A Recommendation Game

Isaac Swift*

September 22, 2023

Updated version at: Link

Abstract

An agent is engaged in a sequential search problem. However, rather than the options arriving randomly the order is chosen by another player. The agent must use the information that the order in which options are presented is being chosen strategically by someone with potentially conflicting preferences. The model is built off the framework of the classic secretary problem for tractability. In equilibrium, either a low-ball option is presented first followed by increasingly attractive options for an unknown amount of time, the best option is presented first followed by each option getting successively worse, or the ordering is a mix of the two. Applications include recommender systems, bargaining, and countless other economic situations where an agent must choose an ordering of objects.

1 The Secretary Problem

Though not the typical approach, allow my to first introduce the classic secretary problem and its elegant solution. I will then present and motivate my problem which builds off of it. There have been many papers working on various extensions to the famous secretary

*Hong Kong Baptist University, isaacswift@hkbu.edu.hk

problem, also known as the marriage problem or the Sultan's dowry problem. As background information, allow me to briefly discuss the secretary problem and its elegant solution before I introduce my recommendation game that builds off of it.

The secretary problem is an optimal stopping problem where you only observe ordinal rankings of the options. Here is the story. You would like to hire a new secretary. There are N candidates for the job. Your objective is to hire the best of the N candidates. There are two twists that make the problem more difficult. First, you interview the candidates sequentially and must decide whether to give them the job immediately. You interview the first candidate. Then you decide if you want to give them the job or not. If not, then you get to interview the second candidate. If you have moved on you cannot later go back and hire the first candidate that you passed up. They have already started working somewhere else. The second twist is that even after interviewing a candidate, you don't know exactly how good they are. All you can tell when you interview a candidate is their ordinal ranking among candidates you have seen so far. Perhaps you don't really have a good understanding of the job or qualifications. For example, the first candidate comes in and says they can type 50 words per minute. Maybe you aren't really familiar with words per minute. You can't tell if that is fast or not. The second candidate comes in and says they can type 60 words per minute. You still don't really know if that is fast or not, but you can tell that it is better than the first candidate. So each time you see a new candidate, you can tell if they are better or worse than each candidate you have seen so far, but that is all you can tell.

There are papers that remove or change one or both of these rules, but this is the core version of the problem. Sometimes the story is told that each candidate is a relationship you are in and you need to decide whether to marry this person or not. The question is thus trying to find the optimal stopping time rule when you only observe the ordinal rankings of the candidates.

1.1 Solution

While the problem is interesting, just as much of the problem's popularity may be driven by the fact that the solution is so elegant. It is optimal to use a cutoff strategy. You reject the first k candidates automatically. Then you will accept any candidate after that that is the best you have seen so far. The probability of getting the best candidate from this strategy is

$$prob = \frac{1}{N} \sum_{i=k+1}^{N} \frac{k}{i-1}. \tag{1}$$

The best candidate has a $\frac{1}{N}$ chance of being in each position. If that position is $k+1$ or later, you will select it as long as you haven't already selected one. You won't have selected one as long as the best up until that point was in the first k positions. The probability of this is $\frac{k}{i-1}$. For any given N, you can find the value of k to maximize that probability.

As N grows, there probability of any particular candidate being the best is going to zero. It might seem hopeless to win with a non-vanishing probability for large N. If there are 20 million candidates, how can you ever be lucky enough to find the right one? The amazing thing is that the simple optimal strategy remains successful for even very large N. For large N, and analytic solution is possible. Call $x = \frac{k}{N}$, the fraction of candidates initially rejected. As N goes to infinity, this probability converges to a Riemann integral

$$prob \to x \int_{x}^{1} \frac{1}{t}\, dt = -x \log(x) \tag{2}$$

This implies as N grows, the optimal fraction to initially reject, x^*, is approaching $\frac{1}{e}$. Coincidentally, the probability of choosing the best candidate with this simple rule is also asymptotically approaching $\frac{1}{e}$. No matter how many candidates there are, you can win the game about 37% of the time by simply rejecting the first 37% of candidates then accepting once you see a candidate that is the best so far.

