Carnegie Mellon University
Research Showcase @ CMU

Heinz College Research | Heinz College

2007

Might Randomization in Queue Discipline Be Useful When Waiting Cost is a Concave Function of Waiting Time?

Jonathan P. Caulkins
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/heinzworks

Recommended Citation
.

This Working Paper is brought to you for free and open access by the Heinz College at Research Showcase @ CMU. It has been accepted for inclusion in Heinz College Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Might Randomization in Queue Discipline Be Useful When Waiting Cost is a Concave Function of Waiting Time?

Jonathan P. Caulkins

Carnegie Mellon University
Qatar Campus and H. John Heinz III School of Public Policy & Management

Abstract
This paper raises the question of whether some degree of randomization in queue discipline might be welfare enhancing in certain queues for which the cost of waiting is a concave function of waiting time, so that increased variability in waiting times may be good not bad for aggregate customer welfare. Such concavity may occur if the costs of waiting asymptotically approach some maximum (e.g., for patients seeking organ transplants who will not live beyond a certain threshold time) or if the customer incurs a fixed cost if there is any wait at all (e.g., for knowledge workers seeking a service or piece of information that is required to proceed with their current task, so any delay forces them to incur the "set up charge" associated with switching tasks).

1. Introduction

It is common to assume that "First Come First Served" (FCFS) is the best queue discipline for queues whose customers are people, possibly modified to reflect different priority classes. "Last Come First Served" (LCFS) might be useful to computer scientists programming stacks or when service time increases when customers are not served immediately. "Service In Random Order" (SIRO) might be fine when the customers are inanimate objects, but (particularly in England and its former colonies) FCFS is the norm for queues comprised of people. This bias is not without basis. Deviating from FCFS leads to "slips" and "skips" that violate a sense of justice, sometimes with severe erosion of customer satisfaction (Larson, 1987).

However, this paper explores the possibility that FCFS may not always be preferred, particularly when not everyone will make it to the front of the queue and/or when there is unusual value in having a very short wait. In those cases, occasionally pulling someone into the server from further back in the queue might be preferable with respect to certain (though not all) performance metrics.

To illustrate, consider the following contrived example. Two people apply to join an (initially empty) organ donation queue moments apart. The two are identical from a medical perspective, and both will die in exactly 12 months without a transplant. A single organ will become available in 12 months. How will the outcome differ with a FCFS vs. SIRO queue discipline?

First note what will not differ. There will be no change in the number of deaths (one of the two will die in 12 months regardless of the queue discipline) or life-years saved or lost. Nor will there be any difference in waiting time; in both cases there will exactly two person-years spent waiting in queue.

One thing that will differ is the perception of time spent in queue. With SIRO, both patients spend a year knowing that they have a 50/50 chance of living or dying at the end of the

year. With FCFS, one person merely waits for a year, knowing with certainty that he or she will get a transplant. The other essentially sits not in queue but on death row, knowing his or her death will come at the appointed hour.

Which is preferable? Arrow (1951) assures us we cannot answer that question in any objective sense by adding the two patients' utilities, but we can consider it using a "veil of ignorance" test (Rawls, 1971). Imagine we get to chose between being born into a world with FCFS or with SIRO discipline given that we are equally likely to be either patient (the one who registered first by a few moments or the one who was second). Essentially that boils down to the following choice, sketched as a decision tree and with the position on the tree's arms indicating passage of time:

Figure 1: FCFS and SIRO Compared When One Organ Will Become Available in One Year and There are Two Patients Who Will Only Live One Year Without a Transplant

To use the jargon of Quality Adjusted Life Years (QALYs), preferences for the FCFS vs. SIRO branches depend on how much life quality is lost during a year of facing certain death vs. spending a year facing a coin toss that will determine whether one lives or dies. If the first is less than twice the second, one prefers the FCFS branch. Otherwise, one prefers the SIRO branch.

There is another way to think about this choice. Suppose you are confronted today with the following terrible coin-toss lottery: Heads you live; tails you die one year from today (irrespective of when the coin is tossed). There is no way to avoid this lottery or to alter its odds or outcomes. All you control is when the coin is tossed. When would you like to toss the coin? Now? In six months? Or not until a moment or a day before the death sentence would be carried out?

I would not argue that most people would or should prefer the SIRO branch. I personally would have a hard time making and defending a choice. However, this example illustrates that the choice of queue discipline can matter, even if it has no effect on the standard performance metrics. Furthermore, there exist utility functions (in this case concerning the likelihood of ever being served/saved) such that SIRO can be preferable to FCFS.

Having motivated the inquiry, this note proceeds in three parts. The next section reviews the standard case for FCFS to remind ourselves what (few) assumptions underpin that preference. This helps make clear why certain queues might be exceptions to the rule. The following section examines a slightly less simplistic depiction of a transplant queue for which some randomness might be appealing. That example is only slightly less simplistic since serious analyses of organ assignment and transplant queues are quite elaborate (e.g., Zenios, 2002; Xuanming and Zenios, 2004, 2006; Alagoz et al., 2004, 2007), whereas this note is meant to be a think piece that puts forward a contrarian idea, not a detailed analysis of any particular application. The final substantive section advances a similarly simplistic example motivated by an entirely different domain, generic symbol processing or knowledge work, to indicate that the contrarian insight might be relevant beyond just transplant queues.

