

The Curse of the First-in-first-out Queue Discipline

Platz, Trine Tornøe ; Østerdal, Lars Peter

Document Version

Accepted author manuscript

Published in:

Games and Economic Behavior

DOI:

[10.1016/j.geb.2017.03.004](https://doi.org/10.1016/j.geb.2017.03.004)

Publication date:

2017

License

CC BY-NC-ND

Citation for published version (APA):

Platz, T. T., & Østerdal, L. P. (2017). The Curse of the First-in-first-out Queue Discipline. *Games and Economic Behavior*, 104, 165–176. <https://doi.org/10.1016/j.geb.2017.03.004>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Jun. 2025



The Curse of the First-in-first-out Queue Discipline

Trine Tornøe Platz and Lars Peter Østerdal

Journal article (Accepted manuscript*)

Please cite this article as:

Platz, T. T., & Østerdal, L. P. (2017). The Curse of the First-in-first-out Queue Discipline. *Games and Economic Behavior*, 104, 165–176. DOI: 10.1016/j.geb.2017.03.004

DOI: [10.1016/j.geb.2017.03.004](https://doi.org/10.1016/j.geb.2017.03.004)

* This version of the article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the publisher's final version AKA Version of Record.

Uploaded to [CBS Research Portal](#): January 2019

© 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

The curse of the first-in-first-out queue discipline*

Trine Tornøe Platz¹

Lars Peter Østerdal²

March 7, 2017

Abstract

We consider a game in which a large number of identical agents choose when to queue up at a single server after it opens. Agents are impatient for service and also incur a cost proportional to time spent in the queue. We show that the first-in-first-out queue discipline and the last-in-first-out queue discipline both lead to a unique equilibrium arrival distribution. However, among all work-conserving queue disciplines, the first-in-first-out performs the worst in terms of equilibrium utility and welfare, while the last-in-first-out performs the best.

JEL codes: C72, D62, R41

Keywords: queue discipline, Nash equilibrium, FIFO, LIFO, welfare, congestion.

*We thank Refael Hassin, Moshe Haviv, an associate editor of this journal, two anonymous referees, and conference and seminar participants at Copenhagen, Odense, Aarhus (NTMD II), Vigo (SING9), Boston (SSCW 2014), Lund, Tokyo (EAGT2015) and Paris for helpful comments. Jesper Breinbjerg has provided excellent research assistance. Financial support from the Danish Council for Strategic Research and The Danish Council for Independent Research | Social Sciences (Grant ID: DFF-1327-00097) is also gratefully acknowledged.

¹Department of Business and Economics, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark.

²Department of Economics, Copenhagen Business School, Porcelænshaven 16A, 1. floor, DK-2000 Frederiksberg, Denmark. Email: lpo.eco@cbs.dk

1 Introduction

In many situations, people wish to be served by a facility that opens at some specified point in time but has limited service capacity. In such situations, each person must decide when to “arrive” at the facility, taking into account at which time he prefers to be served and the waiting time he expects to face upon arrival. Often the person also faces the additional restriction that it is not possible to queue up at the facility before it opens. This is, for example, often the case when patients are calling the doctor in the morning, when passengers at a gate in the airport are boarding a flight, or when an audience is exiting a venue after a concert. A large scale example is the yearly release of the tax assessment notices of 4.5 million Danish residents. The notices are released on the website of the Danish tax authorities on a publicly announced date. A queueing system has been implemented to handle the heavy traffic on the website after the release, but it is not possible to queue up before the release.

In this paper, we consider a facility that serves agents with a fixed capacity from a given point in time. Agents are impatient for service and decide themselves when to queue up at the facility, but they cannot queue up before service begins. We analyze the game of a large number of identical agents choosing arrival times in such situations, and we explore the equilibrium welfare implications of employing specific queue disciplines.

The first-in-first-out (FIFO) queue discipline (also known as first-come-first-served), according to which agents are served in order of arrival and as soon as capacity becomes available, is the most commonly analyzed queue discipline in the literature on queueing with endogenous arrival times. FIFO is considered “fair” by most and is the default discipline in many everyday situations, such as queueing at a grocery store or post office, as well as under more critical circumstances such as in the provision of medical treatment to patients on the waiting list.

It is well known that controlling access to a facility using a queue discipline (such as FIFO) will entail a welfare loss, compared to the socially optimal situation in which people arrive at a rate equal to capacity and there is no queueing. The socially optimal situation is however not obtainable when agents choose themselves

when to queue up. It is therefore of interest to compare equilibrium welfare under different queue disciplines. In particular, as an opposite to FIFO we consider the last-in-first-out (LIFO) queue discipline according to which agents are served in reverse order of arrival to the extent that capacity is available.

We show that when the value of being served is decreasing over time, and the cost of queueing is proportional to time, both the FIFO and the LIFO queue discipline lead to a unique equilibrium arrival distribution. However, among all stochastic work-conserving queue disciplines, FIFO is the *worst* in terms of equilibrium utility and welfare. The LIFO queue discipline is the *best*, though it does not induce socially optimal behavior. Thus, these two queue disciplines provide an upper and a lower bound for equilibrium utility and welfare for a general family of stochastic queue disciplines.

Our setting is related to the classical bottleneck model of Vickrey (1969) that models congestion arising from the existence of a single bottleneck in the context of morning commute and trip timing. The original model was further analyzed and extended by Arnott et al. (1993) and others, see de Palma and Fosgerau (2011) and references therein. Existing literature dealing with Vickrey’s bottleneck model has largely assumed FIFO. An interesting exception is de Palma and Fosgerau (2013) who consider risk-averse agents and a family of stochastic queue disciplines that gives priority to early arrivals (to a vanishing degree), ranging from FIFO to service in random order. They define a “no residual queue” property, which means that there is no queue at the time when the last user arrives at the queue, and they prove that this property holds in equilibrium under all the queueing regimes that they consider. Remarkably, all queue disciplines within this family provide the same equilibrium utility and welfare. In this literature, it has commonly been assumed that the bottleneck facility is open at all times, while agents have a preferred time for passing the bottleneck and will incur a cost from being early or late. In contrast to these models, we consider a setup in which the service facility opens at a specific point in time, and where agents are not able to queue up before service opens.

Another strand of literature on queueing with strategic agents is based on the model from the seminal paper by Naor (1969). In this model, agents with expo-

nentially distributed service times arrive according to a Poisson distribution at a service facility that serves agents according to the FIFO discipline. At arrival they observe the length of the queue and decide whether to join the queue or balk. If they balk, they cannot return. In this somewhat different setting, Hassin (1985) provides arguments to show that the LIFO queue discipline (in contrast to FIFO) results in socially optimal behavior, see Hassin and Haviv (2003, Ch. 2).¹ Within this literature, equilibrium arrival patterns to a server with opening (and closing) times were studied in Glazer and Hassin (1983). Subsequent extensions and variations include Juneja and Jain (2009), Jain et al. (2011), Juneja and Shimkin (2013), Hassin and Kleiner (2011), and Haviv (2013). In line with Hassin and Kleiner and Haviv (2013), we consider a setup in which early arrivals are not allowed.

The paper is organized as follows. In section 2, the model and assumptions are presented. Section 3 contains the results: section 3.1 presents some preliminary findings, and in section 3.2 the FIFO discipline is considered, while 3.3 concerns the LIFO discipline. Section 4 contains concluding remarks.