2 The Recommendation Game

I will now modify the secretary problem to make it a two player game. Another player (recommender or R) gets to go first and choose the order in which the candidates are presented. The second player (decision maker or DM) then plays just like in the original secretary problem except that they know the order has been chosen by the other player.

The decision maker's problem is very similar in setup to the secretary problem. Each round the DM observes the ordinal ranking of another option and chooses whether to accept or reject. Call h_i^n the ranking of option i among the n options seen so far. The first option is automatically the best so far, so $h_1^1 = 1$, $h_1^2, h_2^2 \in \{1, 2\}$, and so forth. A partial history can be represented $h^n = (h_1^n h_2^n \dots h_n^n)$. The set of total histories for the DM, $\mathcal{H}$, is the set of $N!$ permutations of the numbers $\{1, 2, \dots, N\}$. We must note that the partial histories h^n do not converge to the total history h^N as n goes to N. Each time DM observes another option, you don't simply add another number to the end to get h^{n+1}. This new option is better or worse than every option you've seen. So all the elements of h^n, which are relative ranks, could change. DM's strategy is a mapping from partial histories into a accept or reject decision, $\sigma : h^n \to [0, 1]$, with a fractional value representing a mixed strategy. It is also required that $\sum_{n=1}^{N} \sigma(h^n) = 1$ for all histories, h. The DM must select one and only one option.

Call v_i the payoff to DM of selecting their i^{th} best option overall. In the classic secretary problem, $v_1 = 1$ and $v_i = 0$ for $i > 1$. I will simply assume $v_i \geq v_j$ if $i \geq j$, or that you get a weakly higher payoff for selecting a better option. If this weren't the case, it would call into question what it means for and option to be "better" in the first place. The objective can now be rewritten as

$$\max_{\sigma} \sum_{h \in \mathcal{H}} \sum_{n=1}^{N} \sigma(h^n) v_{h_n} \mu(h). \tag{3}$$

$\mu(h)$ is the probability of each possible permutation. In the classic secretary problem, $\mu(h)$ has a constant value of $\frac{1}{N!}$ because every permutation is equally likely. Now, this probability is controlled by the other player.

This other player is the recommender (R). R observes the N options and chooses an ordering. R sees not only their own preference over the N options, but also knows DM's preference ranking. R's information can also be written as a permutation, $p \in \mathcal{H}$. Denote $p = (p_1 p_2 \dots p_N)$, with p_i being the preference ranking for R of DM's i^{th} best option. So if $p_1 = 3$ it means that DM's number 1 choice is R's number 3 choice. We will denote w_i the payoff to R of their i^{th} best option being selected. The objective can be written as

$$\max_{h(p)\in\mathcal{H}} \sum_{p\in\mathcal{H}} \sum_{n=1}^{N} \sigma_{DM}(h^n) w_{p_{h_n}} f(p) \tag{4}$$

with $f(p)$ being the distribution over possible joint preference rankings, p. If R's preference and DM's preference over options are independent of each other, $f(p)$ would simply be a constant $\frac{1}{N!}$. In general, they may be correlated.

While the problem quickly got notationally ugly, it is still conceptually simple. R chooses an order for the options. DM observes the options sequentially in the chosen order and chooses to accept or reject each option. Both players want to maximize their payoff coming from the option selected. If you understand the game, you can still follow the arguments and results without diving into the necessary notation.

2.1 Motivation

Allow me to now very briefly motivate this problem before we jump into searching for a solution. There are numerous papers on the secretary problem. Two good reviews of the literature are Freeman (1983) and Ferguson (1989). In particular, Gilbert and Mosteller (1966) solve the zero-sum version of my all or nothing recommendation game special case using the cyclic Latin square, and my presentation of the secretary problem a background information largely follows their paper.

Complex recommender systems are now commonly used, but no paper studies them in this type of sequential setting. A recent recommender system paper for reference is Che and Hörner (2018). Amazon, or any online shopping site, has many results that match your search and you may be interested in. The site needs to choose the best order in which to present you with the results. Furthermore, the site's preference may not be perfectly aligned with the customer. The customer would like to find the best deal while the site looks to maximize profit. Similarly, Netflix and the like need to choose an order in which to list different options to stream. In addition to recommender systems, you could imagine this game as a bargaining problem. One party has several offers they could make and needs to choose the order in which to present them. Do you start with your best offer or start with a low offer and progressively get higher? What should the second party's strategy be in regards to accepting or rejecting these offers. While there is a whole literature on bargaining (see classics such as Nash (1950), Rubinstein (1982), and Wilson (1987)), no paper looks at the sequencing of offers in this type of way.