2. Reminder of Why First Come, First Served Queues Are Appealing

The most obvious drawback to deviating from FCFS is the creation of "slips" and "skips" in which some people will be served sooner than others who have "seniority" by virtue of having

been waiting longer (Gordon, 1987). Furthermore, under fairly general circumstances, the rules governing which customer next enters service when a server becomes free do not affect *average* customer waiting times. Fuhrmann and Iliadis (1994, p.250-251) outline three conditions that jointly produce such invariance.

- Service is non-preemptive
- Selection of customers for service is independent of their subsequent service times.
- If the queue is non-empty, the next customer enters service as soon as the previous customer completes service.

To the extent that queue discipline does not affect average waiting time and FCFS guarantees no "slips" or "skips", there must be some compelling additional or outside consideration for not wanting to implement FCFS.

The literature describes some such considerations. There are times, particularly in telecommunications applications, when keeping track of queue positions poses a computational burden that is significant relative to other cost considerations, so some simpler discipline, such as SIRO, may be preferred. Likewise, in network switching, the real customers are not associated with individual tasks but rather with streams of packets, so discussions of fairness focus on "flow-fairness" or "stream-fairness" not the fairness of how individual jobs are treated. Also, when service times differ across customers, other rules (such as Shortest Processing Time first, SPT) offer advantages. Finally, if service time increases with waiting time, LCFS might be preferred. Two examples of this would be solving crimes and tracing people who have come in contact with an infected person since in both cases memories and evidence fades over time. However, none of these bases for deviating from FCFS are relevant in applications such as organ transplant queues. (In organ transplant queues, the time until the next patient gets to leave the queue by having an organ become available is determined by the organ supply process, not by the characteristics of the last person who left the queue to get an organ, so SPT is not viable.) Nor are they relevant to the knowledge work application considered in Section 4.

The intuitive notion that FCFS is the most fair can be made rigorous. Avi-Itzhak and Levy (2004) argue from axiomatic principles that for a wide class of service disciplines, the variance of waiting time is a good measure of (departures from) fairness, and Vasicek (1977) showed that under fairly general circumstances, variability is minimized by FCFS, maximized by LCFS, and takes on an intermediate value for SIRO. Indeed, this rank order of FCFS beating SIRO with LCFS bringing up the rear holds for any *convex* function of waiting times (Vasicek, 1977). The qualification concerning convexity is key. The sound bite message of this note is that such convexity may be the norm, but it is not a universal law.

3. Elaboration of a Transplant Queue Example

Given this review of why FCFS is generally preferred, what is special about organ transplant queues that might exempt them from this general preference? Xuanming and Zenios (2004) have advanced one argument predicated on the reality that not all organs are equally desirable and the patient at the front of the queue does not have to accept the first organ that becomes available. With FCFS the person at the front of the queue may be too selective about which organs to accept, generating excessive organ wastage. (Cf. also Alagoz et al., 2007)

Here, the issue is whether conventional measures of efficiency and fairness capture all relevant considerations. Avi-Itzhak and Levy's (2004) axiomatic proof that variance in waiting time is a good measure of (un)fairness applies only to fairness with respect to the *queue discipline*. It does not necessarily apply to the fairness of the overall *queueing system*. Avi-Itzhak et al. (2004) make the distinction clear with a simple example. If an emergency room identifies the next patient to enter service by asking "Who is the sickest?" the result might violate the principle of seniority, but it does not violate everyone's intuitive notion of fairness because it serves another, (some would say higher) sense of justice.

Indeed, transplant queues routinely violate the principle of seniority by giving priority to sicker patients. Alagoz et al. (2007, p.26) describe seniority as being used a second or third-tier tie breaking criteria in the United Network for Organ Sharing (UNOS) rules for allocating cadaveric livers, and even then seniority is not literally time in queue but rather time in queue at a given health state.[1]

The argument here is distinct from this sickest person first exception. In our simplistic examples there are no differences in health status across patients. However, there is a parallel in the sense of wanting to consider sacrificing one fairness principle (not violating seniority) for the sake of some other factors that pertain to overall performance of the queueing system.

In particular, there are three such factors.

1) Utility as a function of waiting time might be concave, not convex, at least over some ranges of times,
2) The queue discipline might directly affect the perceived cost of waiting, and
3) The queue discipline might directly affect objective outcomes associated with waiting.

Waiting cost being concave in time spent waiting stems from the fact that in organ transplant queues death creates a morbid form of reneging that shields customers from extremely long waits in queue. For someone who will die within a year without a transplant, the disutility of what would have been a ten year wait is no worse than the disutility of what would have been a two year wait.

