STABLE MATCHING WITH PREFERENCES DERIVED FROM A PSYCHOLOGICAL MODEL *

John BARTHOLDI, III and Michael A. TRICK

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received May 1986
Revised August 1986

We study a special case of the Stable Roommates problem in which preferences are derived from a psychological model common in social choice literature. When preferences are 'single-peaked' and 'narcissistic', there exists a unique stable matching, and it can be constructed in O(n) time. We also show how to recognize quickly when a set of preferences is single-peaked.

stable matching * single-peaked preferences * consecutive 1's

1. Introduction

The Stable Roommates problem is to match each of $2n$ people so that no two people prefer each other to their assigned partners. This is a generalization of the Stable Marriage problem (match n men to n women). However, unlike the marriage problem, there are instances of the Stable Roommates problem for which no matching is stable (Gale and Shapley [7]). We show that, when preferences are derived from a simple psychological model, a stable matching is guaranteed to exist, and can be constructed in O(n) steps. Examples include when each person prefers a roommate who sets the thermostat as close as possible to his own ideal setting, or when each person prefers a roommate whose hometown is as close as possible to his own. In these and other such cases, two conditions induce special structure on the preferences and enable our results: first, there exists a common 'frame of reference' among the participants by which alternatives are judged (for example, the temperature scale, a map with distances); secondly, each person is 'narcissistic' in that he is his own ideal (i.e., most preferred among all alternatives).

2. Single-peaked preferences

Consider a finite set A of 'alternatives'. Each of a finite set of 'decision makers' D has a strict, transitive preference order P_i on A, with which the decision maker is identified. P_i is said to be *single-peaked with respect to a sequence* S_i of the alternatives iff S_i can be split into two segments X_i and Y_i (one of which may be empty) for which

(i) each alternative in X_i is preferred to any alternative to its left, and

(ii) each alternative in Y_i is preferred to any alternative to its right.

(Here we are slightly more restrictive than Fishburn [5] in that we do not allow indifference.) Roughly speaking, this means that a graph of 'intensity of preference' of P_i versus S_i is unimodal, first increasing along X_i and then decreasing along Y_i.

Let $P = \{P_i\}$ denote the set of preference orders P_i of all the decision makers (duplications allowed). We say that the preferences P are *single-peaked* iff there exists some fixed sequence S with respect to which every P_i is single-peaked. This corresponds to the intuitive idea that there is a

* This research was supported by the National Science Foundation under Grant no. ECS-8351313, and the Office of Naval Research under Grant no. N00014-83-K-0147.

single frame of reference (of one dimension), that is common to all decision makers and along which alternatives are judged. Roughly speaking, alternatives are preferred less as they are 'farther' from each decision maker's ideal. Decision makers may have sharply different ideals, but they share a common world view. A frequently mentioned example from voting theory is the 'liberal–conservative' spectrum often used to explain voting behavior; if all voters base their preferences consistently and exclusively on such a model – and agree on its structure – then the preference orders of the electorate are single-peaked.

'Single-peakedness' of preferences was first studied by Black [2] and has been much discussed in the literature of voting theory since. Its role there has been as an appealing special condition on preference orders which precludes certain paradoxes, especially the famous 'phenomenon of cyclic majorities'. (Black [2] and Fishburn [5] are the most comprehensive references on single-peakedness and social choice; for a survey emphasizing voting systems, see Niemi and Riker [10]; the phenomenon of cyclic majorities is surveyed from a mathematical perspective in Fishburn [6] and Straffin [12].) We will show that single-peakedness also ensures 'good behavior' in the stable matching problem.

3. The Stable Roommates problem

Irving [9] has recently given an ingenious algorithm that, for any instance of the Stable Roommates problem, either constructs a stable matching or else concludes that none exist. Irving also performed computational experiments that suggest that for randomly generated preference orders, the proportion of instances admitting of stable matchings decreases as the number of people increases. For example, less than 70% of the problems with 90 people had stable matchings.

We show that if preferences are (1) single-peaked, and (2) narcissistic, then there always exists a stable matching and it is unique. An example is when each person prefers to be matched with someone whose ideal thermostat setting is as close as possible to his own. Then sequencing everyone from lowest ideal temperature to highest realizes the preferences as single-peaked. Alternatively, we can imagine each person's preferences to be based solely on the volume at which a prospective roommate plays his stereo.

