

Search with Multiple Attributes: Theory and Empirics

Adam Sanjurjo *

March 18, 2017

Abstract

Multiple attribute search is a central feature of economic life: we consider much more than price when purchasing a home, and more than wage when choosing a job. Nevertheless, while single attribute search problems have been studied extensively, little is known about optimal search in multiple attribute environments. I introduce a partial characterization of optimal sequential search in a problem with multiple searchable attributes and alternatives, no order restrictions on search, and full recall. Upon applying the partial rational benchmark to a rich dataset I find that subjects systematically deviate from optimal sequential search by (1) searching too deeply within alternatives and (2) switching too adjacently between alternatives. Finally, I explore how these deviations affect payoffs, and explain why they may constitute a form of boundedly rational search behavior in which subjects re-optimize only occasionally, while “smoothing” search order so as to make costly memory failures less likely.

JEL Classification Numbers: D11; D12; D81; D83.

Keywords: Search Theory; Optimal Sequential Search; Multiple Attributes; Consumer Choice; Procedural Cognition; Bounded Rationality; Cognitive Economics; Behavioral Economics

*Fundamentos del Análisis Económico, Universidad de Alicante, 03690 Alicante, Spain. Email: sanjurjo@ua.es. Phone: (34)965.903.225. Financial support from the Spanish Ministerio de Ciencia y Tecnología and Feder Funds (SEJ-2007-62656), and the Spanish Ministry of Economics and Competition (ECO2012-34928) is gratefully acknowledged. These funding sources had no role in the analysis or interpretations provided in this paper.

1 Introduction

Optimal search policies have been fully characterized for a wide variety of search problems in which a *single* attribute of an alternative, usually its price, determines its desirability, with varying assumptions about value distributions, search costs, number of searchable alternatives, and recall options (Kohn and Shavell 1974; Lippman and McCall 1976; Weitzman 1979). Although these analyses yield invaluable insights, many important applications have alternatives whose values are determined by *multiple* attributes: we consider much more than wage when choosing a job, and much more than price when purchasing a home.

In this paper I study multiple attribute optimal sequential search problems of the following type: Jon will ultimately choose a car from the list of alternatives A, B, and C. He wishes to choose the car with the highest overall sum rating across the attributes price, safety, and fuel economy, but he initially only knows the price ratings of each car, say, 3, 2, and 1, respectively. Given this limited information, he can either choose one of the cars immediately (car A), or he can engage in costly search in order to learn more. Say that Jon decides to learn the safety rating of car A. After acquiring this additional information, he must then decide whether to continue by learning the fuel economy rating of car A, instead learn the safety or fuel economy rating of car B or C, or stop costly search entirely to choose one of the cars, and so on.¹

More precisely, I study problems in which the value of each alternative is equal to the sum of up to many attributes.² For each alternative, one attribute value is initially known, and the rest are unknown. Each unknown attribute value is an independently and symmetrically distributed random variable. In each alternative the attributes are ordered according to variance, by a form of *symmetric second-order stochastic dominance* (henceforth “s.s.o dominance”), in which each dominated attribute can be obtained by a (symmetric) mean-preserving spread of a dominating attribute.³ Any attribute of any alternative can be searched separately, following any search history, at a cost that is constant across attributes. Similarly, any alternative can be chosen following any search history (i.e. there is *full recall*), and even before any search has occurred.⁴ Finally, the

¹Similar real world search problems are now ubiquitous on internet websites, in which suppliers facilitate the comparison of multiple alternatives across multiple attributes, for a wide variety of products from digital cameras to cars and homes, to health care plans and 401k’s (Lee and Lee 2004; Shi, Wedel, and Pieters 2013).

²The separability assumption is standard in the multiple attribute search literature; see, e.g. (Caplin, Dean, and Martin 2011; Gabaix, Laibson, Moloche, and Weinberg 2006; Klabjan, Olszewski, and Wolinsky 2014; Lim, Bearden, and Smith 2006; Payne, Bettman, and Johnson 1993), and is a natural first step. Analogs to the results reported in this paper can also be provided assuming weighted additive utility.

³Symmetric second-order stochastic dominance implies second-order stochastic dominance in the sense of Rothschild and Stiglitz (1970).

⁴The standard assumption in single attribute search problems, for instance, is that an alternative can only be chosen

searcher is rational and risk-neutral.

The optimal sequential search *policy*, which characterizes both the complete optimal order of search and the optimal stopping rule, is known only for a dimensionally small version of this problem, in which there is a single searchable alternative with two unknown attributes, and whose expected value is the same as that of a deterministic outside option (Klabjan, Olszewski, and Wolinsky [2014]; Proposition III).⁵ In the same context, for any number of searchable (binary) attributes in the single searchable alternative, the authors’ Proposition II characterizes the complete optimal order of sequential search: the next attribute searched is the one that is s.s.o dominated by all the others, i.e. the attribute with the most variance.⁶

In Section 2 I introduce several necessary conditions for optimal sequential search. Condition V (for “variance”) is a more general version of Proposition II from Klabjan, Olszewski, and Wolinsky (2014). It characterizes the complete optimal order of sequential search across the attributes within any single alternative, in search problems with: (1) up to any number of searchable alternatives, (2) alternatives that can differ in expected value, (3) each alternative with up to any number of symmetric s.s.o-ordered independent attributes, and (4) attribute distributions that can assume any shape from a large class of symmetric distributions. In particular, the result shows that within any given alternative, search always proceeds with the highest variance attribute first. If there is just one searchable alternative, then Condition V characterizes the complete optimal order of sequential search. This result is reminiscent of Weitzman (1979)’s on the optimal order of sequential search *across* single attribute alternatives.

In Corollary V2 I extend Condition V to characterize the complete optimal order of sequential search *across* alternatives when there are exactly *two* alternatives, each with up to any number of searchable attributes. In particular, the optimal order of search proceeds by searching the attribute with highest variance among the two alternatives, regardless of which alternative it is in; that is, the optimal order does not depend on the expected values of the alternatives, or on their sets of (additional) searchable attributes.

The intuition behind Corollary V2 is that in the case of exactly two alternatives a special symmetry holds: receiving good information from the currently more promising alternative is just once it has been searched. Allowing also for pre-search choice is non-trivial, as explained in Section 2.

⁵This problem is equivalent to one in which there are exactly two alternatives, each with the same expected value and one searchable attribute (see Corollary V2 in Section 2).

⁶Relatedly, the optimal sequential search policy has been characterized analytically when just one of multiple attributes can be searched in each alternative, and alternatives must be searched in a fixed order, though with full recall (Neeman 1995), and numerically for a multiple alternative, two attribute search problem in which the order of search across alternatives and attributes is fixed, and there is no recall (Lim, Bearden, and Smith 2006).

like receiving equally bad information from the currently less promising alternative. Thus, the problem with exactly two alternatives, both searchable, can effectively be reduced to a problem with just one searchable alternative. When moving from two to three or more alternatives, however, this symmetry no longer holds, making it more difficult to obtain results on the optimal order of sequential search *across* alternatives.

Condition T (for “trade-off”) partially addresses the fundamental question in search with any number of multiple attributes and alternatives—how to optimally trade-off depth (within alternatives) against breadth (across alternatives) in search. Roughly speaking, the condition says that alternatives with higher expected values, or more unsearched attributes, all else equal, should be searched before their counterparts. Despite the intuitive appeal of this condition, the dimensionality of the problem makes it generally difficult to demonstrate analytically.⁷ While I am not able to prove Condition T for the general case, I do prove it for the case in which there are any number of alternatives, one of which has any number of unknown attributes, and all of the others of which have at most one unknown attribute. This case allows for the searcher to first partially explore an alternative, then: (1) explore the same alternative further, (2) switch to exploring another alternative, or (3) end search and choose an alternative. The first feature separates this multiple attribute search problem from its single attribute counterparts, e.g. [Weitzman \(1979\)](#). The second feature separates it from the multiple attribute search problem with just one searchable alternative, e.g. [Klabjan, Olszewski, and Wolinsky \(2014\)](#). In Section 2.2 I explain how the techniques that I use to make analysis tractable in the above case no longer work in the case of multiple alternatives with multiple searchable attributes in each. As a consequence, in order to confirm that Condition T holds more generally, I perform numerical computations on the simplest versions of this case, in which there are up to three alternatives and two searchable attributes in each.⁸ Finally, I run simulations on numerically intractable versions of the problem that have many searchable attributes in each of many alternatives. Simulation results corroborate those obtained in the theoretical and numerical analyses (see Section 6).

Conditions V and T can be used to identify potentially important violations of optimality in search data: whereas Condition V identifies instances in which attributes are searched in the incorrect order within an alternative, Condition T identifies instances of *over-search*, which occurs

⁷Condition T is generally more difficult to work with analytically than Condition V, as in the case of the latter alternative policies start search in the same alternative, which relatively facilitates comparison for technical reasons that are explained in Section 2.

⁸To get a sense of the dimensionality issue, in the search problem with three alternatives, and two searchable attributes in each, in principle, there are $\sum_{i=0}^6 \frac{6!}{i!} = 1957$ different possible search paths.

when the same alternative is searched again though another alternative should have been switched to, and *biased switch*, which occurs when the searcher switches alternatives, but to a sub-optimal alternative. Two additional conditions, called R and C, though theoretically trivial, are also of interest for studying search data. Condition R (for “repeat”) requires that each attribute in an alternative be searched at most once, and Condition C (for “choice”) requires that when search ends, the alternative with the highest currently revealed value be chosen. All conditions are stated formally in Section 2.2.

I use the necessary conditions to analyze the data from an especially rich multiple attribute (ten), multiple alternative (eight) search experiment that not only tracks the choices of 390 subjects, but also the complete order and duration of their information search behavior (Gabaix, Laibson, Moloche, and Weinberg 2006; henceforth GLMW).⁹ Although optimal sequential search would be a natural benchmark with which to compare subjects’ behavior, GLMW find that the high dimensionality of their problem makes characterizing optimal search analytically and numerically intractable (p. 1055). As a result, they focus instead on comparing a behavioral model of search—their *Directed Cognition* model—with several simple heuristics taken from the psychology literature (e.g. Tversky [1972]), and find that their model outperforms the others. The Directed Cognition model’s first main departure from rationality is that it is *myopic* in that it intentionally omits option value, which simplifies the search problem considerably by treating each search decision as if it were the last.¹⁰ Its second main departure is that it (myopically) *re-optimizes occasionally*, rather than after every time it searches an attribute in an alternative.¹¹ By contrast, the partial rational benchmark introduced here describes the behavior of a rational searcher who considers the full option value of search, and re-optimizes after every time she searches an attribute in an alternative.