The second factor relates to the value of hope and whether having at least some reasonable probability of eventually getting an organ makes time spent waiting in queue more tolerable. Hope is rarely factored into queueing analyses or welfare analyses more generally. However, hope exists not just in Pollyanna's world; it is empirically measurable in a scientific sense when defined as utility being enhanced by delaying the resolution of uncertainty (Chu and Ho, 1994). Whether or not delaying resolution of uncertainty is a net gain or a net cost for transplant patients is an empirical question because of the complicated ways people process risk-related outcomes, including weighting subjective probabilities nonlinearly (Prelec, 2000) and having certainty effect biases as probabilities become very small (Tversky & Kahneman, 1992). The empirical studies of genetic testing or other screening for disease predisposition reach conflicting conclusions about effects on psychosocial welfare (cf., Arver et al., 2004; Wermer et

[1] According to Alagoz et al. (2007, p.26) "Patients are stratified within Status 1 [most ill] using points, which are assigned based on the compatibility of the patient's blood type with the donor's type and waiting time within Status 1. For non-Status 1 patients with the same MELD [Model for End stage Liver Disease] score, a liver is offered to patients with an exact blood-type match first, compatible patients second, and incompatible patients last. If there are several patients having the same blood-type compatibility and MELD scores, the ties are broken by patient waiting times. For MELD patients, waiting time is calculated as the time accrued by the patient at or above her current MELD score from the date that she was listed as a candidate for liver transplantation."

al., 2005). However, in Chu and Ho's specific sense, SIRO or other randomness can create hope, and it is at least possible that could reduce subjectively experienced waiting costs.

The third argument is that having hope might not only make the wait more tolerable, but also actually improve health outcomes. There is a large and contested literature investigating whether hope or optimism improve health outcomes. It is beyond the scope of this note to review that literature. However, notwithstanding the fact that some people disagree, it is clear that (1) at least some other people believe that having hope can improve health outcomes and (2) there are literature reviews and empirical studies that offer support for that belief (e.g., Mondloch et al., 2001; Allison et al., 2003).

In addition, there is a fourth factor worth mentioning. Organ transplant queues are "blind" in the sense that customers do not observe each other waiting. Hence, patients have no way to know if and when they have been skipped over by someone with less seniority. Avi-Itzhak et al. (2004) address queue blindness directly, arguing that blindness does not make deviations from FCFS any less unfair. Without disputing that, if one is considering giving up FCFS fairness for the sake of gaining something else (higher average utility and/or greater hope), then it is reasonable to ask in a utilitarian sense, how much skip-related outrage will be generated while trying to make progress on these other goals? To the extent that an organ transplant queues' blindness softens some of the outrage because one merely knows in general that the discipline allows skipping, but never knows if and when one has actually suffered a skip, that softening is germane to balancing the consequences of reduced fairness against other gains.

Given these observations, it is useful to compare FCFS and SIRO with respect to a specific model. The introductory example above is unrealistic because it offers choices involving certain life and certain death. In reality, there is no certainty on either side. So suppose that N people register to obtain an organ at essentially the same time and in the same health state, namely one that gives them 12 months and only 12 months to live without a transplant. Suppose further that at the end of the year the number of organs available is a random variable that is binomially distributed with $n = N$ and some probability θ that might be on the order of 0.25 (meaning in expected value terms, one-quarter of those needing a transplant will get one, and three-quarters will die without getting a transplant).

Now the decision tree in Figure 1 looks a bit different. The SIRO option offers a 100% chance of spending a year facing probability θ of surviving beyond the end of that year. With the FCFS branch you are equally likely to spend that year knowing your survival probability is $P\{X \geq n\}$, for $n = 1, 2, \ldots N$, where $P\{X \geq n\}$ is the probability a binomial random variable with parameters N and θ takes on a value of n or larger. Figure 2 shows what this looks like for the case of $N = 4$.

Figure 2: Choosing FCFS or SIRO from Behind a Veil of Ignorance
About One's Queue Position; 4 Patients and X ~ Binomial(4, 0.25) Organs

The expected number of lives saved is the same in either case since for the FCFS case,

$$\sum_{n=1}^{N} P\{X \geq n) \quad = \quad \sum_{n=0}^{N-1} P\{X > n) \quad = \quad \sum_{n=0}^{N} P\{X > n) \quad = \quad N\,\theta.$$

A key question is whether SIRO or FCFS offers better life quality during the year spent waiting. Let $U(x)$ be the quality of life while in queue as a function of the probability of surviving beyond that first year. A Rawl's test would prefer SIRO is

$$U(\theta) \quad < \quad \sum_{n=1}^{N} \frac{1}{N} U(P\{X \geq n\}),$$

which is true for any concave function $U()$. Indeed, SIRO would be optimal since for any concave function $U()$, the maximum over the vector of survival probabilities $\boldsymbol{p} = [p_1, p_2, \ldots p_N]$

$$Max \quad Z \quad = \quad \sum_{n=1}^{N} \frac{1}{N} U(p_i)$$

subject to the constraint

$$\sum_{n=1}^{N} p_i = N\,\theta$$

occurs when $p_i = \theta$ for all i.

