Towards an Optimal Dating Strategy:
Square Root or Linear Trial Size?

Ian Eykamp

January 2023

1 Introduction

In his book *Things to Make and Do in the Fourth Dimension*, Matt Parker suggests that the optimal strategy for finding one's true love among n romantic candidates (assuming each candidate can be quantitatively scored and ranked) is to first "interview" and reject the first $\sqrt{n}$ candidates as trials, and then propose to the next romantic candidate who exceeds the highest trial score. Read more here.

Parker's assertion has been popping into my head occasionally for at least the last four years; finally, I have the discrete math skills to prove him right — or wrong.

2 Solution

The process is best explained through an example. Let us assume our candidate pool is 16 people, each with a rank from 1 to 16. In Parker's terms, this is the number of romantic candidates you can expect to date over your lifetime. According to Parker, the optimal trial size should be $\sqrt{16}$ or 4. Let us also assume that the candidates are lined up in some unknown order such that the order in which we will meet them is predetermined, with the first four being the trials. Let us now calculate the probability that this strategy will result in us picking the #1 ranked candidate to be our true love.

If the #1 ranked candidate is among the first 4 trial candidates, then our search will fail, because we will reject them and never find a better candidate. There is a $\frac{4}{16}$ chance of this happening, because the #1 ranked candidate is equally likely to be in any of the 16 positions, including among the four trials.

If the #1 ranked candidate is not among the trials and the #2 ranked candidate is among the trials, then the search will certainly be successful. This is because we will reject any non-trial candidate who is not better than #2, leaving only the #1 candidate.

There is a $\frac{12}{16} \cdot \frac{4}{15}$ chance of this happening, because the #1 candidate must not be among the trials ($\frac{12}{16}$), meaning they take up one of the non-trial slots; then, the #2 candidate must be located in one of the four trial slots out of the 15 remaining slots ($\frac{4}{15}$).

If the #1 and #2 candidates are not among the trials, the search will be successful if the #3 ranked candidate is among the trials *and* the #1 candidate comes before the #2 candidate in the non-trial ordered list. This is because we will reject any candidate who is not better than #3, and leaving #1 and #2 as viable candidates. Whichever of these we encounter first, we will choose as our true love. If this is the case, there is a $\frac{1}{2}$ chance that we will choose the #1 candidate.

There is a $\frac{12}{16} \cdot \frac{11}{15} \cdot \frac{4}{14} \cdot \frac{1}{2}$ chance of this happening, because candidates #1 and #2 must not be among the trial candidates ($\frac{12}{16} \cdot \frac{11}{15}$), the #3 candidate must be placed in one of the four trial slots among the 14 remaining slots ($\frac{4}{14}$), and the #1 candidate must come before the #2 candidate among the non-trials ($\frac{1}{2}$).

Each successive case is conditional on the previous case not being satisfied (hence the expression for the probability of each successive case resembles the previous case with one extra term). This conditional structure can be visualized using a tree, as in Figure 1.

A pattern is now apparent. Suppose candidates #1 through $\#k$ are not among the trials, but candidate $\#k+1$ is among the trials. Then the search will be successful if and only if the non-trial candidates are arranged such that the #1 candidate comes first among the k worthy candidates, which has a $\frac{1}{k}$ chance of success.

Following the previous pattern, this has a chance of happening equal to

$$\left[\frac{12}{16} \cdot \frac{11}{15} \cdot \frac{10}{14} \cdot \ldots \cdot \frac{12-k+1}{16-k+1}\right] \cdot \frac{4}{16-k} \cdot \frac{1}{k}. \tag{1}$$

In Equation 1, the bracketed expression represents the probability that candidates #1 through $\#k$ are not among the trials; the second term represents the probability that candidate $\#k+1$ is among the 4 trials out of $16-k$ remaining slots; and the last term represents the probability that the #1 candidate comes first among the k worthy non-trial candidates.

Since the above cases are mutually exclusive and exhaustive, the probability of finding your true love $P(\heartsuit)$ is found by summing the probabilities for each case. There are only 12 cases, because the #13 candidate must be one of the trials if candidates #1 through #12 are not among the trials. This ends the tree.