After describing GLMW’s design in Section 3, in Section 4 I provide an overview of the basic features of subjects’ search behavior and performance. GLMW conduct most of their data analysis at the aggregate level, pooling both subjects and tasks, and considering which attributes in which alternatives were searched in each task, i.e. the search *footprints*, but not the order in which the attributes were searched. In Sections 4-6 I build on GLMW’s analysis by studying subjects’ search

⁹I know of no other experimental design with nearly as many searchable attributes and alternatives as GLMW, while also providing sufficient structure for optimal sequential search to have an objective interpretation.

¹⁰By analogy, one might consider how much simpler the game of chess could be if a player were to consider the current move as if it were the last.

¹¹The model further deviates from optimal sequential search by assuming that when multiple attributes are searched at a time, they must all be from the same alternative, and bases the decision of where to search next on the ratio of expected benefits to costs, rather than the difference.

behavior on the more granular subject/task level, and by considering the full order of search in each task.

In Section 5 I compare subjects' behavior to the partial rational benchmark, and find that subjects systematically (and substantially) deviate from optimal sequential search by (1) searching too deeply within alternatives and (2) switching too adjacently between alternatives. The first result suggests an insensitivity to acquired information. I use an econometric analysis to show that the second result is driven by subjects' responding to "irrational" spatial considerations just as much as to rational value considerations.

In Section 6 I explore the economic consequences of subjects' deviations, both with regression and simulation analysis. While I find that subjects' who violate Condition T more often do tend to obtain less valuable information, and earn less, their deviations are, surprisingly, not as costly as what one might expect given their high violation rates, and that simulations show violations of the conditions to generally be quite costly. Also, importantly, I find that the economic losses associated with failing to choose the alternative that has been revealed by search to be the most promising (i.e. a violation of Condition C) overwhelm those associated with violations of Conditions T, V, and R.

In Section 7 I offer an explanation that can accommodate this pattern of findings. Specifically, while subjects do very often violate the order of optimal sequential search, the search footprints that they leave in each task happen to overlap with the footprint of violation-free search more than the high rate of order violations suggests. In particular, the evidence indicates that subjects may effectively commit to searching several attributes at a time, in whatever order they prefer. This behavior would be consistent with an optimal sequential searcher who sometimes lapses into "classic" search (Stigler 1961), which is generally consistent with the proposed occasional re-optimizing of the Directed Cognition model, the model of Morgan and Manning (1985), and field evidence in De los Santos, Hortaçsu, and Wildenbeest (2012). Further, the orders that subjects deviate in happen to systematically reduce working memory load (Sanjurjo 2017), which has been shown to reduce the probability of making (severely) costly choice errors (Sanjurjo 2015). Taken together, this evidence suggests that subjects' systematic deviations from optimal sequential search may be adaptive, and boundedly rational.

2 Necessary Conditions for Optimal Sequential Search

Section 2.1 states the multiple attribute optimal sequential search problem posed in the introduction more formally, and Section 2.2 proposes a partial characterization of optimal sequential search, in the form of several necessary conditions.

2.1 Framework

The value of attribute $a \in A$ in alternative $i \in I$ is an independent random variable \tilde{x}_{ia} with symmetric distribution F_{ia} , such that $E(\tilde{x}_{ia}) = 0$.¹² It is assumed that $F_{ia} = F_a$ for $i \in I$ and $a \in A$, i.e. the distribution of attribute values is independent of the alternative.

Further, $F_a \prec^{s.s.o} F_b$ whenever $a < b$, meaning that F_a can be obtained by a symmetric mean-preserving spread of F_b of the following type: for each realization x_b of \tilde{x}_b there is a unique realization x_a of \tilde{x}_a such that $F_b = F_a$. The correspondence x_a to x_b can be represented by the function $h : \mathbb{R} \rightarrow \mathbb{R}$, and because h is invertible, $h^{-1}(x_b) = x_a$. If $x_b > 0$ then $h^{-1}(x_b) = x_a > 0$ and $h^{-1}(x_b) - x_b > 0$ (by definition of h). Likewise, if $x_b < 0$ then $h^{-1}(x_b) = x_a < 0$ and $h^{-1}(x_b) - x_b < 0$. These differences in payoffs are symmetric around $x_b = 0$, so $h^{-1}(x_b) - x_b = -(h^{-1}(-x_b) - (-x_b))$. Further, let $h^{-1}(x_b) - x_b$ be (weakly) increasing in F_b , and call this the *stretching property*. This property is satisfied by a large class of distributions, including the commonly used binary, uniform, and normal distributions.

The value of each alternative $i \in I$ is $\tilde{V}_i = \sum_{a \in A} \tilde{x}_{ia}$. In order to observe realization x_{ia} of \tilde{x}_{ia} when $a > 1$, the searcher must pay a cost of $c > 0$, whereas x_{i1} is assumed observed ex ante, at no cost, for all $i \in I$. The searcher is rational and risk neutral, can search any attribute at any time, and can terminate search to select an alternative whenever she chooses. Thus, the current state is sufficiently described by, for each alternative $i \in I$, the set of unobserved attributes $S_i \subseteq A$ and the expected value $E(\tilde{V}_i | (x_{ia})_{a \in A \setminus S_i})$, which I will call $E(\tilde{V}_i)$, or sometimes simply v_i , for short, and is equal to the current revealed value of the alternative. Finally, the search policy that next searches \tilde{x}_{ja} , and then proceeds optimally thereafter is $\Psi_{ja}[(E(\tilde{V}_i))_{i \in I}, (S_i)_{i \in I}]$, or Ψ_{ja} for short, and the expected value of following this policy is Φ_{ja} . The stopping rule is that if $\Phi_{ja} < \max_{i \in I} E(\tilde{V}_i)$ for every $a \in S_j$ and $j \in I$ then search ends, and any $j \in \arg \max_{i \in I} E(\tilde{V}_i)$ is chosen.¹³

¹² F_{ia} is symmetrically distributed if $F_{ia}(x) = 1 - F_{ia}(-x)$ for all x .

¹³Notice that the current expected value of an alternative $i \in I$ is equal to the sum of its observed realized attribute values: $\sum_{a \in A} x_{ia}$, where $a \in A \setminus S_i$. Also, the framework allows for any subset of attributes to never have been available for search, simply by treating them as already observed, with realized values of zero. Similarly, a problem in which the searcher knows nothing about any alternative ex ante can be represented by setting $x_{i1} = 0$ for all $i \in I$.

Thus, in the problem a searcher can choose to explore any single attribute, in any alternative, whenever she likes, at some fixed cost. Intuitively, the searcher would like to discover one alternative to have a sufficiently higher expected value than all others, after having searched as little as possible; how much higher the expected value has to be generally depends on the full distribution of the other alternatives' current expected values, and also on the set of remaining unsearched attributes in each alternative. Accordingly, after each time the searcher observes an attribute in an alternative she must decide whether to more deeply explore the currently searched alternative, switch to another alternative and search an attribute there, or terminate search now and choose the currently most promising alternative.

2.2 Necessary Conditions

Throughout the section attribute distributions are assumed to be continuous and unbounded, for simplicity.¹⁴

Condition V (Variance) *Let there be $m \geq 2$ alternatives with expected values $E(\tilde{V}_1), \dots, E(\tilde{V}_m)$, and at least one alternative $i \in I$ with at least two unsearched attributes, i.e. $|S_i| \geq 2$. For any such alternative, if $F_{ia} \prec^{s.s.o} F_{ib}$ then $\Phi_{ia} > \Phi_{ib}$.*¹⁵

Proof: See Appendix A.1.

Condition V says that an attribute cannot be searched next if it has less variance than another attribute in the same alternative, in the sense that it s.s.o dominates the other attribute. The result is reminiscent of (Weitzman 1979)'s on the optimal order of sequential search *across* alternatives when each alternative has just one attribute. Here the order of optimal sequential search is instead characterized *within* each alternative, across its attributes.

Condition V is similar to Proposition II of Klabjan, Olszewski, and Wolinsky (2014), but more general, as it allows: (1) up to any number of searchable alternatives with up to any number of

¹⁴It is straightforward to extend all proofs to the discrete unbounded case. For bounded distributions in some (relatively uninteresting) problems the results hold only weakly. To illustrate, consider the simple search problem in which the values of the two alternatives, respectively, are $v_1 + \tilde{x}_{12} + \tilde{x}_{13}$ and v_2 . If $v_1 > v_2$, and the unsearched attribute distributions \tilde{x}_{12} and \tilde{x}_{13} have sufficiently small bounded supports, then the expected value of searching any attribute is zero, i.e. $\Phi_{12} = \Phi_{13} = 0$, because alternative 1 will be chosen no matter what combination of attribute realizations x_{12} and x_{13} are observed. In this case Condition V holds only weakly.

¹⁵In the case of only one alternative, search has no value, so the lone alternative is chosen immediately. On the other hand, without at least two attributes it is impossible to consider multiple possible search orders within an alternative.

attributes in each, rather than a single searchable attribute, (2) the expected values of alternatives to differ, i.e. $E(\tilde{V}_i) \neq E(\tilde{V}_j)$ for $i \neq j$, and (3) a large class of distributions in addition to the binary distribution.¹⁶

While Condition V addresses which attribute ought to be searched first *within* an alternative, it is also important to consider which attribute ought to be searched first *across* alternatives.¹⁷

Whereas Condition V characterizes the complete optimal order of sequential search within any alternative, Corollary V2 extends this result to characterize the complete optimal order of sequential search across alternatives, when there are exactly two alternatives, each with any current revealed value, and any number of unsearched attributes.

Corollary V2 *Let there be exactly two alternatives $I = \{1, 2\}$, with expected values $E(\tilde{V}_1) \geq E(\tilde{V}_2)$, w.l.o.g., and at least two attributes total across the alternatives, i.e. $|S_1| + |S_2| \geq 2$. If $F_{ia} \prec^{s.s.o} F_{jb}$ for $i, j \in I$ then $\Phi_{ia} > \Phi_{jb}$. If F_{ia} and F_{jb} are the same then $\Phi_{ia} = \Phi_{jb}$.*

Proof: See Appendix A.1.