Numerical calculations for standard, stylized concave utility functions such as $U(x) = ln(x)$ or x^{α} suggest that the gap between FCFS and SIRO in summed individual utilities can be nontrivial.[2] Such calculations are not terribly meaningful because they use toy not empirically validated functions, and the whole notion of summing individual utility has dubious value. However, Figure 3 illustrates the reason for this numerical result in a manner that is independent of the specific utility function $U()$. It shows, as a function of the number of patients N, and for three different values of θ (the ratio of expected number of organs to patients waiting), what proportion of people in a FCFS queue have an intermediate probability of receiving a transplant ($0.05 < p_i < 0.5$). For even modestly large N, most people in a FCFS queue have either a very high probability of receiving an organ (if they are near the front of the queue) or a very low probability (if they are near the back of the queue), whereas for moderate θ values, a SIRO queue gives all individuals a p_i large enough to get fairly far up a concave utility function.

Figure 3: Proportion of Patients Whose Probability of Receiving an Organ Under FCFS is between 5% and 50% as a Function of the Number of Patients, for Three Different θ values (Ratios of Expected Number of Organs to Patients)

4. Knowledge Worker Information Request Example

Randomness in queue discipline might be useful outside the important but specialized domain of organ transplants. This section illustrates this by outlining another stylized queue, one inspired by daily life as a symbol processor in the modern knowledge-based economy.

Many information-age workers juggle multiple tasks that require assembling pieces of information from disparate sources while maintaining high productivity (avoiding down time). Sometimes batches of requests for information can be submitted in parallel, but often the process is sequential. The knowledge worker discovers a need for information and submits an information request immediately, but it is inefficient to proceed further with the current task until

[2] For $N = 20$, $\theta = 0.25$, and $U(x) = ln(1+x)$, $U(x) = x^{0.8}$, or $U(x) = 1 - e^{-x}$, a SIRO queue achieves the same summed utility with a smaller θ of only around 0.2. For $U(x) = x^{\alpha}$ with a smaller α, the required θ for a SIRO queue is even smaller, $\theta = 0.103$ for $\alpha = 0.5$ and $\theta = 0.02$ for $\alpha = 0.2$.

that bit of information is obtained. If the (animate or inanimate) information server has no queue, then the information is obtained without delay so the knowledge worker continues working on the current task. However, if there is a queueing delay of more than a few moments, the knowledge worker does not usually sit in queue; that idle time from the knowledge worker's perspective is unacceptable in a highly competitive work environment. Rather, he or she puts aside the current task and works on a different task until the requested information comes through.

An example familiar to many readers is looking up a piece of information from the literature while working on a research project. Sometimes the library has a full text on-line subscription to the relevant journal, and the article of interest can be obtained without delay. Other times it must be ordered via interlibrary loan and so will not be available until the next day. In that case, the delay sometimes prompts one to put aside the current task (e.g., reviewing the literature) and switch to a different task, whether it is related (perhaps writing computer code for the same project) or not (grading homework or making plans to attend an upcoming conference).

Other examples of task-critical information sought on a just in time basis might include the result of a database query, a manager's approval to proceed with a particular course of action, or a discrete piece of tacit knowledge resident in a colleague's head.

Sometimes shifting from one task (e.g., reviewing the literature) to another (grading or travel planning) can impose a productivity cost. The inefficiency can come from having to close down one set of computer files and open another set or from human cognitive processing limitations – less formally, needing time to change gears and get one's head around the new task. This amounts to a fixed cost paid whenever the waiting time is more than negligible.

Variability in service time can still be undesirable, as in the classic analysis that favors FCFS, so cost might be a convex function of waiting time everywhere beyond this initial step up around 0. However, that initial step function can be enough to make it optimal to introduce randomness into the queueing discipline. We demonstrate this by example. Since it is a proof by example that the converse is false, i.e., that FCFS is not always optimal, we examine an atypical queue whose unusual properties make the analysis particularly convenient.

Consider a D/D/1 queue for which the service and inter-arrival times are the same fixed constant, say 1 minute. For convenience, suppose that the arrival and service completion time epochs are coordinated, so the number of customers in queue never changes, $N(t) = N_0$ for all time t. With a FCFS queueing discipline the waiting time in queue is always exactly N_0 minutes with zero variability in waiting time.

Now consider introducing randomness to the queue discipline in the following way. Flip a coin each time a customer arrives (which also happens to be when the server becomes free). If it comes up heads (with probability p) the customer joins the queue normally; if it comes up tails, the customer enters the server immediately, skipping ahead of everyone waiting in queue.

