. Although no classes correspond to the numerical values of 0 and 6, to simplify notation for latter mathematical development, the following are set $\tau_m^0 = 0$ and $\tau_m^5 = \infty$. Note, it is assumed that individuals do not differ in the way they perceive delay (although later this assumption is tested through empirical analysis). However, it is assumed that the thresholds may be different for individuals using different toll plazas. The bases for these assumptions are as follows.

It is hypothesized that the variability among individuals' responses arises from two distinct sources. One is how they perceive time durations, and the other is what thresholds they use to classify their perceived time into categories such as VS, S, etc. In this work, it is assumed that perception of time duration is more of a cognitive process while the thresholds are influenced by one's expectations and value for time, something that is more a function of socio-economic characteristics of individuals. It is therefore felt that parameters that capture the cognitive aspects, namely, α, β, and the distribution of $\varepsilon_{i,m}$ can be viewed differently from the parameters that are influenced by socio-economic characteristics, namely the thresholds. That is, the former set of parameters (α, β, and the variance of $\varepsilon_{i,m}$) are assumed to be same across individuals[3] while the thresholds are assumed to exhibit heterogeneity due to socio-economic differences, traffic mix and context differences, and differences across toll plazas.

Even though it may be desirable to have a disaggregate model where all parameters are different for different individuals and locations (toll plazas) often it may not be practical to do so from an econometric standpoint or useful for the purposes of engineering design. While developing a model, balance needs to be struck between the representation of variation in a population and the ability to recognize the principles controlling the underlying processes. Often attempts to estimate too many parameters, necessitated by the motivation to represent the population heterogeneity completely, tend to obfuscate the underlying principles as well as the estimation process typically results in estimates that are weak (i.e., not identified empirically). Further, it may be pointed out that in an ordered response modelling setup there is a conceptual equivalence between including explanatory variables as part of the latent model (i.e., including variables in the perceived time function) and including them as a part of the expression describing the thresholds. Therefore, the variables can be included in either the perceived time function or all the threshold functions, but not in both functions; for the parameters of such a specification might be difficult to identify. This is another reason for exploring heterogeneity in only the threshold parameters (more on this later).

Next, the likelihood function is derived for the proposed model. Since it is assumed that the perceived delay is a random variable, one can define the probability that a person's response will be r (where $r \in \{1,2,3,4,5\}$) as follows:

$$P(R_{i,m} = r) = P(\tau_m^{r-1} \le d_{i,m} < \tau_m^r) \tag{8}$$

Or,

$$P(R_{i,m} = r) = P\left(\frac{\tau_m^{r-1}}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}} \le \varepsilon_{i,m} \le \frac{\tau_m^r}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}}\right) \tag{9}$$

[3] In the empirical analysis, this assumption was relaxed to explore heterogeneity of these parameters as well.

Or,

$$P(R_{i,m} = r) = F\left(\frac{\tau_m^r}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}}\right) - F\left(\frac{\tau_m^{r-1}}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}}\right) \tag{10}$$

where, $F(\cdot)$ is the identical cumulative distribution function for each $\varepsilon_{i,m}$. Dividing the numerator and the denominator of the terms in the parentheses of $F(\cdot)$ in Equation (10) by β will not change the probability.

That is,

$$\begin{aligned} P(R_{i,m} = r) &= F\left(\frac{\tau_m^r}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}}\right) - F\left(\frac{\tau_m^{r-1}}{(\alpha e^{-D_{i,m}/k} + \beta)D_{i,m}}\right) \\ &= F\left(\frac{\frac{\tau_m^r}{\beta}}{\left(\frac{\alpha}{\beta} e^{-D_{i,m}/k} + 1\right)D_{i,m}}\right) - F\left(\frac{\frac{\tau_m^{r-1}}{\beta}}{\left(\frac{\alpha}{\beta} e^{-D_{i,m}/k} + 1\right)D_{i,m}}\right) \\ &= F\left(\frac{\tau_m^{r,*}}{(\gamma e^{-D_{i,m}/k} + 1)D_{i,m}}\right) - F\left(\frac{\tau_m^{r-1,*}}{(\gamma e^{-D_{i,m}/k} + 1)D_{i,m}}\right) \end{aligned} \tag{11}$$

In the above set of equations, while changing the ratio $\frac{\alpha}{\beta} = \gamma$ changes the probability, there are infinite combinations of α and β that result in the same ratio γ and the same probability. Therefore, one cannot separately identify the parameters α and β without additional information. Only the ratio γ is identifiable. Similarly, τ_m^r and β cannot be separately identified. One can only estimate $\tau_m^{r,*}$, which is the ratio of τ_m^r and β.

The likelihood, $\mathcal{L}(\cdot)$, of observing a certain sample of responses can be obtained as the product (across the sample) of the relevant probabilities (depending on the response) given in Equation (11). Mathematically, the expression for likelihood can be written as follows:

$$\begin{aligned} &\mathcal{L}(\gamma, \tau_1^{1,*}, \cdots \tau_M^{C,*}, v) \\ &= \prod_{m=1}^{M} \prod_{i=1}^{I_m} \prod_{r=1}^{C} \left\{ F\left(\frac{\tau_m^{r,*}}{(\gamma e^{-D_{i,m}/k} + 1)D_{i,m}}\right) \right. \\ &\left. - F\left(\frac{\tau_m^{r-1,*}}{(\gamma e^{-D_{i,m}/k} + 1)D_{i,m}}\right) \right\}^{\delta_{i,m,r}} \end{aligned} \tag{12}$$

where,

$$\delta_{i,m,r} = \begin{cases} 1 & \text{if } R_{i,m} = r \\ 0 & \text{otherwise} \end{cases}$$

and C, I_m, and M are, respectively, the number of classes or categories, the number of respondents at Toll Plaza m, and the total number of toll plazas; note it is assumed that $F(\cdot)$ is a two-parameter distribution function and therefore can be completely specified by its mean (which is unity in this case) and variance (v). The estimates of the terms in the argument of the likelihood function are obtained using a constrained

maximum likelihood estimation technique with data collected from toll plazas in India. The constraints here are $\tau_m^{C,*} > \tau_m^{C-1,*} > \tau_m^{C-2,*} > \cdots > \tau_m^{1,*}$. The estimation process, the data and the estimates are described in the next section.

Recall that it was mentioned that the impact of traffic mix and context variables like queue length (ahead of the respondent's vehicle), $Q_{i,m}$, number of heavy vehicles, say $H_{i,m}$, variation in service times of vehicles ahead in the queue, $S_{i,m}$, and socio-economic characteristics of individuals, $E_{1,i,m}$, $E_{2,i,m}$, $\cdots E_{J,i,m}$ are incorporated through the thresholds. That is, the threshold parameters can be specified to vary across individuals, toll plazas, and traffic contexts as, $\tau_{i,m}^r = \tau_m^r + h^r\left(Q_{i,m}, H_{i,m}, S_{i,m}, E_{1,i,m}, \cdots E_{J,i,m}\right)$ where $h^r\ (\cdot)$ are threshold specific functions that modify the thresholds depending on values of the argument. Note, however, that the analysis and results presented in the rest of this section and most of the next section deals with thresholds τ_m^r that only vary across toll plazas. Discussion on the analysis with heterogeneous thresholds $\tau_{i,m}^r$ is provided at the end of the next section.

Before leaving this section, a supplementary discussion on the functional form of $g(D)$ and its various implications is provided next.

4.1 More on parameter β and the proposed function g(D)

As discussed, it is apparent that β cannot be estimated. In this section it is argued from two different standpoints that estimation of parameter β is neither necessary nor required.

From an **engineering standpoint** the purpose of this analysis is to determine thresholds between what users consider as very short delay, short delay, etc. on a scale that can be measured and controlled by the engineer / planner entrusted with the duty of designing and operating a toll plaza. The actual delay, D, one experiences at a toll plaza (on an average) can be controlled by building more tollbooths, reducing the service time at the booth, etc. That is, one of the goals of the analysis is to determine thresholds, Δ_m^r on the scale of D. So, one must convert τ_m^r that are in the universe of perceived delay to the corresponding Δ_m^r in the universe of actual delay. Now,

$$\begin{aligned} \tau_m^r &= g(\Delta_m^r) \\ &= \left(\alpha e^{-\Delta_m^r/k} + \beta\right)\Delta_m^r \end{aligned} \tag{13}$$

Therefore,

$$\tau_m^{r,*} = \frac{\tau_m^r}{\beta} = \left(\gamma e^{-\Delta_m^r/k} + 1\right)\Delta_m^r \tag{14}$$

Note (see Equation (12) and the next section) the thresholds $\tau_m^{r,*}$ and γ (and v) can be estimated from the data. Hence, one can estimate Δ_m^r without knowing what β is. That is, from an engineering perspective the knowledge on the value of β is not essential.