It immediately follows from Corollary V2 that when there are exactly two alternatives, the next attribute searched must be the highest variance attribute, regardless of whether it belongs to the alternative with higher or lower current revealed value, and the sets of attributes belonging to each alternative. This may initially seem surprising, as it goes against the intuition that all else equal, one ought to search the alternative with highest currently revealed value first. After all, this is the alternative that is most likely to sufficiently separate itself above the other(s) within any number of searches. However, what this intuition misses is that in the case of exactly two alternatives, good news about the currently more promising alternative is equivalent to equally bad news about the currently less promising alternative. As mentioned in the Introduction, this symmetry means that the problem with two searchable alternatives can be reduced to the problem with one.

¹⁶In principle, analysis of the present problem is relatively complicated by the larger class of possible search paths that is permitted, the complex trade-offs that result from alternatives not sharing the same expected value, and by allowing for continuous distributions the analysis is "... complicated by the absence of a deterministic ranking of the absolute values of the outcomes" (p. 11; Klabjan, Olszewski, and Wolinsky (2014)). The keys to keeping analysis tractable are: (1) finding ways to reduce the dimensionality of the problem so that comparisons can be made on the basis of just two attributes, (2) creating new (sub-optimal) search policies that facilitate comparison, and (3) exploiting the stretching property to stretch (or compress) distributions with different variances so that they align.

¹⁷In principle, this relatively complicates the analysis by further loosening the basis of comparison between different possible search policies, which can now start in different alternatives.

Once a third alternative is added to the problem, even if it has no searchable attributes, the symmetry that allows the optimal order of sequential search to be characterized across alternatives, by Condition V together with Corollary V2, no longer holds. The simplest such problem has two alternatives with one searchable attribute in each, and a third alternative with none, i.e. $|S_1| = |S_2| = 1$ and $|S_3| = 0$. While this problem may initially appear identical to a single attribute search problem, such as those addressed by the elegant reservation rule of Weitzman (1979), there is a subtle, but important difference. In Weitzman’s version of the problem an alternative can be chosen only once its uncertainty has been resolved. By contrast, in the version of the problem that I consider here, the alternative can be chosen once the uncertainty is resolved, or *before* it is resolved. In Weitzman’s setup this would be like having a trivial outside option equal to the current revealed value of an alternative that then disappears once that alternative is searched. Thus, this simple single attribute version of the multiple attribute search problem can be seen as a version of Weitzman’s setup in which there is “incomplete recall” (p. 649), a condition for which Weitzman’s result does not readily extend. Having said that, for the special case in which the deterministic outside value is the highest of the three current revealed values, i.e. $v_3 \geq v_1, v_2$, the difference in the setups disappears, as a rational searcher would never be interested in immediately choosing a lower payoff than that offered by the outside option. Thus, it follows from Weitzman’s result that in this case optimal sequential search starts with the attribute in the alternative with higher current revealed value, so long as the attribute is weakly s.s.o dominated by the other attribute, i.e. $\Phi_{ia} > \Phi_{jb}$ if $v_i > v_j$ and $F_{ia} \preceq^{s.s.o} F_{jb}$.

Obtaining the result for the case of $v_3 \not\geq v_1, v_2$, on the other hand, is non-trivial as in principle the analysis depends on the relative positioning of the current revealed values of the three alternatives, the reservation intervals of the two searchable alternatives (see Appendix A), and the (possibly weak) s.s.o dominance ordering among attributes. Result T.1 in Appendix A proves the result when there is one searchable attribute in both of the searchable alternatives.¹⁸ Further, Result T.2 extends T.1 to allow for up to any number of attributes in the first alternative, and Result T.3 extends T.2 to additionally allow for up to any number of alternatives with a single searchable attribute. I present Result T.3 here alone, as it generalizes T.1 and T.2.

Result T.3 *Let there be $m \geq 3$ alternatives, the first $m - 1$ of which are ordered from highest*

¹⁸The proof also includes an alternate demonstration, relative to simply applying Weitzman’s result, for the case of $v_3 > v_1, v_2$

to lowest expected value, w.l.o.g., i.e. $E(\tilde{V}_i) \geq E(\tilde{V}_j)$ if $1 \leq i < j \leq m - 1$. Assume that the first alternative has any number of unsearched attributes, i.e. $S_1 \subseteq A$, while alternatives 2 through $m - 1$ each have exactly one unsearched attribute, i.e. $|S_i| = 1$ for $2 \leq i \leq m - 1$, and alternative m has none, i.e. $S_m = \{\emptyset\}$. If $F_{1a} \preceq^{s.s.o} F_{ib}$ for all $b \in S_i$ and $i \in \{1, 2, \dots, m - 1\}$, then $\Phi_{1a} \geq \Phi_{ib}$ for any b and i . Further, if $F_{jc} \preceq^{s.s.o} F_{ib}$ for $2 \leq j < i$ then $\Phi_{jc} \geq \Phi_{ib}$.¹⁹

Proof: See Appendix A.2.

Result T.3 says that if the alternative with highest current revealed value has multiple searchable attributes, and the other alternatives have at most one, then the highest variance attribute of this alternative should be searched before any other attribute (in any alternative) so long as it has more variance than each of those attributes (in the sense of s.s.o dominance). Also, among any two alternatives that each have exactly one searchable attribute, it is better to search the alternative with higher current revealed value, so long as its attribute has more variance.

The result contains the fundamental trade-off present in the multiple-attribute search problem with multiple searchable alternatives. In particular, after an attribute is searched in alternative 1, the searcher must then decide whether to 1) search another attribute in the same alternative, 2) switch to search an attribute in another alternative, or 3) end search immediately and choose the highest valued alternative. As mentioned in the Introduction, the first feature is not present in single-attribute search problems, e.g. Weitzman (1979), and the second feature is not present in multiple-attribute problems with a single searchable alternative, e.g. Klabjan, Olszewski, and Wolinsky (2014).

Now, if one adds to this problem even just one searchable attribute to an alternative that previously only contained one, then tractability is compromised, as a dimensional reduction crucial to the proofs of Condition V and Results T.1-T.3 is no longer possible.²⁰

Nevertheless, for the sake of completeness, and because I will use it in my empirical analysis of

¹⁹That any number of alternatives with no searchable attributes are allowed is implicit as $E(V_m) = V_m$ can be interpreted as the maximum value among said alternatives. The proof in Appendix A.2 shows that the result typically holds strictly, detailing the particular (relatively uninteresting) versions of the problem in which it only holds weakly.

²⁰In particular, the dimensional reduction allows the analysis to focus on sequential search across just two attributes (see Appendix A). A similar approach is used in the proofs of related results in Weitzman (1979) and Klabjan, Olszewski, and Wolinsky (2014). To illustrate why the reduction is no longer possible, imagine the simplest problem with multiple searchable attributes in multiple alternatives: the three alternatives, respectively, have values $v_1 + \tilde{x}_{12} + \tilde{x}_{13}$, $v_2 + \tilde{x}_{22} + \tilde{x}_{23}$, and v_3 . Policy Φ_{22} , that starts search with attribute \tilde{x}_{22} , can continue searching with either \tilde{x}_{12} or \tilde{x}_{23} , depending on the realization x_{22} . By contrast, in the absence of \tilde{x}_{23} , search can only proceed with \tilde{x}_{12} . I leave this class of challenging problems for future work.

experimental search data in Section 5, I propose a more general version of Result T.3, Condition T, without proof:²¹

Condition T *Let there be $m \geq 3$ alternatives, ordered from highest to lowest expected value, w.l.o.g., i.e. $E(\tilde{V}_i) \geq E(\tilde{V}_j)$ if $i < j$. Assume that each alternative $i \in I$ has some number of unsearched attributes $S_i \subseteq A$. If $S_i \supseteq S_j$ and $i \leq j$, then $\exists a \in S_i$ such that $\Phi_{ia} \geq \Phi_{jb}$ for all $b \in S_j$.*

Condition T says that among any two alternatives that each have any number of searchable attributes, it is better to search the alternative with higher current revealed value, so long as its set of searchable attributes nests that of the other alternative.

Due to the analytical challenges posed by the general search problem, I switch to numerical and simulation analysis to support that Condition T holds more generally than under the restricted version of the general search problem presented in Result T.3. In particular, numerical computations in Appendix B show that Condition T holds for versions of the search problem with up to three alternatives, and two unsearched attributes in each. Then, in Section 6 these numerical, and the earlier analytical, results are corroborated with simulations from dimensionally large, numerically intractable problems. In particular, the problems considered have many alternatives (eight) and unsearched attributes (nine) in each, which matches the dimensions of the search data studied in Section 5.

The intuition behind Condition T is that because search can only end once an alternative has sufficiently separated itself in value above the others, it is best to first search alternatives with relatively high expected values, and relatively more remaining uncertainty; the worst thing to do is search an alternative with a relatively low expected value and relatively little remaining uncertainty, as it is the least likely to make the searcher “change her mind” about which alternative to choose.

Condition T speaks to the fundamental tension in multiple attribute/alternative search—the trade-off between the depth and breadth of search. When applying the condition to search data, there are two behaviorally distinct ways that Condition T can be violated. As mentioned in the Introduction, the first is that a subject, having decided to search another attribute, continues searching in the same alternative, but should have switched to another (*over-search*). The second occurs when the searcher switches to another alternative, but to a sub-optimal alternative (*biased*

²¹To be precise, this is a generalized version of Result T.3 in which the attributes in alternatives 2 through m are constrained there to be elements of S_1 .

switch).

Isolating the over-search violation of Condition T allows for a partial answer to the question of whether subjects search too many or too few alternatives, and too many or too few attributes within alternatives—a natural extension to the nested question commonly asked in single attribute search—whether subjects search too many or too few alternatives [Camerer \(1995\)](#).²²

Isolating the biased switch violation of Condition T allows for study of another important aspect of subjects’ search behavior. In terms of payoff consequences in the data, violations that occur in alternative switching can be particularly costly because subjects tend to search several attributes in a sub-optimal alternative once they have switched to it, which further compounds the cost. In single attribute search problems [Weitzman \(1979\)](#) characterizes which alternative should be searched next, if one is to be searched at all. The condition presented here is weaker in that it does not always identify which alternative should be searched next, but it does identify most of the cases in which the switch from one alternative to another is sub-optimal. Further, in [Section 5](#) I explain why compliance with Condition T in alternative switches is straightforward, making violations of this type particularly revealing.

Conditions R and C hold for search problems with any number of alternatives and attributes.

Condition R (*Repeat*) *No repeat look-ups.*

The no repeated look-ups condition follows directly from the assumptions of rationality and costly search.