Now with probability $(1 - p)$ the waiting time in queue is zero and the fixed cost of swapping tasks is averted. With probability p, one tossed a head and went to the back of the queue. In that case, the waiting time in queue takes on its standard value of N_0 if and only if all of the next N_0 arrivals tosses a head, an event which occurs with probability p^{N_0}. Otherwise, the randomization increases waiting time in queue. More generally, if one enters the queue, the time spent in queue equals the number of tosses it takes to observe N_0 heads, so it has a negative binomial distribution with parameters N_0 and p. So the overall, unconditional waiting time distribution with randomization is

$$p_W(w) \quad = \quad \begin{cases} 1-p & \textit{for } w=0 \\ p\begin{pmatrix} w-1 \\ N_0-1 \end{pmatrix} p^{N_0}\,(1-p)^{w-N_0} & \textit{for } w = N_0, N_0+1, \ldots \end{cases}$$

The variance of this distribution is

$$\sigma_W^2 \quad = \quad N_0\left(\frac{N_0+1}{p}-1\right) \quad - \quad N_0{}^2.$$

Suppose the overall cost of waiting included a fixed cost c_0 if there is any wait plus a term that is proportional to the variance, with proportionality constant c. Slips and skips are not incorporated explicitly because, as with the transplant example, the queue would usually be "blind" since the knowledge workers would not literally stand in a physical queue where they could observe other people skipping ahead of them. Differentiating this cost function

$$Cost \quad = \quad c_0\, p \quad + \quad c\left(N_0\left(\frac{N_0+1}{p}-1\right) \quad - \quad N_0{}^2\right)$$

with respect to p yields first order conditions that are quadratic in p. They imply a minimum cost when the probability of entering the queue normally, p, is

$$p^* \quad = \quad \begin{cases} \sqrt{\dfrac{N_0\,(N_0+1)\,c}{c_0}} & \textit{provided } \sqrt{\dfrac{N_0\,(N_0+1)\,c}{c_0}} \quad < \quad 1 \\ 1 & \textit{otherwise.} \end{cases}.$$

So if the penalty on increasing the variance in waiting time (c) and the queue length (N_0) are small relative to the fixed cost of having to wait at all (c_0), then a pure FCFS queue discipline is inferior to one that introduces some randomness.

Reflection on typical office etiquette suggests that we may already depart from FCFS queue discipline in ways that mitigate these fixed costs. There can be multiple justifications for answering email in reverse chronological order, but one is so that recent requests can be handled expeditiously. Likewise, the general practice of leaving email turned on throughout the work day can be inconsistent with FCFS. Most office workers have a queue of tasks lined up when they arrive in the morning. A pure FCFS strategy would suggest not reading incoming messages until that queue is cleared. Few people do that. Some who do occasionally refuse to turn on email in order to complete some overdue task can be seen as making an atypical special effort to abide by FCFS for a task that has already suffered a large number of skips.

That is not to say that human servers in office environments literally flip coins to decide whether to put an incoming job at the front or the back of the queue. Rather, the point of the previous paragraph is just that we do not always behave as if FCFS is sacrosanct and we do sometimes behave as if there is a fixed cost associated with any positive waiting time in queue.

Literally introducing randomization in queue discipline is entirely possible in some contexts, however, such as responding to computer help desk inquiries. Many help desks use software to manage service requests. It would be easy for such software to implement a modified FCFS plus randomization scheme like that discussed in the example above.

A similar application could be within email systems. Most already allow users to sort the inbox by ascending or descending date. They could add, as an option, sorting oldest to newest but with some random fraction of new incoming emails jumping to the front of the queue.

A mathematically parallel but substantively different application would be processing a request for government service, such as access to a civil court for contract dispute resolution; processing of a passport, zoning variance, or building permit application; or emergency housing placement. Contract disputes can stall construction and other projects. If the average delay with FCFS is long enough that contractors have to lay-off or redeploy workers, move equipment, and/or take actions to secure a worksite so that it does not become an attractive nuisance, then a partially random queue discipline that offers a 50/50 chance of no delay or twice the average delay might be preferable. Likewise, if a family suddenly becomes homeless, they might prefer a 50/50 chance of getting emergency housing immediately or in two months over a certainty of having to wait a month if a month is long enough to force the family to find another more or less stable arrangements or to lose their job and household possessions if they cannot.

5. Discussion

In queues where jobs are associated with people, as opposed to packets in a telecommunications network, the almost exclusive focus has been on "First Come, First Served" (FCFS) discipline, at least within any given priority class. (High priority individuals might receive expedited service, but for individuals within the same priority class, FCFS is the norm.)

This paper makes the contrarian suggestion that sometimes there may be grounds for introducing randomness into the service discipline, even in contexts such as organ transplantation where fairness is of great importance. The premise underpinning this contrarian suggestion is Larson's (1987) observation that utility as a function of weighting time is often nonlinear (and multi-attribute).

When an objective function is a nonlinear function of its underlying fundamental measure (such as waiting time), it is common to assume a convex function, often justified by some form of diminishing returns argument. Quadratic functions are particularly common (e.g., in minimizing squared error).

However, in certain applications the costs of waiting might be concave in waiting time at least over some ranges. In the case of organ transplantation, once the waiting time is so long that the patient will die before being served, additional waiting time ceases to be relevant. Furthermore, for organ transplant, perceived waiting cost may depend not only on time spent in queue, but also on some integral of time weighted by the probability perceived at that time of ever being served. A year spent waiting in a transplant queue might feel differently for someone who knows an organ will be available in a year as opposed to someone who knows no organ will become available and, hence, that they will die at the end of the year.