From a **psychophysical standpoint** it may not even be required to include β as a parameter. Experiments conducted over many decades clearly indicate that when it comes to estimating waiting times, people, on an average, invariably overestimate (for example, Loehlin 1959, Cottle 1976, Hornik 1984, Antonides et al., 2002, and Fan et al., 2016). Further, it can be inferred from studies that present results for different time durations (for example, see Antonides et al., 2002, and Fan et al., 2016), that the extent of overestimation tends to reduce as the duration being estimated increases. That is, the perceived waiting time is generally greater than the actual value while the fraction by which it is greater reduces with the actual time (duration). Therefore, for a waiting line scenario, like that at a toll plaza, the systematic bias term (see discussion following Equation (5)), $B(D)$, can be written as follows, with β set to 1:

$$B(D) = \left(\alpha e^{-D/k} + 1\right) \tag{15}$$

This implies that on an average the perceived waiting time is greater than the actual and smaller durations are overestimated more than longer ones.

Alternatively, perceived delay as expressed by Equation (6) can be written as

$$d_{i,m} = \left(\alpha e^{-D_{i,m}/k} + 1\right) D_{i,m} \cdot \varepsilon_{i,m} \tag{16}$$

That is, in the scenario where one proposes a psychophysical model of perceived time that accounts for the fact that the users typically overestimate waiting time one would not require the parameter β or it can be specified to be equal to unity and set $\gamma = \alpha$.

Thus, the inability of the proposed estimation process to determine β is not of concern either from engineering or psychophysical standpoints. Of course, this non-determinability of β will pose a problem if one wishes to use similar analysis to fully specify the systematic bias term in other contexts where the physical quantity can be both over- and under-estimated. However, in no situation will this pose a problem if one wishes to only determine the definitions (or thresholds) of the response categories. Finally, since in almost all situations in transportation where perceived time duration plays a role, typically the time duration (for example, travel time or waiting time) is something that users would rather not have spent, a specification of $\beta = 1$ may be acceptable in a wide variety of situations arising in transportation.

4.2 More on Stevens' power function

Use of Stevens' power law for $g(D)$ will require externally specified values for both the parameters (a and b), as they cannot be estimated from categorical response data such as users' rating of waiting times. To understand this, suppose that instead of the proposed function $g(D)$ of Equation (5), Stevens' power function is used in Equation (6). Then Equation (9) would read as

$$P\left(R_{i,m} = r\right) = P\left(\frac{\tau_m^{r-1}}{a\left(D_{i,m}\right)^b} \leq \varepsilon_{i,m} \leq \frac{\tau_m^r}{a\left(D_{i,m}\right)^b}\right) \tag{17}$$

As both τ_m^r and a are variables the probability will not change as long as their ratio remains the same. So, much for the same reason why β and τ_m^r cannot be separately identified τ_m^r and a cannot be separately identified. As for b, note that Equation (6) with Stevens' power law reads as

$$d_{i,m} = a\left(D_{i,m}\right)^b \cdot \varepsilon_{i,m} \tag{18}$$

By taking logarithms on both sides, one can see that for the same reason why variance of error term and the coefficient of the influencing variable cannot be separately identified in an additive error model for the latent propensity function in an ordered response model, here also b and variance of the error cannot be estimated together. In summary, to estimate the parameters of Stevens' power law, data on users' perceived time durations are necessary along with objective time duration data.

Apart from the issue of parameter estimability, the power law function is not as versatile as the functional form proposed in Equation (5) even though both the functions use two parameters to describe perception bias. For example, using the power law one would not be able to represent the type of results reported in Hornik (1984) or observations reported by Fan et al. (2016) where users always overestimated waiting time (in a queue or a line) durations. This is so, because according to the power law, delay will be overestimated for values of D less than the indifference point (where perceived time is veridical) but will necessarily be underestimated for D greater than the indifference point.

5. Application to LOS Analysis at Toll Plazas

This section presents results from the application of the ordered response model (and the time perception model) introduced in the previous sections to the determination of level of service at Toll Plazas. The first subsection presents a detailed description of the data that was collected from three large toll plazas in India. The next section discusses the likelihood function introduced in Equation (12) and the assumed distribution of $\varepsilon_{i,m}$. It also talks about the numerical algorithms used here for obtaining the maximum likelihood estimates. The third section is further subdivided into subsections with each presenting results from various versions of the ordered response model presented earlier.

5.1 Data collection and descriptive statistics

Data was collected from three toll plazas on a 150 km stretch of an expressway in Rajasthan state of India during winter months when daytime temperature ranges from 10 to 20 ℃. Each toll plaza had four or more manual tollbooths per direction. Each tollbooth where data was collected had, inside the booth, a data collector in addition to the person who took the toll. As a driver paid the toll the data collector asks the driver two questions one of which is "How long have you been waiting?"[4] The response to this question is sought in one of five categories introduced earlier and is referred to as $R_{i,m}$.[5] The vehicle of the driver who is interviewed is referred to as the interviewed vehicle (IV). The response is noted along with some details of IV for identification. Each toll plaza was also fitted with high resolution video cameras in addition to the PTZ video cameras already installed at the plaza. From these camera recordings the following data for each IV are tabulated alongside the response: (i) arrival time of the IV to the toll-booth line, $t_{i,m}^{a}$, (ii) departure time of IV from the toll booth, $t_{i,m}^{d}$, (iii) length of IV's queue at $t_{i,m}^{a}$, (iv) vehicle mix in IV's queue at $t_{i,m}^{a}$, (v) service times of IV and all the other vehicles ahead of IV in IV's queue, and (v) IV's vehicle type. Note the actual (objective) delay $D_{i,m} = t_{i,m}^{d} - t_{i,m}^{a}$.

Table (1): Total number of responses and the variation in actual delay faced by the respondents

Toll Plaza (m)	Number of Respondents (I_m)	Descriptive statistics of the actual delay, $D_{i,m}$, faced by the respondents in seconds						
		Mean	Median	St. Dev.	Minimum	Maximum	25th P'tile	75th P'tile
1	545	158.8	132	95.6	27	572	77	220
2	450	148.5	126	96.6	19	556	70	221
3	513	190.7	165	113.6	26	763	104	257

[4] As discussed in Footnote 1, in order to eliminate memory recall errors, questions were posed while the respondent was paying the toll. Besides, asking the questions immediately after the respondent experienced the waiting line helps in avoiding the influence of stochasticity in $D_{i,m}$ due to variability across different instances of such waiting times faced by the respondent. However, this restricted the number of questions we could ask to only two, of which the first question had to be aimed at drawing the respondent's attention to this unusual occurrence of an additional person inside the toll booth asking questions.

[5] For the only relevant question that could be asked, the responses had to be on a category scale as it was necessary to not only capture how people perceived delay but also how they evaluated the delay based on their expectations.

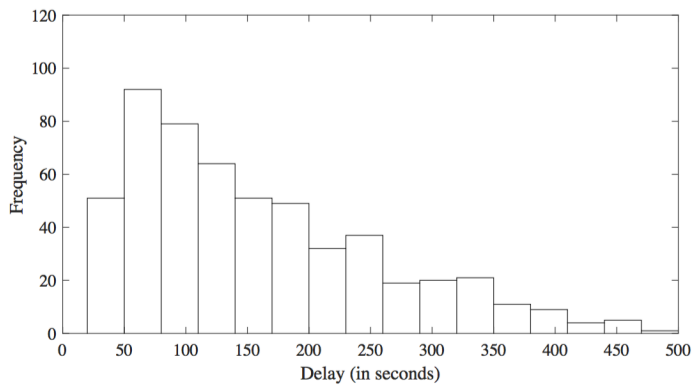

Figure (1): Frequency distribution of actual delay faced by respondents at Toll Plaza 1

For this study, IVs are invariably automobiles. The reason for ignoring other vehicle types is that the number of data points for these is too small. Table (1) presents the total number of respondents and their collective observed (actual or objective) delay characteristics for each of the three toll plazas. As can be seen from the table there are about 500 respondents for each site and a total of 1508 respondents. The delays faced also vary significantly at each toll plaza. As an example, Figure (1) shows the frequency distribution of actual delay faced by the respondents at Toll Plaza 1. The delay distributions at the other two toll plazas are also similar.

Table (2) presents the distribution of responses for each of the toll plazas. As was expected from the variations in the observed delay there is reasonable variation in the responses of individuals. These large spreads in both the explanatory variable (observed delay) and the explained variable (responses) are expected to help generate robust estimates of the model parameters.

That said it can be seen from the table that very few of the total respondents described the delay they experienced as Very Long (VL) or Category 5. Hence, it was decided to merge the last two categories into one and rename it as Long or Very Long (L-VL). For the rest of the analysis Category 4 refers to L-VL. The number of responses in the now modified Category 4 (or L-VL) for Toll Plazas 1, 2 and 3, respectively, is 82, 66, and 100. As C=4 now, the number of threshold parameters at each toll plaza reduces to three.

