# Sequential Search and Weak Signals

Jeong Hyun Lee[*†]

April 2023

## Abstract

Our everyday decisions are often based on incomplete information on alternatives, which we use to formulate our attention. In this study, we examine a specific case under the sequential search model, where the prior information given to the decision maker consists of a weak signal on the true qualities of the alternatives. Under our framework, the results extend to a search scenario where the best and worst alternatives are not that differentiable. Extending Choi and Smith (2022)'s two factor sequential search model, we create a simulation to estimate this phenomenon and find that in fact, signals regarding the search order of the alternatives are more pertinent to the optimal choice that the decision maker makes than the signals concerning their qualities. In addition, we simulate a search environment where we observe the effects of dispersing the observed and unobserved factors under the Choi and Smith model and indeed confirm the results of their paper extends under our framework.

## 1 Introduction

As decision makers face larger and more complex choice sets each day, more decisions are based on the limited information the decision maker perceives regarding his alternatives. Especially with the overflow of information given by digital means, it is often hard for the decision maker to accurately access the true quality of an object given some prior signal about it. That is, the signals, easy-to-access information, regarding alternatives may not be useful in the decision maker's optimization problem.

For instance, suppose a simple Google search looking for the best online image modification service, where the decision maker is looking to change the background of an image. Suppose Google returns a search result that orders websites not necessary in the decision maker's best interest, that is, some online image services that focus on resizing the image and have limited capabilities in changing backgrounds are listed on the first page. However, as it takes time and effort to open a link and examine the functionalities of each service, the decision maker may simply choose one of the results given on the first page and work with it. That is, the decision maker may make his choice not from finding the best service, but because it was listed on the first page.

---

Our model starts from the satisficing model introduced by Herbert Simon (1955), where contrasting to the Standard Model, the decision maker selects his alternatives upon observing an alternative that is above some 'satisficing' level of utility. Hence, in terms of searching through alternatives, it alludes to an item-by-item search, or in as many studies incorporate, a sequential search. Weitzman (1979) formalizes a commonly discussed sequential search model, which we will use as a basis for our theoretical framework.

In the Pandora's box problem that Weitzman describes, the decision maker is given a collection of 'boxes,' each with a random payoff under a known distribution. Each box has some look-to-see cost to open and realize its payoff value. The decision maker at any point can choose to quit for an outside option, open the next box, or recall and select an already opened box. The decision maker wishes to maximize his expected profit, and Weitzman shows that the optimal strategy considers the boxes in decreasing order of the value in which opening it has exactly zero marginal revenue. Let this value be the *reservation value* of the box. The decision maker thus stops the search when either there are no more boxes to open or the maximum value of an opened box is greater than the reservation value of the next box.

In this paper, we choose a slightly modified version of the above model, designed by Choi and Smith (2022), where the value of each box is the sum of a known and a hidden factor. Our decision maker observes the known factor without cost and will pay the information cost to access the value of the hidden factor. Thus, the existence of the known factor allows for the decision maker to receive some signal about the alternatives before the search process begins. We argue, formalizing the Google search example, that the true quality of each search result is not as important as the signal of alternatives the decision maker receives. It will be the case that even if there are minimal distinctions between the best and worst quality alternatives, the decision maker's choice will largely depend on the order in which the alternatives are given, which will be based on the signals.

We design a simulation to show that under the optimal strategy, regardless of the distribution of the unobserved factor, the decision maker will choose from the earlier, if not, the earliest, boxes opened, implying the optimal choice will be determined by the search order, hence the distribution of the signals. Furthermore, we investigate how the model behaves under strict testing conditions where we only vary the dispersion of eithe the observed or the hidden factor. The main findings from this simulation are in line, results of Choi and Smith (2022). Especially, we find that a more dispersed distribution of signals (the observed factor) shortens search and concentrates the optimal choice to the first two alternatives in the search order. In fact, only the dispersion of the observed factor gives some influence in the optimal choice of the decision maker while the unobserved factor gives only a slight inconsistent trend towards this effect.

The paper is organized into six sections. The next section will discuss related literature regarding sequential search and the Pandora's box problem. Section 3 will describe the original Weitzman model, the modified model based on Choi and Smith (2022) that we will incorporate, and the simulation setup based on the previous model. Section 4 will report the results of the simulation outlined in the previous section and Section 5 will further discuss the results in regard to the goals of this study. Section 6 will remark on some implications of the results and Section 7 will conclude.

# 2 Related Literature

As mentioned above, the starting point for our study is Weitzman (1979)'s paper. Our work relates in particular to the discourse of sequential search and its applications in the field of economics.

In a series of literature regarding prior information in sequential search, a paper by Robert and Stahl (1993) explores the role and implications of price advertising under a sequential search model. In this model, the firm's level of advertising is incorporated as the consumer's prior on prices. Choi et al. (2018) show that in an oligopoly where consumers engage in a sequential search model based on partial product information and advertised prices, there exists a unique equilibrium of consumers' shopping outcomes and formalizes this into a discrete-choice problem. In applications, Chu et al. (2017) incorporate Weitzman's model to describe online consumer purchasing behavior.

Our paper also relates to consumer search and discovery, where Ursu (2015), using a sequential search model, examines the causal effects of rankings in online platforms. In line with this topic, Armstrong (2017) studies consumer behavior under different search orders. The model shows how the search can be reformulated as a simpler discrete choice problem without search frictions where the firm with the highest realized effective value is selected. In a slightly more theoretical perspective, Greminger (2022) studies the consumer search problem under the Weitzman model where the decision maker begins with a pool of already known alternatives (boxes already opened). Under this framework, the problem is still tractable, with the same process of solving through reservation values as in Weitzman (1979). Furthermore, in line with our work, under the optimal search policy, a fully predetermined index will determine the choice of the decision maker. The use of signals may imply a framing effect, and Caplin and Martin (2012) create an optimizing model that accounts for framing. Their model incorporates (and does not) various aspects of bounded rationality, and shows in fact, consistency with NIAS.

