from *SIAM News*, Volume 36, Number 6, July/August 2003

How Much Can Matching Theory Improve the Lot of Medical Residents?

By Sara Robinson

Although developed by mathematicians, the theory of matching has thrived in economics departments, which are, after all, closer to the theory's applications. It's not surprising, then, that my informal investigation of ways to add salary competition to the matching algorithm for medical residents (*SIAM News*, April 2003) turned up ideas that had been rigorously addressed by economists nearly two decades earlier.

For readers just tuning in, the earlier article described a lawsuit against the program that, every March, matches graduating medical students to residency programs. The matching procedure, which uses an algorithm designed by an economist in the 1950s, has many desirable features: It ensures that residents and programs are appropriately paired according to their preferences, and it eliminates highly undesirable competition of a different type (more on this below). Yet a group of medical residents has filed a class action antitrust suit against the organizers of the match, alleging that by using an algorithm to match students to programs and bypassing a negotiation process, the match keeps residents' salaries miserably low and working hours long.

Residents indeed work long hours for low salaries (averaging below $40,000). But is the algorithm to blame?

In the earlier article, I began to explore this question, focusing on whether some adjustment to the algorithm might address low residency wages. The article mentions a few off-the-cuff attempts to introduce wage competition into the algorithm, which was formally devised by mathematicians David Gale of the University of California, Berkeley, and Lloyd Shapley of the University of California, Los Angeles.

In this issue's follow-up, I delve into the economics research on matching with price competition, starting with a new paper by Jeremy Bulow and Jonathan Levin. They have come up with a general model for markets like the residency match, and their paper shows that such markets are likely to have salaries below those of markets in which salaries are negotiated, giving credence to the suing residents' claims. I also briefly summarize the combined results of two earlier papers, one by Vincent Crawford and the late Elsie Knoer, and another by Crawford and Alexander Kelso, which devise a method for matching with price competition and prove theorems similar to those of Gale and Shapley for a matching algorithm that takes salaries into account. Perhaps this mechanism, applied to the residency match, would improve working conditions for medical students.

Finally, a new paper, "Matching with Contracts," by Paul Milgrom of Stanford University presents a model that pulls together the approaches of Gale–Shapley and Crawford–Kelso–Knoer, along with auction models, into a single, unified framework. Although beyond the scope of this article, it's well worth a read for those who want to learn more about this topic.

First, a quick review of the history of the match and the algorithm devised by Gale and Shapley.

Gale–Shapley Revisited

In a delightful 1984 article ("The Evolution of the Labor Market for Interns and Residents: A Case Study in Game Theory"), Alvin Roth pointed out that by the 1950s, when the matching algorithm was devised, the residency market had become highly competitive. Many hospitals were competing for relatively few candidates; wages, however, were not going up. Rather, the competition was manifesting itself in the timing of offers; each hospital would try to beat out competitors by giving take-it-or-leave-it offers to residents early in their medical school training. When attempts to set limits on the timing caused other problems, the hospitals eventually banded together and introduced the current centralized system. After some false starts, the hospitals moved toward a scheme that seemed to work well for everyone.

In 1962, mathematicians Gale and Shapley wrote down a stability property that seemed essential for a matching and came up with an algorithm that achieved it. In the 1980s, Roth, a professor of economics at Harvard University, demonstrated that the Gale–Shapley algorithm was essentially the same as the residency matching algorithm.

In their paper, "College Admissions and the Stability of Marriage," Gale and Shapley described their matching goals in terms of a marriage mart: Given n boys and n girls, each armed with a strict preference list of all the participants of the opposite sex, their algorithm can always find a pairing of boys to girls that is "stable," in the sense that no boy and girl would prefer to be with each other than with the people the algorithm paired them with.

In addition to showing that their process always terminates in a stable matching, Gale and Shapley demonstrated that the version of the algorithm that has the boys proposing to the girls terminates in a matching that is "male-optimal," in the sense that the boys are paired with the best girl they could possibly get in any stable matching, and "female-pessimal," meaning that the girls get the worst possible boy in any stable matching. (The situation is reversed, of course, if the girls do the proposing.) The algorithm also has the property that the people receiving the proposals can do better by lying about their preferences, although those doing the proposing cannot. The National Resident Matching Program's version originally had the hospitals proposing to the residents; since a revamping by Roth in the late 1990s, the residents propose to the hospitals.

This algorithm behaves optimally in the sense of providing matchings that are stable, but it does not take wages into account. In practice, the salaries offered by the hospitals are set before the match takes place, and NRMP explicitly forbids a hospital and resident to agree in advance on a salary in the event that they are matched with each other. Thus, the only way hospitals can compete with pricing is by assigning higher prices to their slots without knowing who will take them.

This inability to personalize salary offers can lead to price stagnation, according to the new paper by Bulow and Levin, both of Stanford University.