Condition C (*Choice*) *Once search ends choose any alternative $j \in \arg \max_{i \in I} E(\tilde{V}_i)$.*

Once search ends, the chosen alternative must have a weakly higher expected value than any unchosen alternative. This condition also follows directly from the assumptions of rationality and risk neutrality. Thus, while Conditions R and C are both theoretically trivial, they hold for a broader class of search problems than those described in [Section 2.1](#). In [Section 5](#) I show that the conditions are useful for studying search behavior.

²²[Camerer \(1995\)](#) reports that a robust finding in single-attribute search problems is that subjects systematically search too few alternatives. [Bearden and Connolly \(2007\)](#) find the same result in their two-attribute, fixed order, no recall search problem. When combined with the empirical results from [Section 5](#) what is (tentatively) suggested is that when performing multiple-attribute search subjects may have a tendency to search too few alternatives and explore searched alternatives too deeply.

Necessary but not Sufficient

Consider the four necessary conditions applied to the search problem described in Section 2.1. Condition T (over-search/biased switch) identifies which alternative search should begin with, if it begins at all, and Condition V (Variance) identifies which attribute should be searched within an alternative given that the alternative is being searched. Given that a switch is occurring from one alternative to another, Condition T identifies a list of alternatives that cannot be switched to, which may or may not be exhaustive, depending on the history of search. In addition, Condition T identifies cases in which search must switch from the currently searched alternative to another. Condition C (choice) identifies the optimal alternative choice at the completion of search.

That the partial characterization of optimal sequential search that I propose here falls short of a complete characterization is due to two reasons. First, in terms of the optimal order of sequential search, it is unable to determine which alternative is better to search when, relative to another, one alternative has a lower expected value but is also less searched. As such, the partial characterization cannot always determine when search should switch out of an alternative, nor can it always make a complete prediction of which alternative should be switched to, when a switch does occur. In terms of identifying over-search within alternatives, then, the partial characterization can be thought of as identifying the more egregious violations while being blind to the more subtle ones. Second, independent of the optimal order of search, it says nothing about when search in a problem should stop.²³

Despite the necessary conditions not being sufficient, the partial characterization of optimal sequential search nevertheless allows for a strong test of the degree of optimality in multiple attribute/alternative sequential search behavior; Condition T alone would be violated 70.9% of the time in the average task if the subject were to select an alternative to search at random.²⁴ On the other hand, the corresponding sufficient condition—if it were known—would yield a violation rate of slightly less than 85.8%, meaning that Condition T identifies more than 82.6% of the violations that the corresponding sufficient condition would identify.²⁵ The remaining three necessary condi-

²³In single attribute search problems Weitzman (1979) is able to characterize the optimal order of sequential search and the optimal stopping rule. As mentioned earlier in this section, Klabjan, Olszewski, and Wolinsky (2014) are able to do the same for a problem with one searchable attribute with two unsearched attributes and one deterministic outside option.

²⁴This violation rate is obtained via simulation, taking the number of attributes searched by each subject in each task, and having a robot subject search at random, restricted to obey all necessary conditions other than Condition T (see Section 6). The violation rate is first averaged across tasks within subject, then across subjects.

²⁵This violation rate is computed directly in each subject/task by averaging $(r - 1)/r$ across look-ups, where r is the number of remaining alternatives that have not been searched exhaustively. The violation rate is then averaged across tasks within subject, and finally across subjects. It is a weak overestimate because it assumes that for each look-up

tions are sufficient within their respective categories—that is, they identify 100% of all respective violations.

3 GLMW’s Experimental Design

In GLMW’s experiment, subjects face a series of problems in which they choose one of eight alternatives, with a limited amount of time. The value of each alternative is equal to the sum of its ten individual attribute values, nine of which are initially unknown to the subject (imagine comparing eight different potential employers using information collected on attributes such as wage, benefits, commuting distance, etc.). Attributes are mean zero, normally distributed, independent, and linearly decaying in variance from attribute 1 (σ^2) to attribute 10 ($0.1\sigma^2$). Attributes can be searched in any order and any alternative can be chosen at any moment, to end search. By limiting the amount of time that subjects have to search GLMW create a shadow cost of time in both of their experimental treatments. In the “Endogenous” treatment subjects are given 25 minutes to complete as many different problems as they choose to, given a 20 second buffer screen between problems. In the “Exogenous” treatment subjects are allocated between 10 and 49 seconds, drawn from a uniform distribution, to complete each of 12 different problems.²⁶ The design is within subject, so each subject completes both of the treatments, with half of the subjects starting in the Exogenous, and the other half in the Endogenous, in order to control for order effects.

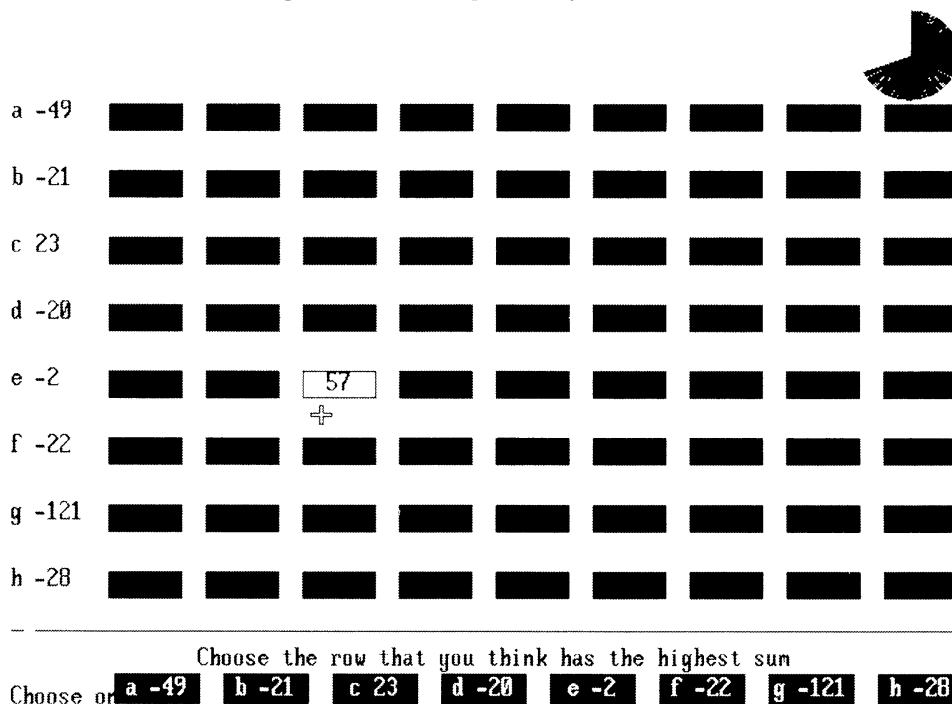
The experiment’s 390 Harvard undergraduate subjects are given complete information of the distribution of all attributes, and the amount of time remaining in each problem. Subjects are paid the sum of attributes 1 through 10 in the alternative they choose, regardless of how many of these attributes they searched.

Building on work by [Johnson, Payne, Bettman, and Schkade \(1989\)](#); [Payne, Bettman, and Johnson \(1993\)](#) GLMW use the MouseLab experimental interface in order to track the complete order and duration of all information acquisitions made by subjects. MouseLab is essentially a mechanical analog to eye-tracking, in which attribute values are contained within “boxes” on the

there is always a unique best attribute to search.

²⁶When analyzing GLMW’s data in Section 5 I follow the reasonable assumption made by GLMW for their endogenous treatment: that there is a constant shadow cost of time. Further, because I find little difference in search behavior between the Exogenous and Endogenous treatments I use the shadow cost assumption when analyzing the pooled dataset. Repeating the analysis without including data from the Exogenous treatment yields similar results. Assuming constant costs is a reasonable approximation of more complicated potential cost structures, in which costs could change over time, spatially over the information display, or even depend on the values of previous attribute realizations. Similarly, one could consider a model of search in which the searcher has a limited time budget and time cost per look-up, with varying search horizons as time decreases.

Figure 1: Search problem from GLMW



Alternatives 1-8 run from top to bottom. Attributes 1-10 run from left to right. Attribute 1 values are observable for the duration of the problem. Of the remaining 72 attributes, only one can be observed at a time. In this particular moment of search the subject is observing attribute 4 of alternative 5. Choice boxes corresponding to each of the alternatives can be seen lining the bottom of the display. The amount of time remaining for each problem is continuously represented by the decaying disc in the top right corner of the display.

computer screen that can be opened, one at a time, by clicking the left button of the computer mouse (see Figure 1). Each box must be closed, with a right-click of the mouse, before the next attribute box can then be opened. In GLMW’s design attribute 1 values for each alternative are openly observable for the entirety of the search problem, whereas attributes 2 through 10 are covered (unless one is opened). At the bottom of the display are choice boxes—one for each alternative. In order to choose an alternative after search is completed, one must left-click once on the corresponding choice box to choose it, and then again to confirm the choice.

The GLMW search problem satisfies all of the conditions of the slightly more general search problem presented in Section 2.1. Thus, the necessary conditions for optimal sequential search presented in Section 2.2 can be applied to the GLMW’s experimental dataset in Section 5.

4 Basic Search Behavior and Performance

I give a descriptive analysis of the basic properties of search behavior in the average problem. All averages are taken first across all of the search problems of a subject, then across subjects. First, search is described in terms of breadth, depth, speed, and the corresponding monetary earnings obtained by subjects, then in terms of the types of search transitions that occur from one searched attribute to the next.

On average, subjects take 1.0 second for each attribute look-up, and perform 31.4 look-ups per search problem, which on average span 4.9 different rows and 7.8 different columns.²⁷ Further, on average, they complete 41.8 problems, and earn \$0.53 per problem. Comparing this to an extreme benchmark, if for each problem they were able to resolve enough uncertainty to always choose the highest-valued alternative, they would earn an average of \$0.75.

Tables 1 and 2 together summarize the characteristics of search transitions from one attribute to the next. Table 1 reports, for the average search problem, that subjects transition within alternative (horizontal) 16 times as frequently as they transition within attribute (vertical). Of the relatively small percentage of search transitions that occur neither within alternative nor within attribute, the majority are identical in one important respect—they are transitions to the first attribute of a transitioned-to alternative as part of a “typewriter” acquisition pattern.²⁸

Table 1: *Search transitions: horizontal, vertical, and other*

Transition Type	Mean
Within Alternative	25.4
Within Attribute	1.5
Other(typewriter)	3.7 (3.4)

Table 2 reports the frequency of search transitions to each of the four cardinal directions, along with the corresponding extent to which these transitions are spatially adjacent. The majority of these transitions occur from left-to-right (88%) and are adjacent (97%).

