A primer on the game theory behind the National Resident Matching Program for the medical educator and student

Authors:

1. Muhammad Maaz[1,2]

   maazm@mcmaster.ca

   ORCID ID: 0000-0002-3869-631X

Affiliations:

1. Faculty of Health Sciences

   McMaster University

   Hamilton, Ontario, Canada

2. Department of Economics

   McMaster University

   Hamilton, Ontario, Canada

This is a nontechnical introduction to the residency matching algorithm used by the National Resident Matching Program (NRMP) and by the Canadian Residency Match Service (CaRMS).

Abstract

Every year, medical students vie for American graduate training through the National Resident Matching Program (NRMP). Some students yet behave in ways that imply persistent misunderstandings about the matching algorithm. This paper explains the economic and mathematical literature underpinning it for a medical audience. The NRMP implements the Roth-Peranson algorithm, finding a stable match by having students propose to residency programs according to their preference ranking. This configuration favors students while disfavoring hospitals. Game-theoretic analysis shows us that students are unequivocally unable to "game the system" by misstating their preferences. Telling the truth is the optimal strategy.

Introduction

Every year, medical students wishing to practice in the United States apply to be matched to graduate training through the National Residency Matching Program (NRMP). Much has been written in the medical literature about what is colloquially termed the “The Match”, from analyses of matching statistics to correlates of matching success. Despite the considerable interest in The Match, the mechanisms of the algorithm and its implications yet remain somewhat a mystery to medical students. Despite educational resources devoted to The Match during medical school, recent empirical evidence by Rees-Jones suggests that medical students participating in the NRMP still engage in behaviour that reflects a misunderstanding of how the algorithm works [1]. Rees-Jones also presents experimental evidence that shows that people behave better in such systems when they are taught the underlying theory of how the algorithm works [1], which is the aim of this paper. A solid understanding of the theory behind the algorithm will be of use to both medical educators and students in preparing for The Match. The algorithm used by The Match was developed by, and has been extensively studied by, researchers in a field called matching theory, inhabited by economists, mathematicians, and computer scientists. Though the matching theory literature is quite extensive and established in answering common concerns about The Match, these questions still abound in medical circles. This is quite possibly due to the language barrier: research in matching theory is communicated in esoteric mathematical symbolisms and formal logic, leading to a disconnect between what is known in matching theory and what is known by those in the medical field. In this paper, I aim to cross that divide by elucidating the underlying matching theory and game theory for a non-economist/non-mathematician audience in the hopes of clearing up the confusion surrounding The Match.

How does the NRMP matching algorithm work?

The NRMP[1] uses an algorithm called the Roth-Peranson algorithm [2], a modification [2] of the deferred acceptance (DA) algorithm developed in 1962 by Gale and Shapley [4], which gave birth to the field of matching theory. Matching theory is concerned with finding the best match between members of two separate groups: say, workers to firms. Gale and Shapley developed the framework and terminology of matching theory by looking at the stable marriage problem, which roughly asks: how can a set of men and a set of women be best matched to each other for marriage? They develop the concept of a match being *stable* as a desirable property. Stability in a match can be best defined by its opposite: namely, an *unstable* match, in our marriage example, means there exist "a man and woman who are not married to each other but prefer each other to their actual mates" [4]. Clearly, one can imagine that this is an undesirable property for marriages. Therefore, a stable match is defined as a match in which there are no unstable pairs. This concept of stability is critical to matching theory, as it ensures the long-term viability of such a system: Roth and Sotomayor give many examples throughout history, including in prior conceptions of the American medical residency match, where unstable matching systems lead to widespread resentment and eventually arrangements begin to be made outside of the formalized clearinghouse, causing chaos and confusion [5].

In their paper, Gale and Shapley are interested in finding a stable match between the men and women. They propose the DA algorithm and prove that this algorithm always yields a stable match [4]. The DA algorithm asks firstly for everyone to construct an ordered list of their

[1] A point of international concern: the same algorithm is used in other residency matches outside of the United States. Both the Canadian [6] and Japanese [7] residency matches use the Roth-Peranson algorithm (however, I do not profess this to be an exhaustive list). The properties and results of the algorithm described in this paper apply to those countries as well.

preferences: men rank the women and women rank the men. It is important that preferences are *strict* – that is, given the choice between two people, there is no indifference between who one would choose. Next, have every man propose to his 1st choice. Women with multiple proposals choose the man that they ranked highest among their proposers and reject the others. Rejected men then go on to propose to their 2nd choice, and the same process continues until everybody is matched. However, a woman's acceptance is tentative: if, in a later round, she receives a proposal from a man who she ranks higher, she will renege on her current suitor and take the newer one instead [4]. This crucial property is why it is termed *deferred* acceptance. It is somewhat intuitive to imagine why there can be no unstable pairs with this method: if a man would rather have another woman that is not his, then he would have proposed to her somewhere prior in the algorithm. However, the fact that they are not matched means that she rejected him in favour of her present husband, and so she prefers her present husband – thus there is no instability. I have told this story with the men proposing, but it would work just as well if the women were the ones proposing to the men. Indeed, both man-proposing and woman-proposing implementations yield stable matchings [4], albeit possibly different ones – note that it is possible for multiple stable matches to exist.