Table (2): Spread of response categories among the responses

Toll Plaza (m)	Number of Responses for Response Category				
	1 (or VS)	2 (or S)	3 (or A)	4 (or L)	5 (or VL)
1	60	263	141	70	11
2	80	186	118	49	17
3	80	191	142	78	22

5.2 Maximum likelihood estimation of the model parameters

With C=4 the proposed model has 11 parameters, namely, $\gamma, \tau_1^{1,*}, \tau_1^{2,*}, \tau_1^{3,*}, \tau_2^{1,*}, \tau_2^{2,*}, \tau_2^{3,*}, \tau_3^{1,*}, \tau_3^{2,*}, \tau_3^{3,*}, v$.[6] These parameters are estimated from the data by maximizing the logarithm of the likelihood $\mathcal{L}(\gamma, \tau_1^{1,*}, \cdots \tau_M^{C,*}, v)$ introduced in Equation (12). It is assumed that $\varepsilon_{i,m}$ are independent and identical lognormal distributions with an expected value of unity and variance of v (to be estimated from the data). That is, $F(\cdot)$, is the cumulative lognormal distribution.

All estimations were carried out by coding and maximizing the log-likelihood functions in R programming platform. The next section presents and analyses the results obtained here.

5.3 Parameter estimates and their implications

This section starts with presenting the results from the 11-parameter model. This model is referred to as the "Location Specific Model" as the thresholds are assumed to depend on the toll plaza site. Based on the results from the location specific models, other versions are proposed and analyzed. The results from these analyses are presented in the first three subsections of this section. The last subsection discusses the experiences with estimating thresholds as functions of other possible influencers as described in Section 4 and earlier.

5.3.1 Location Specific Model

Table (3) presents the estimates obtained from the maximum likelihood estimation of the location specific model. The threshold estimates are in seconds. The table also presents the Δ_m^r values derived from the estimated $\tau_m^{r,*}$ and γ values using Equation (14). Each estimated value is followed by two numbers in parentheses. The first number refers to the standard error of the corresponding estimate and the second refers to the corresponding *t*-statistic. As can be seen from the table all the estimates are statistically significant.

The estimates of the Δ_m^1 values indicate that toll plaza users think delay is very small if the actual delay is less than about half a minute. Similarly, Δ_m^2 values indicate, users consider actual delay less than about 2 minutes to be Small (S). When actual delay exceeds about 5 minutes (see Δ_m^3 values) users consider it to be long or very long. Other than giving the engineer a feel for how users evaluate toll plazas the results reinforce the premise (see section on Problem Statement) that delay expectations of users waiting at toll plazas are very different from those waiting at more frequent flow interrupters like signalized intersections.

Another outcome from this analysis is the observation that the threshold estimates are not that different across toll plazas. On testing the null hypothesis, for a given r and $m_1 \neq m_2$, that $\tau_{m_1}^{r,*} = \tau_{m_2}^{r,*}$ (with the alternate hypothesis of $\tau_{m_1}^{r,*} \neq \tau_{m_2}^{r,*}$) for all three combinations of m_1 and m_2 it is observed that the null hypothesis cannot be rejected except for when $r = 1$. Thus, it appears that drivers, as a group, differ from location to location only when deciding whether a waiting time is Very Short (VS) or Short (S). For the other categories (A and L-VL), where the thresholds are larger, this distinction based on location vanishes. The observation that some of the thresholds are possibly indistinct leads to a modified location specific model described in the next section.

[6] In addition to the 11-paramter model, a 13-paramter model that allowed the γ parameter to be different for each of the three toll plazas was estimated. The location specific estimates of γ in the 13-paramter model were not statistically different from each other. For this reason and other reasons discussed later, a location-abstract γ parameter is used for subsequent analysis.

Table (3): Parameter estimates with relevant statistics for the Location Specific Model

Parameters		Toll Plaza (m)		
		1	2	3
Estimated	$\tau_m^{1,*}$ (seconds)	54.5 (14.6, 3.7)	63.6 (15.8, 4.0)	73.1 (17.2, 4.3)
	$\tau_m^{2,*}$ (seconds)	171.0 (22.9, 7.5)	164.0 (22.8, 7.2)	174.5 (23.0, 7.6)
	$\tau_m^{3,*}$(seconds)	327.1 (23.3, 14.0)	314.3 (24.2, 13.0)	337.9 (22.5, 15.0)
	γ	1.34 (0.52, 2.57)		
	v	0.51 (0.15, 3.33)		
Derived	Δ_m^1 (seconds)	26.9	32.3	38.2
	Δ_m^2 (seconds)	122.8	115.2	126.7
	Δ_m^3 (seconds)	308.2	293.4	320.5

Note: γ and v were assumed to be the same across all three toll plazas.

Before leaving this section, however, (i) the implications of the estimate of γ, (ii) the implications of the estimate of v, and (iii) the lack of impact of the value of β on the estimation of any of the parameters are discussed with the aid of the results obtained here. $\gamma = 1.34$ implies a $B(D)$ whose variation with D is shown by the solid line in Figure (2). (The figure also shows $B(D)$ with $\gamma = 1.41$; this line will be referred to in a latter section). As expected, the systematic bias term $B(D)$ is always more than unity but reduces with D. Also plotted are the ratios of perceived waiting time to the actual waiting time (or, as the study refers to them, objective waiting time) computed using data on these quantities obtained from experiments on a waiting-to-receive-service scenario conducted by Antodines et al. (2002). As can be seen, the $B(D)$ line from the present study is a good match with the values reported earlier. There are other earlier studies in waiting situations that also report results that are in close conformity with the present results. For example, Hornik (1984) reports on an average the ratio of estimated waiting time to the actual waiting time is about 1.36 while it can be seen from Fan et al. (2016) that for waiting time durations less than 5 minutes the median of the ratio is around 1.5. For $\gamma = 1.34$, the average of $B(D)$, over the 1508 observed waiting time values of this study, is 1.38 (Hornik (1984) reports 1.36) and the median is 1.32; if the median is determined with the data where delay is less than 5 minutes (like in Fan et al., 2106) then the value increases to 1.4 (Fan et al. (2016) reports 1.5).

The results of Hornik (1984), Antonides et al. (2002) and Fan et al. (2016) referred to here, were obtained or inferred using data from direct retrospective measurements of perceived time durations from various waiting time scenarios. In the present case, however, the responses are obtained in a universe of ordered, mutually exclusive, discrete classes that are described to the respondent through linguistic labels like short, long, etc. Here, γ, the parameter of the systematic bias, $B(D)$ is embedded within the latent variable and is estimated using an ordered response modelling framework representation of the response mechanism. It is reassuring to note that the estimated value of γ indicates a behavior for the systematic bias that is in close conformity with reported direct observations of the same from other waiting line scenarios. These similarities in behavior implied by the estimated γ parameter (which is significantly different from zero) with that from other studies involving disparate waiting line scenarios provide support for: (i) the assumption that there is systematic bias in perception in addition to random errors, (ii) the reasonableness of the proposed algebraic expression for $B(D)$ and (iii) the assumption that the perception of the waiting

time duration is likely an outcome of deep-seated properties of human cognition and should not change from one experiment site to another.

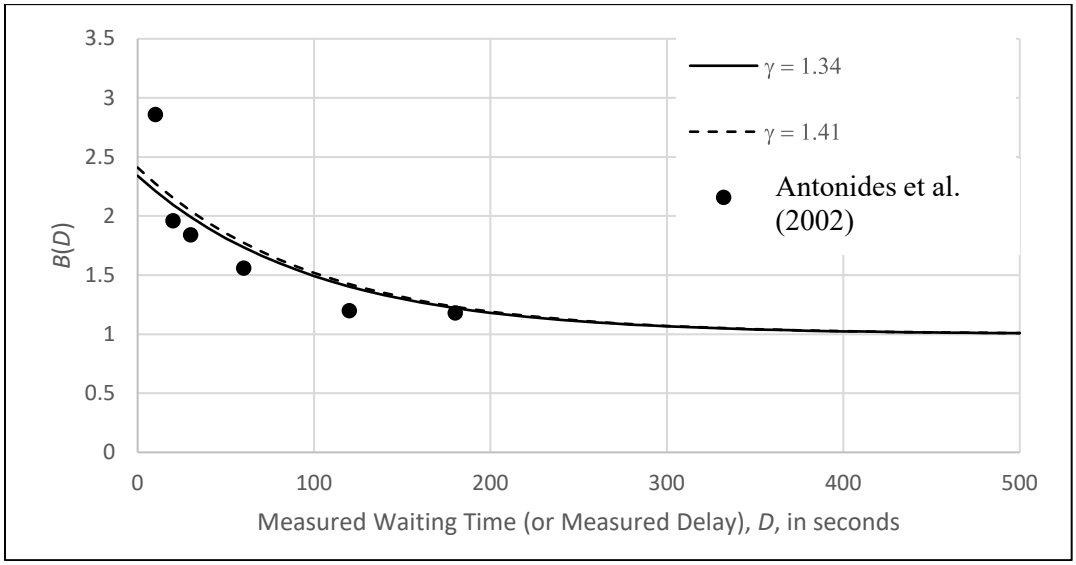

Figure (2): Variation of $B(D)$ with D for estimated γ values.

