Repository Entry
Embedded EthiCS @ Harvard Teaching Lab

Overview

Course:	CS 222: Algorithms at the End of the Wire
Course Level:	Graduate
Course Description:	"Covers topics related to algorithms for big data, especially related to networks and database systems. Themes include sketch-based data structures, compression, graph and link information, and information theory. Requires a major final research-based project."[1]
Module Topic:	Fair Queuing
Module Author:	Samuel Dishaw
Semesters Taught:	Fall 2020
Tags:	Fair queuing [CS], optimization [CS], fairness [phil], aggregation [phil]
Module Overview:	The module discusses issues of fairness as they arise in the context of allocating resources through queues. Two kinds of queues are distinguished: queues for non-exhaustible goods (e.g. server throughput) and queues for exhaustible goods (e.g. ICU beds). The module introduces the notion of pairwise comparison between outcomes as a way of determining whether a distribution of burdens or of benefits is fair.
Connection to Course Material:	A week prior to the module students were introduced to algorithms for scheduling policies, and reflected on ways to optimize (i.e. minimize) average waiting time in a queue by using predictions about job length for jobs in the queue. The module thus discusses an ethical desideratum within a design problem that students were already familiar with.	One reason for choosing this topic is that there exists a body of literature in computer science dedicated to the topic of fairness in queues, some of which briefly considers philosophical work on fairness. The topic is thus a natural place to explore philosophical considerations of fairness currently being considered in that literature. Another reason is that the topic is of broad ethical significance, given the many different uses of queues as a resource-allocation mechanism.

Goals

Module Goals:	1. Distinguish two kinds of queues and what issues each kind of queue raises with respect to fairness.

[1] https://canvas.harvard.edu/courses/74186/assignments/syllabus

	2. Introduce a framework for thinking about fairness, i.e. pairwise comparison of outcomes. 3. Apply this framework to the question of fair queues for non-exhaustible goods. 4. Compare the merits of queues and lotteries in the context of the allocation of exhaustible goods	
Key Philosophical Questions:	1. What is a queue? 2. What are the harms of waiting? 3. What is fairness? 4. Should we always maximize the good? 5. What is the fairest way of distributing the burden of waiting? 6. Are lotteries fairer than queues?	Questions 2 and 5 pertain in particular to queues for non-exhaustible goods. Question 6 was discussed in relation to queues for exhaustible goods.

Materials		
Key Philosophical Concepts:	● Exhaustible vs. non-exhaustible goods ● Fairness vs. moral aggregation ● Pairwise Comparison ● Lotteries as allocation mechanisms ● Fair chances	The distinction between queues for exhaustible and non-exhaustible goods is important because each raises different issues with respect to fairness. In the case of queues for exhaustible goods (e.g. ICU beds), the primary issue is securing fair access, or fair opportunities of access, to some good that not all can have. In the case of queues for non-exhaustible goods (e.g. boarding a plane), the primary issue is distributing the burden of waiting in a way that is fair. The distinction between fairness and doing the most good, as well as the method of pairwise comparison, is illustrated by way of a discussion of T.M. Scanlon's argument against moral aggregation in *What We Owe to Each Other*. The notion of 'Pairwise comparison' (from Nagel 1979: chapter 8) is another name for what Scanlon calls the 'individualist restriction' on moral justification.
Assigned Readings:	● Avi-Itzhak, B. et. al. (2008). Quantifying Fairness in Queuing Systems: Principles, Approaches and Applicability. *Probability in the engineering and informational sciences* 22 (4), 495-517.	This paper provides a useful overview of the fair queuing literature in computer science, and does an especially good job motivating some different but incompatible principles of fair

		queuing (namely: first-in-first-out and shortest-job-first).

Implementation

Class Agenda:	1. Distinguish two kinds of queues 2. Summarize the problem of fair queuing in the computer science literature 3. Introduce a philosophical concept on fairness by way of Scanlon's World Cup example and the notion of pairwise comparison 4. Apply the philosophical tool of pairwise comparison to a specific queuing problem (call centers with larger and smaller customer job times) 5. Discuss the relative merits of lotteries and queues for exhaustible goods, using two examples from bioethics (vaccine distribution and ICU beds)	Although we covered all of these items on the agenda, it would have been nice to spend more time on the fifth item, since it raises ethical questions at a different and more fundamental level (whether queues are a fair mechanism at all). The discussion of the Scanlon example ate up a surprising amount of time.
Sample Class Activity:	In the module's main activity, students were presented with a specific queuing problem. A single queue in a call-center has nine short jobs (five minutes to process) and one long job (thirty minutes to process). As it stands, the longer job is at the head of the queue. How much time this longer job has already waited was a variable in this problem. Students were then asked whether it would be fair to re-order the queue—in particular whether it mattered how much time the longer job had already waited—and encouraged to use the pairwise comparison metric to answer this question. After students worked in small groups, they reconvened to discuss their answers with the rest of the class.	Although the students took to using the notion of pairwise comparison, they all used it in a way that collapsed it into the shortest-job-first principle. This is because students used pairwise comparison of *marginal* burdens (i.e. how much *longer* one would have to wait, relative to the status quo). Using that notion means that even if the long job has already waited an hour to be served, the next shortest job in line should still be bumped ahead of them, because that way the longer job will only be made to wait five *more* minutes, whereas the second will otherwise have to wait thirty more minutes. But this isn't the only way to apply the notion of pairwise comparison to the case. We can instead compare the total waiting time of each job. This way of using the metric results in effect in a maximin principle, where the fairest distribution of waiting time is the one on which the waiting time of whoever is made to wait the most is the shortest. One reason why this option was perhaps not as salient was the formulation of the prompt, which

		asked students what way of *re-ordering* the queue would be fair. This prompt invites thinking in terms of marginal burdens. A better way of framing the problem might be to ask students directly what queuing algorithm should be operative if everyone in this particular queue is to be treated fairly.
Module Assignment:	Discussion board prompt: The authors of the paper discuss two scheduling policies: First-In-First-Out and Shortest-Job-First. Which do you think is more fair and why?	This question was posted on the online course website three days ahead of the module. Most students answered the prompt before the class but discussion continued on the course website after class as well.
Lessons Learned:	Scanlon's anti-aggregation argument is a good pedagogical tool to get students thinking and talking about constraints on maximizing the good and more generally about the difference between doing the most good and fairness.	Partly because the World Cup example generates so much student engagement, the discussion can also go in less fruitful directions (e.g. a debate about how much pleasure people derive from watching sports). In the original text, Scanlon asks two questions: whether we should rescue Smith or wait until the match is over, and whether the right thing to do depends on how many people are watching. To get the most out of the example, it may be best to highlight the second question and make it the focal point of discussion.