2 Model

2.1 Basics

Suppose that at time 0 a facility opens where agents can be serviced by a single server, and service continues until all agents have been served and have exited the system. We assume that there is a continuum of agents with a total “mass” of 1, and that the capacity use of each agent is negligible. The service capacity of the server is k per unit of time. In what follows, we let T denote the earliest time within which it is possible to serve all agents. Thus, $T = 1/k$.

A *cumulative arrival distribution* (CAD) is a cumulative distribution function R on $[0, \infty[$, such that for each t , $R(t)$ is the share of agents that have arrived at the service facility at latest at time t . The CAD thereby summarizes the arrival behavior of a population of agents who choose arrival times according to (non-symmetric)

¹See also the discussion in the Jerusalem Report article from July 6, 2009, <http://math.tau.ac.il/~hassin/J-Rep.pdf>.

pure strategies.² R is non-negative, non-decreasing, and $\lim_{t \rightarrow \infty} R(t) = 1$. Let $I_t = R(t) - \lim_{s \rightarrow t^-} R(s)$ and note that I_t is positive if R jumps at t and 0 otherwise. We assume that R is piecewise differentiable.³ At any time t where R does not jump, we interpret the right-derivative $R'_+(t)$ as the rate (“speed”) at which agents arrive. If R jumps at t , we interpret I_t as the share of the entire population that arrives at t . Figure 1 shows an example of a CAD reflecting an arrival distribution where one third of the agents arrive at time zero, there is a second jump at time t_1 , and the arrival rate increases at t_2 during a period of smooth arrivals. By t_3 all agents have arrived.

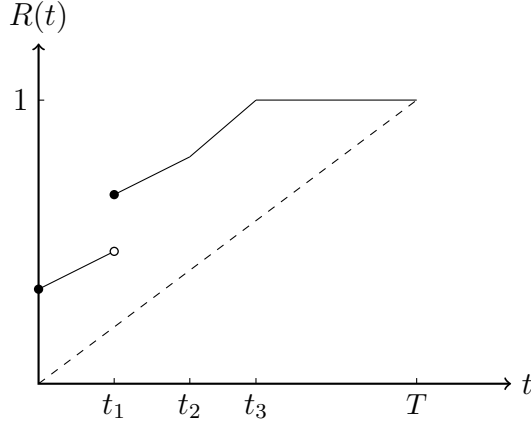


Figure 1: Example of a CAD

The *cumulative exit probability for time s arrival* is a function S^s on $[0, \infty[$ such that for any t , $S^s(t)$ is the probability that an agent arriving at time s is served at latest at time t . We assume that S^s takes values in $[0, 1]$, is non-decreasing, and that $S^s(t) = 0$ for all $t < s$. A profile of cumulative exit probabilities (one for each possible arrival time $s \geq 0$) is denoted $\{S^s\}_{s \geq 0}$.

For a CAD R and a profile $\{S^s\}_{s \geq 0}$, we define a *cumulative exit distribution* C . For any t , $C(t)$ provides the cumulative share of agents served by time t .⁴

$$C(t) = \int_0^t S^s(t) dR(s) = \sum_{0 \leq s \leq t} I_s \cdot S^s(t) + \int_0^t R'_+(s) S^s(t) ds.$$

²For a review of the general theory of large games, we refer to Khan and Sun (2002).

³Since R is non-decreasing, it is differentiable almost everywhere. However, to ensure an intuitive interpretation of the CAD, we limit ourselves to consider only piecewise differentiable CADs.

⁴All integrals in this paper are Lebesgue integrals.

2.2 Feasibility and queue disciplines

A profile $\{S^s\}_{s \geq 0}$ is *feasible* with respect to a CAD R , if the total mass of agents served during any time interval does not exceed serving capacity, that is, if $C(0) = 0$, and $C(y) - C(x) \leq k(y - x)$ for any $0 \leq x < y$.

A *queue discipline* is a mapping that associates with any given CAD a feasible profile $\{S^s\}_{s \geq 0}$ of cumulative exit probabilities.

Given a CAD R , a queue discipline *empties* the queue at time t if $S^s(t) = 1$ for all $0 \leq s \leq t$. A queue discipline is *work-conserving* if, for any R , it empties the queue whenever possible. That is, for any R , the set of times for which the queue is empty is maximal. Note that since emptying the queue at some point in time never makes it more difficult to empty the queue at a later time, such a maximal set is uniquely defined.⁵ In particular, every work-conserving queue discipline will, for a given CAD, empty the queue at the same points in time. Throughout the paper, we limit ourselves to consider work-conserving queue disciplines.

2.3 Agent preferences

All agents have identical preferences. We assume that there is a *queueing cost* c for each unit of time spent waiting in the queue. The *value of being served* at time t is given by the bounded, continuous, and strictly decreasing function w on $[0, \infty[$. The utility of an agent arriving at time $s \geq 0$ and being served at time $t \geq s$ after queueing for $t - s$ units of time therefore equals $w(t) - c(t - s)$. In general, exit time (i.e., the time where an agent is served and exists the queue) is stochastic depending on R and the choice of queue discipline, and we assume that agents want to maximize expected utility.

We denote the expected utility of an agent arriving at time $s \geq 0$ and exiting

⁵To see this, let T_1 and T_2 be maximal sets under disciplines 1 and 2, respectively. These sets are left-closed, so at each point in time we can identify which discipline is emptying the queue most quickly. Define a discipline 3 which combines 1 and 2 such that for each arrival time s , it assigns the agents arriving at s the cumulative exit probability that corresponds to the discipline that empties the queue most quickly after or at s . Since the capacity use, as measured by the change in C , is not higher than under disciplines 1 or 2 for any interval of time, discipline 3 is feasible, and it empties the queue at $T_1 \cup T_2$. Thus, if T_1 and T_2 are both maximal, then $T_1 = T_2$.

according to S^s by $U[S^s]_R$, or when no confusion arises, simply by $U[S^s]$. If t denotes the point in time when an agent is served, then

$$U[S^s] = E[w(t) - c(t - s)|s] = E[w(t) - ct|s] + cs,$$

where the support of t under S^s is $[s, s + \frac{1}{k}]$. Note that the expectation exists since the discipline is work-conserving, implying that the agent has been served with certainty at time $s + \frac{1}{k}$ the latest.

2.4 Equilibrium and welfare

Recall that for a given CAD R , a queue discipline induces for each arrival time s a cumulative exit distribution S^s (known to all agents) and hence, expected utility $U[S^s]_R$. Under a given queue discipline, a CAD R , is a (*pure strategy Nash*) *equilibrium* if no agent can unilaterally improve his expected utility by choosing another arrival time s . That is, R is an equilibrium if $U[S^s]_R \geq U[S^{s'}]_R$ for all $s > 0$ where R either jumps or $R'_+(s) > 0$ and all $s' \geq 0$, $s' \neq s$. Note that in equilibrium, expected utility must be the same for all arrival times s for which R has a jump or $R'_+(s) > 0$. Thus, in equilibrium we can evaluate the welfare associated with a given queue discipline by the equilibrium utility of a representative agent under this queueing regime.

3 Results

3.1 Preliminary results

First we state some general observations about work-conserving queue disciplines and equilibrium cumulative arrival distributions.

We start by noting that for any CAD that constitutes an equilibrium under some queue discipline, agents arrive faster than they can be served, i.e., the share of agents that have arrived at any time $0 < t < T$ exceeds the maximum share of agents that can be served by time t .

