Pricing Equilibrium - Inherently Trustworthy, Approximately Fair

Whitepaper for Workshop on Trustworthy Algorithmic Decision-Making

Last spring my spouse and I filled in an online form with our preferences for our son's kindergarten placement. We ranked Ms. A's class first, having heard she was an exceptional teacher, and Ms. B's class second. This summer we found out our son was placed in Ms. B's class. While our son loves his new class and so do we, we have no way of verifying that the placement decision process was fair. Our only choice is to trust our local municipal system, and its algorithm of choice for matching children to schools.

Deciding how to allocate scarce resources – e.g., seats in good public schools, affordable housing, over-demanded higher-education courses, shared scientific resources, etc. – constitutes a challenging algorithmic task. It requires soliciting preferences from those competing for the resources, possibly combining data to predict who will utilize the resources well, and computing an efficient allocation based on the reported preferences and data. The algorithm for computing an allocation can be quite complicated and opaque, and may resist provable performance guarantees.[1] How can we make such an algorithm trustworthy and fair?

In this note I wish to discuss one case study from the field of *market design* [1], a vibrant research area in economics and computer science (the latter increasingly gaining in relevance as markets become more reliant on algorithms). The simple setting described below raises theoretical questions of how to define fairness and how to verify it. I shall present preliminary answers to these questions from the market design literature [2] and from my own recent work [3]. A meta-goal is to indicate how recent research on computational market design can inform, as well as benefit from, the emerging debate on trustworthy algorithmic decision making.

Case study: Allocation via Pricing Equilibrium

Consider the following stylized setting: There are different items, possibly with multiple copies of each (e.g., seats in different higher-education courses). Participants (students) have preferences over different combinations of items (different schedules). The crucial decision here is *who gets what*, and the question is which algorithmic procedure to use to reach this decision.

One possibility is the following naïve algorithm: (1) Assign a random order to the students; (2) Give each student in turn their favorite schedule among the ones still available. It is not hard to see that this algorithm can be quite unfair – if there are two "star" professors, one student (e.g., the first one) can get both, while another student (e.g., the last one) gets neither. Moreover, this algorithm is non-verifiable and so not trustworthy,[2] in the sense that a student who wishes to verify her own allocation must possess knowledge of the allocations of all students before her in order to reconstruct the naïve algorithm's decision process.

Instead, consider an alternative algorithm that publishes a "price" for each course [2]. The students have budgets of fake "money", and each student "buys" her favorite schedule that fits within her budget. The algorithm sets the prices such that at the end, all seats in courses for which there is more demand than supply are filled, and so there is no waste. Unlike the naïve algorithm, the pricing algorithm is inherently verifiable, since the allocation decision is made by the student herself based on public knowledge (the published prices). The algorithm also "feels" fair, as long as students have either the same budgets or different budgets based on a relevant difference between them (such as year of study). This is because all students face the same prices, whereas in the naïve algorithm students face diminishing sets of available schedules.

[1] Such an algorithm is often referred to as a *heuristic*.

[2] The question of defining the *trustworthiness* of an algorithm is challenging and we do not fully address it here; however, we assume that verifiability by those affected by its decisions is a major component of trustworthiness, in the spirit of "*Not only must Justice be done; it must also be seen to be done*" [4]. See also the role of verification in [5].

There are still a few gaps to fill to complete the picture: (1) How can prices guarantee that all students are simultaneously able get their most preferred schedule within their budget, and all seats in over-demanded courses are filled? (2) Can the intuitive fairness of the algorithm be formalized?

The answer to the first question hinges on the notion of a *pricing equilibrium* from microeconomics. A pricing equilibrium is precisely a set of prices that coordinate the choices of buyers such that the total set of items they demand is equal to the set of available items (the supply) on the market. Notice that the prices need to be set just right – if they are too high, classes will not fill up, and if too low, classes will be "sold out" and some students will not get the favorite schedule they can afford. Thus, the question of a fair and trustworthy algorithm for deciding how to allocate courses boils down to the existence of a pricing equilibrium.