In expanding the Pandora's box problem, perhaps the most pertinent paper to our study is the aforementioned work by Choi and Smith (2022). They implement the two factor search model and show a variety of results regarding this new model, in which the most important to our work is that search duration increases with a more dispersed payoff. However, where they use the perspective of an outside observer oblivious to both the known and hidden factors, our model will remain in the perspective of the decision maker. To study the problem in a more generalized setting, Olszewski and Weber (2015) examine the model where optimization is over all the discovered prizes. They show that under a threshold-based rule, the general problem is simplified into the regular Pandora's box problem. In a different generalized setting, Doval (2018) introduces a model where the decision maker may select a box without inspecting it. The model identifies sufficient conditions for which Weitzman's search rule still holds, yet under a different stopping condition. In a separate study, Boodaghians et al. (2020) study a model of sequential search with a constraint on the order in which the boxes are opened. They show that even with such constraints, like the original Weitzman model of 1979, a greedy optimal strategy exists which can be efficiently computed for tree-like order constraints and can still be described as a threshold-based rule.

# 3 The Model

## 3.1 Weitzman (1979)

As a precursor, we first briefly describe Weitzman's original Pandora's box problem. The decision maker faces a collection of $N$ boxes. Each box $i \in \{1, \ldots, N\}$ will contain prize $x_i$, which is distributed under $F_i(x_i)$, with each $F_i$ known and independent of each other. Also at known cost $c_i$, the decision maker can open box $i$ and observe the realized value $x_i$. The decision maker can open the boxes in any order, and stop at any point. The decision maker wishes to maximize his expected profit, that is, choose the highest value box without incurring a high cost. Formalizing, the optimization problem, if $S$ is the subset of opened boxes, is to maximize the following payoff.

$$P = \max_{i \in S} x_i - \sum_{i \in S} c_i \tag{1}$$

The problem has two aspects, the optimal order to search through the boxes and a stopping rule. Weitzman proposed the optimal strategy using *reservation values.* The reservation value $z_i^*$ of each unopened box $i$ is the value for which the decision maker will be indifferent between taking a prize with that value and opening box $i$. To calculate $z_i^*$, Weitzman supposes the following hypothetical: the expected profit from opening box $i$. The decision maker must incur a cost $c_i$, and the decision maker will choose the opened box $i$ if $x_i > z_i^*$, otherwise, choose the outside option, $z_i^*$. Hence, the expected profit.

$$z_i^* \int_{-\infty}^{z_i^*} F_i(x_i) dx_i + \int_{z_i^*}^{\infty} x_i F_i(x_i) dx_i - c_i \tag{2}$$

Hence, $z_i^*$ is the value that satisfies $z_i^* = (2)$.

Weitzman states that the decision maker, under the optimal strategy, will open the boxes in decreasing order of the reservation value of each box. The decision maker will then terminate search once the opened prize exceeds the reservation value of the rest of the unopened boxes, $x_i > z_{i+1}^*$. Afterwards, the decision maker will choose the highest valued opened box.

## 3.2 The Two Factor Model

Our model setup receives a slight modification from Choi and Smith (2022), which introduces the two factor model. From the Weitzman model, we include an observed and unobserved factor, $w_i$ and $v_i$ respectively, to the payoff of each box.

$$x_i = w_i + v_i \tag{3}$$

Both $w_i$ and $v_i$ come from known distributions $W_i$ and $V_i$ and the cost to open $c_i$ is same as before. Not to a surprise, with this modification, the searching algorithm requires a slight modification. The reservation value is now calculated with the observed factors, $\{w_i\}$.

$$z_i^* = z_i^* \int_{-\infty}^{z_i} V_i(z_i - w_i) dz_i + \int_{z_i^*}^{\infty} z_i V_i(z_i - w_i) dz_i - c_i \tag{4}$$

The searching and stopping rule is therefore equivalent to Weitzman as proved by Choi and Smith (2022).

4

## 3.3    Simulation

Under this model, we conducted a computerized simulation to study the effects of decreasing costs and the spread of payoffs on the optimal choice.

To simplify the simulation, we fixed the cost of opening each box and fixed a Normal distribution for the observed and unobserved factors across all boxes. Hence, there was a singular cost $c_i = c$, and distributions $w_i \sim N(30, \sigma_w^2)$ and $v_i \sim N(0, \sigma_v^2)$ for each factor. Furthermore, we conducted each simulation condition across six different cost values, $\{0.01, 0.3, 0.6, 0.9, 1.2, 1.5\}$, where we added 0.01 to approximate the case of zero cost, and the remaining costs were selected to be varied from one to five percent of the mean payoff.

The simulation consisted of two groups of treatments. In the first group, we conducted treatments varying either $\sigma_w$ or $\sigma_v$. Hence, holding $\sigma_v = 3$, there were 30 treatments varied across five different $\sigma_w$ values, $\{1, 2, 3, 4, 5\}$, and six different costs, and another 30 treatments holding $\sigma_w = 3$, and varying $\sigma_v$ across $\{1, 2, 3, 4, 5\}$, for a total of 60 treatments.

The second group of treatments varied the ratio $\sigma_v/\sigma_w$. Hence, for different reference values of $\sigma_w$, selected from $\{0.5, 1, 2, 3\}$, we examined six different ratios, $\{0.5, 1, 2, 3, 4, 5\}$, for a total of 144 different treatments. The purpose of these treatments was to bring a controlled environment where we vary the dispersion of both the observed and the unobserved factor. We wished to explore the optimal choice under varying relation between the dispersion of the observed and the unobserved factor.

Given the parameters of the treatment, the simulation ran 1000 trials each with 10 boxes in each trial. The simulation followed the below process, with the detailed code attached in appendix A.

1. Create 10 observed values for each box.

2. Calculate the reservation value for each box.

3. Open each box in the order of decreasing reservation value.

4. Stop the search once opened value is greater than the next reservation value, or if opening the next box leads to negative payoff.