The results of this unstructured analysis are straightforward. In general, search sequences are smooth and highly systematic. On average, 99 out of every 100 look-up transitions occur either within alternative, within attribute, or as an example of deliberate typewriter alternative switching.

²⁷In the GLMW problem there are 72^n possible search paths for every n boxes searched.

²⁸This typewriter search pattern typically performs multiple sequential look ups in one alternative going left-to-right, then switches to the left-most attribute of any other alternative, followed by another run of left-to-right look-up transitions, and so on.

Table 2: *Search transitions: direction and degree of adjacency*

Transition Type	Mean
Right	23.7
(adjacent)	(23.3)
Left	1.6
(adjacent)	(1.4)
Down	1.1
(adjacent)	(1.0)
Up	0.4
(adjacent)	(0.3)

5 Violations of the Necessary Conditions

I compare the search behavior observed in GLMW’s experiment to the partial rational benchmark for optimal sequential search introduced in Section 2.2. I find behavior to deviate from optimal sequential search at high rates. In particular, 94% of all search problems, and 57% of all search actions within search problems violate at least one of the four necessary conditions.²⁹ More importantly, these deviations are systematic. Consistent with the finding in Section 4 that subjects’ search patterns tend to be “smooth” and highly systematic, this section reports that they sequentially search too deeply in alternatives and switch too adjacently between alternatives. Interpretations of violations of necessary conditions will, for the most part, be postponed until Sections 6 and 7.

Table 3 reports subjects’ summary statistics, averaged first across tasks within subject, then across subjects. From left to right, columns contain (N) the number of subjects, (Time) end time of search task, (LU’s) number of attribute look-ups; also, the violation rates of (1) Condition T—Trade-off, (T↔) Condition T violated by over-search in an alternative, (T↕) Condition T violated by a biased switch between alternatives, (2) Condition V—Variance, (3) Condition R—Repeat, (4) Condition C—Choice, (TUVUR) attribute look-ups in which at least one of Conditions T, V, or R is violated, and (Payoff) the expected payoff (in dollar-cents). For Conditions T, V, and R (%) represents the percentage of violations on all look-ups that permit a violation, and for Condition C the percentage of violations on all final choices.

²⁹Search actions consist of all attribute look-ups, and choice of alternative. Averages are taken first across tasks within subject, then across subjects.

Table 3: *Violations of necessary conditions in search behavior*

N	Time secs.	LU's	T %	T \leftrightarrow %	T \updownarrow %	V %	R %	C %	TUVUR %	Payoff \$-cents
390	32.4	31.3	48.6	47.2	54.9	5.6	12.0	30.8	57.0	53.4

5.0.1 Condition V

Because all attributes are symmetric second-order stochastically dominated by any attribute located to their right in GLMW’s design, a subject should never search an attribute that has an unsearched attribute to its left, in the same alternative. A small fraction of all attribute look-ups, roughly 6%, on average, violate Condition V.³⁰

5.0.2 Condition T

A violation of Condition T occurs in 49% of subjects’ look-ups, on average.³¹ To put this violation rate into perspective, searching alternatives at random would yield an average violation rate of 71%, thus subjects’ 49% violation rate is substantial.³² All violations occur in the form of either over-search within an alternative, or an incorrect (“biased”) switch from one alternative to another. Because these two forms of violation are behaviorally distinct, I explore each separately.

Within Alternative Transitions

A violation of Condition T occurs on a within alternative transition if the currently searched alternative is explored more deeply when the subject would have been better off switching to explore a (weakly) less searched alternative instead. Table 3 reports that this type of violation occurs on 47% of all within alternative transitions.³³

An additional test of the degree of sensitivity to acquired information, i.e. contingency, in within alternative search is whether subjects manage to switch out of an alternative when their current look-up is not a violation of Condition T, but their next look-up would be.³⁴ Subjects fail to switch out of the alternative at an alarming rate of 94% of these look-ups, further indicating

³⁰The fact that most subjects are used to reading left to right, which is consistent with complying with Condition V, may contribute to the low violation rates.

³¹All averages are taken first across tasks within subject, then across subjects.

³²The 71% violation rate is obtained by simulated search that randomly selects an alternative/attribute without adhering to Condition T, but adhering to Conditions V and R, and with the number of simulated look-ups yoked to the number actually performed by each subject in each task (see Section 6).

³³This constitutes 79% of all violations of Condition T.

³⁴In particular, this occurs when the current attribute realization has made its alternative both weakly more searched than, and have an expected value that is weakly less than, with one of these inequalities strict, another alternative.

their lack of optimal responsiveness to obtained information.

Alternative Switching

Table 3 reports that subjects' average violation rate of Condition T, on alternative switches, is 55%. As mentioned in Section 2, switching violations are of special interest because once an alternative is switched to subjects tend to search several of its attributes in sequence, which exacerbates the cost of having switched to a sub-optimal alternative.

The remaining analysis of alternative switches will limit consideration to those that occur to previously unsearched alternatives. By doing so, I avoid comparing these switches, which allow the searcher a completely transparent view of the alternative's expected value, with switches to alternatives whose (revealed) expected value is not currently observable, thus not transparent. Violations on this type of switch are particularly revealing, given how straightforward it is to avoid them in GLMW's design. In particular, the subject need only switch to the alternative with the "largest number in front of it"—its fully observable attribute value (see Figure 1 in Section 4). Empirically, this is a weak restriction, as the selected sub-sample accounts for 92% of all alternative switching violations of Condition T.³⁵

Despite the relative ease, and importance, of not violating Condition T when switching between alternatives, 53% of subjects' alternative switches are violations, on average. Moreover, as the ex post number of alternatives searched within a problem increases, the average relative frequency of violations increases dramatically. This pattern can be seen in Table 4, where the (ex post) number of alternatives searched in a problem are listed as rows, and the switch numbers (first, second, . . . , eighth) within the problem, are listed as columns. The average violation rate of Condition T for each switch number and ex post number of alternatives searched is reported in columns labeled 1-8, while the final two columns give the number of subjects (N) with at least one problem in which the corresponding (ex post) number of alternatives is searched, and the number (# Opt.) of subjects that in these problems perfectly comply with Condition T when switching alternatives. The probability of violating the condition by chance, for each switch number, is shown in the bottom row of the table.

The average violation rate for each pairing of alternative switch number and ex-post number of alternatives searched in Table 4 is significantly larger than zero ($< .001$, Wilcoxon signed rank test), except for the eighth switch, which has to be equal to zero. Furthermore, the average violation

³⁵In general, 82% of all alternative switches occur to previously unsearched alternatives; subjects do not seem to like switching back to alternatives that they have partially searched previously.

Table 4: Average relative frequencies of violations of Condition T on alternative switches, as a function of the ex post number of alternatives searched in a task, and whether it is the first, second, . . . , eighth switch

Alternatives Searched	Switch Number								N	#Opt.	
	1	2	3	4	5	6	7	8			
1	0.32									286	104
2	0.44	0.35								333	31
3	0.49	0.51	0.38							344	8
4	0.50	0.55	0.52	0.41						351	1
5	0.52	0.58	0.58	0.56	0.41					359	3
6	0.56	0.63	0.59	0.60	0.55	0.42				348	3
7	0.54	0.62	0.65	0.65	0.58	0.57	0.33			339	3
8	0.66	0.72	0.72	0.72	0.65	0.62	0.41	0.00		329	1
Chance	0.87	0.86	0.83	0.80	0.75	0.67	0.50	0.00			

rate increases consistently as the ex post number of alternatives searched in a task increases. While tasks with only one searched alternative yield a 32% violation rate, when eight alternatives are searched the first alternative switch violates Condition T 66% of the time. The first average violation rate in each column is significantly different from the last ($< .01$, Wilcoxon signed rank test), whenever there are at least two rates to compare.³⁶ In addition, very few subjects maintain compliance to Condition T throughout all alternative switches, for search problems with any number of alternatives searched. In particular, only 36% of subjects do so for all search problems in which they search only one alternative, whereas when three or more alternatives are searched the rate of compliant subjects drops to close to zero.

That compliance rates decrease systematically with the ex post number of alternatives searched indicates that search behavior becomes increasingly non-contingent as more alternatives are searched. In order to discern whether these violations of optimal sequential search contain any systematic patterns, I next conduct an econometric analysis.

Econometric analysis of alternative switching

An analysis performed in Appendix C using conditional logit regressions strongly suggests that violations of Condition T on alternative switches occur because subjects are influenced by (“irrational”) spatial preferences roughly as much as by (rational) value preferences. In particular, when choosing which alternative to search first, searchers respond roughly as much to the spatial *height* of an alternative as to its expected *value*. Then, in subsequent alternative switches, searchers respond

³⁶In general, within-column differences in average violation rates of around .05 and up are statistically significant at the 5% confidence level (two-sided test).

roughly as much to the degree of spatial *adjacency* (i.e. “closeness”) of a candidate alternative as to its expected value.³⁷ Thus, subjects exhibit a strong spatial bias when choosing which alternative to search. Section 7 contains a discussion of how such a spatial bias may be due primarily to subjects adapting in order to reduce memory load. See Section 6 for an analysis of the costs of these systematic deviations from optimality.

5.0.3 Condition R

Due to the scarcity of time in both experimental treatments, a subject should never repeat an attribute look-up within a problem. Nevertheless, 12% of all look-ups, on average, are repeats.

5.0.4 Condition C

When done searching subjects should always choose the alternative with the highest expected value, but fail to do so in 31% of problems.³⁸

6 The Costliness of Violations

I perform an analysis of the effects of violations of necessary conditions on the (1) value of information obtained in search, and (2) economic performance of experimental subjects. Results suggest that despite the high violation rates reported in Section 5, subjects do search relatively valuable information, just often not in the order prescribed by optimal sequential search. Further, the costs of failing to choose the highest-valued observed alternative at the completion of search (i.e. violations of Condition C) are large relative to the costs associated with violations of Conditions T, V, and R.

6.1 Effect of Violations on Value of Information

The more valuable the information that one searches, the greater the tendency to observe more promising alternatives, and *visa versa*. Thus, a natural way to measure the value of information

³⁷The GLMW dataset is sufficiently rich to run regressions on the individual subject level. Effects on spatial independent variables (height, distance) are significant for around 290/390 subjects. The effect of value is significant for roughly the same number of subjects. Average effect sizes reflect considerable trade-offs between spatial variables and value, e.g. for the first alternative searched, a subject is willing to search an alternative one position lower in the display if its observable attribute 1 value is at least \$.07 higher.