The world is full of economic problems that involve choosing an ordering of options or objects. It could be where to locate goods in your store, how to write a good paper, etc. Attention paid to such problems has been sparse. Part of the reason is that these problems are difficult to solve. The discreteness of the problem makes maximizing more difficult and makes analytic solutions more rare. The size of the problem makes computation difficult. The recommender is choosing a mapping from preferences (cardinality is $N!$) to an ordering (also $N!$). This gives $(N!)^2$ possible actions for one of the players, and the other players action space is also far from small. For even moderate values of N, brute force search for an equilibrium is impossible. I hope that others will find insights in this paper that will help them to solve similar discrete problems.

2.2 Results

I am looking for the Nash equilibria. First I will put bounds on possible equilibrium payoffs. Using those bounds, I characterize when no pure strategy equilibrium exists and construct a mixed strategy equilibrium that exists for all parameter values.

The first bound on payoffs is easy.

Lemma 1. *The DM can guarantee themselves a value of at least* $\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i$.

If DM simply randomized the round in which they accept with equal probabilities, this strategy would guarantee them a payoff of $\bar{v}$ regardless of R's strategy. This typically isn't the optimal strategy, it it assures that DM must receive a payoff of at least $\bar{v}$ in any equilibrium.

The second bound is a little bit trickier to see.

Lemma 2. *R's best response to any pure strategy of DM gives R a value of* w_1.

No matter what pure selection strategy DM is using, there is a way for R to arrange the sequence to completely take advantage. R can always assure their top choice is selected, holding fixed the strategy of DM. The proof can be found in the appendix.

We still don't know if an equilibrium exists in pure strategies, but we have a lower bound on the possible payoffs of each player. The existence issue is going to hinge on the correlation of preferences between R and DM. Call ρ_i the expected value to DM of R's i^{th} best option. While ρ_i as defined is not the a standard correlation coefficient, it captures the necessary features of the distribution of joint preferences, $f(p)$, for understanding equilibria. If R and DM have independent preferences, ρ_i will equal $\bar{v}$ for all i. I will refer to R and DM's preferences as aligned if $\rho_1 > \bar{v}$ and as opposed if $\rho_1 < \bar{v}$ with the word weakly being added if the inequality is weak. The zero-sum version of the game would have $\rho_i = -w_i$ for all i.

Two immediate results come out of the preceding lemma.

Proposition 1. *The payoffs for R and DM in any pure strategy equilibrium must be* (w_1, ρ_1).

Corollary 1. *If preferences are opposed, there cannot be any pure strategy equilibrium.*

The first result is a consequence of Lemma 2 and the second result follows from the first result and Lemma 1.

The key to finding an equilibrium is a strategy for R that will consistently offer the average payoff to DM while revealing no useful information. Suppose R has a strategy where DM can do no better than randomly pick a round to accept. This would guarantee the average payoff for both R and DM. Since DM is accepting randomly, no strategy for R can get a better than average payoff. Both players would have everything in their best response. Thus we would have an equilibrium. Note that R just randomly choosing with equal probability on all orderings does not achieve this objective. We already saw in section 2 that against a random order DM can do significantly better than picking a random round with equal probability. With the zero-one payoffs, a pure random guess gets an expected payoff of $\frac{1}{N}$ while the optimal strategy can get 37% even for very large N.

The strategy that R needs to use is what I will call the maximal risk ordering strategy (MRO). In each round R will randomize between submitting DM's best remaining option with probability q and DM's worst remaining option with probability $1 - q$. Let

$$q = \frac{\tilde{v} - v_{min}}{v_{max} - v_{min}} \tag{5}$$

where $\tilde{v}$, v_{max}, and v_{min} are the average, the best, and the worst of the remaining options respectively. For the first round, $\tilde{v} = \bar{v}$, $v_{max} = v_1$, and $v_{min} = v_N$.