Even though, as is apparent from the discussions and results presented in this and the earlier sections, there is considerable evidence pointing to the existence of a systematic bias term like $B(D)$, the parameters of the ordered response model (Equation (12)) were also estimated while ignoring systematic bias. That is, it was assumed $d_{i,m} = D_{i,m} \cdot \varepsilon$. This is akin to writing the likelihood of Equation (12) with $\gamma = 0$. In this case the estimate of variance, v of the error term ε shoots up to 1.29 from 0.51 (see Table (3)). This 2.6 times increase occurs because the variations that otherwise could be explained by the systematic bias term now have to be attributed to random error. The behavior of variance, v further reinforces the need for including and the correctness of a systemic bias term like $B(D)$.

Other than acting as an indicator of the existence of a systematic bias in time duration estimation, recall that the variance, v of the error term $\varepsilon_{i,m}$ can also be interpreted as the square of the Weber fraction. Thus, the estimated Weber fraction, w for the present study is $\sqrt{0.51} = 0.71$. As mentioned earlier in Section 3, Getty's (1975, 1976) idea of a generalized Weber's law was also implemented in this study by appropriately specifying the variance of the error term as a function of D. Even with a generalized Weber's law representation, statistical analysis on the estimates strongly suggested that for the present study the classical form of the Weber's law, where the Weber fraction is a constant, holds.

Earlier it was shown as well as argued why β may not be necessary in this model even though it serves the purpose of representing the systematic bias term in an easy-to-comprehend manner. Here results are presented from the estimation process using the likelihood function of Equation (12) written using $P(R_{i,m} = r)$ given in Equation (10) and not with the $P(R_{i,m} = r)$ given in Equation (11). The estimation for all the eleven parameters (with γ being replaced by α) is carried out for three specified values of β; $\beta = 1$, $\beta = 2$, and $\beta = 3$. The results are shown in Table (4); the values of τ_m^r, $\tau_m^{r,*}$ and Δ_m^r are in seconds. As expected, the value of β has no impact on Δ_m^r; it remains constant for a given r and m irrespective of the value of β. For every value of β the maximum likelihood (which, by the way, is the same for all values of β) occurs at a point where the α estimate adjusts itself in a way that the ratio α/β remains constant at a value of 1.34; not surprisingly, this is equal to the estimate of γ obtained in Table (3). Interestingly, the

estimated variance for the error term remains independent of β. This indicates that the systematic bias term, or at least the amount of variability it explains, depends, not on α and β separately but on the ratio γ. And since, for different values of β, α always adjusts itself so as to maintain a single γ it can be said that as far as the systematic bias term's purpose is to explain a part of the variation in the data, γ is the only parameter that matters. Physically what it means is that in an ordered response model, where thresholds between response categories are determined, the relative difference in the bias term across different values of the argument (in this case, observed delay) is what matters rather than the actual bias.

In any case, results in Table (4) again highlight that the value of β is inconsequential in the present analysis.

Table (4): Impact of value of β on the parameter estimates of the Location Specific Model.

β values	Estimated (Derived) Parameter(s)	Toll Plaza (m)		
		1	2	3
β = 1	τ_m^1 (seconds) $(\tau_m^{1,*} = \tau_m^1/\beta, \Delta_m^1)$	54.5 (54.5, 26.9)	63.6 (63.6, 32.3)	73.1 (73.1, 38.2)
	τ_m^2 (seconds) $(\tau_m^{2,*} = \tau_m^2/\beta, \Delta_m^2)$	171.0 (171.0, 122.8)	164.0 (164.0, 115.2)	174.5 (174.5, 126.7)
	τ_m^3 (seconds) $(\tau_m^{3,*} = \tau_m^3/\beta, \Delta_m^3)$	327.1 (327.1, 308.2)	314.3 (314.3, 293.4)	337.9 (337.9, 320.5)
	α $(\gamma = \alpha/\beta)$	1.34 (1.34)		
	v	0.51		
β = 2	τ_m^1 (seconds) $(\tau_m^{1,*} = \tau_m^1/\beta, \Delta_m^1)$	109.0 (54.5, 26.9)	127.1 (63.6, 32.3)	146.1 (73.1, 38.2)
	τ_m^2 (seconds) $(\tau_m^{2,*} = \tau_m^2/\beta, \Delta_m^2)$	341.9 (171.0, 122.8)	327.8 (163.9, 115.2)	348.9 (174.5, 126.7)
	τ_m^3 (seconds) $(\tau_m^{3,*} = \tau_m^3/\beta, \Delta_m^3)$	654.2 (327.1, 308.2)	628.6 (314.3, 293.4)	675.8 (337.9, 320.5)
	α $(\gamma = \alpha/\beta)$	2.68 (1.34)		
	v	0.51		
β = 3	τ_m^1 (seconds) $(\tau_m^{1,*} = \tau_m^1/\beta, \Delta_m^1)$	163.5 (54.5, 26.9)	190.6 (63.5, 32.3)	219.1 (73.1, 38.2)
	τ_m^2 (seconds) $(\tau_m^{2,*} = \tau_m^2/\beta, \Delta_m^2)$	512.9 (171.0, 122.8)	491.8 (164.0, 115.2)	523.5 (174.5, 126.7)
	τ_m^3 (seconds) $(\tau_m^{3,*} = \tau_m^3/\beta, \Delta_m^3)$	981.3 (327.1, 308.2)	942.9 (314.3, 293.4)	1013.8 (337.9, 320.5)
	α $(\gamma = \alpha/\beta)$	4.02 (1.34)		
	v	0.51		

Note: α, β, and v were assumed to be the same across all three toll plazas.

5.3.2 Modified Location Specific Model

In this model, the following modifications to the earlier 11-parameter model are made (or restrictions are imposed): $\tau_1^{2,*} = \tau_2^{2,*} = \tau_3^{2,*} = \tau^{2,*}$ and $\tau_1^{3,*} = \tau_2^{3,*} = \tau_3^{3,*} = \tau^{3,*}$. Thus, in this version there are only 7 parameters. The only parameter that is location specific is that of the threshold between VS and S. The estimates obtained for this model are shown in Table (5). The units of thresholds and the meaning of the terms in parentheses are the same as in Table (3).

The likelihood ratio between the 11-parameter location specific model and the 7-parameter modified location specific model comes out to be 1.65. Since the number of restrictions that need to be enforced to convert the 11-parameter model to the 7-parameter model is 4 the null hypothesis that the restrictions are acceptable (or that the 7-parameter model performs similarly to the 11-parameter model) cannot be rejected at the 95% confidence level.

Table (5): Parameter estimates with relevant statistics for the Modified Location Specific Model

Parameters		Toll Plaza (m)		
		1	2	3
Estimated	$\tau_m^{1,*}$ (seconds)	53.8 (14.7, 3.7)	63.7 (15.9, 4.0)	71.7 (17.2, 4.2)
	$\tau^{2,*}$ (seconds)	169.5 (22.5, 7.6)		
	$\tau^{3,*}$ (seconds)	327.6 (19.7, 16.6)		
	γ	1.34 (0.53, 2.5)		
	v	0.52 (0.16, 3.3)		
Derived	Δ_m^1 (seconds)	26.6	32.3	37.3
	Δ^2 (seconds)	121.1		
	Δ^3 (seconds)	308.7		

Note: γ and v were assumed to be the same across all three toll plazas. $\tau^{2,*}$ and $\tau^{3,*}$ were fixed to be the same across all three toll plazas. Therefore, Δ^2 and Δ^3 were also same across all three toll plazas.

Another point is that the parameter estimates of γ and v in Tables (3) and (5) are quite close to each other. This reasonable constancy of the estimates of γ and v also points to the fact that they do not depend on how one treats the response category definitions (i.e., the threshold parameters). This is expected as γ and v relate to how humans perceive waiting times and not how they classify them. Also, as in the previous case, the $\tau_m^{1,*}$ remains statistically different across the various locations (or toll plazas). At the present time no reason is being put forward as the cause of this difference for the low-end threshold. More research into the demographics of the population who predominantly use these toll plazas needs to be carried out. In conclusion it can be said that for the proposed modelling framework the modified location specific model is a reasonable representation.

Given that this is identified as the best model of the responses of users at toll plazas, an analysis is conducted to see whether the parameter estimates obtained here can be retrieved from a simulated data closely representing the observations. First, data on actual delay is randomly generated using the empirical frequency distribution of actual delays (like the one shown in Figure (1)) at each of the three toll plazas; for

each toll plaza 10000 delay data are generated. At the same time random error terms are generated from a lognormal distribution with a mean of unity and variance equal to the estimated variance of 0.52. Next, using the generated actual delay, the generated error and the estimated γ (= 1.34), 10000 perceived delays are determined. These perceived delays and the estimated threshold values are then used to determine the simulated responses. Finally, the set of simulated responses and corresponding simulated actual delays is used as "observations" from which the maximum likelihood estimates of the seven parameters (namely, $\tau_1^{1,*}, \tau_2^{1,*}, \tau_3^{1,*}, \tau^{2,*}, \tau^{3,*}, \gamma$, and v) are obtained. The process of generating simulated "observations" is repeated to obtain 100 different simulated sets. For every set, the seven parameters are estimated and hence at the end, there are 100 estimates for each parameter.