5. Choose box with the highest opened value and save the data.

The simulation code was written and run in a Python engine version 3.10.x, with the data saved as a Python Pickle file.

# 4    Choice Probability Results

## 4.1    Treatment Group 1

Table 1 reports a summary of the simulation results when we fix either variance. For both panels, column (1) reports the choice probability of the first alternative and also the cumulative choice probability of the first three alternatives inspected. In column (2), we report the average differences in the true value (observed + unobserved) between the worst and best alternative for each treatment. This is to estimate if the alternatives are in fact, similar in quality. As expected, most of our choice probabilities across all treatments were concentrated toward the earlier alternatives inspected. That is, in

most cases, the optimal choices were made from the first few alternatives that were inspected, and this will be discussed further in Section 5.

Table 1: Summary of Choice Probabilities for Treatment Group 1

| Cost | | (a) Fixed $\sigma_v = 3$ | | | | (b) Fixed $\sigma_w = 3$ | | |
| | $\sigma_w$ | (1) Choice Prob.[1] | | (2) Mean Diff. | $\sigma_v$ | (1) Choice Prob.[2] | | (2) Mean Diff. |
| | | First | Top 3 | | | First | Top 3 | |
| 0.01 | 1 | 0.305 | 0.631 | 6.1853 | 1 | 0.625 | 0.917 | 6.1748 |
| 0.3 | | 0.518 | 0.885 | 6.1589 | | 0.804 | 0.993 | 6.1331 |
| 0.6 | | 0.627 | 0.950 | 6.1591 | | 0.864 | 0.996 | 6.0891 |
| 0.9 | | 0.712 | 0.981 | 6.2867 | | 0.934 | 1.000 | 6.1480 |
| 1.2 | | 0.790 | 0.996 | 6.1285 | | 0.953 | 1.000 | 6.2238 |
| 1.5 | | 0.849 | 0.998 | 6.1491 | | 0.976 | 1.000 | 6.1520 |
| 0.01 | 2 | 0.400 | 0.725 | 6.8858 | 2 | 0.565 | 0.912 | 6.8259 |
| 0.3 | | 0.564 | 0.917 | 6.9380 | | 0.738 | 0.983 | 6.8078 |
| 0.6 | | 0.677 | 0.970 | 6.9256 | | 0.829 | 0.995 | 6.8691 |
| 0.9 | | 0.744 | 0.987 | 6.9219 | | 0.867 | 0.999 | 6.8485 |
| 1.2 | | 0.809 | 0.998 | 6.8432 | | 0.902 | 0.999 | 6.9214 |
| 1.5 | | 0.880 | 0.999 | 6.8977 | | 0.932 | 1.000 | 6.9212 |
| 0.01 | 3 | 0.428 | 0.765 | 7.5118 | 3 | 0.59 | 0.915 | 7.6277 |
| 0.3 | | 0.595 | 0.928 | 7.5382 | | 0.720 | 0.974 | 7.4621 |
| 0.6 | | 0.701 | 0.982 | 7.5811 | | 0.772 | 0.992 | 7.5016 |
| 0.9 | | 0.750 | 0.991 | 7.5289 | | 0.822 | 0.999 | 7.5506 |
| 1.2 | | 0.849 | 0.996 | 7.5653 | | 0.883 | 0.998 | 7.5077 |
| 1.5 | | 0.870 | 0.997 | 7.4880 | | 0.922 | 1.000 | 7.5372 |
| 0.01 | 4 | 0.457 | 0.806 | 8.1605 | 4 | 0.586 | 0.931 | 8.2216 |
| 0.3 | | 0.636 | 0.954 | 8.1486 | | 0.669 | 0.977 | 8.0037 |
| 0.6 | | 0.739 | 0.985 | 8.0564 | | 0.765 | 0.987 | 8.1332 |
| 0.9 | | 0.806 | 0.997 | 8.1835 | | 0.825 | 0.997 | 8.0578 |
| 1.2 | | 0.856 | 0.997 | 8.1954 | | 0.846 | 0.997 | 8.1159 |
| 1.5 | | 0.880 | 1.000 | 8.0360 | | 0.892 | 0.999 | 8.1147 |
| 0.01 | 5 | 0.494 | 0.849 | 8.6979 | 5 | 0.547 | 0.919 | 8.7302 |
| 0.3 | | 0.634 | 0.96 | 8.6596 | | 0.673 | 0.978 | 8.6472 |
| 0.6 | | 0.762 | 0.986 | 8.8528 | | 0.731 | 0.99 | 8.6442 |
| 0.9 | | 0.815 | 0.996 | 8.7279 | | 0.805 | 0.994 | 8.7417 |
| 1.2 | | 0.866 | 0.998 | 8.6838 | | 0.828 | 0.997 | 8.6363 |
| 1.5 | | 0.895 | 1.000 | 8.7423 | | 0.865 | 0.996 | 8.6352 |

[1] We find $p < 0.01$ for all treatments for a $\chi^2$-test against a uniform distribution.
[2] $p < 0.01$ for all treatments.

Furthermore, along with Figure 1, we confirm that as cost increases, the probability of selecting the first alternative increases. It becomes more costly to inspect the next alternative, and hence, the decision maker is incentivized to stick to the earlier alternatives that have already been inspected.