The Roth-Peranson algorithm is DA with some modifications; however, the crux of the matter is the same. Students submit preference lists of residency positions, and hospitals similarly rank applicants to their programs [2]. In the NRMP, it is students that do the proposing [3]. In the first round, students propose to their first-choice hospital. Hospitals accept the top students from their proposers according to their own ranking of students, up to their quota, and reject the others. Rejected students go on to propose to their next preferred hospital, and so on. Acceptances are tentative until the termination of the algorithm, such that a hospital may let go of a previously accepted student to make room for a higher ranked student in a later round of proposals. This process is best seen in an example.

Example of matching in the NRMP

Assume there are 6 students (A through F) and 3 hospitals (H1 through H3), each with 2 spots. Next, assume that students' and hospitals' preferences over each other are as given in Table 1.

Table 1. Example of student and hospital preferences

Student preferences	**Hospital preferences**
(most to least preferred)	**(most to least preferred)**
A: H1 > H2 > H3	H1: A > B > E > F > C > D
B: H2 > H1 > H3	
C: H3 > H2 > H1	H2: B > A > F > D > E > C
D: H2 > H3 > H1	
E: H1 > H3 > H2	H3: F > D > C > E > A > B
F: H1 > H2 > H3	

In the first round, all students propose to their first choice. A, E, and F propose to H1; B and D propose to H2; and C proposes to H3. H1 only has 2 spots but 3 proposals, so it takes the top two, which are A and E, and rejects F. H2 accepts both B and D, and H3 accepts its only proposal, C. F, having been rejected, now proposes to his next preferred hospital, H2. According to H2's ranking, F outranks one of its current students, namely D, and so it lets D go (this is that concept of tentative acceptances) and takes F instead. Now, D must participate in another proposal and proposes to its next ranked hospital, H3. With one spot still open, H3 accepts the proposal. Our final matching is therefore: H1 with A and E, H2 with B and F, and H3 with C and D. This process is summarized in Table 2.

Table 2. Example of run-through of the Roth-Peranson algorithm (student-proposing deferred acceptance) per preferences from Table 1

Round	**Proposals**	**Responses to proposals**	**Allocations at end of round**
1	A, E, F propose to H1. B, D propose to H2. C proposes to H3.	H1 accepts A and E, rejects F. H2 accepts B and D. H3 accepts C	H1: A, E H2: B, D H3: C None: F
2	F proposes to H2.	H2 reneges on D and accepts F instead.	H1: A, E H2: B, F H3: C None: D
3	D proposes to H3.	H3 accepts D.	**H1: A, E** **H2: B, F** **H3: C, D** **Algorithm terminates**

This is a stable matching: there is no such pair that would mutually rather be with each other over this assignment. For example, H1 prefers B to its student E, but B does not prefer H1 to its hospital H2, so this is not an instability.

This process described above is exactly what the NRMP does every year, writ large: with many times more students and hospitals and much longer ranked order lists.

Does the algorithm favour students or hospitals?

Just as there are man-proposing and woman-proposing configurations in the stable marriage problem, so too there are student-proposing and hospital-proposing versions in the context of the medical residency match. The NRMP uses the student-proposing version of the algorithm [2]. This turns out to be quite an important property, as the student-proposing stable match is the *best outcome* for the students [3]. Note that in any such two-sided matching problem, there may be many possible stable matchings. One of these stable matchings is given by the student-proposing algorithm, while another one of these stable matchings is given by the hospital-proposing algorithm (in general, these are different, but may be the same stable matching in particular cases). Among all these stable matchings, the one yielded by the students proposing is the *optimal* match from the perspective of the students [5]. Intuitively, this is because when the students propose, they are the ones "in control", and will be matched to the highest hospital on their list that will accept them. Put another way, there is no other stable matching in which they would be matched to a higher-ranked hospital than the one they are matched to when they propose [5]. Oppositely, the hospital-proposing algorithm would yield the optimal stable matching for the hospital [5].