Table (6) presents the mean and other statistics related to the estimates obtained from the simulated data. The values used to obtain the simulated (delay and responses) data are given parenthetically below the parameter name in the second row. As can be seen from the mean of the estimated values, the parameter values used to simulate the "observed" data can be retrieved as estimates (of the same parameters) using the simulated "observations." The median and mean of each estimate are almost identical indicating symmetry in the frequency distribution of each of the estimates. The third and fourth rows show that for any given parameter the frequencies of estimates within one standard deviation on either side of the mean are almost equal. This again indicates symmetric density function for each of the parameters. Finally, the observations that for any parameter, (i) almost 67% to 72% of the estimates lie within one standard deviation of the mean and (ii) about 93% to 95% lie inside two standard deviations of the mean further indicate that each parameter estimate is normally distributed. In fact, the hypotheses that a parameter estimate is distributed normally could not be rejected at the 95% confidence level using a χ^2 test for any of the parameters. Also, the estimates were consistent, that is as the number of sets of observations were increased the standard deviation of the estimates reduced monotonically.

Table (6): Parameter estimates from simulated data and their associated statistics

Estimated parameter statistic	Model Parameters (Value used to obtain simulated data)						
	$\tau_1^{1,*}$ (53.8)	$\tau_2^{1,*}$ (63.7)	$\tau_3^{1,*}$ (71.7)	$\tau^{2,*}$ (169.5)	$\tau^{3,*}$ (327.6)	γ (1.34)	v (0.52)
Mean	53.8	63.6	71.4	169.2	327.4	1.34	0.53
Median	53.9	63.8	71.5	169.6	328.4	1.35	0.52
Std. dev.	5.16	5.54	5.91	8.03	6.85	0.19	0.06
$\xi_{1,-}$ #	0.34	0.32	0.34	0.30	0.31	0.32	0.39
$\xi_{1,+}$ ##	0.36	0.37	0.35	0.39	0.41	0.35	0.33
ξ_2 $	0.93	0.94	0.94	0.95	0.94	0.94	0.95

$\xi_{1,-}$ refers to the fraction of estimates that lie between the mean and mean – Std. dev.
$\xi_{1,+}$ refers to the fraction of estimates that lie between the mean and mean + Std. dev.
$ ξ_2 refers to the fraction of estimates that lie between (mean – 2Std. dev.) to (mean + 2Std. dev.).

5.3.3 Location Abstract Model

From the empirical analysis presented in the previous section, it is clear that except for the threshold between the VS and S all other parameters are location abstract; that is, they do not vary with the location.

The question that has so far been stepped around is how many categories are necessary for analyzing level of service at toll plazas. This question and associated analysis form the subject matter of this section.

Level of service analysis, as also mentioned earlier, has two distinct purposes. First, it helps the engineer or planner gauge the health of a system. Second, this analysis helps engineers design traffic facilities that guarantee, to the extent possible, a minimum service quality. From the standpoint of the first purpose, depending on the level of granularity in which you want the results, any number of categories may be desirable. At one level of coarseness knowing what is acceptable and what is not may be enough. While at another level one may want categories to be so fine so as to give an almost continuous idea of the changing performance levels of a facility. Of course, given the uncertainties associated with human responses it may neither be prudent nor feasible to have more than a certain number of categories. What is the correct number from this standpoint has not been answered here; it was assumed that four to five classes are both discernible from data and enough to understand human characterization of the service levels at a toll plaza.

However, from the standpoint of the second purpose – that of designing traffic facilities – the answer to how many categories is necessary is easier to arrive at. Traffic facilities are typically designed for peak or close to peak loads. Therefore, such facilities remain under-loaded for large portions of the day. Given that performance of a facility improves when the system operates with loads lesser than the design load, traffic facilities are typically not designed for stringent (or very good) level of service conditions. So, from a facility design standpoint a facility is often designed for "good" or "fair" conditions. For toll plazas, waiting times that are classified as "Short (S)" or "Acceptable (A)" may be enough from the design standpoint.

Hence, in this section the observed data is analyzed by assuming that responses are in three categories instead of the earlier four. These categories of responses are, "Very Short or Short (VS-S)," "Acceptable (A)," and "Long or Very Long (L-VL)." In order to maintain parity in the numbering of categories, VS-S is referred to as Category 2. The likelihood function for the present three-category model has only four parameters, namely, $\tau^{2,*}, \tau^{3,*}, \gamma,$ and v and since none of them depend on the location this model is referred to as the "Location Abstract Model."

Table (7): Parameter estimates with relevant statistics for the Location Abstract Model

Parameters		Toll Plaza (m)		
		1	2	3
Estimated	$\tau^{2,*}$ (seconds)	175.8 (25.0, 7.0)		
	$\tau^{3,*}$ (seconds)	327.2 (20.4, 16.1)		
	γ	1.41 (0.6, 2.4)		
	v	0.44 (0.15, 2.9)		
Derived	Δ^2 (seconds)	125.4		
	Δ^3 (seconds)	307.2		

Note: γ and v were assumed to be the same across all three toll plazas. $\tau^{2,*}$ and $\tau^{3,*}$ were fixed to be the same across all three toll plazas. Therefore, Δ^2 and Δ^3 were also same across all three toll plazas.

Table (7) provides the maximum likelihood estimates. To improve readability, the format of the table has been kept the as the same as Tables (3) and (5). As can be seen from the table, the estimates are statistically

significant. The threshold values are reasonably stable across models and from an engineering sense repeat the same story – the threshold is about 2 minutes (of actual delay) between VS-S and A and about 5 minutes between A and L-VL. Therefore, the following prescription can be given to engineers: (i) design a toll plaza offering an average (actual) delay of less than 2 minutes if the design is for S or better, and (ii) design it for an average (actual) delay of less than 5 minutes if the design level of service is A (acceptable) or better.

The change in γ value from 1.34 (in the location specific models) to 1.41 here is not statistically significant. Also, as can be seen from Figure (2), this 0.07 difference in γ does not give rise to any meaningful or discernible difference in $B(D)$. Similarly, the difference in v estimates (between this and the location specific models) is not statistically significant. Plots (not included here) of the lognormal distributions with mean of unity and variance of 0.52 (or 0.51) and with mean of unity and variance of 0.44 also reveal that the differences are not quite discernible. Again, as before, this relative constancy of the estimates of γ and v in Tables (3), (5) and (7) points to the fact that they do not depend on how one treats the response category definitions (i.e., the threshold parameters). This is expected as these two parameters relate to how humans perceive waiting times and not how they classify them.

5.3.4 Estimating Variable Thresholds, $\tau_{i,m}^{r}$

In the previous section it was proposed that the possible impact of socioeconomic variables $E_{j,i,m}$, queue length, $Q_{i,m}$, number of heavy vehicles, $H_{i,m}$, variability in service times of vehicles, $S_{i,m}$ (measured as the standard deviation of the service times of the vehicles ahead in the queue for respondent i at toll plaza m) etc. will be incorporated through their effect on the thresholds. As indicated earlier, it is assumed that

$$\tau_{i,m}^{r} = \tau_{m}^{r} + h^{r}\left(Q_{i,m}, H_{i,m}, S_{i,m}, E_{1,i,m}, \cdots E_{J,i,m}\right)$$

In this study, however, data on socioeconomic characteristics for respondents could not be collected. Recall, responses were obtained as they were paying toll. It had to be ensured that the data collection did not in any way impact the service time at the tollbooth. So only a few seconds were available for obtaining responses during which it was not possible to obtain data on socioeconomic characteristics of the respondents. (It may be pointed out here that gender differences could not be studied as the number of women drivers/respondents was negligible). However, data on other variables were available with enough variations to allow for meaningful econometric analysis. Data on $Q_{i,m}$ ranged from nil to 17 with a mean of 3.71 and standard deviation of 3.03; $H_{i,m}$ ranged from nil to 7 with a mean of 1.03 and standard deviation of 1.14; $S_{i,m}$ ranged from 0.50 seconds to 204.97 seconds with a mean of 19.07 seconds and standard deviation of 17.46 seconds. Linear models of $h^{r}(\cdot)$ were used to study the effect of these variables on the thresholds of the modified location specific model. While non-linear functions for $h^{r}(\cdot)$ as in the generalized ordered response models reported in Eluru et al. (2007) and Greene and Hensher (2010) can be attempted, these explorations are not in the scope of this study.