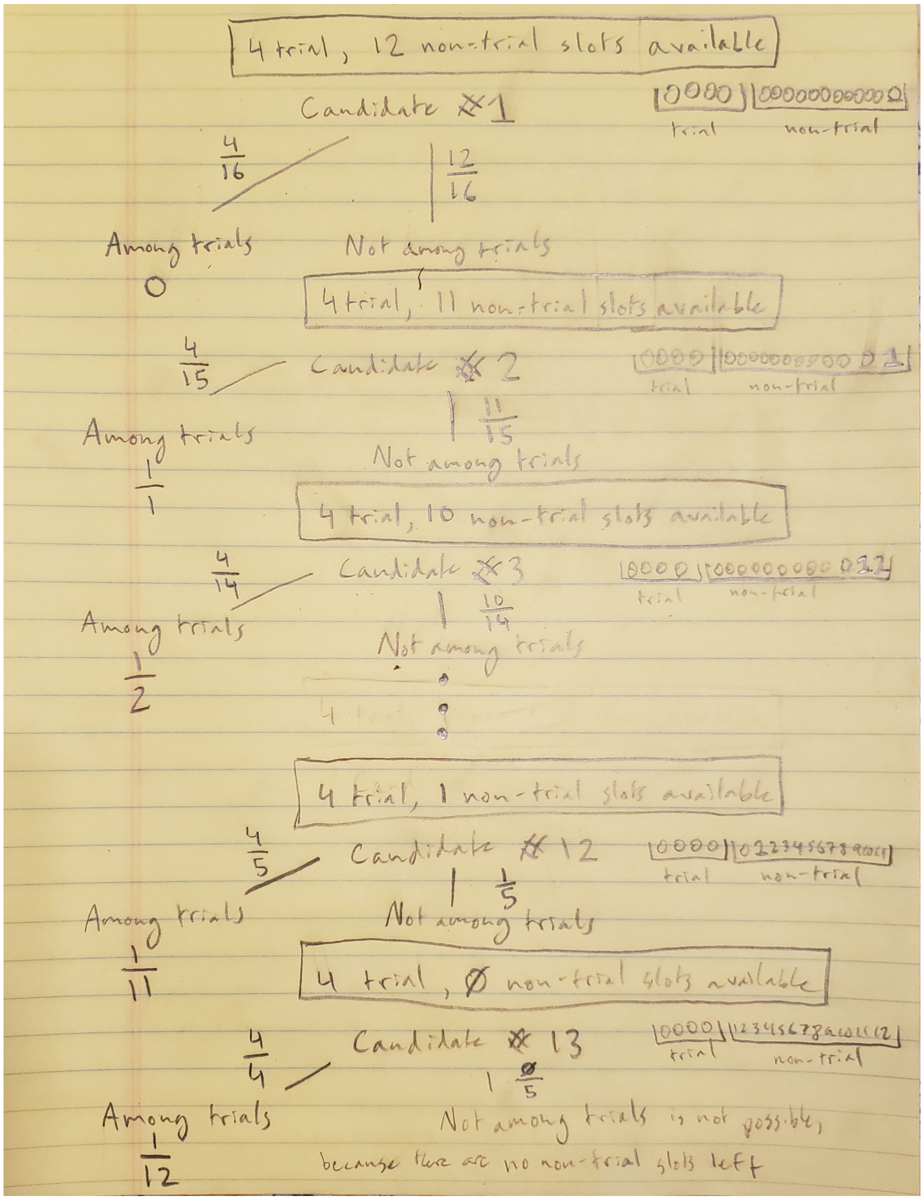

Figure 1: Tree diagram representing the probability of each case based on the previous cases. The trunk of the tree runs vertically, starting with the #1 ranked candidate; if they are not among the trials, then the #2 ranked candidate is considered; if they are also not among the trials, the #3 ranked candidate is considered, and so on. The branches to the left of each case represent the probability of success if that candidate is among the trials. The probabilities for each case are multiplied along the branches of the tree. The cases are mutually exclusive and exhaustive.

$$P(\heartsuit) = \sum_{k=1}^{12} \left[\frac{12}{16} \cdot \frac{11}{15} \cdot \frac{10}{14} \cdot \ldots \cdot \frac{12-k+1}{16-k+1} \right] \cdot \frac{4}{16-k} \cdot \frac{1}{k}. \tag{2}$$

Or, more concisely:

$$P(\heartsuit) = \sum_{k=1}^{12} \frac{12!}{(12-k)!} \cdot \frac{(16-k-1)!}{16!} \cdot \frac{4}{k}. \tag{3}$$

Generalizing to an arbitrary size of candidate pool *total* and an arbitrary trial size *trial*, yields:

$$P(\heartsuit) = \sum_{k=1}^{total-trial} \frac{(total-trial)!}{(total-trial-k)!} \cdot \frac{(total-k-1)!}{total!} \cdot \frac{trial}{k}. \tag{4}$$

3 Optimizing Trial Size

The question remains: for a given pool size of expected candidates, what is the optimal trial size to produce the highest chance of finding the best candidate? For this, I turned to MATLAB. The code is available on GitHub.