In other contexts there may be fixed costs associated with suffering any non-negligible wait, potentially yielding concave-convex cost functions. An example is knowledge workers seeking information that is necessary in order to proceed with a given task rather than switching to some different task that entails some "set up charge".

The traditional rule with respect to queue discipline might be summarized as, "bits and bytes may suffer injustice, but when the queues involve people the discipline better be FCFS within a given priority level." This note argues that a better rule of thumb might be, "When people are literally standing on line, FCFS is the default, but for spatially distributed or virtual queues comprised of people and/or their processing requests, strategies that blend randomness

and FCFS might have a role in creating the hope of being lucky, whether that is the existential hope of getting an organ transplant or the everyday hope of getting a quick answer to an information request.

Acknowledgements
Ed Kaplan and Cindy Bryce made useful comments on an earlier draft. This work was funded in part by the Qatar Foundation.

References

Alagoz, Oguzhan, Lisa M. Maillart, Andrew J. Schaefer, and Mark S. Roberts (2004) The Optimal Timing of Living-Donor Liver Transplantation. *Management Science*, 50(10):1420-1430.

Alagoz, Oguzhan, Lisa M. Maillart, Andrew J. Schaefer, and Mark S. Roberts (2007) Determining the Acceptance of Cadaveric Livers Using an Implicit Model of the Waiting List. *Operations Research*, 55(1):42-36.

Paul J. Allison, Christophe Guichard, Karen Fung, and Laurent Gilain (2003) Dispositional Optimism Predicts Survival Status 1 Year After Diagnosis in Head and Neck Cancer Patients. *J Clin Oncol 21:543-548.*

Arver, Brita, Aina Haegermark, Ulla Platten, Annika Lindblom, and Yvonne Brandberg (2004) Evaluation of psychosocial effects of pre-symptomatic testing for breast/ovarian and colon cancer pre-disposing genes: a 12-month follow-up. *Familial Cancer* 3: 109–116.

Avi-Itzhak, B. and H. Levy (2004). On measuring fairness in queues. *Advances of Applied probability*, 36(3):919-936.

Avi-Itzhak, B., H. Levy, and D. Raz (2004). *Quantifying Fairness in Queueing Systems: Principles and Applications*. RUTCOR Research Report RRR 26-2004, July, Rutgers University, Piscataay, NJ.

Andorno, J *J Med Ethics.*2004; 30: 435-439.

Arrow, Kenneth J. (1951) *Social Choice and Individual Values*, Wiley, New York.

Chew, Soo Hong & Ho, Joanna L, 1994. "Hope: An Empirical Study of Attitude toward the Timing of Uncertainty Resolution," *Journal of Risk and Uncertainty*, Vol. 8(3), 267-88.

Gordon, Ethel-Sherry (1987) New Problems in Queues: Social Injustice and Server Production Management. Doctoral Dissertation in Operations Research, MIT, Cambridge, MA.

Larson, Richard C. (1987) Perspectives on Queues: Social Justice and the Psychology of Queueing. *Operations Research*, 35(6):895-905.

Mondloch, Michael V. Donald C. Cole, and John W. Frank (2001) Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. *CMAJ*; 165(2):174-179.

Oguzhan Alagoz, Lisa M. Maillart, Andrew J. Schaefer, and Mark S. Roberts (2007) Determining the Acceptance of Cadaveric Livers Using an Implicit Model of the Waiting List *Operations Research*, 55: 24-36

Oguzhan Alagoz, Lisa M. Maillart, Andrew J. Schaefer, and Mark S. Roberts (2004) The Optimal Timing of Living-Donor Liver Transplantation. *Management Science*, 50: 1420-1430

Prelec, D. (2000). Compound invariant weighting functions in prospect theory. In D. Kahneman & A. Tverky (Eds.), *Choices, Values, and Frames.* New York: Cambridge University Press.

Rawls, John (1971) *A Theory of Justice*, The Belknap Press of Harvard University Press, Cambridge, MA.

Tversky, A. & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*. 5, 297-323.

Wermer, M.J.H., I.C. van der Schaaf, P. Van Nunen, P.M.M. Bossuyt, PhD, C.S. Anderson, G.J.E. Rinkel (2005) Psychosocial Impact of Screening for Intracranial Aneurysms in Relatives With Familial Subarachnoid Hemorrhage. *Stroke*. 36: 836-840.

Xuanming Su and Stefanos Zenios (2004) Patient Choice in Kidney Allocation: The Role of the Queueing Discipline. *MSOM*, 6: 280-301.

Xuanming Su and Stefanos A. Zenios (2006) Recipient Choice Can Address the Efficiency-Equity Trade-off in Kidney Transplantation: A Mechanism Design Model. *Management Science*, 52: 1647-1660.