Interestingly, what is best for the students and what is best for the hospitals seem to be at odds with one another. When there are two or more different possible stable matchings, the student-proposing algorithm yields the student-optimal, but the hospital-proposing algorithm yields the hospital-optimal match. Even more strikingly, a well-known result in matching theory states that the student-optimal match is in fact the hospital-pessimal (ie. the worst stable match for the hospitals), while the hospital-optimal is the student-pessimal match [5]. The proof of this statement is mathematical, but it can be demonstrated using our prior example. In Table 2, a run-through of an example of the student-proposing algorithm is shown. In contrast, Table 3 shows what would occur if instead the hospitals did the proposing.

Table 3. Hospital-proposing version of the example in Table 2 per preferences in Table 1

Round	**Proposals**	**Responses to proposals**	**Allocations at end of round**
1	H1 proposes to A, B. H2 proposes to B, A. H3 proposes to F, D.	A rejects H2 and accepts H1. B rejects H1 and accepts H2 F and D accept H3.	H1: A H2: B H3: D, F None: C, E
2	H1 proposes to E. H2 proposes to C.	E accepts H1. C accepts H2.	**H1: A, E** **H2: B, C** **H3: D, F** **Algorithm terminates.**

While both matchings given in Table 3 and Table 2 are stable, they have allotted different students to the hospitals. In both instances, H1 is allotted A and E. However, H2 and H3 are given different students. H2 is given B and F from the hospital-proposing, which it prefers to B and D (as per the preferences given in Table 1), which is given in the student-proposing match. Similarly, H3 prefers D and F from the hospital-proposing outcome to D and C from the student-proposing version. On the other hand, any student that is given different hospitals in either configuration prefers the hospital in the student-proposing match to the one it is given in the hospital-proposing match; for example, C prefers H3 – its allotment when it proposes – to H2, which it gets when the hospitals propose. So, in general, there is a trade-off between what the hospitals prefer and what the students prefer: when the students propose, it is the best stable match for them, while being the least preferred stable match for the hospitals, and vice versa.

Thankfully for students, the NRMP has chosen to take the side of the students [3] rather than the hospitals by having the students do the proposing, which was not always the case historically – indeed, it used to be that the NRMP used the hospital-proposing configuration [2-3].

How does the algorithm deal with couples?

A modification to DA that Roth-Peranson makes is allowing students to participate in The Match as couples. This is critical from a historical perspective; prior to allowing couples to match using Roth-Peranson, the American residency match experienced a relative lack of participation from couples compared to singles [5]. Roth-Peranson allows couples to submit a preference list over *pairs* of positions. For example, a couple might specify that they would like to be matched to A/A, A/B, A/C, and B/C, in that order. The NRMP also allows couples to specify a “No match” [8], such that they can specify that one of them is willing to not match if the other matches somewhere, if it comes down to it. This is entered in the list along with the preferences, as in:

A/A, A/B, A/C, B/C, A/”No match”, “No match”/B. The algorithm then treats the couple as one unit when it performs the usual process of proposing matches to residencies [2].

While this is an improvement over some older, more chaotic ways of incorporating couples [5], it is not a complete solution. Indeed, a significant result in matching theory is that there may be no stable matches possible by Roth-Peranson due to couples [5]. The crux of the issue is that hospitals still have preferences over individual students, not couples. DA, when only talking about singles, always ensures a stable match because the hospital never regrets rejecting a candidate, as they only reject them if there is another candidate that they prefer more. However, in the Roth-Peranson case, consider what would happen if a hospital rejects one member of a couple and wishes to accept the other member, but that couple prefers to match to that hospital or not at all. This is unstable individualistically: the hospital would like to have that student, and that student would like to match there, if only the hospital would take their spouse as well. The hospital therefore regrets the rejection. Thus, while the incorporation of couples is inclusive, it comes at some cost for the stability of The Match. Indeed, the couples’ problem is an active area of research in matching theory, and recent work, especially by algorithmic computer scientists, has been done to investigate how to minimize that regret that hospitals could experience [9 - 11]. It will remain an interesting question as to if and how The Match will change in the future in this regard.

However, it is worth noting that the couples’ problem is perhaps, practically speaking, a rarity. The Roth-Peranson algorithm is now sold as a service by the National Matching Services Inc. (owned by the algorithm’s namesake Elliott Peranson), and their data shows there have only been a few instances of no stable matches found “over the last decade in several dozen annual markets” that use Roth-Peranson [12]. As well, American data shows that couples comprise only

1.9% of those who submit to the residency match [12], and it has been recently shown that as the proportion of couples participating in The Match tends to zero, the probability of a stable match existing approaches one [12], which seems intuitive in some sense.

Can students "game the system" by misstating preferences?