³⁸It is reasonable to assume GLMW’s subjects are risk neutral given that GLMW encourage risk neutrality by reminding subjects in each problem to “Choose the row you think has the highest sum,” as seen in Figure 1, above the choice boxes.

obtained in search is with the highest revealed (expected) payoff across alternatives. Table 5 reports the results of an ordinary least squares regression in which the dependent variable is the average highest-valued alternative (in expected \$-cents) observed by the subject, and the independent variables are the subject’s average number of violations of Condition T (over-search or biased switching), Condition V (Variance), Condition R (repeats), and number of look-ups, respectively, where the averages are taken across all of a subject’s tasks. The natural predictions are that violations of any of the necessary conditions will lead the value of the highest-valued observed alternative to decrease, while performing additional look-ups will have the opposite effect.

Table 5: *OLS regression results for the effect of necessary condition violation rates (and the number of look-ups) on the highest-valued alternative observed (in \$-cents)*

	Coefficient	Standard Error	p-value
Constant	51.424***	1.347	0.000
Look-ups	1.091***	0.122	0.000
Condition T	-1.131***	0.178	0.000
Condition V	0.044	0.124	0.722
Condition R	-1.341***	0.214	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$n = 390$, $\bar{R}^2 = 0.23$

Consistent with predictions, the results indicate that the larger the number of look-ups, the higher the highest-valued alternative, and the larger the number of look-ups that violate Condition T, or Condition R, the lower the highest-valued alternative. These effects are highly significant, with considerable magnitudes: in the average task, subjects perform roughly 14 violations of Condition T, and each violation is associated with a decrease in the highest-valued observed alternative of more than a cent.³⁹ The magnitude of the effect of repeating a look-up is similar, but subjects only do this around 3.6 times per task, on average. The coefficient on violations of Condition V, which subjects perform around 1.6 times per task, on the other hand, is insignificant. Finally, each additional look-up is associated with an increase in the highest-valued alternative of a bit more than a cent.⁴⁰

³⁹One might be concerned about multicollinearity, especially between the violation frequencies for each necessary condition, and the total number of look-ups. By converting the frequencies to relative frequencies, the degree of multicollinearity is greatly reduced. This yields no qualitative difference in regression results, other than perhaps the sign on the effect of Condition V switching to negative (though still insignificant), as initially predicted. Thus, the independent variables are left as frequencies in order to facilitate interpretation.

⁴⁰An alternative specification in which the total number of search problems the subject faced is also included as a regressor yields virtually identical results, with the coefficient on number of games roughly zero and insignificant, whether pooling the data across treatments or only including data from the Endogenous treatment.

While the regression results reported in Table 5 indicate that subjects who violate Conditions T and R obtain less valuable information from search, there is reason to caution against a causal interpretation. In particular, each subject’s violation rates of each condition depend on the unobservable search rule(s) that she employs, as does the highest-valued alternative that is observed.⁴¹ Also, for a brief discussion on multicollinearity see footnote 39.

In light of the possible endogeneity issues just described, I use a simulation approach in order to allow for causal interpretations of the effects of violations of necessary conditions (and additional look-ups) on the value of information obtained by search. First I simulate search for each subject in each task, according to how much the subject searched in that task, but varying the search rule used from completely random, to random but without violating: Condition R, Conditions R or V, Conditions R, V, or T (along with a variant), respectively. This reveals the effect of violating each condition on the value of information obtained, subject to the same general conditions of search faced by the experiment’s subjects in each task. 100 complete simulated search patterns are run for each subject/task pairing, with the highest-valued observed alternative first being averaged across simulations within task, then across tasks for each subject, and finally across subjects, creating one measure of each variable of interest, and allowing direct comparison with the averages reported in Table 3.⁴²

Table 6: *The highest-valued observed alternative (in \$-cents) as a result of various random search rules that incrementally (and increasingly) obey necessary conditions for optimal sequential search, along with corresponding values for an idealized omniscient searcher and experimental subjects.*

Search Rule	Highest-Valued Observed Alternative
No Conditions	49.0
Condition R	52.3
Condition R&V	58.3
True Search	65.5
Condition R&V&T	66.1
Condition R&V&T*	68.3
Omniscient	75.2

The results for each of these simulated search behaviors can be seen in Table 6, where the (average) highest-valued observed alternative is reported. For each search rule, this value can then be compared to the extreme of the true highest-valued alternative if *all* information in each task

⁴¹This may be related to the estimated effect of violating Condition V being found to be insignificant.

⁴²Because simulated search patterns are highly contingent (i.e. depend on search history), and I generate 100 for each of the nearly 16,000 search tasks in the dataset, results for a particular search rule can take up to around 20 days of computation time.

were to be observed, which I call *Omnipotent*. Given that roughly 31 of 72 possible look-ups are performed in the average task, one can see that completely random search (*No Conditions*) yields a highest-valued observed alternative of 49.0, as compared to the 75.2 of Omnipotent search. Between these two extremes, as random search becomes incrementally more restricted to obey necessary conditions for optimal sequential search, the highest observed alternative value can be seen to increase incrementally. In addition, the *True Search* of experimental subjects yields only slightly lower highest-valued observed alternatives than random search restricted to obey Conditions R, V, and T (*Condition R & V & T*).⁴³

That subjects observe information that is almost as valuable as that observed by the simulated search rule that complies with Conditions R, V, and T may at first glance seem surprising, given: (1) the high violation rates of necessary conditions in search behavior reported in Table 3, (2) the regression results reported in Table 5 indicating that these violations are costly, and (3) the simulation results of Table 6 now reporting the same. How might we reconcile these seemingly contradictory results? First, we can recall that the necessary conditions for optimal sequential search fall short of sufficiency, meaning that fully optimal search would result in more valuable information being obtained than that by random search restricted to obey Conditions R,V, and T. To illustrate this point, one can consider, for example, a minor tweak to random search obeying Conditions R,V, and T, in which it must also obey the following restriction: the searcher continues in the presently searched alternative until this is no longer possible. This search rule, *Condition R&V&T** in Table 6, results in an average relative increase in the highest-valued observed alternative of around 2 cents per task.⁴⁴

A second reason why the difference between the highest-valued alternative observed by true searchers, and simulated random search obeying Conditions R,V, and T is perhaps not as large as expected, is that when search is completed in each task subjects’ ex post search patterns, i.e. “footprints,” overlap with simulated footprints more than what is suggested by the violation rates reported in Table 3. For example, when simulating random search in each task of each subject, in the absence of violations of Conditions R,V, and T, the *relative overlap* with true search is roughly 77%, though the fact that around half of look-ups violate Condition T alone—as compared to the roughly 71% chance violation rate—may appear to suggest otherwise. By contrast, the relative

⁴³I use the Mann-Whitney Signed Rank test (5%) to test for equality of means. All pairwise differences are significant.

⁴⁴The intuition behind why this tweak results in more valuable information being obtained is that there are instances in which Condition T allows search to continue in the current alternative, but also allows a switch to an alternative with relatively (very) low observed expected value, but that is relatively less searched. The random search rule (unless otherwise restricted) chooses between these two options with equal probabilities.

overlap with fully random search is roughly 44%.⁴⁵ What this suggests is that while subjects do severely violate the *order* of optimal sequential search, the search footprint that they leave is more similar to violation-free search than what it initially appears.

So far in the simulation analysis I have compared the value of information obtained by different search rules when they each perform precisely the number of look-ups made by experimental subjects. However, one may also wonder how the value of information for each simulated search rule evolves with the number of look-ups. Accordingly, I run another battery of simulations in which I take each of the 160 different search tasks used in GLMW’s experiment, and simulate search in each 1000 times, allowing zero look-ups, then one, then two, . . . , all the way up to 72 look-ups, i.e. exhaustive search.⁴⁶ I do this for each search rule, which allows for a comparison of the value of information revealed, all else equal, when necessary conditions are violated or not, as a function of the extent of search performed. In Figure 2 one can see that, as in Table 6, the costs due to violating each of the necessary conditions for optimal sequential search are considerable for a large range of number of look-ups.⁴⁷

6.2 Effect of Violations on Economic Performance

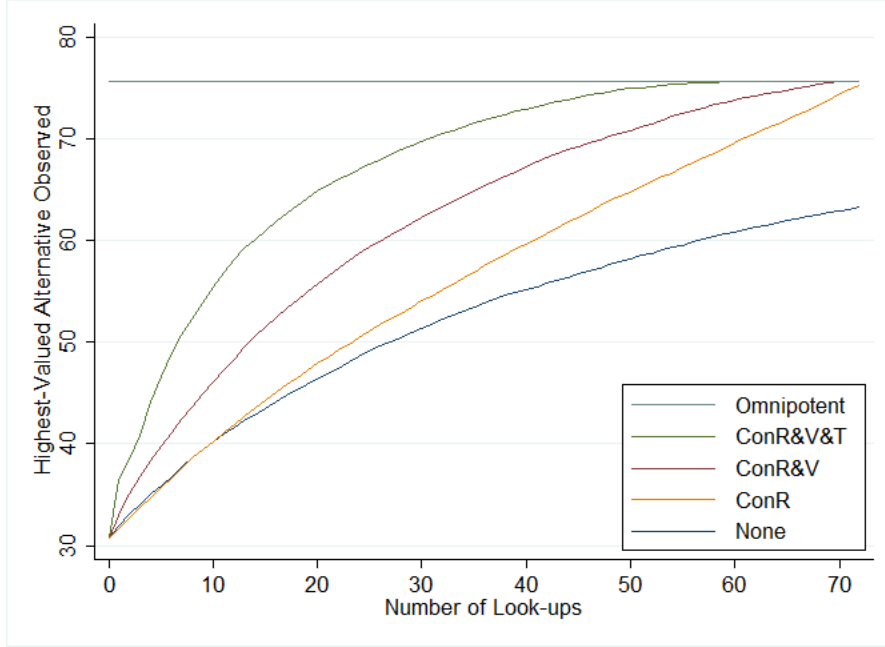
The highest valued alternative observed in a search problem is equal to the searcher’s expected payoff so long as she successfully chooses this alternative. However, as reported in Section 5, subjects fail to choose this alternative in approximately one out of every three tasks, on average. These violations of Condition C are costly; whereas Table 6 shows that subjects’ search, on average (across all tasks), reveals a highest valued alternative of \$0.66, the average score obtained by

⁴⁵The way that I calculate these relative overlap percentages starts with a measure of absolute overlap between two search “footprints”: $overlap = \frac{\#(a \cap b)}{\min\{\#a, \#b\}}$, where $i \in \{a, b\}$ is the set of look-ups in a task according to search rule i , and $\#i$ the number of look-ups in the set. Notice that this measure equals zero if there is no overlap, and 1 if the set with (weakly) fewer look-ups is entirely contained in the other. Then, I use each search rule of interest to simulate a pattern of search for each of the nearly 16,000 tasks searched by experimental subjects. For each task, I compute the overlap between the true subject search pattern and the simulated search pattern (with the number of simulated look-ups equal to those of the corresponding subject/task), and then average this overlap across tasks within subject, and finally across subjects. I find the average absolute overlap between completely random search and true search to be 0.38, and when random search is restricted to obey Conditions T,V, and R the overlap goes to 0.67. Finally, to obtain the relative overlaps I use the absolute overlap between simulated random search restricted to obey Conditions T,V, and R, and itself—roughly 0.87— as the divisor, which yields the relative overlaps of 44%, and 77%, respectively.