Lemma 1. *Let R be an equilibrium CAD under some work-conserving queue discipline. Then (a) $R(t) \geq kt$ for all $0 \leq t \leq T$, with strict inequality for all $0 < t < T$, and (b) $C(t) = kt$, for $0 \leq t \leq T$.*

Proof. Let R be an equilibrium CAD under the work-conserving queue discipline A. Since R is an equilibrium, there does not exist a point in time $t < T$ where the queue is emptied. Indeed, if such a t exists, then $U[S^t] = w(t) > U[S^{t'}]$ for all $t' > t$, and an agent arriving after t would be better off by arriving at t , where he would be served immediately with certainty. Then, for R to be an equilibrium, $R(t) = 1$ must hold. However, since the queue is emptied at t , this implies that all agents must be served by $t < T$, and this contradicts that queue discipline A assigns a feasible profile of cumulative exit probabilities to R . Therefore, if R is an equilibrium, there exists no $t < T$ such that the queue is emptied at t .

Next, suppose there is a time $0 < t < T$, such that $R(t) \leq kt$. We show that this would imply the existence of an $0 \leq s \leq t$ such that the queue is emptied at s , again contradicting that R is an equilibrium. To see this, consider the FIFO discipline under which agents are served in order of arrival as soon a capacity becomes available. Under FIFO, we know that if the queue is not empty for any $0 \leq s \leq t$, then $C(t) = kt$. Thus, when applying FIFO to R , then since $R(t) \leq kt$, there must exist an $0 \leq s \leq t$ such that FIFO empties the queue at s . Since queue discipline A is work-conserving too, it also empties the queue at time s . This contradicts, however, that R is an equilibrium under A. Hence, in equilibrium $R(t) > kt$ for all $0 < t < T$.

For the second part, let $0 \leq t \leq T$. By feasibility, we have $C(t) \leq kt$. Furthermore, since $R(t) \geq kt$ for $0 \leq t \leq T$, it is possible to serve all agents with certainty at latest at time T (for example using the FIFO discipline). Thus, since the queue discipline is work-conserving, all agents will be served with certainty at latest at time $t = T$. This implies that $C(t) \geq kt$, and we are done. \square

We are interested in comparing equilibrium welfare under different queue disciplines. Two observations regarding comparison of equilibrium welfare under arbitrary work-conserving queue disciplines are stated in the lemma below. First, if the same CAD R is an equilibrium under two different queue disciplines, then equilibrium welfare is the same under both disciplines. Second, if two different CADs R and Q are equilibriums under different queue disciplines, and R is first order stochastically dominated by Q , then welfare is higher under the discipline

associated with Q as under the discipline associated with R . Figure 2 illustrates the type of situation considered in the second part of Lemma 2.

Lemma 2. *Let R and Q be equilibrium CADs under the work-conserving queue disciplines A and B respectively. Then (a) If $R(s) = Q(s)$ for all $s \in [0, T]$, equilibrium utility is the same under the two queue disciplines. (b) If $R(s) \geq Q(s)$ for all $s \in [0, T]$, and the inequality is strict for some time interval, then equilibrium utility is higher under discipline B than under discipline A .*

Proof. Let $\{S^s\}_{s \geq 0}$ be the profile of cumulative exit probabilities that queue discipline A associates with R . For the proof of (a), we show that the average expected utility over all agents is independent of the queue discipline and depends on R alone. As noted in section 2.3, the expected utility of an agent arriving at time $s \in [0, T]$ equals $U[S^s] = E[w(t) - ct|s] + cs$. Next, if we take the expectation of this expression with respect to s , thereby taking into account the distribution of arrivals, we can by using the law of total expectation write the overall average expected utility as:

$$\begin{aligned} & E[E[w(t) - ct|s]] + E[cs] \\ &= E[w(t) - ct] + cE[s] \\ &= \int_0^T (w(t) - ct) dC(t) + c \int_0^T s dR(s), \end{aligned}$$

From Lemma 1, $C(t) = kt$, so we can rewrite the average expected utility over all agents as:

$$\int_0^T (w(t) - ct) k dt + c \int_0^T s dR(s). \quad (1)$$

In equilibrium, expected utility is the same for all agents and equals average expected utility. Thus, (1) provides an expression for equilibrium utility given the equilibrium CAD R . We conclude that if $Q(s) = R(s)$ for all $s \in [0, T]$, then equilibrium utility is the same for both CADs. In fact, equilibrium utility is independent of the queue discipline and equal to the expected utility of any agent who arrives at time s where R either jumps or $R'_+(s) > 0$.

For the proof of (b), note that since $R(s) \geq Q(s)$, and the inequality is strict for some interval, we have $\int_0^T s dR(s) < \int_0^T s dQ(s)$. It then follows from (1) that equilibrium utility is higher with Q and hence, under discipline B . \square

Note also that from (1), the average cost of queueing corresponds to c times the area between the R curve and the capacity line kt .

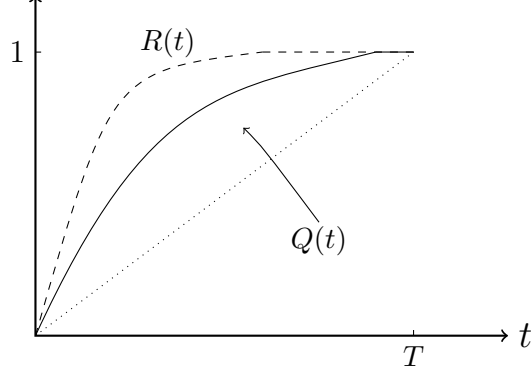


Figure 2: $R(t) \geq Q(t) \geq kt$

In the following subsections we consider more specifically the FIFO and LIFO disciplines.

3.2 First-in-first-out (FIFO) queue discipline

Consider the FIFO queue discipline under which agents are served in order of arrival and as soon as capacity becomes available.

For an agent arriving at a time $s \geq 0$ where $R(s)$ has no jump, exit time is deterministic. He will be served as soon as every agent in the existing queue has been served, i.e., at $\frac{R(s)}{k}$. (Mathematically speaking, the cumulative exit probability for time s arrival is a unit step function). For agents arriving at a time s where R jumps, exit time is uniformly distributed over an interval, $\left[\frac{R(s)-I_s}{k}, \frac{R(s)}{k}\right]$, starting from the point in time, where all the agents that arrived previously have been served.

Next, we turn to the analysis of equilibria under FIFO.

Lemma 3. *Let R be an equilibrium under the FIFO queue discipline. Then R is Lipschitz continuous.*

Proof. We will first show that R does not jump for any $s > 0$, and then we show that this implies that $|R(s_2) - R(s_1)| \leq k|s_2 - s_1|$ holds for all $0 < s_1 < s_2 \leq T$.

First, since R is an equilibrium, it follows from Lemma 1 that $R(T) = 1$. Also, note that an R such that all agents arrive at the same time $s > 0$ cannot be an equilibrium. To see this, assume on the contrary that $I_s = 1$ for some $s > 0$. Then there exists an $\epsilon > 0$ such that $U[S^{s-\epsilon}]_R = w(s - \epsilon) > U[S^s]_R$, since an agent arriving at $s - \epsilon$ is served immediately with certainty. Thus, all agents arriving at $s > 0$ is not an equilibrium.