Interestingly, MLE estimates (obtained from constrained maximization of the log likelihood function; with constraints ensuring a linear order in the thresholds) of none of the coefficients associated with the variables, $H_{i,m}$ and $S_{i,m}$ were statistically significant. The coefficients associated with $Q_{i,m}$ (when the analysis is done only with $Q_{i,m}$ as none of the other variables have a statistically significant effect) for the lowest thresholds (between VS and S) were not significantly different from zero[7] and those associated with the two higher

[7] The reason for this could be one of the following: (i) since delays that are classified as VS are small they typically arise when queue lengths are negligible and hence the impact of queue length on the lowest threshold (upper bound on D that define VS) is not discernible, or (2) since the section of the data which ultimately influence the lowest threshold have small variability in queue lengths (around 0 to 2) the influence of the latter cannot be identified in a statistically significant manner.

thresholds did not vary (statistically significantly) from one toll-plaza site to another and between themselves. This statistically significant coefficient is equal to – 4.6 seconds per vehicle in the queue with a t-statistic value 2.54.

The above result indicates that queue length, which is perceivable by the respondent, as he/she joins a queue, for the most parts, affects the way an individual evaluates the waiting time. A plausible reason could be that longer queues create an expectation of longer delays (i.e., creates a negative predisposition) and hence one tends to classify even short perceived delays as long[8]. Although, from a behavioral standpoint this result is important, the magnitude of the coefficient is small and does not make a practically meaningful change to the thresholds in the D universe. For example, the range of the thresholds Δ^2 and Δ^3 that incorporate the impact of queue lengths are $[105.3s, 132.3s]$ and $[292.3s, 326.1s]$, respectively. The same thresholds, when estimated without incorporating the effect of $Q_{i,m}$ (see Table (5)) are 121.1 s and 308.7 s; thus, the effect of $Q_{i,m}$ was to introduce an approximately ± 15 seconds range to these estimates. Effectively, the interpretation that the second threshold is around 2 minutes and the third around 5 minutes does not change even with the introduction of queue length as an influencing variable.

6. Discussions and Conclusions

How humans perceive and respond to physical quantities like time and distance impacts a variety of traffic systems as well as transport demand. Mathematical models of driver behavior, like car-following, overtaking, gap-acceptance, etc. and those of transport demand traditionally incorporate these physical quantities without accounting for the fact that humans must necessarily be using perceived values of the physical quantities to react or respond. Even in those works where authors attempt to incorporate the perceived values, the method has been to introduce a latent variable, where an additive (and in very few cases a multiplicative) stochastic error term is used to represent the gap between perception and measurement. Little or no effort has been directed to include the psychophysical aspects of perception as gleaned from the vast literature in psychology and cognitive sciences.

In this research, taking a cue from the psychology literature, a unified model of time duration perception is proposed that, possibly for the first time, subsumes within one expression, Vierordt's and Weber's laws. Vierordt's law is used to represent systematic bias in human perception and Weber's law is used to represent random variability in perception. While incorporating these principles into a single model it is apparent that multiplicative errors are a natural choice while handling random variability in perceptions of not only time but also of other physical quantities. Earlier sections present examples from other authors who seem to suggest, purely from statistical standpoints, that multiplicative error terms work better than additive ones while handling perceptions. The current study provides a psychology theoretical grounding for such a specification.

Another notable contribution is the proposed form for $B(D)$, a function used to represent bias in perception, in place of the more popularly used (in psychology literature) Stevens' power law. In earlier sections, the inability of the power law to represent only overestimation or only underestimation have been pointed out. Further, the fact that parameters of the power law cannot be estimated (while the proposed form's parameters can be) in an ordered response setting has also been demonstrated. Note that in transportation

[8] Curiously, Hornik (1984) reports that he did not find any impact of queue length on how respondents perceived delay. The present proposal that queue lengths do not impact perceptions of time durations but affect the way humans classify them (i.e., they affect the linguistic class definitions) is one way to reconcile Hornik's (1984) somewhat counter-intuitive results with the general expectations on how queue lengths impact human description of delay in waiting lines.

(and in many other fields that deal with human behavior) one often encounters problems where humans are assumed to work with discrete ordered categories defined on a universe of perceived quantities on time, distance, or speed. In such systems, like that of determining levels of service definitions, a natural modelling framework is the ordered response model. Thus, the shortcoming of the power function in terms of non-estimable parameters in ordered response setting is a serious one.

The proposed model is applied to understand, using carefully collected data from toll plazas in north India, how users rate their experience. From the results here it is clear that, at least as far as time duration is considered, humans exhibit both systematic bias and random variations in perceptions. What increases the weight of the assertion that time perceptions in transport settings, like in other spheres, have systematic bias is the fact that the bias estimated from the present data is in good consonance with bias reported in earlier literature from different waiting line scenarios. The bias in perception, however, has generally been ignored in most transportation related work and this omission may have caused parameter estimates in those studies to become weak or erroneous.

This research endeavor also achieved its engineering goal to determine the definitions of the level of service categories (based on threshold values on actual delay) that the study area population might be using to classify their experience at toll plazas. In the process it demonstrated the feasibility of using the proposed methodology for such purposes. It also demonstrated, as the category definitions were very different from another waiting line scenario often encountered in transportation facilities – that of signalized intersections, that humans evaluate similar delay experiences in different facilities (settings) differently. The premise of this work that the existing practice of drawing toll plaza level of service definitions from definitions used for signalized intersections may be flawed, is supported by the data. The fact that users encounter toll plazas more infrequently (for a trip of a given length) than they encounter intersections significantly changes their expectations from these facilities and consequently what they consider as very short, short, etc.

This study indicates that the definition of categories depends on the type of facility, while how one perceives time duration is steady across waiting line scenarios. This may be because human perception of time is a fundamental cognitive process and therefore tends to be similar across individuals and across normally encountered situations. However, the thresholds used to classify waiting times into level of service categories are more an outcome of one's expectations and therefore change from one type of facility to another, depend on context variables (like queue length in the case of toll plazas), and may also depend on socio-economic characteristics of the population.

The current methodology of embedding the bias and random error present in perceived quantities within an overarching ordered response modelling framework to determine level of service definitions in terms of measurable variables can be effectively used for all traffic facilities. In this sense the proposed methodology goes beyond the present application domain of toll plazas.

Further, it is worthwhile to note that the methodology outlined here has the prospect of becoming a tool that can be used to study how humans perceive time durations (and possibly other physical quantities such as distance) from experiments where respondents provide insights into their perceptions through linguistic categorical variables. Such responses can be obtained more naturally and easily from humans than the various other methods often employed to study perceptions (see Grondin, 2010). This provides opportunities to design feasible field experiments in a variety of situations.

It is also worth noting that the model formulated in this study for perceived time does not consider stochasticity due to variability in the objective time duration (i.e., stochasticity in $D_{i,m}$). Therefore, in the current study, it was a conscious decision to design the data collection effort to elicit user responses

immediately after they experienced the waiting; for their responses can now be assumed to be based on the single experience of delay thereby helping to circumvent the question of stochasticity in $D_{i,m}$. However, in several situations such as travel mode choice and route choice, the objective travel time duration ideally should be treated as stochastic since people make such choices based on anticipated or experienced variability in travel time. In these situations, it might be important to consider stochasticity in objective durations in addition to considering how users perceive stochastic time durations. Although a good amount of literature exists on users' risk attitudes toward variability in attributes such as travel time, much of this literature assumes that users are fully knowledgeable of the objective travel time and its variability (i.e., no perception errors). Not much exists on how users perceive stochastic attributes while analyzing their response to such stochasticity. In this context, there is a need to formulate approaches that accommodate bias and variability in perception while recognizing underlying stochasticity in objective time duration. Several important questions emerge in this context, such as: (a) how to represent users' perception of stochastic objective time durations? and (b) how to disentangle the stochasticity in objective time durations from the stochasticity in user's perception? In the context of representing user perceptions, additional questions emerge – (a) should the users' perceptions be represented as a stochastic distribution of their perceptions of different experiences (i.e., is it a distribution of perceptions)? or (b) do the users form a perception of the stochastic objective time duration (i.e., for example, is it a perception of the metrics describing the distribution such as: (i) perceptions of average time duration and variance in time duration, or (ii) perceptions of usual and worst case scenario times)? Future research will benefit from delving into these questions through appropriately designed experiments.

Finally, most of the model parameters estimated in this study (except the first threshold parameter) did not turn out to be statistically different across the three toll plazas. Further, the empirical data did not have the information to explore parameter heterogeneity due to socio-demographic and other factors. In this context, there is a need to assemble datasets from a wide variety of geographical and demographic contexts, including information on socio-demographic characteristics of the sampled individuals and context-specific factors. This would require carefully designing experiments which, while continuing to minimize the impact of memory recall-related problems, will allow collection of such detailed information. This remains a potentially important avenue for future research.

Author Contributions

Partha Chakroborty: Conceptualization, Methodology, Formal analysis, Writing – original draft. Abdul Rawoof Pinjari: Conceptualization, Methodology, Formal analysis, Writing – review & editing. Jayant Meena: Software, Validation, Investigation, Data curation. Avinash Gandhi: Data collection and curation.

Declaration of Competing Interest

None.

Acknowledgement

The authors acknowledge the help provided by Mr. Saurabh Srivastava of the Department of Civil Engineering, IIT Kanpur during data collection. Two anonymous reviewers and the handling editor provided valuable comments addressing which enriched this manuscript.