Lemma 1. *If among all available choices person x most prefers person y, and person y most prefers person x, then in any stable matching x and y must be matched.*

Proof. If x and y are not matched together, they would prefer each other to their assigned partners. □

Theorem 1. *If preferences are single-peaked and narcissistic, then there exists a unique stable matching.*

Proof (by induction). The claim is obviously true for $k=1$ (two people) – the only matching is stable. Assume that it is true for $2(k-1)$ people and consider the case of $2k$ people with single-peaked preferences. Let S be a sequence which realizes single-peakedness.

By the hypotheses of the theorem, each person's first (feasible) choice must be adjacent to him in S. If we ask each person to point to his favorite, then, since there are $2k-1$ adjacencies and $2k$ people pointing, by the 'pigeon-hole principle' two people adjacent in S must mutually point to each other. By Lemma 1 these two people must be matched in any stable matching; remove them from the sequence and expunge them from everyone else's preferences. The remaining people are $2(k-1)$ in number, and their preferences remain single-peaked, since single-peakedness is hereditary on subsets. By the induction hypothesis there exists a unique stable matching on this subset, and adjoining the removed, matched pair cannot introduce any instability. □

The recursive algorithm implicit in the proof requires $O(n)$ worst-case time to construct a stable matching. It achieves this efficiency since, for each pair matched, at most two other people must revise their choice of favorites. By way of comparison, for the more general problem of arbitrary preference orders, $O(n^2)$ worst-case time seems to be required to determine whether a stable matching exists (Irving [9]).

It is worth noting that the above algorithm need not match each person with someone who is adjacent in S. In fact, no more than the first-matched pair need have been adjacent in S. Thus, in the interests of stability, it is possible that the person who prefers his stereo the loudest might be assigned to room with the poor person who prefers his the quietest.

Finally we observe that single-peaked, narcissistic preferences confer an additional robustness. Dubins and Freedman [4] showed that stable marriage is susceptible to strategic manipulation: it is sometimes possible for an individual to improve his outcome by misrepresenting his true preferences. (See also Gale and Sotomayor [8].) In contrast, the $O(n)$ matching algorithm for single-peaked, narcissistic preferences is immune to this kind of manipulation because there is no incentive: each person is matched with the first person who will have them. However, there is susceptibility to the more powerful manipulation of misrepresenting one's place in the frame of reference. For example, to get the roommate of choice, one might lie about one's preferred thermostat setting. This kind of manipulation is more powerful in that it changes the preferences of others.

Example. Consider the following set of preferences:

$1 > 2 > 3 > 4 > 5 > 6 > 7 > 8$,
$2 > 3 > 4 > 5 > 1 > 6 > 7 > 8$,
$3 > 4 > 2 > 5 > 1 > 6 > 7 > 8$,
$4 > 3 > 5 > 6 > 2 > 7 > 1 > 8$,
$5 > 4 > 3 > 2 > 6 > 7 > 8 > 1$,
$6 > 5 > 4 > 3 > 2 > 7 > 1 > 8$,
$7 > 6 > 5 > 4 > 3 > 2 > 8 > 1$,
$8 > 7 > 6 > 5 > 4 > 3 > 2 > 1$.

These are narcissistic since each person is their own highest preference; they are also single-peaked since labels have been chosen so that for the sequence (1 2 3 4 5 6 7 8), person j's preferences increase on $X_j = \{1, \ldots, j\}$ and decrease on $Y_j = \{j+1, \ldots, 8\}$. The behavior of the matching algorithm on this instance is shown below, where '$i \to j$' means that i holds j favorite among the alternatives available to him, and '$i = j$' means '$i \to j$ and $i \leftarrow j$'.

iteration 1: $1 \to 2 \to 3 = 4 \leftarrow 5 \leftarrow 6 \leftarrow 7 \leftarrow 8$,
iteration 2: $1 \to 2 \quad = \quad 5 \leftarrow 6 \leftarrow 7 \leftarrow 8$,
iteration 3: $1 \quad \to \quad 6 = 7 \leftarrow 8$,
iteration 4: $1 \quad = \quad 8$,

stable matching: 3–4, 2–5, 6–7, 1–8.

In the first iteration 3 and 4 were mutual favorites and so were matched. In the next iteration 2 and 5 revised their choices and chose each other. Then 1 and 6 revised their choices, and 6 and 7 chose each other. Finally, 1 and 8 revised their choices and chose each other.