Next, consider $s_1, s_2 \in [0, T]$ where R jumps or $R'_+ > 0$ and such that $s_1 < s_2$. Since R is an equilibrium, $U[S^{s_1}]_R = U[S^{s_2}]_R$, implying that

$$\begin{aligned} U[S^{s_1}]_R &= E_{S^{s_1}}[w(t_1) - c(t_1 - s_1)] = E_{S^{s_2}}[w(t_2) - c(t_2 - s_2)] = U[S^{s_2}]_R \\ \Rightarrow cE_{S^{s_1}}[(t_1 - s_1)] - cE_{S^{s_2}}[(t_2 - s_2)] &= E_{S^{s_1}}[w(t_1)] - E_{S^{s_2}}[w(t_2)] > 0, \end{aligned}$$

where the inequality follows since $s_1 < s_2$ implies $t_1 < t_2$ under FIFO. Thus for R to be an equilibrium, and since $c > 0$, we must have $E_{S^{s_2}}[t_2] - E_{S^{s_1}}[t_1] < s_2 - s_1$. We show by way of contradiction that for this inequality to hold, R cannot contain any jumps.

To arrive at a contradiction, assume that there is an arrival time s_2 where R jumps, that is $I_{s_2} > 0$. The expected exit time for an agent arriving at s_2 is then $E_{S^{s_2}}[t_2] = \frac{2R(s_2) - I_{s_2}}{2k}$. For an agent arriving at s_1 , $0 < s_1 < s_2$, where R does not jump, exit time is deterministic and equals $\frac{R(s_1)}{k}$. Therefore, $E_{S^{s_2}}[t_2] - E_{S^{s_1}}[t_1] = \frac{R(s_2)}{k} - \frac{I_{s_2}}{2k} - \frac{R(s_1)}{k}$. However, for $s_1 \rightarrow s_2$, $R(s_1) \rightarrow R(s_2) - I_{s_2}$ implying that $E_{S^{s_2}}[t_2] - E_{S^{s_1}}[t_1] \rightarrow \frac{R(s_2)}{k} - \frac{I_{s_2}}{2k} - \frac{R(s_2) - I_{s_2}}{k} = \frac{I_{s_2}}{2k}$. But then for s_1 sufficiently close to s_2 , $E_{S^{s_2}}[t_2] - E_{S^{s_1}}[t_1] < s_2 - s_1$ no longer holds, contradicting that R is an equilibrium. We therefore conclude that since R is an equilibrium, R does not jump for any $s > 0$.

Next, consider s_1, s_2 with no jumps. In this case, exit times for agents arriving at s_1 and s_2 are deterministic and equal to $\frac{R(s_1)}{k}$ and $\frac{R(s_2)}{k}$ respectively, and since R is an equilibrium, it follows that $\frac{R(s_2)}{k} - \frac{R(s_1)}{k} < s_2 - s_1$ implying that $R(s_2) - R(s_1) < k(s_2 - s_1)$. Hence, R is Lipschitz continuous. \square

Let $R'_+(t)$ denote the right-derivative of R . Since R is Lipschitz continuous (and hence absolutely continuous), we have $R(t) = R(0) + \int_0^t R'_+(s)ds$.

Considering the FIFO discipline, note that although Lemma 3 implies that R

has no jumps for $t > 0$, we necessarily have $R(0) > 0$, i.e., a non-zero fraction of the agents arrive at time 0.

We may even have an equilibrium in which all agents arrive at time 0, implying $R(0) = 1$. However, for R with $R(0) = 1$ to be an equilibrium, the expected utility of an agent arriving at time 0 must be higher than the utility from arriving at time T and being served immediately. That is $U[S^0]_R > w(T)$ must hold.

In case not all agents arrive at time 0, the following result holds:

Lemma 4. *Suppose R is an equilibrium under the FIFO queue discipline and $R(0) < 1$. Then the last agent who arrives is served immediately (i.e., the “no residual queue property” holds).*

Proof. From Lemma 1, there exists a time t such that $R(t) = 1$. Let $r = \min\{t | R(t) = 1\}$ be the earliest point in time, where all agents have arrived. Let $R(0) < 1$, and assume that $r < T$. Then since $R(0) < 1$, and there are no jumps for $t > 0$, exit time is deterministic and equal to T for a player arriving at r . This player could therefore increase expected utility by postponing arrival until $t = T$, in which case he would be serviced at time $t = T$ while avoiding waiting time entirely. That is, $U[S^T]_R = w(T) > w(T) - c(T - r) = U[S^r]_R$ for all $r < T$, which contradicts that R is an equilibrium. Therefore, we must have $r = T$, and the last agent that arrives is served immediately. \square

Thus, due to Lemma 4 and FIFO being work-conserving, we have $R(T) = 1$, and if $R(0) < 1$, then we have $R(t) < 1$ for all $t < T$.

Before we proceed, we establish existence of an equilibrium CAD under FIFO.

Lemma 5. *There exists an equilibrium CAD under FIFO.*

Proof. Suppose the queue discipline is FIFO. Let Q be the CAD for which $Q(0) = 1$. Then, an agent that arrives at time $t > 0$ will be served at $\max\{t, T\}$. Thus, either the expected utility of arriving at time 0 is greater than the utility from arriving at $t = T$, in which case Q is an equilibrium since no agent can profitably deviate, or the opposite is true, and Q is not an equilibrium under FIFO. If Q is an equilibrium, we are done. Assume therefore that Q is not an equilibrium. In this case, we provide a constructive argument for the existence of an equilibrium.

From Lemma 4, the expected utility of an agent arriving at time T is $w(T)$. Let I be the fraction of agents that would have to arrive at time 0 in order for the expected utility of these agents to equal $w(T)$. Since Q is not an equilibrium, note that $I < 1$ exists and is uniquely determined.

Note also that an agent arriving immediately after time 0 will at best be served at time $\frac{I}{k}$. For t sufficiently small, an agent arriving in the interval from 0 to t could therefore increase expected utility by arriving at time 0 instead. This implies that the jump at 0 will be followed by a period of time where the CAD is flat, and no mass of agents arrives. Next, let t^* between 0 and I/k be the point where $w(I/k) - c(I/k - t^*) = w(T)$, i.e., we choose t^* such that an agent obtains the same expected utility from arriving at t^* and being served at time I/k as from arriving at time 0.

Now, for any s with $t^* \leq s \leq T$, define $x(s)$ such that $s \leq s + x(s) \leq T$ and

$$w(s + x(s)) - cx(s) = w(T),$$

i.e., for an agent arriving at time s , $x(s)$ is the queueing time that gives the agent the same expected utility as the agents arriving at time 0. Note that $x(s)$ exists and is uniquely determined for each s . Moreover, $x(s)$ is strictly decreasing and continuous, $s + x(s)$ is strictly increasing, and $x(s) \rightarrow 0$ for $s \rightarrow T$.

Now, define $R(s)$ such that $R(s) = I$ for $0 \leq s \leq t^*$, and $R(s) = k(s + x(s))$ for $t^* < s \leq T$. Then by construction $R(s)$ is an equilibrium CAD.⁶ \square

In equilibrium under FIFO, either every agent chooses to arrive at time 0, or some fraction of the agents arrive at time 0 followed by a period (from 0 to t^*) where the CAD is flat (no arrivals) and finally a period where agents arrive smoothly until time T where $R(t) = 1$, see Figure 3. Though the two models differ, the equilibrium CAD just described corresponds to the arrival profile found in Hassin and Kleiner (2011).