References

Antonides, G., Verhoef, P. C. and van Alast, M., 2002. Consumer perception and evaluation of waiting time: A field experiment. Journal of Consumer Psychology, 12 (3), 193 – 202.

Allan, L. G., 1979. The perception of time. Perception and Psychophysics, 26 (5), 340 – 354.

Allan, L. G., 2001. Time perception models. International Encyclopedia of Social and Behavioral Sciences, 15696 – 15699.

Aycin, M. F., 2006. Simple methodology for evaluating toll plaza operations. Transportation Research Record: Journal of Transportation Research Board, 1988, 92 – 101.

Burnett, P., 1978. Time cognition and urban travel behavior. Geografiska Annaler: Series B, Human Geography 60 (2), 107–115.

Brown, S. W. and Stubbs, D. A., 1988. The psychophysics of retrospective and prospective timing. Perception, 17, 297 – 310.

Carrion, C., 2013. Travel Time Perception Errors: Causes and Consequences. Ph.D. Thesis, University of Minnesota, Minneapolis.

Castro, C., Martínez, C., Tornay, F.J., Fernández, P.G., and Martos, F.J., 2005. Vehicle distance estimations in nighttime driving: a real-setting study. Transportation Research Part F, 8, 31 – 45.

Choocharukul, K., Sinha, K. C. and Mannering, F. L., 2004. User perceptions and engineering definitions of highway level of service: an exploratory statistical comparison. Transportation Research Part A, 38, 677 – 689.

Clark, J. E., 1982. Modeling travelers' perceptions of travel time. Transportation Research Record 890, 7–11.

Cottle, T. J., 1976. Perceiving Time: A Psychological Investigation with Men and Women, John Wiley, New York.

Fan, Y., Guthrie, A. and Levinson, D., 2016. Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security. Transportation Research Part A, 88, 251 – 264.

Fosgerau, M. and Bierlaire, M., 2009. Discrete choice models with multiplicative error terms. Transportation Research Part B, 43(5), 494 – 505.

Fraisse, P., 1984. Perception and estimation of time. Annual Review of Psychology, 35, 1 – 36.

Getty, D. J., 1975. Discrimination of short temporal intervals: A comparison of two models. Perception and Psychophysics, 18 (1), 1 – 18.

Getty, D. J., 1976. Counting processes in human timing. Perception and Psychophysics, 20, 191 – 197.

Greene, W. H. and Hensher, D. A., 2010. Modelling Ordered Choices: A Primer. Cambridge University Press, New York.

Grisolia, J.M. and Ortuzar, J. de. D., 2010. Forecasting vs. observed outturn: studying choice in faster inter-island connections. Transportation Research Part A, 44(3), 159-168.

Grondin, S., 2010. Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. Attention, Perception and Psychophysics, 72 (3), 561 – 582.

Grondin, S. and Plourde, M., 2007. Judging multi-minute intervals retrospectively. The Quarterly Journal of Experimental Psychology, 60 (9), 1303 – 1312.

Henley, D.H., Levin, I.P., Louviere, J.J., Meyer, R.J., 1981. Changes in perceived travel cost and time for the work trip during a period of increasing gasoline costs. Transportation, 10(1), 23 – 34.

Highway Capacity Manual,1965. Special Report 87. Highway Research Board, National Research Council, Washington D.C., U.S.A.

Highway Capacity Manual, 1985. Special Report 209. Transportation Research Board, National Research Council, Washington D.C., U.S.A.

Highway Capacity Manual, 2000. Transportation Research Board, National Research Council, Washington D.C., U.S.A.

Highway Capacity Manual, 2010. Transportation Research Board, National Research Council, Washington D.C., U.S.A.

Hornik, J., 1984. Subjective and objective time measures: A note on the perception of time in consumer behavior. Journal of Consumer Research, 11, 615 – 618.

Indian Highway Capacity Manual (Indo-HCM), 2017. CSIR – Central Road Research Institute, New Delhi, India.

Kikuchi, S. and Chakroborty, P., 2006. Frameworks to represent uncertainty when level of service is determined. Transportation Research Record: Journal of Transportation Research Board 1968, 53 – 62.

Klodzinsky, J. and Al-Deek, H. M., 2002. New methodology for defining level of service at toll plazas. Journal of Transportation Engineering, 128 (2), 173 – 181.

Lejeune, H. and Wearden, J. H., 2009. Vierordt's *The Experimental Study of the Time Sense* (1868) and its legacy. European Journal of Cognitive Psychology, 21 (6), 941-960.

Li, Y., 2003. Evaluating the Urban Commute Experience: A Time Perception Approach. Journal of Public Transportation, 6(4), 41 – 67.

Lin, F.-B. and Su, C.-W., 1994. Level-of-service analysis of toll plazas on freeway main lines. Journal of Transportation Engineering, 120 (2), 246 – 263.

Loehlin, J., 1959. The influence of different activities on the apparent length of time. Psychological Monographs, 73 (4), 1 – 27.

Malokin, A., Circella, G., Mokhtarian, P.L., 2019. How do activities conducted while commuting influence mode choice? Using revealed preference models to inform public transportation advantage and autonomous vehicle scenarios. Transportation Research Part A, 124, 82 – 114.

Navandar, Y. V., Singh, M., Dhamaniya, A. and Patel, D. A., 2019. Empirical analysis of level of service at toll plaza using ordered probit model. Transportation Letters, 12 (10), 692 – 700.

Nilsson, R., 2000. Drivers' impressions of front and rear gaps in queues. Ergonomics, 43, 1985 – 2000.

Nirmale, S. K. and Pinjari, A. R., 2020. Discrete choice models with multiplicative stochasticity in choice environment variables: application for accommodating perception errors in driver behavior models. Working paper. Indian Institute of Science.

Obelheiro, M. R., Cybis, H. B. B. and Ribeiro, J. L. D., 2011. Level of service method for Brazilian toll plazas. Procedia Social and Behavioral Sciences, 16, 120 – 130.

Ogden, R. S., Simmons, R.F., Wearden, J. H., 2020. Verbal estimation of the magnitude of time, number and length. Psychological Research, https://doi.org/10.1007/s00426-020-01456-4.

Othayoth, D., Krishna Rao, K. V., Bhavathrathan, B. K., 2020. Perceived level of service at signalized intersections under heterogeneous traffic conditions. Transportmetrica A: Transportation Science 16 (3), 1294– 1309.

Peer, S., Knockaert, J., Koster, P., Verhoef, E. T., 2014. Overreporting vs. overreacting: commuters' perceptions of travel times. Transportation Research Part A 69, 476–494.

Saifuzzaman, M., Zheng, Z., 2014. Incorporating human-factors in car-following models: a review of recent developments and research needs. Transportation Research Part C, 48, 379-403.

Stevens, S.S., 1957. On the psychophysical law. Psychological Review, 64(3), 153 – 181.

Stevens, S. S.,1975. Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. Wiley, New York.

Tenenboim, E. and Shiftan, Y., 2018. Accuracy and bias of subjective travel time estimates. Transportation, 45, 945 – 969.

Treiber, M., Kesting, A., Helbing, D., 2006. Delays, inaccuracies and anticipation in microscopic traffic models. Physica A: Statistical Mechanics and its Applications, 360, 71 – 88.

Treiber, M., Kesting, A., 2013. Traffic Flow Dynamics: Data, Models and Simulation. Springer-Verlag, Berlin Heidelberg.

Wu, X., Levinson, D. M. and Liu, H. X., 2009. Perception of waiting time at signalized intersections. Transportation Research Record 2135, 52 – 59.

Varela, J. M. L., Borjesson, M., Daly, A., 2018. Quantifying errors in travel time and cost by latent variable. Transportation Research Part B, 117, 520 – 541.

van Lint, J.W.C. and Calvert S.C., 2018. A generic multi-level framework for microscopic traffic simulation—Theory and an example case in modelling driver distraction. Transportation Research Part B, 117, 63 – 86.

van Lint, H., Calvert, S., Schakel, W., Wang, M., Verbraeck, A., 2017. Exploring the Effects of Perception Errors and Anticipation Strategies on Traffic Accidents-A Simulation Study. AHFE 2017: Springer.

Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M., Longo, G., 2017. Mitigating the impact of errors in travel time reporting on mode choice modelling. Journal of Transport Geography, 62, 236 – 246.

Vierordt, K., 1868. Der Zeitsinn nach Versuchen. H. Laupp'sche Buchhandlung, Tübingen.

Wiedemann, R., 1974. Simulation des StraBenverkehrsflusses Proceedings of the Schriftenreihe des tnstituts fir Verkehrswesen der Universitiit Karlsruhe, Germany.

Wu, X., Levinson, D., Liu, X., 2009. Perception of Waiting Time at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, 2135, 52 – 59.