Figure 1: Choice Probabilities for the First Alternative (Treatment 1)



(a) Fixed $\sigma_v = 3$        (b) Fixed $\sigma_w = 3$
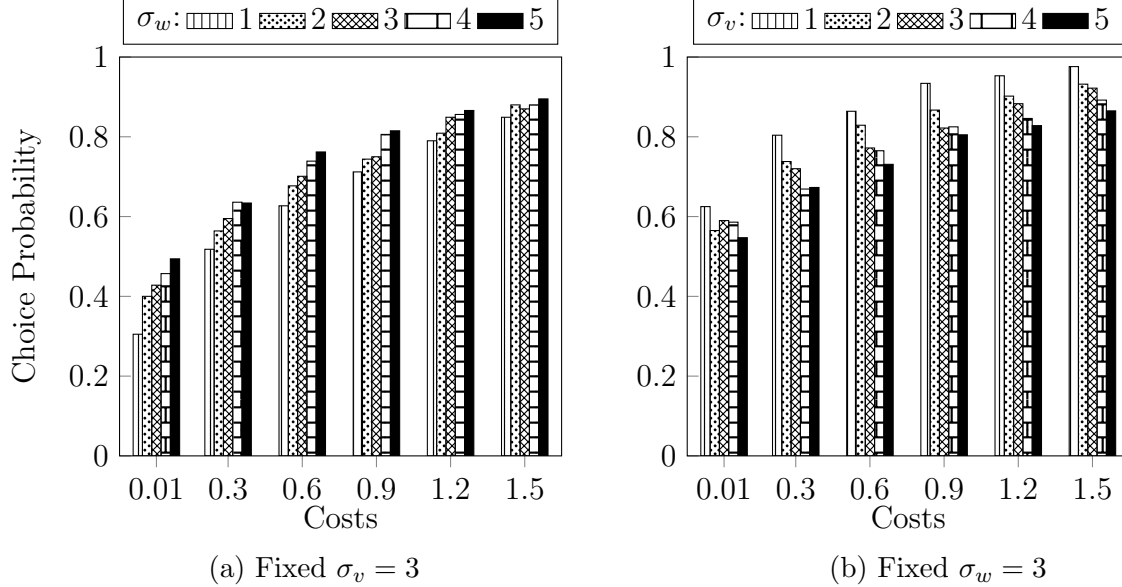
The most striking result here may be the sudden drop in choice probabilities for the near-zero cost. If we inspect this from a practical setting, having zero cost implies no incentive to stop the search earlier. Hence, a potential optimal strategy in this case is to open all boxes and then choose the highest valued box. However, if this were the case, we should observe an even selection of choices across all alternatives, yet the hypothesis test returns statistical significance at the strictest level of confidence across all treatments. In fact, the search order still contains some information about the true value of each alternative, and hence, even with zero cost, the decision maker displays satisficing behavior.

We also note that across Panel (a) and (b), the choice probability of selecting the first alternative is always higher when we vary $\sigma_v$ than $\sigma_w$. At the minimal level of variance and near-zero cost, where our model is closest to having alternatives of similar qualities, the probability of choosing the first alternative is more than twice when varying the dispersion of the hidden factor than the observed factor. Furthermore, at the lower variance levels, this effect is observed across higher levels of costs as well. Even though the difference in varying either $\sigma_w$ or $\sigma_v$ should not influence the true quality of the alternatives, a dispersed observed factor rather than a dispersed hidden factor leads to more versatile choice selection. Furthermore, in Panel (a), we note the increasing choice probability of choosing the first alternative as the observed factor gains more variance while in Panel (b), increasing the variance of the hidden factor rather decreases the choice probabilities across all cost levels. Thus, this calls for a stronger exploration on the effect of the relation between the two variances of the payoff

factors, prompting our results in Section 4.2.

## 4.2  Treatment Group 2

Similar observations to Section 4.1 regarding each individual treatment case were recorded in this group, as the nature of the decision making process is not altered at large in these treatments. At higher cost levels, the probability of choosing the first alternative increased drastically. We include a limited summary table of the results at the near-zero cost level only in Appendix B. However, it is our main objective to investigate the effects of changing $\sigma_v/\sigma_w$ at different cost levels. Hence, Figure 2 shows the different choice probability of the first alternative based on the reference values $\sigma_w$ for each level of $\sigma_v/\sigma_w$. The higher the value of the reference, the absolute magnitude of the dispersion on both observed and unobserved factor is greater. Then, the different ratios control for the relative dispersion between the two factors.