This is one of the most prevalent concerns that students have. There seems to be a perennial misguided belief by some medical students that they can "game" The Match by strategically choosing their rank order list. In the paper by Rees-Jones which motivated the writing of this paper, he shows that "some students [participating in the NRMP] pursue futile attempts at strategic misrepresentation" [1]. An intuitive elaboration of the game theory at work in The Match could dispel this notion. The Roth-Peranson algorithm has been mathematically proven to be what is called "strategy-proof" [5]: it is in the best interests for the students to present their beliefs truthfully, and there is unequivocally no benefit to lie about preferences in a bid to improve one's matching outcome. Therefore, misstating preferences is suboptimal behaviour, which has been empirically demonstrated to be yet present in about a fifth of students in the NRMP [1].

Based on our prior discussion of the algorithm, it follows that misstating one's preferences is suboptimal. When students propose, they are allocated to their highest ranked hospital that will have them. A misstating of preferences (putting a lesser preferred hospital higher on the list) will either be inconsequential or will hurt the student. In our example in Table 2, if, for example, student F had ranked H3 as its top choice, based on a belief that H3 ranks them highly (a prevalent line of reasoning amongst match participants [1]), then they would have been matched to H3, and not H2, which they actually prefer to H3. It is always in the best interest for the students to present their rankings to reflect their true preferences, and not on notions of how the hospitals would rank them, or on other strategic ideas [5].

Interestingly, on the other hand, Roth proved it is possible for hospitals to lie about preferences in order to achieve a better outcome when students propose [5]. While there is the possibility of manipulation by hospitals, to the author's knowledge, no observation of such manipulation has been seen or confirmed in the NRMP, and such manipulation is theoretically and practically exceedingly difficult to perform successfully given the sheer size of the set of students, and lack of knowledge about students rank order lists [13].

Conclusion

The medical residency match is a significant accomplishment of algorithmic design. The Roth-Peranson algorithm used by the NRMP is an extension of the deferred acceptance algorithm that Gale and Shapley developed to solve the stable marriage problem. In the NRMP configuration, students propose to residencies sequentially down their ranked order list. This delivers the best outcome possible for the students, while disfavoring the hospitals. Students can apply to match as couples by submitting preferences over pairs of positions: while their incorporation presents a threat to the stability of The Match, it seems to be working so far. Students cannot benefit from lying about preferences; basing their ranked order list on anything other than their true preferences is suboptimal behaviour. Understanding and appreciating the implications of these results would serve medical students and their educators well in preparing for graduate medical training[2].

[2] For those readers interested in a more mathematical treatment of the algorithm and its results discussed in this paper, Roth and Sotomayor's book "Two-sided matching" [5] is an exceptional resource. The book also discusses the history of the NRMP and the interesting interplay between game theory and the real world in the evolution of the NRMP over the decades.

Compliance with Ethical Standards

Funding (Optional): NA

Conflict of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval: NA

Informed consent: NA

References

1. Rees-Jones A. Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match. Games and Economic Behavior. 2018 Mar 1;108:317-30.
2. Roth A, Peranson E. The Redesign of The Matching Market for American Physicians: Some Engineering Aspects of Economic Design. American Economic Review. 1999;89(4):748–80.
3. Roth AE, Peranson E. The effects of the change in the NRMP matching algorithm. JAMA. 1997 Sep 3;278(9):729-32.
4. Gale D, Shapley LS. College Admissions and the Stability of Marriage. The American Mathematical Monthly. 1962;69(1):9–15.
5. Roth AE, Sotomayor MAO. Two-sided matching: study in game-theoretic modeling and analysis. Cambridge: Cambridge University Press; 1992.
6. Canadian resident matching service [Internet]. CaRMS. CaRMS; 2018 [cited 2019Aug11]. Available from: https://www.carms.ca/
7. Kamada Y, Kojima F. Stability and strategy-proofness for matching with constraints: A problem in the japanese medical match and its solution. American Economic Review. 2012 May;102(3):366-70.
8. Couples in The Match. The Match, National Resident Matching Program. http://www.nrmp.org/couples-in-the-match/. Accessed November 1, 2019.
9. Drummond J, Perrault A, Bacchus F. SAT is an effective and complete method for solving stable matching problems with couples. InTwenty-Fourth International Joint Conference on Artificial Intelligence 2015 Jun 22.

10. Klaus B, Klijn F. Paths to stability for matching markets with couples. Games and Economic Behavior. 2007 Jan 1;58(1):154-71.

11. Biró P, Klijn F. Matching with couples: a multidisciplinary survey. International Game Theory Review. 2013 Jun 28;15(02):1340008.

12. Kojima F, Pathak PA, Roth AE. Matching with couples: Stability and incentives in large markets. The Quarterly Journal of Economics. 2013 Aug 31;128(4):1585-632.

13. Kojima F, Pathak PA. Incentives and stability in large two-sided matching markets. American Economic Review. 2009 Jun;99(3):608-27.