⁴⁶In order to vastly reduce computation time, rather than conducting simulations independently for each number of look-ups, an approximation is used in which each of the 160 different tasks is first searched exhaustively in 1000 random orders, then the last look-up is removed from each, then another, etc., until all simulated search patterns perform just one look-up, then zero. Clearly, as the number of simulated orders grows arbitrarily large, the output of the approximation converges to that of the alternative method. Estimates for 100 and 1000 simulated orders are nearly identical.

⁴⁷The results of Table 6 and Figure 2 corroborate the costliness of violations of Condition T, with the theoretical results of Section 2 as well as with the numerical computations performed in the Appendix A.

Figure 2: *Highest-valued alternative observed as search rule and amount searched vary*



Average highest-valued alternative observed as a consequence of simulated search 1000 times on each of the 160 different search tasks in GLMW, as random search incrementally and increasingly obeys the necessary conditions for optimal sequential search presented in Section 2, and as the total number of look-ups vary from 0 to 72.

subjects is shown in Table 3 to be considerably less: \$0.53.

Table 7: *OLS regression results for the effect of necessary condition violation rates (and the number of look-ups) on subjects' average payoffs per task (in \$-cents)*

	Coefficient	Standard Error	p-value
Constant	57.715	1.811	0.000
Look-ups	0.889***	0.145	0.000
Condition T	-0.699***	0.217	0.001
Condition V	0.046	0.147	0.755
Condition R	-1.154***	0.254	0.000
Condition C	-60.437***	3.954	0.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$n = 390$, $\bar{R}^2 = 0.44$

Table 7 reports the results of an OLS regression that is similar to that presented in Table 5, but with a dependent variable of average payoffs per task, rather than highest-valued observed alternative, and with the addition of an independent variable indicating whether a violation of Condition C, i.e. an incorrect choice, has occurred. One can see that the size and significance of the effects on look-ups and violations of Conditions T and R are similar to those presented in Table

5, while the effect of incorrect choice is also significant, but substantially larger. In particular, failing to choose the highest-valued observed alternative is associated with a loss of \$0.60, which is larger than the average payoffs per task.⁴⁸

7 Discussion

At first glance, subjects' high rates of within alternative over-search, and their tendency to switch too adjacently between alternatives, suggests that they deviate substantially from searching the most valuable information. However, the analysis performed in Section 6.1 suggests that subjects' search "footprints" are actually quite similar to the footprints left by violation-free search. This in turn suggests the possibility of some form of boundedly rational search behavior.

In order to investigate what might be causing the observed deviations it is useful to begin by considering the standard candidate explanations: myopia and costly-reoptimization. In particular, because the GLMW search space is so large, the option value calculations necessary to search optimally are intractable, even for modern computers.⁴⁹ Myopia makes for an intriguing candidate explanation because myopically optimal sequential search is computationally much less demanding than its non-myopic counterpart, due to its complete disregard of option values. In addition, it yields a sharp prediction of search behavior. Nevertheless, it does not predict the alternative switching violations, which occur on 55% of alternative switches, and in general seems more consistent with within alternative under-search than the high rates of over-search observed in behavior, as (1) myopia leads to less search in general, as it sets to zero the option value of search, which is usually positive and never negative, and (2) myopic searchers are more often indifferent about which of multiple alternatives to search (because they use less information to discriminate between different possibilities), so they should be expected to switch more often between alternatives. In addition, myopia would not predict repeat look-ups, incorrect alternative choice, or violation of the variance condition.

On the other hand, optimal sequential searchers must re-optimize after each attribute is searched, which may be quite cognitively taxing. Thus, searchers may choose to instead re-optimize less fre-

⁴⁸In principle, it may be better to use as the dependent variable average highest observed payoff divided by the average highest possible payoff, in order to adjust for the fact that some subjects face tasks with more promising alternatives than other subjects. Nevertheless, because regression results with this dependent variable are similar, the dependent variable of average payoffs is used instead, in order to facilitate interpretation of regression coefficients.

⁴⁹Of course, it is conceivable that there is an equivalently optimal reservation rule (as in many single attribute search contexts), but none have been found in any problem with multiple searchable attributes and alternatives, let alone in this rich of one.

quently, preferring to do so only after searching chunks of several attributes at a time. This conjecture is generally consistent with the results reported in Table 4 of Section 5, that as the ex post number of alternatives searched increases, subjects increasingly violate optimal alternative switching. That is, they appear, to some extent, to pre-commit to searching multiple alternatives at a time, which would make search order less relevant. As mentioned in the introduction, this type of behavior would be consistent with an optimal sequential searcher who sometimes lapses into “classic” search (Stigler 1961), which is generally consistent with the proposed occasional re-optimizing of the Directed Cognition model, the model of Morgan and Manning (1985), and field evidence in De los Santos, Hortaçsu, and Wildenbeest (2012). Nevertheless, the fact that the frequency of re-optimization is not observable makes it difficult to use this approach to predict search behavior. In particular, with a sufficiently high frequency of re-optimization, the occasional re-optimizer would not perform the observed within alternative over-search or alternative-switching violations, and with a sufficiently low frequency of re-optimization there are no restrictions placed on search order whatsoever, so “anything goes,” which does not make for a useful descriptive model of search behavior.

An alternative conjecture is memory considerations, which is suggested by the fact that subjects repeat look-ups 12% of the time, and fail to choose the highest expected value alternative at the completion of search in 31% of tasks. Accordingly, in Sanjurjo (2017) I build a simple model of working memory load, as a function of search behavior, and find that both within alternative over-search and adjacency bias in alternative switching greatly reduce experienced memory load. These results demonstrate that cognitive load is not only a function of what attributes are searched, or how many, but what *order* said attributes are searched in.⁵⁰ Intuitively, the more smooth and systematic is search, the less the experienced load. Further, in an experiment designed to test this model of memory load in search (Sanjurjo 2015) I find that high memory load search orders lead to substantially higher rates of choice error than low memory load orders, holding the information searched constant. Given that Section 6 shows the costs of choice error violations to be substantially larger than those of Conditions T,V, and R, it makes sense that subjects could be willing to violate optimal order of search in order to smooth and systematize search behavior, which in turn reduces the probability of memory failure, and thus costly choice error.

⁵⁰By contrast, the information overload literature typically focuses on the number of alternatives or the number of attributes searched. See, e.g. Iyengar, Huberman, and Jiang (2004); Jacoby (1984); Malhotra (1982); Reutskaja, Nagel, Camerer, and Rangel (2011)

8 Conclusion

I study rich multiple attribute search problems, similar to those ubiquitous in modern economic life. I introduce a partial characterization of optimal sequential search, in the form of several necessary conditions. When taking the partial rational benchmark to the data from a rich search experiment, I find that subjects systematically violate optimal sequential search by searching too deeply within alternatives, and switching too adjacently between alternatives. While, in principle, these violations are costly, subjects seem to mitigate the costs by searching in a way that nonetheless leads the ex post observed information to be valuable. The order of search that they opt for reduces experienced memory load, which has been found in recent research to make it less likely to commit a very costly error at the completion of search, when choosing between alternatives.

9 Acknowledgements

Many thanks to Nageeb Ali, Coralio Ballester, Neil Bearden, Carlos Cueva, Karim Chalak, John Conlisk, Richard Carson, David Eil, Patrick Fitzsimmons, Mark Machina, David Miller, Joshua Miller, Thomas Palfrey, Giovanni Ponti, Justin Rao, Collin Raymond, Ricardo Serrano-Padial, Joel Sobel, William Thomson, Yixiao Sun, and anonymous referees for insightful comments, to Xavier Gabaix, David Laibson, Guillermo Moloche, and Stephen Weinberg for providing their data, and to Vince Crawford for invaluable guidance.

References

- BEARDEN, J. N., AND T. CONNOLLY (2007): “Multi-attribute sequential search,” *Organizational Behavior and Human Decision Processes*, 103(1), 147–158.
- CAMERER, C. (1995): *Handbook of experimental economics; Individual decision making chapter*. Princeton University Press.
- CAPLIN, A., M. DEAN, AND D. MARTIN (2011): “Search and satisficing,” *The American Economic Review*, 101(7), 2899–2922.
- DE LOS SANTOS, B., A. HORTAÇSU, AND M. R. WILDENBEEST (2012): “Testing models of consumer search using data on web browsing and purchasing behavior,” *The American Economic Review*, 102(6), 2955–2980.
- GABAIX, X., D. LAIBSON, G. MOLOCHE, AND S. WEINBERG (2006): “Costly information acquisition: Experimental analysis of a boundedly rational model,” *The American Economic Review*, pp. 1043–1068.