Figure 2: Choice Probabilities for the First Alternative (Treatment 2)
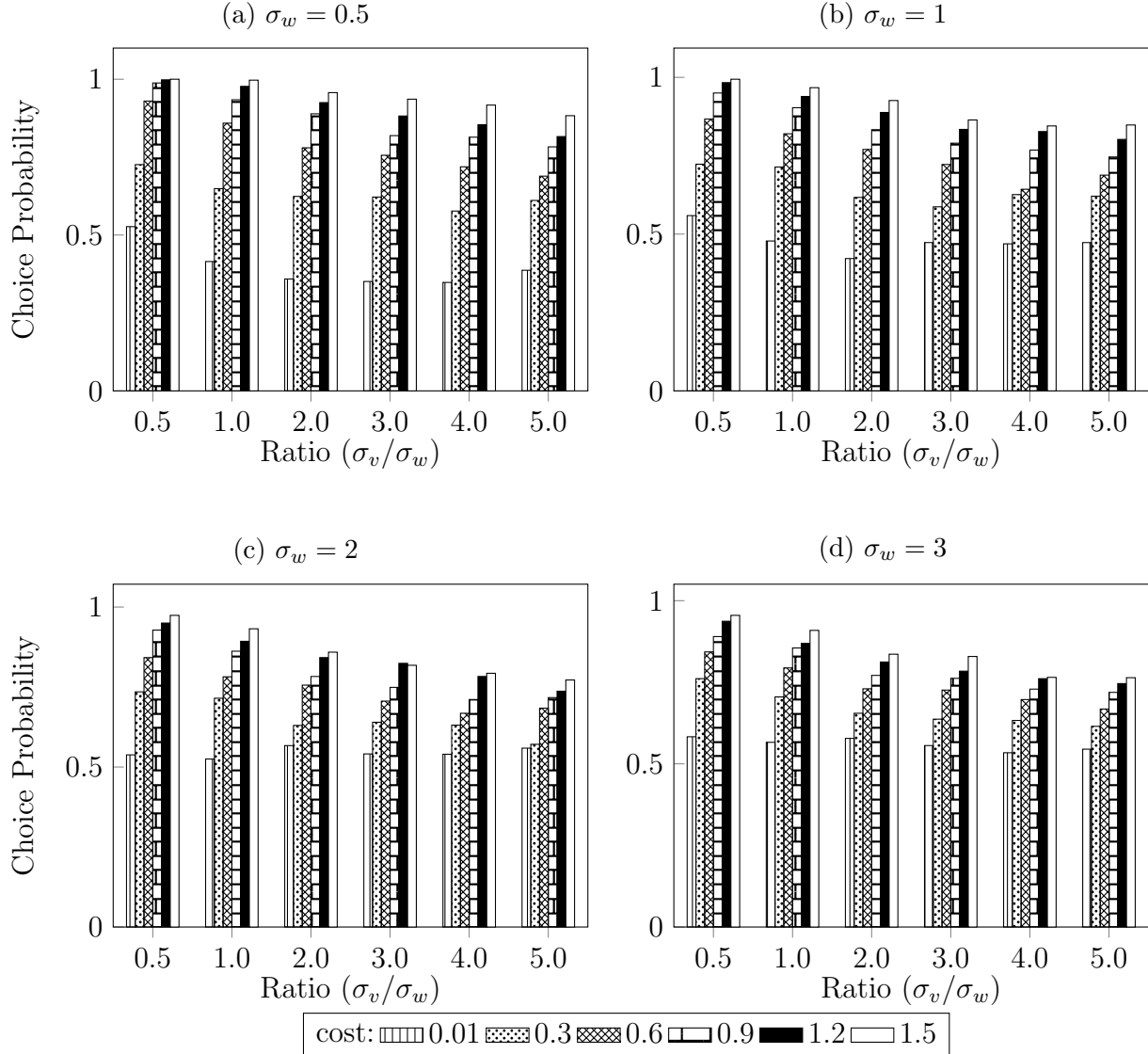
Across different absolute magnitudes of dispersion, we see little difference in choice probabilities. However, in each panel, there is a greater chance of selecting the first alternative when the dispersion of observed factor is relatively greater than that of the unobserved factor. We note a visual difference between when the ratio is below 1 and above 1. Yet for the higher magnitudes, panel (c) and panel (b), for the near-zero cost, we see that even then, this proposition does not hold. This may be due to the greater effect of the higher absolute magnitude of dispersed information the decision maker faces. As the ratio increases beyond 1, this trend no longer seems to hold across higher cost levels as well. A visually weak decreasing trend can be spotted but is neither consistent nor strict.

# 5    Sequential Search Under Weak Signals

Our hope in conducting this simulation was to estimate the effect of signals on similar valued alternatives on optimal choice. In the setup of our simulation, this was done by weakening the dispersion of the observed factor, hence, a weak signal. Hence, our interests lie in mostly the treatments with low $\sigma_w$ and $\sigma_v$. We see its first indication in Figure 3. This shows how many times the $n$th searched alternative was chosen of the 1000 trials. Each bar represents the choice probability for each cost in the treatment. Hence, for instance, $c = 0.01$, we see that the first alternative was chosen about 55% of the time.

Figure 3: Choice Probabilities for $\sigma_w = 0.5$, $\sigma_v = 0.25$
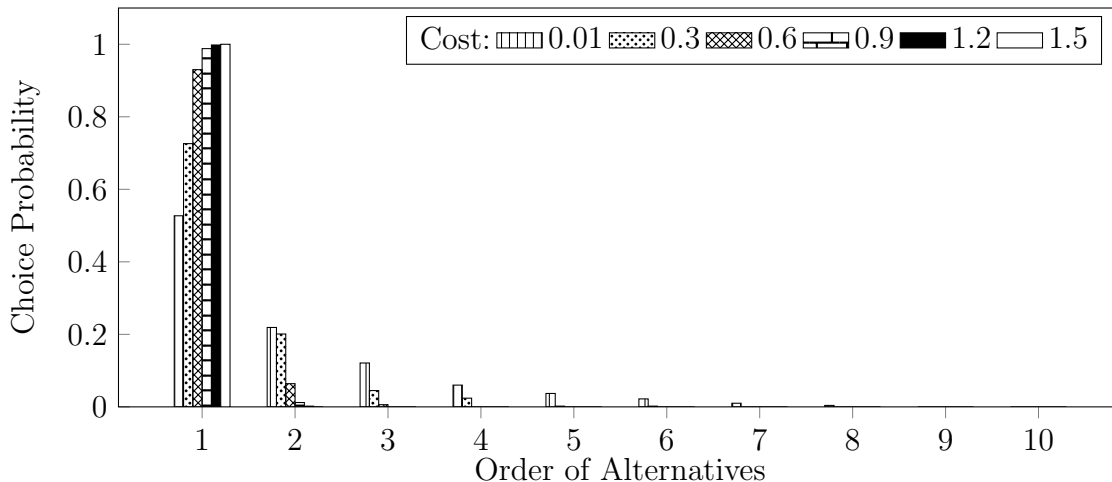


More importantly, Figure 3 strongly suggests an affinity to look towards search order rather than the quality of alternatives. Focusing on positive costs, which is likely the case in the practical setting, across the cost levels, a significant portion of choices are concentrated on the first alternative, and almost all choices are selected within the first two alternatives inspected by our simulated decision maker. With low variability in the true quality of the alternative, there is less variability among the reservation values, implying less chance of a box with a high reservation value occurring later, and thus, encourages the decision maker to stop the search earlier and be satisfied with the current high alternative he has. Even at the zero cost, while relatively lower

9

than the other costs, we still see a concentration of choices in the first two or three alternatives inspected. Thus regardless of the quality of the alternatives, the first presented alternatives are more likely to be selected. In fact, observing the average realized value of the alternatives, the difference between the worst and best alternative was only 2.19 across all cost treatments for $\sigma_w = 0.5$, $\sigma_v = 0.25$.