Appendix A. More on the Functions $B(D)$ and $g(D)$

In Section 3, when discussing mean of perceived time durations, μ_d, for a given actual (objective) duration, D two functions were introduced; namely, the systematic bias term $B(D)$ and $g(D)$; recall, $g(D) = \mu_d$. The expressions for $B(D)$ and $g(D)$ are reproduced for ready reference.

$$B(D) = \alpha e^{-D/k} + \beta$$

$$g(D) = B(D).D = \left(\alpha e^{-D/k} + \beta\right)D$$

Recall, the parameters of this function have the following restrictions: $\alpha \geq 0, \beta > 0$ and $k > 0$. Also, since D is actual delay, $D \geq 0$. In the two subsections of this appendix some of the properties of the two functions and the monotonicity property of $g(D)$, for $D \in [0, \infty)$, are discussed in greater details.

A.1 Variation of $B(D)$ and $g(D)$ with D

As can be seen from the functional form, $B(D)$ is a monotonically decreasing function ranging from $B(0) = \alpha + \beta$ to $\lim_{D\to\infty} B(D) = B(\infty) = \beta$. If $\alpha > 1$ and $0 < \beta < 1$ then $B(D)$ crosses unity at $D = k \ln\frac{\alpha}{1-\beta}$ from a higher value. Note, for values of $D < k \ln\frac{\alpha}{1-\beta}$, since $B(D) > 1$, on an average the perceived value is greater than the actual value of delay, i.e., $\mu_d = g(D) > D$. Similarly, for values of $D > k \ln\frac{\alpha}{1-\beta}$, since $B(D) < 1$, on an average the perceived value is less than the actual value of delay, i.e., $\mu_d = g(D) < D$. As an example of the behavior of these two functions Figure A.1 is plotted with $\alpha = 1.2$, $\beta = 0.8$ and $k = 100$. The top figure shows the variation of $B(D)$; as can be seen it starts from $\alpha + \beta = 2$ and is asymptotic to $\beta = 0.8$. It crosses unity from a higher value at $D = k \ln\frac{\alpha}{1-\beta} = 179.2$ seconds. The bottom figure plots the average perceived value $g(D)$. In the figure the dash-dot line is a $g(D) = D$ or a 45° line. Again, as explained earlier, for $D < 179.2$ seconds on an average the perceived delays are higher than the actual delay while for $D > 179.2$ the latter is higher.

It can be seen from the plot of $B(D)$ or $g(D)$ that the systematic bias in the perception of delay varies with the actual delay. Also, the proposed functional form can easily accommodate, by choosing $\beta = 1$, past observations that for waiting times, people, on an average, overestimate even though the extent of overestimation may reduce with increasing (actual) delays.

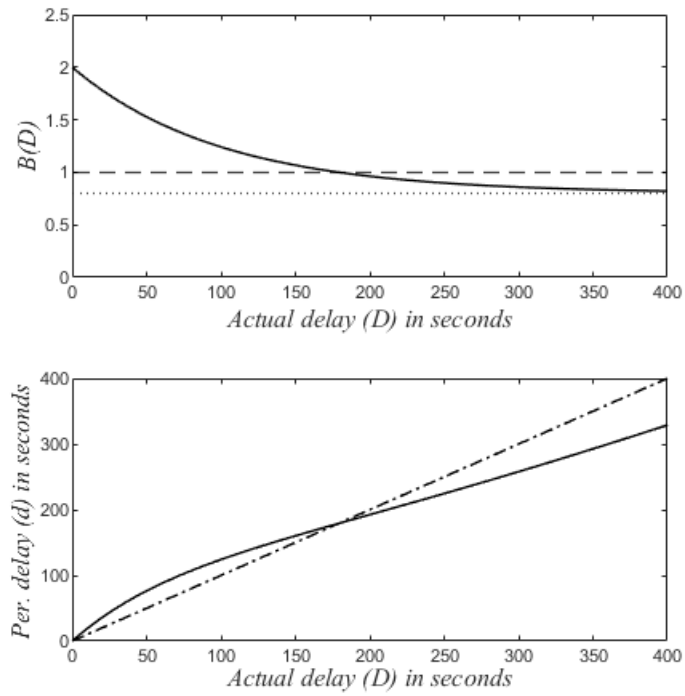

Figure A.1: Plot of $B(D)$ versus actual delay, D (top) and plot of $g(D)$ versus actual delay, D (bottom) for $\alpha = 1.2$, $\beta = 0.8$ and $k = 100$.

It was explained in Section 3 that $g(D)$ should be a monotonically increasing function for $D \geq 0$ and is so as long as $\gamma \in (-1, e^2)$; recall, $\gamma = \alpha/\beta$. In this section these limits are derived. Note,

$$g'(D) = \alpha e^{-\frac{D}{k}} - \frac{\alpha}{k} D e^{-\frac{D}{k}} + \beta$$

For $g(D)$ to be monotonically increasing, $g'(D) > 0$. Hence,

$$\alpha e^{-\frac{D}{k}} - \frac{\alpha}{k} D e^{-\frac{D}{k}} + \beta > 0$$

For, $\frac{D}{k} = 1$, $g'(D) = \beta$ and hence always positive. Now the behavior of $g'(D)$ is investigated when $\frac{D}{k} \neq 1$ (note that $\frac{D}{k}$ is always greater than or equal to zero).

As $\beta > 0$ and $\gamma = \frac{\alpha}{\beta}$, the earlier inequation implies

$$\gamma e^{-\frac{D}{k}} \left(1 - \frac{D}{k}\right) + 1 > 0$$

or

$$\gamma e^{-\frac{D}{k}}\left(\frac{D}{k}-1\right)<1$$

For $g'(D)$ to be greater than zero when $\frac{D}{k}-1>0$ (or $\frac{D}{k}>1$), the earlier expression implies the following limit on γ,

$$\gamma<\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}$$

and for $g'(D)$ to be greater than 0 when $\frac{D}{k}-1<0$ (or $\frac{D}{k}<1$), the following limit on γ is implied,

$$\gamma>\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}$$

The limits on γ are dependent on *D*. However, from these and the earlier discussion on the behavior of $g'(D)$ at $\frac{D}{k}=1$ it can be readily seen that if the following modified limits on γ are satisfied then $g'(D)$ is always positive irrespective of the value of *D*.

For $\frac{D}{k}>1$,

$$\gamma<\min_{D/k}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)},$$

for $\frac{D}{k}<1$,

$$\gamma>\max_{D/k}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)},$$

and for $\frac{D}{k}=1$

$$\gamma\in(-\infty,\infty).$$

Thus, if $\gamma\in\left(\max_{\frac{D}{k}<1}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)},\min_{\frac{D}{k}>1}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}\right)$ then $g'(D)$ is always positive irrespective of the value of *D*. Simple differential calculus based analysis shows that $\max_{\frac{D}{k}<1}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}=-1$ at $\frac{D}{k}=0$ and $\min_{\frac{D}{k}>1}\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}=e^2$ at $\frac{D}{k}=2$. Thus, if $\gamma\in(-1,e^2)$ or (since $e^2\approx 7.39$) if $\gamma\in(-1,7.39)$ then $g'(D)$ is always positive irrespective of the value of *D*.

Figure A.2 shows the determination of the modified limits graphically. It plots $r(D)=\frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}$ versus $\frac{D}{k}$ for $\frac{D}{k}\geq 0$. The plot also shows a dashed line at $r(D)=e^2$ and a dotted line at $r(D)=-1$. As can be seen

from the plot, if γ remains within these two lines then since γ never crosses $r(D) = \frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}$, the value of $g'(D)$ is always positive irrespective of the value of D.

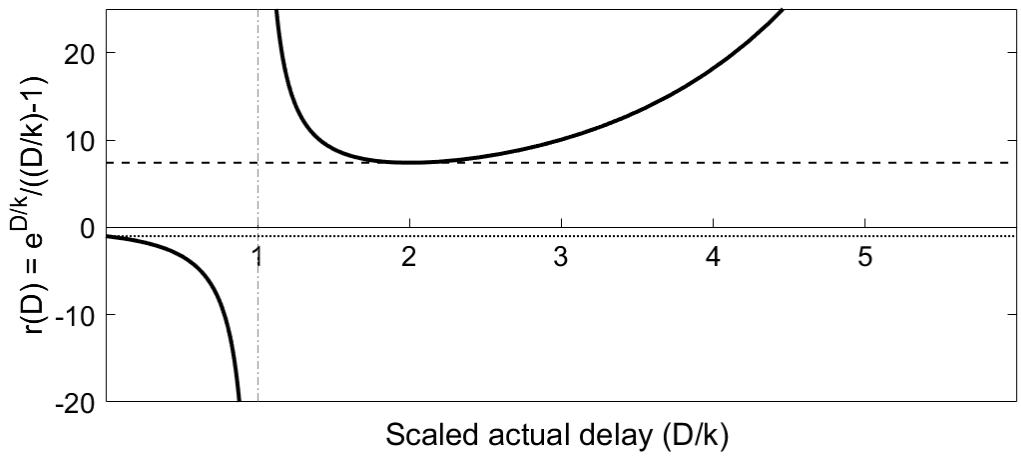

Figure A.2: Plot of $r(D) = \frac{e^{\frac{D}{k}}}{\left(\frac{D}{k}-1\right)}$ versus scaled actual delay, *D/k*.