First note that in each round, the expected value of accepting is equal to the average of all remaining values.

$$v_{max}q + v_{min}(1 - q) = \tilde{v} \tag{6}$$

If the payoffs (v_i) were all evenly spaced (order statistics from a uniform distribution), then the q would always be equal to $\frac{1}{2}$. When the payoffs are not evenly spaced the value of q is adjusting to that randomizing over the best and worst remaining option is the same in expectation as randomizing over every remaining option.

Second, see that this ordering strategy reveals no "useful" information to DM. Information is revealed each round. In fact, in round n DM knows that exact overall ranking of each of the $n-1$ options that came before. However, this order ensures that none of this information is useful in distinguishing the current option from any other remaining option. Among the previous options, some were the highest and some were the lowest. All remaining options are going to in between the high options that have been seen and the low options that have been seen. Thus, the remaining options are indistinguishable. At every point the remaining options are all grouped together in the ranking. Nothing from the middle of them has been observed to separate them.

This turns out to be the critical condition.

Theorem 1. *It is always an equilibrium for R to play the maximal risk ordering and for DM to randomize with equal probability of which round to accept.*

I will refer to this as equilibrium A.

2.3 Zero-Sum Game

I have constructed an equilibrium to the game that always exists. However, there is no guarantee of uniqueness. In the zero-sum special case of the game, at least the payoff is unique. The zero-sum case is when the distribution $f(p)$ is degenerate on $(N, N-1, \ldots, 2, 1)$ and $v_i = -w_{N-i+1}$. On other words, the players have a completely opposite preference ranking and the magnitudes are the same.

Now equilibrium A gives the unique equilibrium payoffs $(-\bar{v}, \bar{v})$. We already know from Lemma 1 that DM must get a payoff of at least $\bar{v}$ in every equilibrium. Furthermore, we can see from the construction of the maximal risk ordering that R can keep DM's payoff from going any higher than $\bar{v}$. Since the game is zero-sum, these two results put an upper and lower bound on DM's payoffs of $\bar{v}$. This payoff is obtained by equilibrium A.

3 All or Nothing

We have found an equilibrium of the game, but we have no assurance that it is the only equilibrium or that it is the best equilibrium in any sense of the word. Let us now turn our attention to a special case of the game in line with the classic secretary problem where we can determine much more. Suppose now that the game is all or nothing. Each player gets a payoff of one if their top choice is chosen and zero otherwise. That is, $v_1 = w_1 = 1$ and $v_i = w_i = 0$ for all $i \neq 1$. Keep in mind that DM's number one choice and R's number one choice are not necessarily the same option.

There is now a large amount of indifference. Every option except the top choice is equally good. Note that I'm not saying these options are indistinguishable, only that they give the same payoff. This indifference is going to give us a lot of freedom to modify the maximal risk ordering and still have an equilibrium. Since payoffs are only zero or one, in round n $\tilde{v}$ is equal to $\frac{1}{N-n+1}$ if the top choice is among the remaining options and 0 if it is not. This means that the maximal risk ordering would present the top remaining option with probability $\frac{1}{N-n+1}$ and the bottom remaining option with probability $\frac{N-n}{N-n+1}$. I will now construct a new equilibrium strategy that doesn't present the bottom choice with probability $\frac{N-n}{N-n+1}$ but rather just presents one of the choices that give a payoff of zero with that probability. The key to this equilibrium strategy is the following cyclic Latin square.

$$\begin{array}{ccccc} 5 & 4 & 3 & 2 & 1 \\ 4 & 3 & 2 & 1 & 5 \\ 3 & 2 & 1 & 5 & 4 \\ 2 & 1 & 5 & 4 & 3 \\ 1 & 5 & 4 & 3 & 2 \end{array}$$

Each row of the square is a sequence of numbers that decrease by one, until reaching 1, then they reset at N. Think of these numbers as the rank ordering of options for DM. The options are in increasing order. Each option is better than the previous one, except that it wraps around after the best one. Each row represents one such sequence of options with a different starting point.