4. Recognizing single-peaked preferences

Since single-peakedness is of independent interest, we consider it on its own, without the additional assumption of narcissism.

It has been suggested that single-peakedness largely pertains in popular elections. However it is difficult to test such a suggestion for at least two reasons. First and most obvious is the difficulty of determining preferences; this task is complicated by the costs of sampling and the unreliability of the data. Secondly, and at issue here, is that even if preferences were accurately known, it may still be difficult to establish single-peakedness, since this is equivalent to showing that some one among the $|A|!$ permutations of the alternatives constitutes a single dimension along which the decision-makers effectively rank alternatives. (This spectrum is only a formal structure and need not correspond to the perceptions for the decision-makers.) Furthermore this difficulty could be more than theoretical: *The New York Times* of 1 April 1986 reported 20 candidates for mayor of Tulsa, Oklahoma!

We show that single-peakedness of preferences is recognizable in polynomial time, so that even for a large number of alternatives, a set of preferences can be tested quickly.

We transform the problem of recognizing single-peakedness to a problem on matrices, and then observe the solution for the matrix problem. Let each decision maker i represent his preferences as follows: to each a_j in A is assigned an integer Rank(a_j) between 0 and $|A|-1$ (inclusive) so that if a_j is preferred to a_k then Rank(a_j) > Rank(a_k). Thus Rank(a_j) is the preference ordinal for the alternative a_j. Now the preferences of any individual i can be summarized by an $|A| \times |A|$ matrix, the *individual preference matrix*, which unary encodes the preference ordinals in the following manner: column a_j consists of $|A|$ − Rank(a_j) 0's, followed by Rank(a_j) 1's, as illustrated below.

Example. This individual preference matrix encodes the preference order $4 > 3 > 5 > 6 > 2 > 7 > 1 > 8$. Under the permutation of alternatives (1 2 3 4 5 6 7 8) this preference order is single-peaked.

$$P_i = \begin{matrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{matrix},$$

alternatives: 1 2 3 4 5 6 7 8.

Now consider the $(|D| \times |A|) \times |A|$ matrix P formed by permuting the columns of the individual preference matrices to any common permutation of A, and then arranging the individual preference matrices one above the other. Then this matrix expresses the preferences of all the decision-makers,

$$P = \begin{matrix} P_1 \\ \vdots \\ P_i \\ \vdots \\ P_{|A|} \end{matrix}, \quad \text{where columns of } P \text{ correspond to elements of } A.$$

Lemma 2. *The preferences of the decision-makers are single-peaked iff there exists a permutation of the alternatives (columns of P) such that P has the property of consecutive* 1*'s in rows.*

Proof. If, for a fixed sequence S, some row of some P_i violates consecutive 1's, then that row must contain the sequence $\ldots 1\ 0 \ldots 0\ 1 \ldots$ But then the preferences P_i are first decreasing and then increasing in this region, and this violates the definition of single-peakedness. The converse can be argued similarly. □

Booth and Lueker [3] give an algorithm whereby consecutive 1's in rows can be tested in time $O(m + n + f)$, where the matrix is of dimension $m \times n$, and has f non-zero elements. In this case P is $(|D| \times |A|) \times |A|$, and has $O(|D| \times |A|^2)$ non-zero elements, and so can be tested within time $O(|D| \times |A|^2)$. Furthermore, the Booth and Lueker algorithm makes available as a by-product all sequences of the elements of A which realize the property of single-peakedness. This may be useful in formulating hypotheses about possible spectra along which alternatives might be judged.

Unfortunately, because of similar results for the property of consecutive 1's, it is NP-complete to find a permutation of alternatives for which the preferences are 'as single-peaked as possible'; that is, as few rows of P as possible violate consecutive 1's. It is this approximation problem, rather than recognizing 'pure' single-peakedness, that seems the more pertinent to practical study of voting populations.

Finally we observe that it is even simpler to test for single-peakedness under the additional conditions that the alternatives are the decision makers themselves (as in the Stable Matching problem), and everyone is narcissistic. For preferences to be single-peaked, there must exist some sequence of alternatives which realizes this. Furthermore, each person's least preferred alternative is an extreme of such a sequence. Thus we can identify an extreme element by choosing anyone's least preferred alternative. By narcissism and single-peakedness, the preference ordering of this alternative must be such a sequence. It requires $O(|A|^2)$ time to check this.