Needed for School-Choice Fix: Honest Students

Economists Atila Abdulkadiroglu of Columbia University and Tayfun Sonmez, visiting Columbia from KOC University in Istanbul, have drafted a paper applying the ideas of matching theory to a practical problem plaguing the school systems in many major cities. Many districts now offer "school choice," giving high school students (or their parents) the opportunity to ask for slots in schools outside their immediate neighborhoods or even outside their local school districts. Because it's not possible to assign all students to the schools of their choice, the local education departments have to design mechanisms for assigning students to schools. Many have designed their mechanisms quite unwisely.

In their paper, Abdulkadiroglu and Sonmez cite several examples of bad mechanisms, including the following one, used by the city of Boston:

Each student submits a preference ranking of the schools, and the schools submit rankings of students. The schools determine their preferences by first dividing students into four priority classes and then ordering them within the classes based on a lottery.

In round 1 of the matching, only the students who have listed a given school as their first choice are considered. They are assigned slots in order of priority, until either no slots or no students who have ranked that school first are left. In round 2, only the second choices of the students are considered, and so on.

The problem with this approach is the strong incentive it gives students not to express their real preferences. If a student's real first-choice school is likely to be the first choice of lots of other students as well, it's in the student's interest to list another, less competitive school first. This makes the match obtained unstable, although the procedure does produce stable matches when the students give their real preferences.

Abdulkadiroglu and Sonmez offer two possible fixes to the school-choice problems, one of which is based on the Gale–Shapley algorithm. If school districts used Gale–Shapley to assign students to schools with the students doing the proposing, they would be guaranteed stability of matching and the students would have no incentive to lie about their preferences.

So far, Abdulkadiroglu says, the school districts don't seem to be interested in obtaining his help.

Impersonal Offers and Their Consequences

In their paper, "Matching and Price Competition," Bulow and Levin describe a model that seems to be similar to the residency match and show that it leads to lower salaries than would be found in a corresponding idealized market with negotiated wages.

In their model, N residents compete for single slots at N hospitals. The hospitals fix the salaries ahead of time. Hospital n's *surplus*, or benefit, from hiring resident m is $v(n,m) = \delta_n \cdot m$, where $\delta_N \geq \ldots \geq \delta_1 \geq 0$. Thus, all the hospitals rank the students in order, from 1 (least preferred) to N (most preferred). If hospital n hires a resident at salary p, the hospital's net utility, or gain, is $v(n,m) - p$ and m's utility is p (so that from the students' perspective, salaries encompass all the desirable aspects of a given hospital). A hospital or resident that doesn't find a match has a utility of 0. This information is available to all the hospitals.

The match is then quite simple: The hospitals make simultaneous salary offers; the one bidding the highest gets the best student, and so on. If several hospitals offer the same salary, the one that values talent the most (has the highest value of δ) gets its preferred student. Notice that a hospital pays the salary it offered regardless of the resident it gets in the matching.

In a *pricing equilibrium* for this model, each hospital chooses its offer to maximize its expected surplus, taking into account the strategies of the other hospitals. Bulow and Levin find a unique pricing equilibrium and then compare the equilibrium salaries with those that would be found in an idealized *competitive equilibrium* model, where prices are negotiated rather than set at the beginning. They show that the salaries in their model are both lower overall and closer to one another than negotiated salaries would be.

The element of their model (and the residency market) that is mainly responsible for this effect, Levin says, is that the hospitals cannot tailor their offers to specific residents. A hospital can compete by setting an unusually high salary for all its slots, or by setting a small number of slots aside at a higher salary and entering them in the match as a separate entity. What it cannot do is state that it would be willing to pay a higher salary for, say, Mary Smith, but would otherwise prefer to pay less.

Bulow (who happens to be a Yankees fan) adds: "If you're the Yankees and you want the best free agent pitcher, maybe you'll say, 'Okay, I'll pay $28 million for him.' But what if you end up not with that person but with someone not as good?" The Yankees can choose to pay less for a lesser pitcher or hire no pitcher at all, he points out, but in the residency match, the hospitals have to take what their offers attract.

In a real market with negotiated salaries, by contrast, a student who received a very lucrative offer from a lower-tier school would show the offer to the top-ranked schools, Bulow explains. As a result of the competition, the higher-ranked schools would then increase their offers slightly. In the residency match, this effect is damped out.

Thus, it seems that any fix to the residency match would have to include a mechanism for tailoring salaries to individuals, perhaps as part of the matching process. An algorithm devised in the 1980s does exactly this.

Matching by Algorithm with Built-in Salaries

In a 1981 paper, "Job Matching with Heterogeneous Firms and Workers," Vincent Crawford and Elsie Knoer sketched out a procedure much like the Gale–Shapley algorithm, but in the framework of firms and workers rather than hospitals and residents, and with price negotiation built into the algorithm.

Crawford and Knoer assume, for simplicity, that each firm hires exactly one worker, but their arguments easily extend to the general case under the assumption that preferences are *separable across pairs*, meaning that the job satisfaction and productivity for a given worker at a given firm do not depend on the other workers hired by the firm.