Imagine that the sequence of options presented was a randomly selected row of the cyclic Latin square. DM cannot get a score higher than $\frac{1}{N}$ with any selection strategy. This is because no useful information comes until it is too late. All you see is that every option is the best one so far, until the point when you've passed the top option. At any point in time, the best option is equally likely to be the current option or any one of the future options. We can use backward induction to see why this means there is no profitable strategy or just state that it follows from Lemma 3.

We now have another equilibrium where R chooses a random row from the cyclic Latin square and DM random selects which round to accept. I will call this equilibrium B. This new equilibrium isn't very different from the one we already constructed in the general case, but we can expand it to look at all the interesting equililbria in this special case. Particularly, I would like to find the best equilibrium for R. The best equilibrium is going to be dependent of the correlation of preferences between R and DM. Take first the case where R and DM have weakly aligned preferences ($\rho \geq \frac{1}{N}$). In fact, we will need a slightly stronger assumption. Call μ the distribution over DM's ranking ($\{1, 2, \ldots, N\}$) of R's top choice. Let us assume that $\mu_1 \geq \mu_i$ for all $i \in \{1, 2, \ldots, N\}$. This implies that preferences are weakly aligned. The assumptions is that R's top choice is at least as likely to be DM's top choice as any other ranking. The case of independent preferences is still allowed by the assumption.

With this assumption, the best equilibrium for R is easy to find. Rather than randomly selecting a row from the cyclic Latin square, R can select the row that begins with their own most preferred option. This first option is at least as likely to be DM's top choice as any other choice. Thus, DM gets a payoff of at least $\frac{1}{N}$ from accepting the first option. If DM rejects the first option, they will simply be facing a random row of the cyclic Latin square. They will be able to do no better than the average of the remaining options, which is less than or equal to $\frac{1}{N}$. Thus is is an equilibrium for R to choose the row of the cyclic Latin square that begins with their own top choice and for DM to accept the first option presented.

The payoffs in this equilibrium are $(1, \rho_1)$.

3.1 Opposing Preferences

Consider now the case where preferences are opposed. Equilibrium B is still an equilibrium (as is equilibrium A). The payoffs to the two players in equilibrium B is $(\frac{1}{N}, \frac{1}{N})$ because it is completely random which option ends up being selected. The important thing to notice is that the payoffs to both players are decreasing in the number of available options, N. Both players would be made better off if there were fewer options. Imagine they could just throw out one option that isn't either player's top choice, then play the same equilibrium with the remaining options. Everyone would be happy with this.

In fact, the best equilibrium for either player in this game only involves two options. R will randomize with equal probability over putting their own top choice first and DM's top choice second or putting DM's top choice first and their own top choice second. The order of all options after the second is unimportant. DM will randomize with equal probability over accepting the first or second option. This is an equilibrium. As DM's top choice is among the first two options and could equally likely be either one, any strategy that puts probability one on selecting one of the first two is in the best response. If DM is randomizing equally of selecting one of the first two, and strategy for R that puts their own top choice in one of those first two spots is in their best response.

This equilibrium gives a payoff of $\frac{1}{2}$ to both players. Neither player is going to get higher that $\frac{1}{2}$ in any equilibrium with preferences are opposed. To get hire would require DM to accept with a probability above $\frac{1}{2}$ under some condition. Following the logic of Lemma 2, R can put their own top choice in this spot. Accepting in this spot now gives DM a payoff of ρ_1 which is below average. Thus, it is contradictory for DM to accept in this spot with any probability. Interestingly, the payoff that DM receives in R's preferred equilibrium with opposed preferences ($\frac{1}{2}$) is higher than DM's payoff in R's preferred equilibrium with independent or slightly aligned preferences (ρ_1).