In Treatment Group 1 and especially in Treatment Group 2, we have results that allow us examine the individual effects of magnitude of dispersion in the observed and unobserved factors. Figure 4 shows the choice probabilities for the first alternative for varying $\sigma_w$ and $\sigma_v$ when the other is fixed. As before, Panel (a) highlights an inconsistent trend in an increasing probability for selecting the first alternative as the dispersion in the observed factor increases. Though, the inconsistency lies mostly in the higher cost results, and thus, the results of Panel (a) provides some evidence in showing weaker signals lead to higher selection of the first alternative. Panel (b) confirms the inconclusive results of Section 4.2. However, revisiting Figure 2, in the specific cases of high absolute magnitudes (panels (c) and (d)) and high $\sigma_v$ to $\sigma_w$ ratios, the probability of choosing the first alternative under higher costs is much lower than its average selection rate. The higher absolute magnitude of dispersion may imply higher levels of uncertainty felt by the decision maker, especially in the unobserved factor, which raises the reservation value, and incentivizes the decision maker to continue inspecting the boxes. Thus, despite the high cost, the expected return of opening more boxes may increase in this case as there is a higher chance of a much higher quality alternative existing in the latter section of the order.

Figure 4: Choice Probabilities for the First Alternative under varying factor variances



(a) Fixed $\sigma_v = 3$        (b) Fixed $\sigma_w = 3$

Cost: 0.01 0.3 0.6 0.9 1.2 1.5
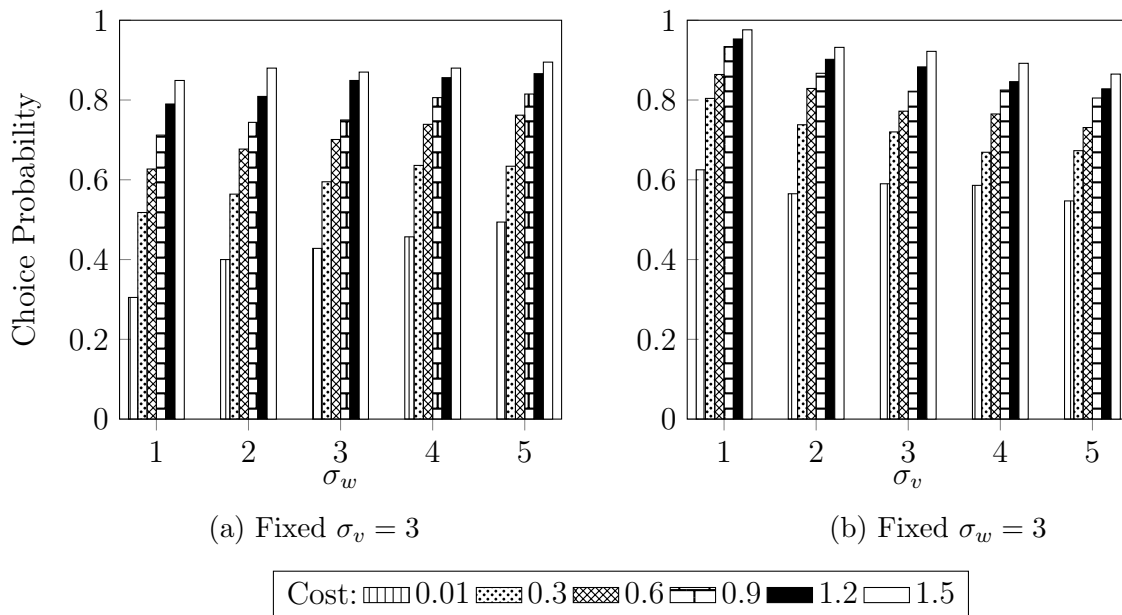
We do not delve deep into higher cost results because precisely the effect that higher cost has. Figure 1 and 2 highlight that as costs increases, the effect of largely dispersed observed and unobserved factors is minimized. The decision maker, regardless of the signal and prior information he receives, cannot afford to open as many boxes as desired and will likely stick to the first few alternatives. In fact, we see that for almost all the

treatments with a high cost, the decision maker almost always chooses from the first two alternatives inspected.

From our model in Section 3.2, we noted the search order was based on the reservation values calculated from the observed factor. By design of our simulation, the order of reservation values will be in line with the order of the observed factors $\{w_i\}$. This is consistent with the lemma proved by Choi and Smith (2022), where they mention the search order will be in decreasing order of known factors if done through calculating the reservation values. Hence, even if the signal is weak, the search order will be in some correlation with the most optimal order (best to worst alternative), and thus, at some point before the last alternative, the current best value observed will be better than the remaining reservation values. Extending this thought process, we can reach a conclusion that under weak signals, the decision maker will still choose from earlier rather than later alternatives that are inspected.

# 6 Discussion

The implications of the simulation results may be straightforward in the sense that a brute way to look at the above results is to conclude that only search order matters in deciding the optimal choice in a sequential search model. While we need not be this extreme, the message is clear: as long as there is some correlation between the prior information given to the decision maker before the search, the optimal choice made will be heavily influenced by the order of the search given the prior. From the decision maker's perspective, however weak the signal may be, having the signal can't make the decision maker worse off than not having the signal, and thus the decision maker has no incentive not to incorporate the signal in the consideration process.

In a practical setting, this may imply that even if some search engine gives a search result that is close to random, the decision maker will be choosing from the first page, even if the best result is on the third page. Outside the online platform, suppose a decision maker in a supermarket. He is looking for protein bars and enters the aisle with protein bars. Assuming the different brands of protein bars provide more or less the same nutrition and health effects, the decision maker will most likely select one of the first three protein bars he examines, and he will examine each protein bar in the order based on the prices and the brand names that he easily observe from the aisle.