Figure 3 shows an example of an equilibrium CAD under FIFO. Note that for an agent arriving at time $s > t^*$, time spent in the queue is given by the horizontal

⁶If assuming e.g. a linear w function, it is straightforward to find explicit expressions for I , t^* and $x(s)$ and thereby for the equilibrium CAD, by following the steps outlined in the proof.

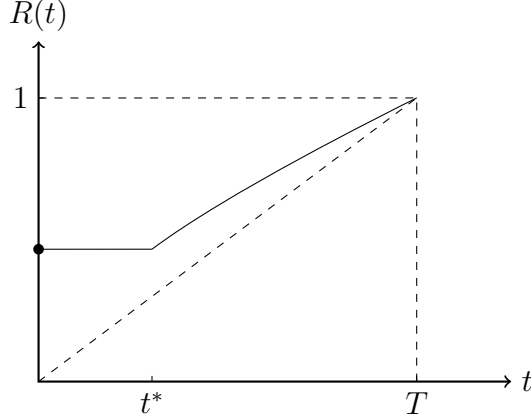


Figure 3: Equilibrium under FIFO

distance between the $R(t)$ curve and the kt line that shows the cumulative number of agents that have been served up until time t . The figure therefore also illustrates how the queueing time decreases with t (from t^*) until it reaches zero at T .

Next, we address the question of whether there is a unique equilibrium CAD under FIFO.

Lemma 6. *There exists at most one equilibrium CAD under FIFO.*

Proof. Let R and Q be two equilibrium CADs under FIFO. We need to show that $R(t) = Q(t)$ for all t .

To arrive at a contradiction, we assume on the contrary that $Q(0) < R(0) \leq 1$ for $t = 0$. From Lemma 4 and the queue discipline being work-conserving, the equilibrium utility of an agent arriving at time 0 is at least $w(T)$ if $R(0) = 1$ and exactly $w(T)$ if $R(0) < 1$. Therefore, since both Q and R are equilibria, and $Q(0) < R(0) \leq 1$, we must have $U[S^0]_Q = w(T) \leq U[S^0]_R$. However, $Q(0) < R(0)$ implies that $U[S^0]_Q > U[S^0]_R$, i.e., that the expected utility of an agent arriving at 0 is greater for Q than for R , a contradiction. Therefore, we must have $R(0) = Q(0)$. Now, if $R(0) = Q(0) = 1$, then we are done. Otherwise, given that $R(0) = Q(0) < 1$, it readily follows that t^* is the same for the two arrival profiles. Next, for agents arriving at time s between t^* and T , there is no jump at s but a positive arrival rate, so exit time is deterministic and equal to $\frac{Q(s)}{k}$ and $\frac{R(s)}{k}$ for Q and R respectively. Furthermore, for both Q and R the expected utility of every agent equals $w(T)$, and

therefore it must be that $R(t) = Q(t)$ for all t . Otherwise, two agents arriving at time s between t^* and T would experience different exit times, and hence, different expected utilities. \square

Having established that there exists at most one equilibrium CAD under FIFO, we move on to state the following negative result regarding the welfare properties of the FIFO queue discipline.

Proposition 1. *FIFO minimizes welfare among all work-conserving queue disciplines.*

Proof. We distinguish between two cases for FIFO equilibrium: (a) all agents arrive at time 0, and (b) some fraction of the agents arrive at time 0 followed by a period where nobody arrives and then a period where agents arrive smoothly until time T when the last agent arrives.

In case (a) total queueing time is the highest possible among work-conserving queue disciplines. Thus, no other queue discipline can do worse, and FIFO therefore minimizes welfare among work-conserving queue disciplines.

In case (b), $r = \min\{t | R(t) = 1\} = T$, and it follows from Lemma 4 that equilibrium utility is equal to $w(T)$. Under every work-conserving queue discipline, all agents are served at latest at time T , in equilibrium. Thus, a player can always choose to arrive at $t = T$ and obtain utility $w(T)$. If, for some queue discipline $r < T$, then this implies that equilibrium utility is at least $w(T)$, since otherwise a player arriving at r could increase expected utility by postponing arrival to $t = T$. Thus, for no queue discipline can equilibrium utility be strictly lower than $w(T)$, and FIFO therefore minimizes equilibrium utility among work-conserving queue disciplines. \square

Intuitively speaking, the problem with the FIFO discipline is that the strict queue discipline gives the agents incentives to join the queue early, which in the end will hurt all agents in equilibrium.

3.3 Last-in-first-out (LIFO) queue discipline

As stated above, the FIFO queue discipline will induce people to queue up “too early” which in turn leads to lower equilibrium utility. In this section, we consider the last-in-first-out (LIFO) queue discipline, which can be considered as an opposite extreme to the FIFO discipline. We will show that not only is the LIFO queue discipline better than the FIFO discipline, it is in fact welfare optimal among all queue disciplines that do not waste capacity. In our setting, the LIFO queue discipline works as follows. It always gives highest priority to those agents who have arrived the latest. When agents arrive continuously at a lower rate than capacity, they are all served immediately. If they arrive continuously at a higher rate than capacity, a fraction of agents are served immediately corresponding to capacity, while the rest must wait to be served until those arriving later have all been served. In that case, an arriving agent faces a lottery over exit times with two possible outcomes: either the agent is served immediately at arrival, or he is served when everyone who arrives later has been served. In case an agent arrives at a jump, he will face a probability distribution over exit times, where the period of potential exit times starts when capacity becomes available after all agents that have arrived (but arrived later than the agent in question) have all been served.

Before turning to consider equilibrium utility under LIFO, we provide some preliminary observations on equilibrium CADs under LIFO and prove uniqueness and existence of an equilibrium under LIFO.

Lemma 7. *Let R be an equilibrium CAD under LIFO, and let $r = \min\{t | R(t) = 1\}$. Then (a) $R'_+(t) > k$ for all $0 \leq t < r$ where R does not jump, (b) R is Lipschitz continuous on $[0, r[$, (c) $R(0) = 0$, and (d) $r < T$.*

Proof. (a) First, note that R cannot jump at time r . If R jumps at time r such that $I_r > 0$, then an agent arriving at r has an expected utility of $U[S^r] = E_{S^r}[w(t^*) - c(t^* - r)]$ where $t^* \in [r, r + \frac{I_r}{k}]$. But then for $\epsilon > 0$ sufficiently small, an agent arriving at r could increase his expected utility by arriving instead at $r + \epsilon$, where he would be served immediately with certainty obtaining an (expected) utility of $U[S^{r+\epsilon}] = w(r + \epsilon) > U[S^r]$. Then, since $R(r) = 1$, and there is no jump at r , an agent arriving at time r is served immediately with certainty.

Now, if there exists a $0 \leq t < r$ such that $R'_+(t) \leq k$, then an agent arriving at time t is also served immediately with certainty, whereby $U[S^t] = w(t) > w(r) = U[S^r]$, and the agent obtains higher utility than an agent arriving at time r . This contradicts that R is an equilibrium CAD. Therefore, $R'_+(t) > k$ for all $0 \leq t < r$ where R does not jump.