In the Crawford–Knoer model, there are n workers and firms and the ith worker's job satisfaction, productivity, and salary at the jth firm are a_{ij}, b_{ij}, s_{ij}, respectively. All quantities are integers, with $a_{ij} + b_{ij} \geq 0$. s_{ij} is actually a (discrete) function of time, so that $s_{ij}(t)$ is the salary the jth firm can offer the ith worker at round t in the adjustment process.

A matching of workers to firms is considered *stable* if there is no worker/firm/salary such that the worker and firm would prefer to be paired with each other at that salary rather than with the firm and worker assigned to them by the matching. (Each worker also has to have positive job satisfaction plus salary, of course, and the productivity minus salary for each hospital had better be positive, too.)

Such a matching can be achieved by the following adjustment procedure, which begins by setting $s_{ij}(0) = -a_{ij}$ for all i,j. In the first round, each firm makes an offer to its favorite worker at these salaries (with the favorite defined as the worker who maximizes productivity minus salary). Each worker who receives more than one offer rejects all but his/her favorite (measured as salary plus satisfaction), which he/she tentatively accepts. Workers can break ties in any way they like.

Can you imagine medical students required not only to rank the hospitals but also to affix price differentials to their rankings?

Any offer not rejected in previous rounds is still in effect in subsequent rounds. With each rejected offer, the salary is increased by one. Firms continue to make offers to their favorite workers, and the procedure ends when no rejections are issued in a given round.

Crawford and Knoer go on to prove that this process converges in finite time to a stable matching (as defined above) of workers and firms, and they extend their proofs to the case of multiple workers per firm and to continuous salaries.

Crawford and Knoer's proofs rely on the assumption of what Alvin Roth calls *responsive preferences* of firms and workers. What it means is that the decision maker's preferred set of workers from any available pool can be deduced using the rankings of individual workers and the number of slots he wishes to fill and that every worker has a non-negative net contribution.

This condition turned out to be stronger than what Crawford and Knoer needed for their theorems. Another paper, "Job Matching, Coalition Formation, and Gross Substitutes," published a year later by Alexander Kelso and Crawford, shows, among other things, that the Crawford–Knoer results hold under a weaker condition that economists call *substitutability*. A firm's (or worker's) preferences are substitutable if whenever w_1 is one of a firm's most preferred set of workers $P(S)$ from a pool S containing both w_1 and w_2, and w_2 is removed from S, w_1 is still part of $P(S - w_2)$. Without substitutability, stability breaks down.

To see that substitutability is weaker than responsiveness, consider the following example. Suppose that two firms, F_1 and F_2, and three workers, w_1, w_2, and w_3, have the following preferences:

$$P(F_1) = \{w_1, w_2\}, \{w_1, w_3\}, \{w_2, w_3\}, \{w_3\}, \{w_2\}, \{w_1\}$$
$$P(F_2) = \{w_3\}$$
$$P(w_1) = P(w_2) = P(w_3) = F_1, F_2.$$

Firm 1's preferences aren't responsive, in that it prefers $\{w_1, w_2\}$ over $\{w_1, w_3\}$ even though it prefers worker $\{w_3\}$ to worker $\{w_2\}$, yet all the preferences are substitutable and there does exist a stable matching:

$$F_1 \leftrightarrow \{w_1, w_2\} \text{ and } F_2 \leftrightarrow \{w_3\}.$$

The Kelso–Crawford approach does come up with matches that are stable and take wages into account, but it lacks at least one aspect of a free market. This framework has no mechanism for ensuring that a hospital stays within a fixed budget, a feature that might make it impractical in the real world.

But this flaw is intrinsic to the system. Achieving stable matchings while keeping a hospital within a fixed budget is not possible—placing budget limits on the matching process will make substitutability fail, in which case it's possible that no stable matching will exist.

Conclusion

The inability to impose budget restrictions is just one reason the elegant Kelso–Crawford–Knoer algorithm is unlikely to help the residents in court, assuming that the algorithm is indeed responsible for some of the wage depression. A major hurdle to practical implementation, as Milgrom of Stanford points out, is that the algorithm requires unrealistic information from the participants. Can you imagine medical students required not only to rank the hospitals but also to affix price differentials to their rankings?

When they go to court to argue their case, the residents might be advised to bring along the Bulow–Levin paper, with its demonstration that a model similar to the match leads to salary stagnation. But Bulow denies that the model implicates the hospitals in anti-competitive behavior. The match was developed for efficiency, he says, and his and Levin's model shows that the results are indeed efficient. He also points out that other markets, such as the one for lawyers fresh out of law school, have similar properties despite not being centralized. He speculates that the wage depression in the medical market may result not just from the match, but also from the limited number of good slots available and the lack of fluidity in the resident labor market.

Matching theory does not provide an easy fix, and residents may just have to choose: the stability and fairness of the match with lower salaries, or open competition with the risk of a return to the chaos of the early 1950s.

For an example of a market that would clearly benefit from matching theory, check out the sidebar on matching and school choice.

Sara Robinson is a freelance writer based in Pasadena, California.