Zenios, Stefanos A. (2002) Optimal Control of a Paired-Kidney Exchange Program *Management Science*. 48: 328-342

Figure 1: FCFS and SIRO Compared When One Organ Will Become Available in One Year and There are Two Patients Who Will Only Live One Year Without a Transplant

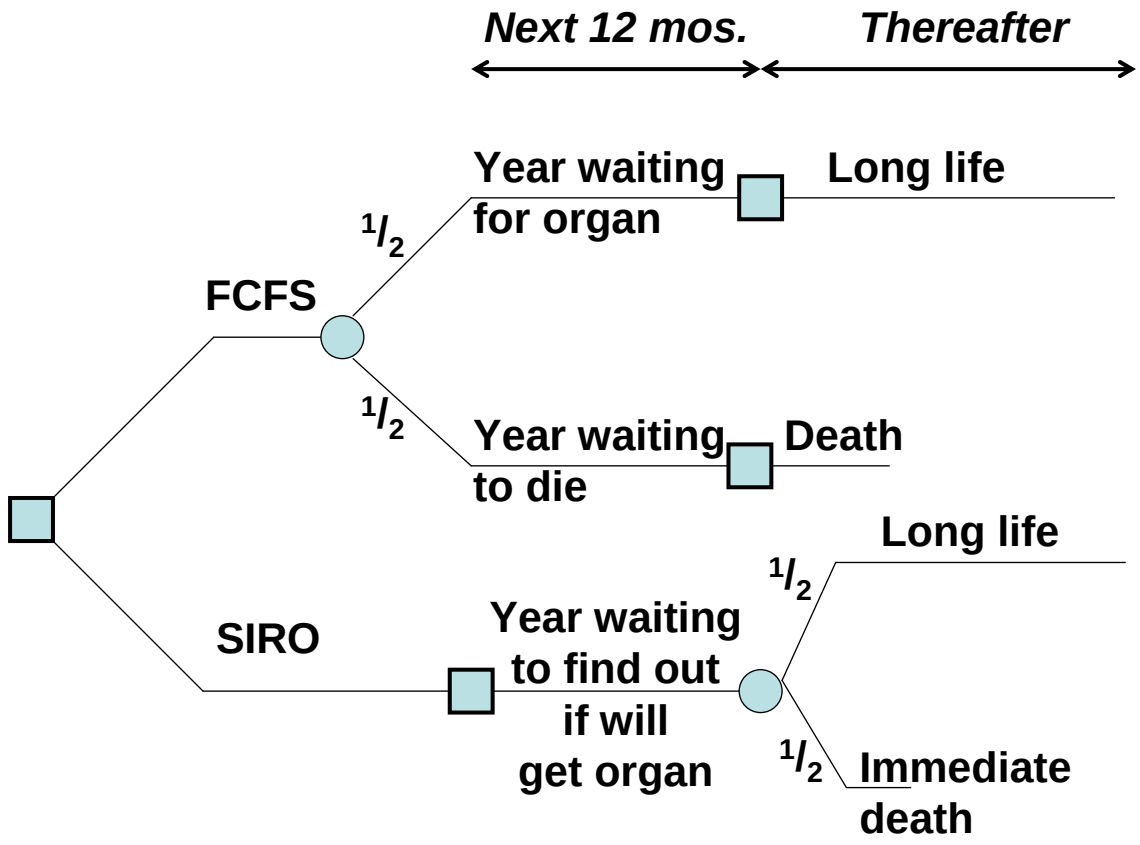

Figure 2: Choosing FCFS or SIRO from Behind a Veil of Ignorance
About One's Queue Position; 4 Patients and X ~ Binomial(4, 0.25) Organs