(b) First, observe that R is continuous on $[0, T]$. To see this, assume on the contrary that the equilibrium CAD R has a jump at time t , such that $I_t > 0$. Then, since $R'_+(t) > k$ for all $0 \leq t < r$ where R does not jump, an agent arriving at t must wait for service until all agents that arrive after him have been served, implying that he will be served sometime during the interval $[t + \frac{1-R(t)}{k}, t + \frac{1-R(t)}{k} + \frac{I_t}{k}]$ and at $t + \frac{1-R(t)}{k}$ the earliest. However, if we consider some $\varepsilon > 0$ such that $R'_+(t + \varepsilon) > k$, then we see that an agent arriving at $t + \varepsilon$ will be served immediately with probability $p_{t+\varepsilon} = \frac{k}{R'_+(t+\varepsilon)} > 0$ and at time $t + \frac{1-R(t+\varepsilon)}{k}$ with probability $1 - p_{t+\varepsilon}$, implying that for $\varepsilon > 0$ sufficiently small $U[S^{t+\varepsilon}]_R > U[S^t]_R$, which contradicts that R is an equilibrium. Therefore, R is continuous on $[0, T]$.

Next, let b be a fixed constant, where $0 < b \leq r$, and let $\theta > 0$ be such that $0 < b - \theta < b$. We show that R is Lipschitz continuous on $[0, b - \theta]$. For this, note that since R is continuous, $R'_+(s)$ exists for each $s \in [0, T]$. It is therefore sufficient to show that $R'_+(s)$ is bounded. From (a) we know that $R'_+(t) > k$ for all $0 < t < r$, so it remains to be shown that $R'_+(s)$ is bounded from above, i.e., there is a $K > 0$ such that $R'_+(s) < K$ for all $s \in [0, b - \theta]$. For this, suppose on the contrary that there is a sequence s_1, s_2, \dots in $[0, b - \theta]$ such that $R'_+(s_1) < R'_+(s_2) < \dots$ and $R'_+(s_h) \rightarrow \infty$ for $h \rightarrow \infty$. Since $s_h \leq b - \theta$, an agent arriving at s_h who is not served immediately must wait for a period of time of at least θ to be served at some time after r . Since the probability of being served immediately goes to 0 as $h \rightarrow \infty$, the expected utility of an agent arriving at time s_h falls to a level below that of an agent arriving at time r (who is being served immediately with certainty). This contradicts that R is an equilibrium. Thus, $R'_+(s)$ is bounded from above on $[0, b - \theta]$, and the conclusion follows.

(c) Follows readily from (b) by applying the argument under the first part of (b) to the case of $t = 0$.

(d) Since R is Lipschitz continuous (and hence absolutely continuous), we have $R(t) = \int_0^t R'_+(s)ds$. By (a), the desired conclusion follows. \square

Uniqueness and existence of an equilibrium is established in the following lemmas.

Lemma 8. *There exists at most one equilibrium CAD under LIFO.*

Proof. The proof is by contradiction. Let R and Q be equilibrium CADs, $R \neq Q$. Let $r = \min\{t | R(t) = 1\}$ and $q = \min\{t | Q(t) = 1\}$. We distinguish between three cases:

(i) $r < q$: Since the agents arriving at time r in R and time q in Q are served immediately, the equilibrium expected utility is $w(r)$ and $w(q)$ when agents arrive according to R and Q respectively, and equilibrium utility is therefore greater with R , since $r < q$. Now, let $s = \max\{t | R(t) = Q(t), 0 \leq t < q\}$ be the latest point in time at which the share of agents that have arrived is the same for the two CADs. Since $R(r) = 1 > Q(r)$, and Q and R are continuous functions, s is well defined. Moreover, because $r < q$, and s is the latest point in time (before q) such that the curves of the two CADs intersect, it follows that $R'_+(s) > Q'_+(s)$, implying that the probability of being served immediately if arriving at time s is lower with R than with Q . However, an agent arriving at time s , who is not served immediately, will be served at the same later time for both R and Q , thus contradicting that expected utility is higher with R .

(ii) $r = q$: Since the agents arriving at time r in R and at time $q(=r)$ in Q are both served immediately, the expected utility in equilibrium is the same for both CADs. However, since $R \neq Q$, $R(0) = Q(0) = 0$, and $R(r) = Q(q)$, there is some t such that (a) $R(t) > Q(t)$ and $R'_+(t) < Q'_+(t)$ or (b) $Q(t) > R(t)$ and $Q'_+(t) < R'_+(t)$. If (a), then an agent arriving at t has a higher probability of immediate service with R than with Q , and if the agent is not served when arriving at t , then he will be served earlier with R than with Q since less people will arrive later than him. This contradicts that R and Q provide the same ex ante utility in equilibrium. A symmetric argument holds in case (b).

(iii) $r > q$: The case is symmetric to (i) and thus omitted. \square

Lemma 9. *There exists an equilibrium CAD under LIFO.*

Proof. The method of proof is to define a family of CADs, $\{\bar{Q}_b\}_{0 < b < T}$, where for each b , \bar{Q}_b is constructed as the limit of a converging sequence of CADs, and show that a member of this family is an equilibrium CAD. We start by providing some useful notation.

Let b be a constant, $0 < b < T$. For each $s \in [0, b]$, let $p^1(s)$ be the (uniquely determined) probability such that the expected utility of an agent who arrives at time s and is served immediately with probability $p^1(s)$ and otherwise served at time T with probability $1 - p^1(s)$ is equal to the utility of an agent arriving at time b who is served immediately with certainty. Let $r^1(s) = k/p^1(s)$ denote the corresponding rate of arrival. Since $w(s)$ is continuous, $r^1(s)$ and $p^1(s)$ are continuous too. Note that $r^1(s) \geq k$ for all s . Furthermore, let l denote the straight line with slope k that goes through the point $(b, 1)$. Then we define $\sigma_b(1)$ as the smallest number s where the graph of $\int_0^s r^1(t)dt$ hits the line l if it hits the line for $s < b$, and let $\sigma_b(1)$ be b otherwise.

We are now ready to define a sequence of CADs Q_b^1, Q_b^2, \dots . We start by defining Q_b^1 and then define Q_b^h recursively for $h > 1$.

First, define Q_b^1 such that

$$Q_b^1(s) = \int_0^s r^1(t)dt$$

for $s \in [0, \sigma_b(1)[$; if $\sigma_b(1) < b$, then $Q_b^1(s)$ is identified with the segment of the line l for $s \in [\sigma_b(1), b[$; finally, $Q_b^1(s) = 1$ for $s \geq b$, see Figure 4. By the Fundamental Theorem of Calculus, Q_b^1 is differentiable (and hence Lipschitz continuous) on $]0, \sigma_b(1)[$.

Now, suppose that a CAD $Q_b^{h-1}(s)$ and a point $\sigma_b(h-1)$ have been defined such that $0 < \sigma_b(h-1) \leq b$, $Q_b^{h-1}(s)$ is differentiable on $0 \leq s < \sigma_b(h-1)$, and the derivative on this domain is everywhere greater than or equal to k . Furthermore, $Q_b^{h-1}(s)$ is identified with the line l for $\sigma_b(h-1) \leq s \leq b$, and $Q_b^{h-1}(s) = 1$ for $s \geq b$.

Next, for each $s \in [0, b]$, let $\beta^{h-1}(s)$ denote the point in time where the straight line from $(s, Q_b^{h-1}(s))$ with slope k meets the horizontal line connecting $(0, 1)$ and

$(T, 1)$. Further, let $p^h(s)$ be the (uniquely) determined probability where the expected utility of an agent who arrives at time s and is served immediately with probability $p^h(s)$ and otherwise served at time $\beta^{h-1}(s)$ with probability $1 - p^h(s)$ is equal to the utility of an agent arriving at time b who is served immediately with certainty. Let $r^h(s) = k/p^h(s)$.