As the observed factor is the sole component in deciding the search order, from the perspective of the 'box provider,' one can alter the observed factors to influence the search order of the decision maker. Thus, another way to look at this problem is in terms of framing the decision maker. In a less natural search setting, where there is another party that gives the decision maker the observed factor information, the information then can be systematically altered to produce a specific result in some party's favor. While we did not discuss the case in this study as it seemed unnatural, the prior information given may be exactly the opposite order of the most optimal order of alternatives. Even then, as any information is better than none, the decision maker will take this into account and the search order will thus be constructed around this false prior information.

Extending the example of the search engine, it is a basic example of a model with a high $\sigma_v$ relative to $\sigma_w$. Despite the signal indicating the traits of the results website, the quality of the actual website may differ drastically. Think of a simple academic search,

whose answers are often found in various forums. Most likely, the same question has been asked before on such forums and hence, a Google search references the question asked on a forum. Based on the question header that Google provides, consumer receives a strong signal due to the similarity to his own question. However, the forum result may contain useless leads or may be even unanswered by attracted popularity due to the essential nature of the question. Hence, the quality of the result itself is more versatile than the signal given by the results. On the other hand, picking an item of the grocery shelf constitutes a model with a lower $\sigma_v$. Since marketplace items, especially food items which undergo various restriction and processed items, cannot drastically differ in quality, the consumer, without much thought, can form a steady expectation for the quality of the item.

In regards to Choi and Smith (2022) whose study was very similar to ours, we find consistent results. Choi and Smith show that if the observed factor is nonexistent, search duration increases with payoff dispersion, i.e. the dispersion of the unobserved factor directly associated with search duration. What we found is that only with the dispersion of the observed factor, i.e. weaker signal, does search duration decrease, but with increasing dispersion of the hidden factor, given a constant observed factor dispersion, we actually see a slight decreasing, yet inconsistent, trend in the choice probability for the first alternative, implying a longer search. We note that Choi's model is much more sophisticated however and we have in fact simplified a lot of the assumptions regarding costs and distributions of payoffs. Choi and Smith work with fully variable distributions of both factors and cost. However, we still find significant that under our stricter assumptions, we can closely reproduce some of the theoretical results proved by Smith and Choi.

# 7    Conclusion

In this paper, we have extended an increasingly important sequential search framework of Choi and Smith (2022) and simulated it under a weak signal in the observed factor to capture practical economic and real-world applications, such as grocery shopping and online searches. We found that under both scenarios, the order of search is more pertinent to the decision maker's choice rather than the qualities of the alternatives. Ultimately staying within the framework of Weitzman (1979), the results lead to an important finding: optimal choice is more related to the signal given about the alternatives than the quality of the alternatives themselves.

While our simulation has instituted stronger assumptions Choi and Smith (2022)'s, these are hard to view as completely unrealistic. Especially on online platforms, search costs are often uniform and with the increasing value of online services, the gap between the worst and best services is closing. Given empirical support, the implications are also valuable, especially in consumer research and online marketing. It may lead the way to systematic exploitation of search results, which if we see Google, we already find the advertised services on the top, but also can further consumer quality of life on these online platforms. But as we only based our results on a simulation, an empirical follow-up may lead most fruitful in continuing the line of research.

# References

[1]     Mark Armstrong. "Ordered Consumer Search". In: *Journal of the European Economic Association* 15.5 (June 2017), pp. 989–1024. ISSN: 1542-4766. DOI: 10.1093/jeea/jvx017. eprint: https://academic.oup.com/jeea/article-pdf/15/5/989/20599706/jvx017.pdf. URL: https://doi.org/10.1093/jeea/jvx017.

[2]     Shant Boodaghians et al. "Pandora's Box Problem with Order Constraints". In: *Proceedings of the 21st ACM Conference on Economics and Computation.* EC '20. Virtual Event, Hungary: Association for Computing Machinery, 2020, pp. 439–458. ISBN: 9781450379755. DOI: 10.1145/3391403.3399501. URL: https://doi.org/10.1145/3391403.3399501.

[3]     Andrew Caplin and Daniel Martin. "Framing effects and optimization". In: *Manuscript, New York Univ.* (2012). URL: https://pdfs.semanticscholar.org/1853/a29d1bf93b7%202f0b5be649071479c737b4e46.pdf.

[4]     Michael Choi, Anovia Yifan Dai, and Kyungmin Kim. "CONSUMER SEARCH AND PRICE COMPETITION". In: *Econometrica* 86.4 (2018), pp. 1257–1281. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/44955991.

[5]     Michael Choi and Lones Smith. "Optimal Sequential Search with Recall". In: *SSRN Electronic Journal* (2022). DOI: 10.2139/ssrn.4164425. URL: https://doi.org/10.2139/ssrn.4164425.

[6]     Leon Yang Chu and Heng Zhang. "Position Ranking and Auctions for Online Marketplaces". In: *SSRN Electronic Journal* (2017). DOI: 10.2139/ssrn.2926176. URL: https://doi.org/10.2139/ssrn.2926176.

[7]     Laura Doval. "Whether or not to open Pandora's box". In: *Journal of Economic Theory* 175 (2018), pp. 127–158. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2018.01.005. URL: https://www.sciencedirect.com/science/article/pii/S0022053118300073.

[8]     Rafael P. Greminger. "Optimal Search and Discovery". In: *Management Science* 68.5 (May 2022), pp. 3904–3924. DOI: 10.1287/mnsc.2021.4085. URL: https://doi.org/10.1287/mnsc.2021.4085.

[9]     Wojciech Olszewski and Richard Weber. "A more general Pandora rule?" In: *Journal of Economic Theory* 160.C (2015), pp. 429–437. URL: https://EconPapers.repec.org/RePEc:eee:jetheo:v:160:y:2015:i:c:p:429-437.

[10]    Jacques Robert and Dale Stahl. "Informative Price Advertising in a Sequential Search Model". In: *Econometrica* 61.3 (1993), pp. 657–86. URL: https://EconPapers.repec.org/RePEc:ecm:emetrp:v:61:y:1993:i:3:p:657-86.