Figure 2 shows a plot of $P(\heartsuit)$ versus total candidate pool size and trial size. The top projection uses the color scale to represent $P(\heartsuit)$. It appears that $P(\heartsuit)$ is maximized along a narrow band that is a linear function of the total number of candidates, with a slope of approximately 0.4. The other projections attempt to show the shape of the 3D surface, including its curvature on either side of the optimal line and its harmonic cross section along the $trials = 0$ plane. (This is expected, because if there are no trials, we will simply select the first candidate, which has a $\frac{1}{total}$ probability of success). The maximum value of $P(\heartsuit)$ is achieved in the case with 2 candidates and either 0 or 1 trial, in which case $P(\heartsuit) = 0.5$. The trivial case with only one candidate is omitted. The probability of success decreases asymptotically towards ≈ 0.3710 with increasing pool size.

For each candidate pool size, the trial size with the greatest $P(\heartsuit)$ is found. The optimal trial size is plotted against the pool size in Figure 3. It is clear from the figure that the optimal trial size is in fact a linear function of pool size, and not a square-root relationship as Parker suggested. The slope of the line of best fit is ≈ 0.3674 (which is curiously close to $\frac{1}{e}$). This suggests that if you want to have the best shot at finding your true love, you should "interview" 37% of your total possible candidates, and then settle down with the next candidate who exceeds the best candidate from the trial round. This strategy should yield no worse than a 37.1% success rate for candidate pool sizes smaller than 100.

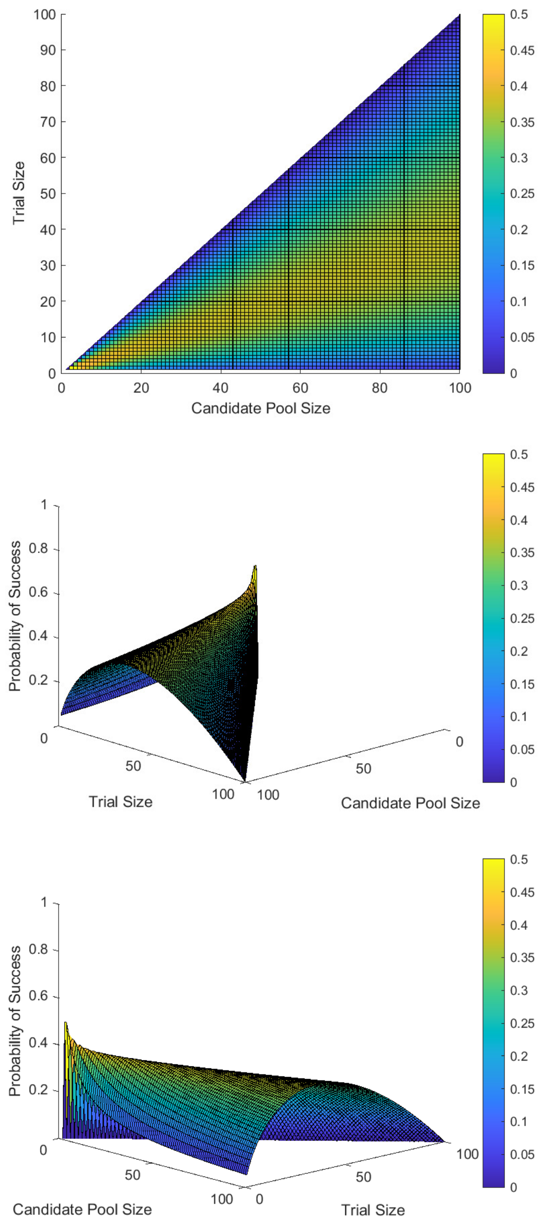

Figure 2: Projections of the 3D plot of $P(\heartsuit)$ versus total candidate pool size and trial size.

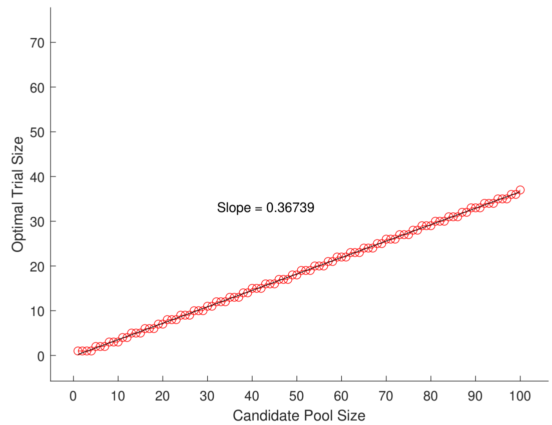

Figure 3: Optimal trial size to maximize $P(\heartsuit)$ versus total candidate pool size. The optimal trial size is a linear function of candidate pool size, with a slope of $0.3674 \approx \frac{1}{e}$.

4 Conclusion

The article linked to in the Introduction explains why my result differs from Matt Parker's. In this paper, the only metric of success is $P(\heartsuit)$, the probability that the #1 candidate will be selected. However, it is likely that the #2 candidate is really not that bad either. The square root rule applies when the metric of success is some function of the rank of the candidate that you ultimately end up with; even if it yields a slightly lower chance of finding your true love, it does a better job of ensuring you will at least end up with someone tolerable.