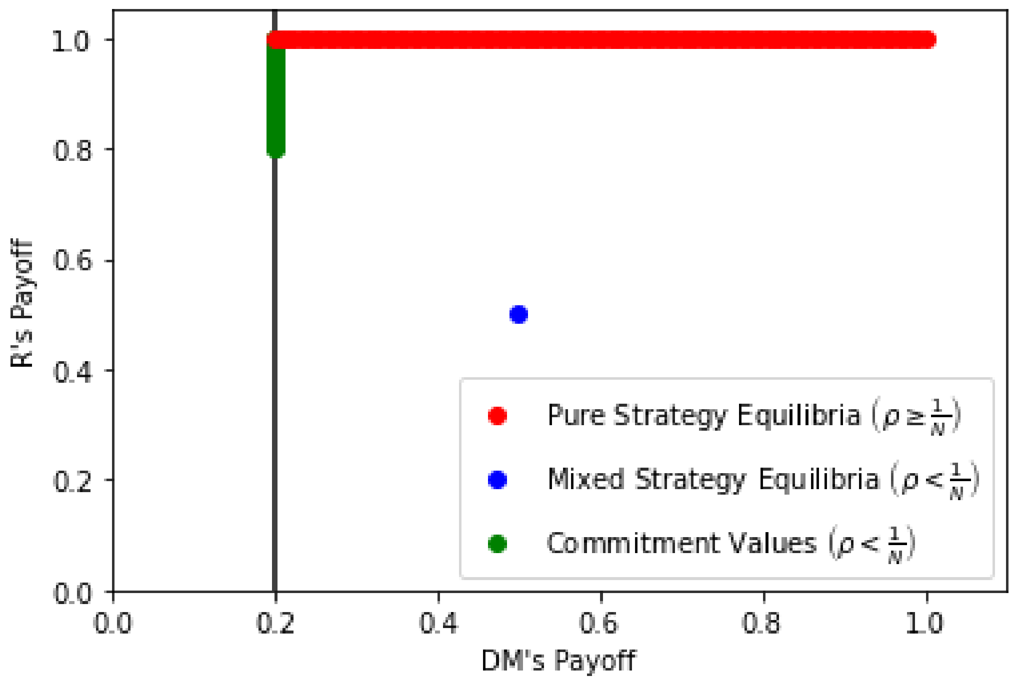

Figure 1: This graph shows the payoffs in R's preferred equilibrium. The horizontal red line on top is when preferences are positively correlated. The blue dot in the center is when preferences are negatively correlated. The vertical green line in the upper left is when preferences are negatively correlated and R has commitment power.

One final note in this section is that the equilibrium is still not efficient. Each player is getting their top choice with probability $\frac{1}{2}$. However, as long as we aren't in the zero-sum case, there is still some probability that the players agree on the top choice. Thus, in an efficient allocation these probabilities would add up to more than one. There are times when both players agree on the top choice, it gets placed in one of the first two spots, but DM chooses the other of the first two spots. Herein lies the inefficiency.

3.2 Commitment

In this section I consider how the game would change if R could commit to an ordering strategy and find their best commitment equilibrium. In the weakly aligned preference case described above, R is already getting their top choice and thus gets no benefit from being able to commit. So, let us suppose that preferences are opposed. The best equilibrium without commitment gives payoffs of $\frac{1}{2}$ to both players. With commitment R will do much better, but this will come mostly at the expense of DM.

R has the threat of the cyclic Latin square which limits DM to $\frac{1}{N}$ and can start with any option. Using this threat, R can induce any distribution of options that gives DM an expected payoff of at least $\frac{1}{N}$. The best such distribution is as follows. With probability $\frac{1}{N}$, select DM's top choice. With probability $\frac{N-1}{N}$ select R's top choice. Play the row of the cyclic Latin square that begins with the selected option. DM can do no better than to accept the first option. As long as the $\frac{1}{N}$ fraction of the time DM's top choice is selected is done in a way that contains the ρ_1 fraction of the time the two players agree on a top choice, this is the best distribution of options R can induce. The payoff to R is $\frac{N-1}{N} + \rho_1$.

4 Feeling Lucky?

When you go go Google's website and type your search query into the box, there are two buttons you can click. One says "Google Search", with will present an ordered list of search results similar to the model presented here. The other button you can click says "I'm Feeling Lucky." This but will take you directly to one of the search results rather than presenting you with the list. Similarly Netfilx has a "Surprise Me" button you can click on start-up. It will instantly start playing something rather than presenting options for you to pick from.

The beauty of these buttons is that they transfer all the great equilibria in the all-or-nothing game into the general game. Consider the general game but at an additional stage at the beginning. First R selects one option to be the "Lucky" option. Then, DM can choose take the Lucky option (of which they have no information) or just play the regular recommendation game.