[11]  Herbert A. Simon. "A Behavioral Model of Rational Choice". In: *The Quarterly Journal of Economics* 69.1 (Feb. 1955), pp. 99–118. ISSN: 0033-5533. DOI: 10.2307/1884852. eprint: https://academic.oup.com/qje/article-pdf/69/1/99/5413375/69-1-99.pdf. URL: https://doi.org/10.2307/1884852.

[12]  Raluca Mihaela Ursu. "The Power of Rankings: Quantifying the Effects of Rankings on Online Consumer Search and Choice". In: *SSRN Electronic Journal* (2015). DOI: 10.2139/ssrn.2729325. URL: https://doi.org/10.2139/ssrn.2729325.

[13]  Martin L. Weitzman. "Optimal Search for the Best Alternative". In: *Econometrica* 47.3 (1979), pp. 641–654. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1910412.

# A    Simulation Code

```python
import numpy as np
import pandas as pd
import sympy as sp
from sympy.stats import density, Normal, cdf
from scipy.optimize import root_scalar
import pickle


# Reservation value
z = sp.symbols("z")



# The equation that defines the reservation value.
# Simplification of equation (4).
def f(z: float, cost: float, w: float, v_std: float) -> float:
    F = Normal("f", w, v_std)
    F1 = Normal("f1", 0, v_std)
    result = sp.N(z * cdf(F)(z) + v_std * density(F)(z) + w *
        (1 - cdf(F)(z)) - z - cost)
    return result

# Function that calculates the reservation value given the
    cost, observed factor,
# and the standard deviation of the unobserved factor
    distribution
# 'a': a threshold parameter for the search area of the
    optimization
def calc_res_value(cost, w, v_var, a):
    v_std = np.sqrt(v_var)
    u0 = w
    bracket = (w - a * v_var, w + a * v_var)
```

14

```python
        solution = root_scalar(f, args = (cost, w, v_std), x0=u0,
            bracket=bracket)
        return solution.root

# The main function that runs the simulation given the
    parameters of the treatment.
def choice(n, x_mean, x_var, v_var, cost):
    x_vals = np.random.normal(x_mean, np.sqrt(x_var), n)
    opening_order = np.argsort(x_vals)[::-1]
    true_vals = x_vals + np.random.normal(0, np.sqrt(v_var), n
        )

    highest_true_val = 0
    highest_true_val_index = 0

    a = 5

    for i in range(len(opening_order)):
        index = opening_order[i]
        true_val = true_vals[index]

        if true_val > highest_true_val:
            highest_true_val = true_val
            highest_true_val_index = i + 1

        if i < n - 1:
            next_res_val = calc_res_value(cost, x_vals[
                opening_order[i + 1]], v_var, a)
            if highest_true_val > next_res_val:
                return highest_true_val_index, true_vals[
                    opening_order[0]], true_vals[opening_order[
                    n - 1]], max(true_vals), min(true_vals)
        else: return highest_true_val_index, true_vals[
            opening_order[0]], true_vals[opening_order[n - 1]],
            max(true_vals), min(true_vals)

N = 1000
n = 10
x_mean = 30

# Setting for Treatment Group 1, fixed sigma_v = 3
w_vars = [1, 2, 3, 4, 5]
v_vars = [3]
costs = [0.01, 0.3, 0.6, 0.9, 1.2, 1.5]

collections = dict()
```

```python
for w_var in w_vars:
# Progress marker for output
#     print("Starting x_var = ", w_var)
    for v_var in v_vars:
#         print("\tStarting v_var = ", v_var)
        for cost in costs:
#             print('\t\tStarting cost = ', cost)
            data = np.zeros((N, 5))
            for i in range(N):
#                 print('\t\t\t\tCurrent Index =', i)
                data[i] = choice(n, x_mean, w_var, v_var, cost
                    )
            with open('collections_w' + str(w_var) + 'v' + str
                (v_var) + 'c' + str(cost) + '.pkl', "wb") as
                file:
                pickle.dump(data, file)
#             print('\t\t\tFinished!')
```

# B

Table 2: limited Summary of Choice Probabilities for Treatment Group 2 (cost = 0.01)

| $\sigma_w$ | $\dfrac{\sigma_v}{\sigma_w}$ | (1) Choice Prob. | | (2) Mean Diff. |
|---|---|---|---|---|
| | | First | Top 3 | |
| 0.5 | 0.5 | 0.527 | 0.867 | 2.6988 |
| | 1 | 0.415 | 0.762 | 3.0929 |
| | 2 | 0.359 | 0.702 | 3.7167 |
| | 3 | 0.351 | 0.704 | 4.3804 |
| | 4 | 0.348 | 0.731 | 4.7904 |
| | 5 | 0.387 | 0.764 | 5.3605 |
| 1 | 0.5 | 0.559 | 0.874 | 3.8155 |
| | 1 | 0.478 | 0.815 | 4.3662 |
| | 2 | 0.422 | 0.818 | 5.2705 |
| | 3 | 0.473 | 0.851 | 6.1735 |
| | 4 | 0.469 | 0.839 | 6.9339 |
| | 5 | 0.473 | 0.869 | 7.6369 |
| 2 | 0.5 | 0.538 | 0.893 | 5.3052 |
| | 1 | 0.525 | 0.873 | 6.1564 |
| | 2 | 0.567 | 0.893 | 7.5103 |
| | 3 | 0.541 | 0.912 | 8.601 |
| | 4 | 0.54 | 0.891 | 9.6665 |
| | 5 | 0.559 | 0.896 | 10.6729 |
| 3 | 0.5 | 0.583 | 0.915 | 6.4612 |
| | 1 | 0.566 | 0.917 | 7.5919 |
| | 2 | 0.578 | 0.919 | 9.1809 |
| | 3 | 0.556 | 0.927 | 10.6017 |
| | 4 | 0.534 | 0.921 | 11.7596 |
| | 5 | 0.545 | 0.922 | 13.0317 |