Define $\sigma_b(h)$ as the smallest number s such that the graph of $\int_0^s r^h(t)dt$ hits the line l if it hits the line for $s < b$. Otherwise, let $\sigma_b(h) = b$. We then define $Q_b^h(s)$ such that

$$Q_b^h(s) = \int_0^s r^h(t)dt,$$

and if $\sigma_b(h) < b$, then $Q_b^h(s)$ is identified with the segment of the line l from the point $\sigma_b(h)$ up to b . Finally, $Q_b^h(s) = 1$ for $s \geq b$. Since $w(s)$ is continuous, and $Q_b^{h-1}(s)$ is continuous on $[0, b]$, $r^h(s)$ and $p^h(s)$ are continuous on $[0, b]$. Note that $Q_b^h(s)$ is non-decreasing and by the Fundamental Theorem of Calculus it is differentiable (and hence Lipschitz continuous) on $]0, \sigma_b(h)[$.

It follows from the recursive construction that $\sigma_b(1) \geq \sigma_b(2) \geq \dots$, $Q_b^1(s) \leq Q_b^2(s) \leq \dots$ for all $s \in [0, T]$, and $r^1(s) \leq r^2(s) \leq \dots$ for all $s \in [0, b]$. We have $\lim_{h \rightarrow \infty} \sigma_b(h) > 0$, since if $\lim_{h \rightarrow \infty} \sigma_b(h) = 0$, the highest value of the derivative of $Q_b^h(s)$ on $[0, \sigma_b(h)]$ would go to infinity as $h \rightarrow \infty$, implying that the expected utility of a h -sequence of agents in $[0, \sigma_b(h)]$ at some point in the sequence would fall below the utility of an agent who is arriving at b and served immediately, contradicting the construction of Q_b^1, Q_b^2, \dots . Also, for each $0 \leq s \leq \lim_{h \rightarrow \infty} \sigma_b(h)$, $\lim_{h \rightarrow \infty} r^h(s)$ is finite, for the same reason as above. We now define $\bar{Q}_b(s)$ by $\bar{Q}_b(s) = \lim_{h \rightarrow \infty} Q_b^h(s)$ for each $s \in [0, T]$. Let $\bar{\sigma}_b = \lim_{h \rightarrow \infty} \sigma_b(h)$.

So far, b has been fixed. We now make the following observations:

- (i) For b sufficiently close to T , $\bar{\sigma}_b < b$. This follows from the fact that $Q_b^h(s)$ is increasing in b .
- (ii) For b sufficiently close to 0, $\bar{\sigma}_b = b$. This follows (again) from the fact that $Q_b^h(s)$ is increasing in b (since it means it decreases when b goes to zero).
- (iii) $\bar{\sigma}_b$ is continuous in b . This follows since $w(s)$ is continuous and by the construction of $\bar{\sigma}_b$ and \bar{Q}_b .

Combining (i), (ii) and (iii), we get that there exists $b = b^*$ such that $\bar{\sigma}_b = b^*$ and $\bar{Q}_{b^*}(s)$ is continuous on $[0, T]$. By construction, with the CAD $\bar{Q}_{b^*}(s)$ the expected utility for an agent arriving at any time s with $0 \leq s < b^*$ is equal to the utility of an agent arriving at b^* (who is served immediately with certainty), and thus, $\bar{Q}_{b^*}(s)$ is an equilibrium CAD under LIFO. \square

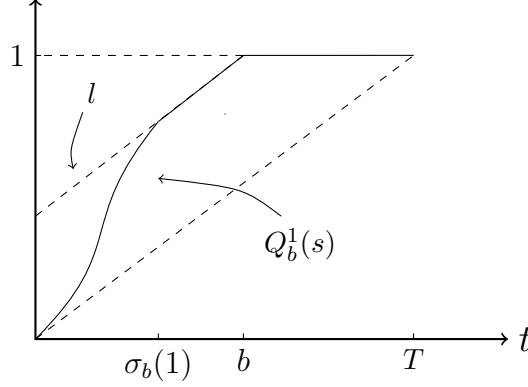


Figure 4: Illustration of a possible Q_b^1

Having established existence and uniqueness of LIFO, we now move on to analyzing the welfare properties.

Lemma 10. *Let R be an equilibrium CAD under LIFO, and let Q be an equilibrium under some other queue discipline that gives higher welfare. Let $q = \min\{t | Q(t) = 1\}$ and $r = \min\{t | R(t) = 1\}$. Then $q < r$, i.e., the latest arriving agent according to Q arrives earlier than the latest arriving agent according to R .*

Proof. An agent arriving at r under LIFO is served immediately. Then, since LIFO gives lower welfare by assumption, the expected utility of an agent arriving at q under the alternative discipline must be strictly greater than $w(r)$. Therefore, this agent must arrive earlier than r , as illustrated in Figure 5. \square

Lemma 11. *Let R be an equilibrium CAD under LIFO, and let Q be an equilibrium under some other queue discipline, such that $Q(\bar{t}) = R(\bar{t})$ for some $\bar{t} < T$, and $Q(t) \geq R(t)$ for all $\bar{t} \leq t \leq T$ with strict inequality for some interval of time. Then the equilibrium utility (and welfare) is higher under LIFO.*

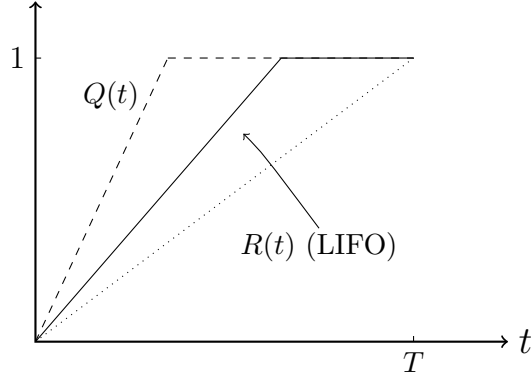


Figure 5: LIFO and a possible equilibrium under another discipline giving higher welfare

Proof. Since $Q(\bar{t}) = R(\bar{t})$, the share of agents arriving at \bar{t} or later is the same for the two CADs, see Figure 6. Under LIFO, all these agents will get priority over the agents already in the queue at time \bar{t} , and since no queue discipline can do better than this, average utility for this group of agents under LIFO must be higher than the average utility for the same group of agents under the queue discipline for which the distribution Q is an equilibrium. \square

The situation from Lemma 11 is illustrated in Figure 6.

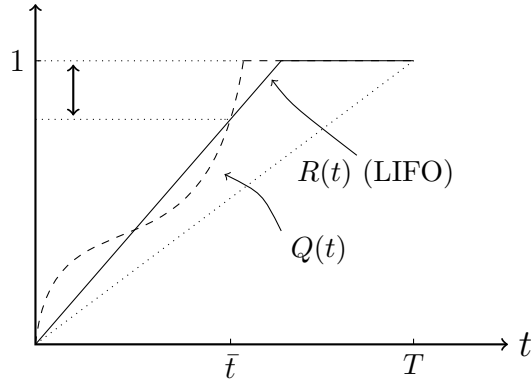


Figure 6: Higher welfare under LIFO

Now, if there exists an equilibrium CAD under an alternative queue discipline that fulfills the requirement of Lemma 10, then by Lemma 11 the equilibrium welfare

will be higher under LIFO. Thus, it follows from Lemmas 10 and 11 that no queue discipline induces an equilibrium giving higher welfare, and we can therefore state the following proposition.