4.1 Results

Consider the case of positively correlated preferences. It is an equilibrium for R to put their most preferred option in the Lucky position and for DM to take the Lucky option. The Lucky option gives a payoff of ρ_1 while playing the regular game would only give a payoff of $\bar{v}$. As long as payoffs are positively correlated, DM prefers this Lucky option.

It's the same idea as in the all-or-nothing game. In the all-or-nothing game R could signal their own top choice by putting it first without messing up the equilibrium ordering. The equilibrium ordering already has any option showing up first with some probability. In the general game this isn't possible. To get this good equilibrium payoff, R would need to signal their own top choice somehow without giving away information. The Maximal Risk Ordering does allow this. The MRO equilibrium works because it only presents the most extreme options that don't reveal any information about other options. If R put their own top choice first and it isn't one of the extreme options, DM can reject it then use what they've observed to distinguish the high and low options in the next round. The Lucky spot allows R to signal their top choice without giving away any additional information.

However, the Lucky spot is not of any use when the players have negatively aligned preferences. If DM accepts the Lucky spot with any probability, R's best response is to put their own top choice in that spot. R's top choice gives DM a payoff of ρ_1, which is worse that randomly guessing. Thus, DM cannot ever accept the Lucky option.

4.2 Commitment

If R is able to commit ex ante to a strategy, then the Luck spot becomes valuable even with negatively correlated preferences. R can commit to any distribution of options in the Lucky spot that give DM at least $\bar{v}$. The best of these distributions will give R a payoff at least as large as $\bar{w}$. $\bar{w}$ is obtained by randomizing over each option equally which is one of the allowable distributions here. As long as the preferences aren't perfectly opposite of each other, the best distribution is going to be strictly better.

References

Che, Y.-K. and Hörner, J. (2018). Recommender systems as mechanisms for social learning. *The Quarterly Journal of Economics*, 133(2):871–925.

Ferguson, T. S. (1989). Who solved the secretary problem? *Statistical Science*, 4(3):282–289.

Freeman, P. R. (1983). The secretary problem and its extensions: A review. *International Statistical Review*, 51(2):189–206.

Gilbert, J. P. and Mosteller, F. (1966). Recognizing the maximum of a sequence. *Journal of the American Statistical Association*, 61(313):35–73.

Nash, J. (1950). The bargaining problem. *Econometrica*, 50:155–162.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109.

Wilson, R. (1987). Game theoretic analyses of trading processes. *Advances in Economic Theory*, pages 33–70.

Appendix

Additional details of proofs are here.

Lemma 2: First, see that for any strategy of DM, there is always an ordering where DM gets stuck with their worst option. Suppose that in DM's strategy, at some partial history DM accepts an option that is the worst option they've seen so far. If this is the case, there is an ordering where they get their worst overall option. Take an ordering that generated that history. If that option they accept is already the worst overall, you're done. Otherwise, swap the option being accepted (worst so far) with the worst overall option (must be later in the sequence). We now have another valid sequence. Furthermore, this sequence is indistinguishable from the original (generates the same partial history). Thus, DM must accept the worst option overall. Suppose now that DM's strategy would never accept an option that is the worst so far. Consider the ordering of options from best to worst. At every partial history, the current option is always the worst one seen so far. DM cannot accept any of them. DM will then get stuck with the last option which is the worst overall.

Next see that for any strategy, DM can always get stuck with the second (or i^{th}) worst option. This time split the possible strategies into three ($i+1$) categories: DM accepts an option that is the worst so far, DM accepts and option that is the second worst so far, or DM never does either of those (DM accepts and option that is the j^{th} worst so far with $j \leq i$ in round $N-j+1$ or earlier). In the first two (i) cases, you can rearrange the sequence to make the selected option the second (i^{th}) worst overall. In the last case you have the options ordered from best to worst but with the last two options switched (the order reversed on the last i options). Then DM will be stuck with the last option, which is the second (i^{th}) worst overall.

As we can see, if R's top choice is DM's i^{th} best option and DM is playing a pure strategy, R can always find an ordering to guarantee it is chosen.