5. Further remarks

The results of Section 3 can be generalized somewhat without losing their savor. The essential requirement is that the algorithm match a pair of roommates at each step. Thus we must be sure that at each step there exist at least two people who are each other's favorites among the choices available to them. Furthermore this property must be hereditary on each subgroup as individuals are matched and removed from the problem. These conditions pertain when, for example, narcissism holds but single-peakedness is replaced by a different psychological model: preferences based on nearness to an ideal in some multidimensional 'attribute space', where we additionally require that distances be symmetric (so that, in a sense, people agree on their differences). An example is when each person prefers to room with someone whose hometown is close to his own (where, since we are not allowing indifference, all distances between hometowns are distinct). For this example the algorithm still requires only $O(n)$ steps, since after two people are matched, less than 12 others will have lost their favorites and must choose others. (This follows since in Euclidean space no more than six points can have a given point as their closest.)

The assumption that preferences are based on distances in some attribute space is a standard model in public choice theory (see for example Aranson [1] or Riker and Ordeshook [11]). It generalizes one aspect of single-peakedness in that it allows preferences to be based on more than a single dimension (i.e., attribute); but since it is based on a notion of distance, it is more restrictive than single-peakedness, for which only the order of the alternatives along the line is important.

A more abstract statement of these results – which, however, has the disadvantage of not suggesting how individuals form preferences – is as follows. We consider the Stable Matching problem to be this: given an undirected graph G with $2n$ vertices, together with strict preference orders on those vertices, find a perfect matching that is stable. Here the graph need not be complete, and may include loops. Also, stability means that there do not exist two vertices who are not matched to each other *but could have been* (because they are connected by an edge in G), and who prefer each other to their assigned partners. Now we formalize a condition for which the algorithm will produce a stable matching if there exists any matching at all. Construct a new graph G' by creating nodes corresponding to the edges of G. Represent all preferences by directing an arc in G' from node (i, j) to node (i, k) iff i prefers j to k. (Note that we also direct an arc from (i, j) to (k, j) iff j prefers i to k.) Then the algorithm has this interpretation: find a node (i, j) in G' of indegree 0: match i with j in G and delete node (i, j) and all incident edges in G'; repeat. The algorithm succeeds if, at each step, there exists a node of indegree 0 in G'. Equivalently, the derived graph must be free of directed cycles.

Under these conditions, more general than single-peakedness with narcissism, the algorithm can require $O(n^2)$ effort, since after two people have been matched all of the remaining people may have to update their favorites. However, a stable matching exists and is unique.

Finally we observe that all of the above results hold for the Stable Marriage problem when both the men and the women share a common frame of reference.

Acknowledgement

The authors thank Charles Blair for helpful comments.

References

[1] P. Aranson, *American Government: Strategy and Choice*, Winthrop, Cambridge, MA, 1981.

[2] D. Black, *The Theory of Committees and Elections*, Cambridge University Press, Cambridge, MA, 1958.

[3] K.S. Booth and G.S. Lueker, "Testing for the consecutive 1's property, interval graphs, and graph planarity using PQ-tree algorithms", *J. Comput. System Sci.* **13**, 335–379 (1976).

[4] L.E. Dubins and D. Freedman, "Machiavelli and the Gale-Shapley algorithm", *American Mathematical Monthly* **88**, 485–494 (1981).

[5] P.C. Fishburn, *The Theory of Social Choice*, Princeton University Press, Princeton, NJ, 1973, esp. ch. 9.

[6] P.C. Fishburn, "Discrete mathematics in voting and group choice", *SIAM J., Alg. Disc. Meth.* **5** (2), 263–275 (1984).

[7] D. Gale and L.S. Shapley, "College admissions and the stability of marriage", *American Math. Monthly* **69**, 9–14 (1962).

[8] D. Gale and M. Sotomayor, "Ms. Machiavelli and the stable matching problem", *American Math. Monthly* **92**, 261–268 (1985).

[9] R.W. Irving, "An efficient algorithm for the 'Stable Roommates' problem", *J. Algorithms* **6**, 577–595 (1985).

[10] R.G. Niemi and W.H. Riker, "The choice of voting systems", *Scientific American* **234** (6), 21–27 (1976).

[11] W. Riker and P. Ordeshook, *An Introduction to Positive Political Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

[12] P.D. Straffin, *Topics in the Theory of Voting*, Birkhauser, 1980.