Proposition 2. *LIFO maximizes welfare among all (work-conserving) queue disciplines.*

Note that the LIFO queue discipline maximizes welfare not only among work-conserving queue disciplines but on the class of stochastic queue disciplines in general (work-conserving or non-work-conserving).⁷

3.4 Welfare properties

We are now ready to state the main result on the welfare effects of work-conserving queue disciplines.

Theorem 1. *Within the family of work-conserving queue disciplines, FIFO gives the lowest equilibrium utility while LIFO gives the highest equilibrium utility. Hence, the equilibrium utility (and welfare) of any work-conserving queue discipline is bounded from above by the LIFO and from below by the FIFO queue discipline.*

Proof. Follows immediately from Propositions 1 and 2. □

While LIFO gives higher equilibrium welfare than other queue disciplines it nevertheless does not impose socially optimal behavior. This is caused by agents choosing arrival time without taking into account the adverse effect their arrival has on those already waiting in line.

4 Concluding remarks

The results in this paper suggest that the commonly used FIFO queueing discipline employed for example when impatient users wait in on-line queues, or when airport passengers are waiting to board a flight, is welfare inferior to disciplines that gives less strict priority to those who arrive first. Complementing the findings of Hassin

⁷To see this, observe that Lemmas 10 and 11 also hold if the queue discipline being compared to LIFO is non-work-conserving.

(1985), these results show that the traditional use of the FIFO queue discipline may, counterintuitively perhaps, be a curse rather than a blessing.

However, while the LIFO discipline has some theoretically appealing properties, logistical limitations and acceptability issues play an important role in practical queueing discipline design, and in this respect implementation of LIFO features some obvious challenges.

In terms of acceptability, LIFO might be considered less fair than FIFO due to the cultural and traditional habit of using FIFO as the default queueing discipline. However, as no agent is discriminated on the grounds of individual characteristics, LIFO is arguably fair from an ex-ante perspective. From an ex-post perspective, agents are treated differently (some will be served immediately at arrival while others are forced to wait in the queue for longer periods of time), which may lead to envy. People's attitude toward - and actual arrival decisions under - the LIFO discipline remains a behavioural issue for future research.⁸

In terms of implementing the LIFO discipline, some types of facilities would obviously face difficulties in connection with the practical execution of service. These would include low-tech queueing facilities where agents have to physically line up like when people are exiting a building. In some situations, a compromise between FIFO and LIFO, like service in random order, might be a preferred alternative. For other types of service facilities LIFO is more easily implemented. Take for example on-line queueing systems, as in the example of the release of tax assessment notices, and telecommunication queues, as when customers call a support center, or patients call their doctor to make an appointment in the period after the telephones are opened. In such cases, the computer or phone support software could easily be modified to serve the latest arriving customer rather than the first arriving customer.

In the present model, we assume identical agents and a cost of queueing that is proportional to queueing time. Therefore, an area for future research will be to consider heterogeneous utility functions, for example, allowing for different agent types in terms of differences in preferred time of service or queueing cost. The latter

⁸Breinbjerg et al. (2016) takes a first step towards uncovering such issues.

approach was taken in Jain et al. (2011). Furthermore, generalized cost functions could be considered, and in particular it seems relevant to consider cost functions with increasing marginal costs of waiting in the queue. Existence and uniqueness would be preserved for both FIFO and LIFO in such instances.⁹ However, the comparison of welfare under different equilibrium CADs becomes more complex, since Lemma 2 no longer holds. This implies that while LIFO will still perform at least as good as FIFO in terms of equilibrium welfare, the two queue disciplines may no longer provide a bound for equilibrium welfare.

Finally, it is worth bearing in mind that our model assumes a continuum of agents and a continuous strategy space to approximate a queuing situation with a fixed and very large number of identical agents. As discussed by Otsubo and Rapoport (2008), it is common practise in the transportation science and economics literature to use continuous models for analyzing phenomena that are essentially discrete. They compare Vickrey’s (continuous) bottleneck model to a discretized version, and their numerical (FIFO) equilibria computations indicate that as the population size grows the difference between the continuous and discrete models of traffic congestion diminishes and can safely be ignored. Likewise, Juneja and Shimkin (2013) in their finite-user queueing model with tardiness costs, exponentially distributed service time, early arrivals and no closing time, show convergence of the (FIFO) equilibrium solution to that of the associated continuous model as the number of users is increased. It is beyond the scope of this paper to compare our results to those from a discretized model with a growing number of agents, but a similar conclusion might be anticipated.¹⁰

References

- [1] Arnott, R.A., de Palma, A. and R. Lindsey, 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 83, 161-179.

⁹This follows immediately from inspecting the relevant proofs.

¹⁰Breinbjerg et al. (2016) show for their experimental setting with three agents and linear utility that the best equilibrium under LIFO is always at least as good as any equilibrium under FIFO and SIRO (Service in Random Order). However, they do not analyze more general discrete settings.

- [2] Breinbjerg, J., Sebald, A. and L.P. Østerdal, 2016. Strategic behavior and social outcomes in a bottleneck queue: experimental evidence. *Review of Economic Design*, 20, 207-236.
- [3] de Palma, A. and M. Fosgerau, 2011. Dynamic traffic modeling, in A de Palma, R Lindsey, Quinet, E. and R. Vickerman (eds), *A Handbook of Transport Economics*. Edward Elgar Publishing, Incorporated.
- [4] de Palma, A. and M. Fosgerau, 2013. Random queues and risk averse users. *European Journal of Operational Research*, 230, 313-320.
- [5] Glazer, A. and R. Hassin, 1983. $M/M/1$: On the equilibrium distribution of customer arrivals. *European Journal of Operational Research*, 13, 2, 146-150.
- [6] Hassin, R., 1985. On the optimality of first come last served queues. *Econometrica*, 53, 1, 201-202.
- [7] Hassin, R. and M. Haviv, 2003. *To queue or not to queue: equilibrium behavior in queueing systems*. Kluwer, Dordrecht.
- [8] Hassin, R. and Y. Kleiner, 2011. Equilibrium and optimal arrival patterns to a server with opening and closing times. *IEEE Transactions*, 43, 3, 164-175.
- [9] Haviv, M., 2013. When to arrive at a queue with tardiness costs? *Performance Evaluation*, 70, 387-399.
- [10] Jain, R., Juneja, S. and N. Shimkin, 2011. The concert queueing problem: to wait or to be late. *Discrete Event Dynamic Systems*, 21, 103-138.
- [11] Juneja, S. and R. Jain, 2009. The concert/cafeteria queueing problem: a game of arrivals. *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- [12] Juneja, S. and N. Shimkin, 2013. The concert queueing game: strategic arrivals with waiting and tardiness costs. *Queueing Systems*, 74, 4, 369-402.

- [13] Khan, M.A. and Sun, Y.N., 2002. Non-cooperative games with many players, in: R. Aumann and S. Hart (eds), *Handbook of Game Theory with Economic Applications* Volume III, Elsevier Science, Amsterdam.
- [14] Naor, P., 1969. The regulation of queue size by levying tolls. *Econometrica*, 37, 1, 15-24.
- [15] Otsubo, H. and A. Rapoport, 2008. Vickrey's model of traffic congestion discretized. *Transportation Research Part B: Methodological*, 42,10, 873-889.
- [16] Vickrey, W.S., 1969. Congestion theory and transport investment. *The American Economic Review*, 59, 2, 251-260.