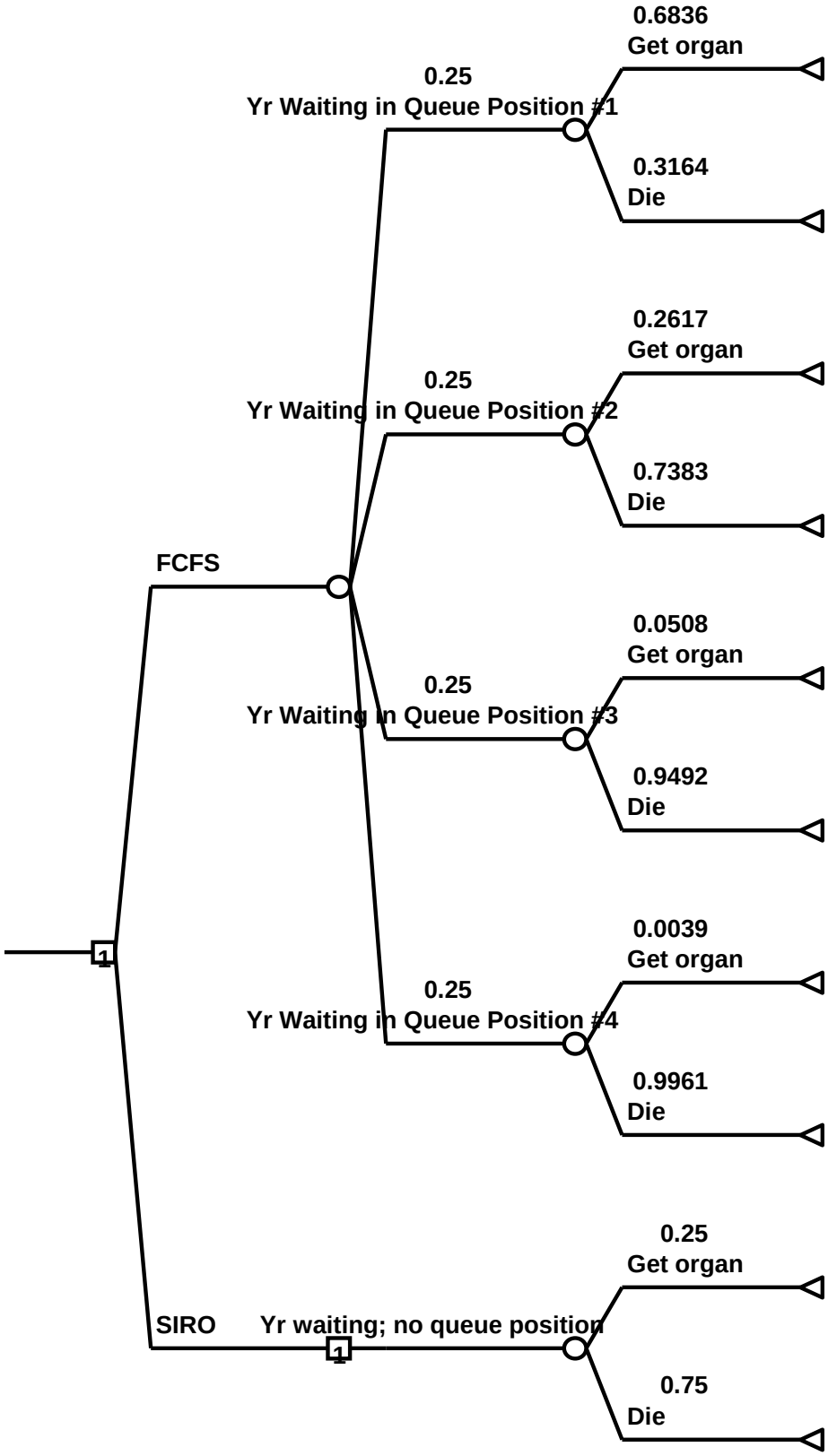
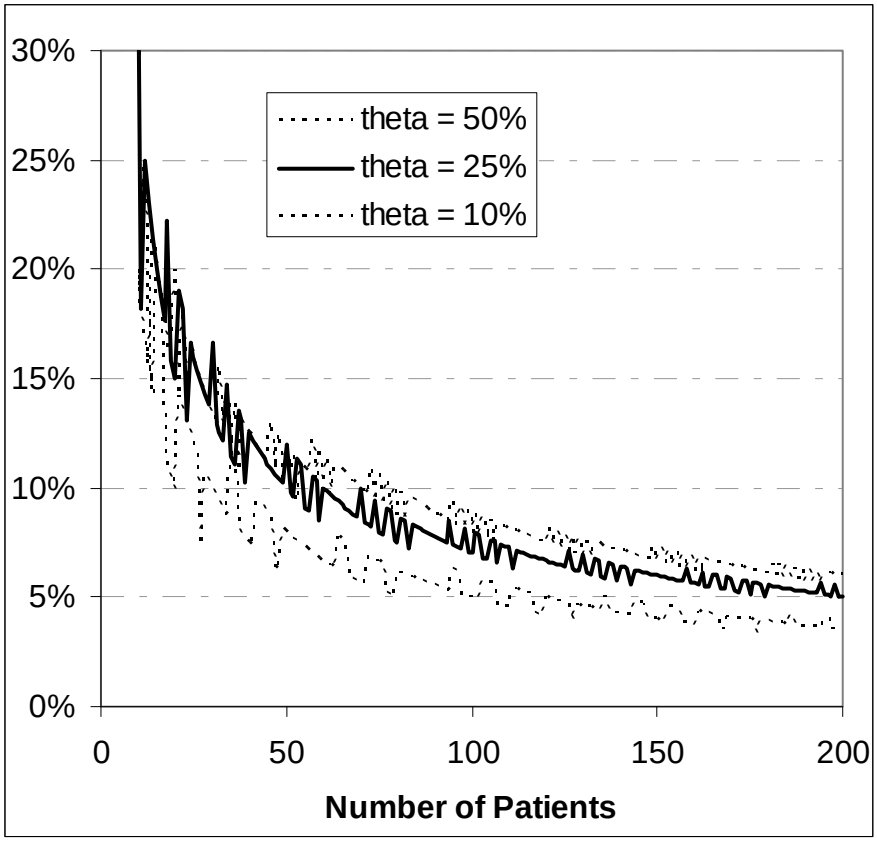
Figure 3: Proportion of Patients Whose Probability of Receiving an Organ Under FCFS is between 5% and 50% as a Function of the Number of Patients, for Three Different *p* values (ratios of organs to patients)