At first glance, this seems like a non-starter: if there is a single seat in a course that two students with the same budget are interested in, the price is necessarily too high (above the joint budget, leading to waste) or too low (within budget, leading to over-demand). But if the budgets are slightly perturbed, the student with the slightly higher budget will have precedent over the other student and will get the course. More generally, slightly perturbing the budgets can provably ensure existence of a pricing equilibrium in several cases of interest [3], and it is an open question how far this phenomenon generalizes. Of course, perturbed budgets introduce certain unfairness, but only to the extent that seems absolutely necessary given the limitations of the setting. We now turn to how this "approximate fairness" intuition can be made more precise.

In [2] and [3], several fairness properties are proved for perturbed pricing equilibria. Here we hint at one such formalization through the following example. Consider a freshman with budget X and a senior with budget 2X (i.e., the freshman has 1/3 of the total budget while the senior has 2/3). If we were allocating (say) dorm space between them, a fair way to do this is to have the freshman divide the space into three thirds, and have the senior choose two of the three thirds.[3] An allocation is *fair* if it each student likes her part at least as much as what she can guarantee for herself through the divide-and-choose procedure. While we cannot apply this fairness definition directly to course allocation (since the number of courses may not be divisible by three), we can apply an appropriately modified definition in the same spirit. A pricing equilibrium can be shown to be fair according to this definition.

We conclude our discussion of this case study by mentioning another nice property of the price-based algorithm for deciding how to allocate. When the market is large enough, participants tend to see themselves as "price takers" (as if responding to exogenous prices), and do not worry about reporting their preferences strategically to the algorithm so as to influence prices. This levels the playing field for sophisticated vs. unsophisticated participants, and thus also plays a role in the algorithm's trustworthiness and fairness.

Conclusion and future research

Decision-making by humans is rife with biases and chance [e.g., 7]. In this sense, replacing a human decision-maker with an algorithm has huge potential for *improving* trustworthiness and fairness, by enabling the application of formal and provable standards. The main argument in this note is that markets and prices are a useful tool in the algorithm designer's toolkit for defining and achieving these standards, alongside "the usual suspects" like machine learning and statistics.

How far can we push the pricing approach? I.e., for which additional decision-making domains can we use pricing equilibria, their variants [e.g., 8] and other forms of pricing [see, e.g., 9] to achieve trustworthiness and fairness? An additional direction for future research is whether pricing equilibria are "obviously" fair in the sense that bounded-rationality participants can immediately grasp [cf., obvious strategy-proofness in 10], or is there a need to strengthen the definitions of fairness to achieve this.

[3] This is a generalization of a method going back to Abraham and Lot [6].

Bibliography

[1] The Royal Swedish Academy of Sciences. Stable allocations and the practice of market design. Scientific background on the Nobel Prize in economics, 2012.

[2] Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. Journal of Political Economy 119 no. 6, 1061-1103, 2011.

[3] Moshe Babaioff, Noam Nisan, and Inbal Talgam-Cohen. Competitive equilibria with indivisible goods and generic budgets. In submission, 2017.

[4] R v Sussex Justices, Ex parte McCarthy ([1924] 1 KB 256, [1923] All ER Rep 233).

[5] Tim Roughgarden, and Inbal Talgam-Cohen. Why Prices Need Algorithms. Proceedings of the 16th ACM Conference on Economics and Computation (EC), 2015.

[6] Benjamin Plaut, and Tim Roughgarden. Almost envy-freeness with general valuations. To appear in the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2018.

[7] Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. Extraneous factors in judicial decisions. PNAS 108 no. 17, 6889–6892, 2011.

[8] Shahar Dobzinski, Michal Feldman, Inbal Talgam-Cohen, and Omri Weinstein. Welfare and Revenue Guarantees for Competitive Bundling Equilibrium. Proceedings of the 11th Conference on Web and Internet Economics (WINE), 2015.

[9] Ilan Reuven Cohen, Alon Eden, Amos Fiat, and Lukasz Jez. Pricing Online Decisions: Beyond Auctions. Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 73-91, 2015.

[10] Shengwu Li. Obviously Strategy-Proof Mechanisms. To appear in American Economic Review, 2017.