- IYENGAR, S. S., G. HUBERMAN, AND W. JIANG (2004): “How much choice is too much? Contributions to 401 (k) retirement plans,” *Pension design and structure: New lessons from behavioral finance*, pp. 83–95.
- JACOBY, J. (1984): “Perspectives on information overload,” *The Journal of Consumer Research*, 10(4), 432–435.
- JOHNSON, E. J., J. W. PAYNE, J. R. BETTMAN, AND D. A. SCHKADE (1989): “Monitoring information processing and decisions: The mouselab system,” Discussion paper, DTIC Document.
- KLABJAN, D., W. OLSZEWSKI, AND A. WOLINSKY (2014): “Attributes,” *Games and Economic Behavior*, 88, 190–206.
- KOHN, M. G., AND S. SHAVELL (1974): “The theory of search,” *Journal of Economic Theory*, 9(2), 93–123.
- LEE, B.-K., AND W.-N. LEE (2004): “The effect of information overload on consumer choice quality in an on-line environment,” *Psychology & Marketing*, 21(3), 159–183.
- LIM, C., J. N. BEARDEN, AND J. C. SMITH (2006): “Sequential search with multiattribute options,” *Decision Analysis*, 3(1), 3–15.
- LIPPMAN, S. A., AND J. MCCALL (1976): “The economics of job search: A survey,” *Economic Inquiry*, 14(2), 155–189.
- MALHOTRA, N. K. (1982): “Information load and consumer decision making,” *Journal of Consumer Research*, pp. 419–430.
- MORGAN, P., AND R. MANNING (1985): “Optimal search,” *Econometrica: Journal of the Econometric Society*, pp. 923–944.
- NEEMAN, Z. (1995): “On determining the importance of attributes with a stopping problem,” *Mathematical Social Sciences*, 29(3), 195–212.
- PAYNE, J. W., J. R. BETTMAN, AND E. J. JOHNSON (1993): *The adaptive decision maker*. Cambridge University Press.
- REUTSKAJA, E., R. NAGEL, C. F. CAMERER, AND A. RANGEL (2011): “Search dynamics in consumer choice under time pressure: An eye-tracking study,” *The American Economic Review*, 101(2), 900–926.
- ROTHSCHILD, M., AND J. STIGLITZ (1970): “Increasing risk: 1: A definition,” *Journal of Economic Theory*, 2, 225–243.
- SANJURJO, A. (2015): “Search, memory, and choice error: An experiment,” *PloS one*, 10(6), e0126508.
- (2017): “The role of memory in search and choice,” *Available at SSRN 2479561*.
- SHI, S. W., M. WEDEL, AND F. R. PIETERS (2013): “Information acquisition during online decision making: A model-based exploration using eye-tracking data,” *Management Science*, 59(5), 1009–1026.

- STIGLER, G. J. (1961): "The economics of information," *The Journal of Political Economy*, 69(3), 213–225.
- TVERSKY, A. (1972): "Elimination by aspects: A theory of choice.," *Psychological review*, 79(4), 281.
- WEITZMAN, M. L. (1979): "Optimal search for the best alternative," *Econometrica: Journal of the Econometric Society*, pp. 641–654.

A Proofs

In Subsection A.1 I prove Condition V and Corollary V2, and in Subsection A.2 Results T.1, T.2, and T.3, all under the assumptions described in Section 2.1.

For simplicity, in the proofs all attributes \tilde{x}_{ij} are assumed to satisfy the minimal condition that $\int_0^\infty \tilde{x}_{ij} dF_{ij} > c$, which ensures that there exists a unique *reservation distance* r_i , such that $v_i + r_i = -c + \int_{-\infty}^{r_i} (v_i + r_i) dF_{ij} + \int_{r_i}^\infty (v_i + \tilde{x}_{ij}) dF_{ij}$, which can be re-expressed as $c = \int_{r_i}^\infty (\tilde{x}_{ij} - r_i) dF_{ij}$.⁵¹ Call $[v_i - r_i, v_i + r_i]$ the *reservation interval* of alternative i .

In order to keep the length of the section down, and minimize the repetition of arguments, later proofs to some extent build on earlier proofs.

A.1 Condition V

Condition V (Variance) *Let there be $m \geq 2$ alternatives with expected values $E(\tilde{V}_1), \dots, E(\tilde{V}_m)$, and at least one alternative $i \in I$ with at least two unsearched attributes, i.e. $|S_i| \geq 2$. For any such alternative, if $F_{ia} \prec^{s.s.o} F_{ib}$ then $\Phi_{ia} > \Phi_{ib}$.*

Proof. Consider a search policy that follows any given history of search by searching the lower variance attribute \tilde{x}_{ib} of alternative i instead of \tilde{x}_{ia} , where $F_{ia} \prec^{s.s.o} F_{ib}$, then proceeds optimally thereafter. Call the search policy Ψ_b and its expected payoffs Φ_b , setting $i = 1$, then suppressing the 1 in notation, without loss of generality.

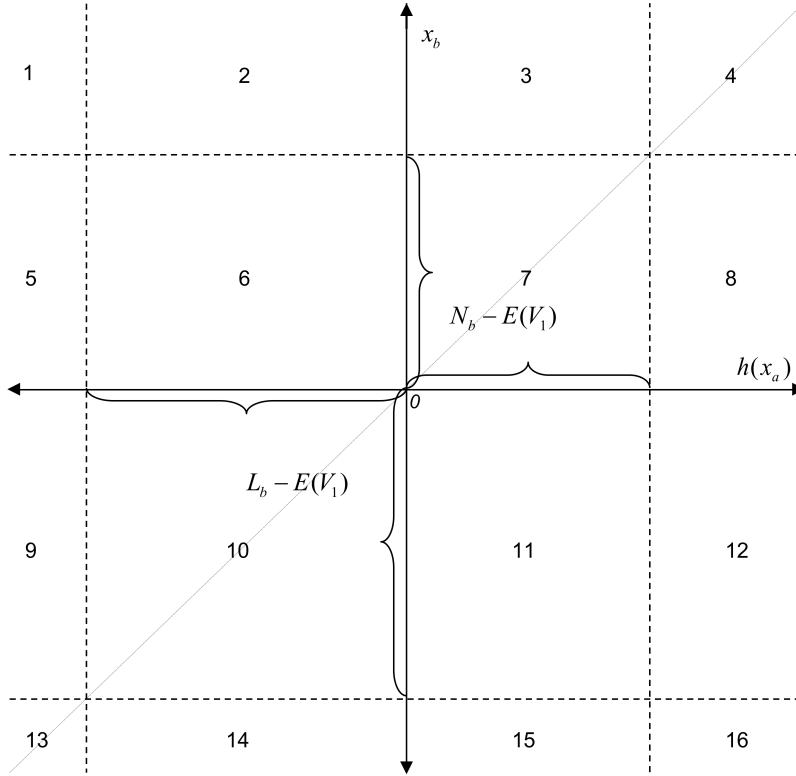
Now consider a search policy Δ_a , with expected payoffs Λ_a , that instead begins by searching the higher variance attribute \tilde{x}_a . Δ_a strictly follows the decision rules of Ψ_b , but uses $h(x_a)$ in place of x_b , until it reaches \tilde{x}_b in search (if it does). Recall from Section 2.1 that h effectively “compresses” attribute \tilde{x}_a so that $h(\tilde{x}_a)$ has the same distribution as \tilde{x}_b . In the case that \tilde{x}_b is reached, Δ_a switches to perfectly mimicking the decision rule of Ψ_b for the remainder of search. Notice that, in principle, Δ_a is not an optimal policy, i.e. $\Lambda_a \leq \Phi_a$.

The remainder of the proof consists of showing that when integrating over the joint probability density function $f(h(x_a), x_b)$, Λ_a is at least as large as Φ_b . This will be done by comparing Λ_a and Φ_b across regions that are formed by exploiting the symmetry of the joint density function, i.e. $f(h(x_a), x_b) = f(x_b, h(x_a))$.

In order to build these symmetric regions of integration, begin by considering policy Ψ_b , which

⁵¹The concept of a reservation distance is well-known and widely used in the search literature; see, e.g. [Weitzman \(1979\)](#).

Figure 3



searches x_b first. By Lemma 1 (below) there exists some N_b such that if $x_b \geq N_b$ then search ends immediately and alternative 1 is chosen. Recall that the current revealed value of alternative 1 is $E(V_1)$. Then for realizations of x_b such that $x_b \geq N_b - E(V_1)$ alternative 1 is chosen immediately. Figure 3 depicts the $(h(x_a), x_b)$ plane with critical distance $N_b - E(V_1)$ bounding regions 1, 2, 3, and 4 from below. By construction, the respective critical distance when first searching x_a , under policy Δ_a , is identical, and shown in Figure 3 to bound regions 4, 8, 12, and 16 from the left. On the other hand, while in some search problems there exists an L_b such that if $x_b \leq L_b$ then search ends immediately and the alternative with highest revealed sum is chosen (i.e. any $j \in \arg \max_{i \in I} E(\tilde{V}_i)$), in other search problems there does not exist such an L_b . The proof for the latter case, in which there are relatively fewer regions to consider, follows directly from the proof of the former. Thus, consider the critical distance $L_b - E(V_1)$ for negative realizations of x_b , which is shown in Figure 3 to bound regions 13, 14, 15, and 16 from above. It then follows that the same critical distance bounds regions 1, 5, 9, and 13 from the right. Given the horizontal and vertical axes, this makes 16 regions in total. Finally, in Figure 3 a dotted line corresponding to $x_b = h(x_a)$ reflects symmetric equiprobable regions on either side. What follows is to compare the expected

payoffs of Δ_a and Ψ_b across these regions.

For pairs of attribute realizations in Regions 4 and 13 both policies Ψ_{22} and Δ_{12} choose the same alternative after the first attribute is searched, yielding the same expected payoffs.

On the other hand, in the remaining regions search decisions differ across Δ_a and Ψ_b , and comparing Λ_a and Φ_b is tedious.

For Regions 6 and 11, call a pair of attribute realizations $(h(x_a), x_b)$ from Region 6 $(h(x_a^6), x_b^6)$, and the equiprobable pair of realizations from Region 11 $(h(x_a^{11}), x_b^{11}) := (x_b^6, h(x_a^6))$. Notice that $h(x_a^6) < x_b^6$. Either policy Ψ_b or Δ_a can observe either $h(x_a^6)$ or x_b^6 first, with equal probabilities. Given that each policy continues with search after observing the first attribute realization, there are two possibilities: either alternative 1 is (eventually) chosen, or another alternative is chosen. Clearly, the probability that policy Ψ_b chooses alternative 1 is higher with x_b^6 than with x_b^{11} , as $x_b^6 > h(x_a^6) = x_b^{11}$. Let p_h be the probability that Ψ_b chooses alternative 1 after observing x_b^6 , so that with probability $(1 - p_h)$ an alternative other than 1 is chosen. Call the expected payoffs of this other alternative v_{oh} . Similarly, let p_l be the probability that Ψ_b chooses alternative 1 after observing x_b^{11} , so that with probability $(1 - p_l)$ an alternative other than 1 is chosen. Call the expected payoffs of this other alternative v_{ol} . As mentioned, $p_h > p_l$. Now consider the expected payoffs of Ψ_b for each pair of attribute realizations in Regions 6, and 11, respectively: $\Phi_b^6 = p_h(v_1 + h(x_a^6) + x_b^6) + (1 - p_h)(v_{oh})$, and $\Phi_b^{11} = p_l(v_1 + h(x_a^{11}) + x_b^{11}) + (1 - p_l)(v_{ol})$. The expected payoffs for Δ_a are similar, but notice that whereas Ψ_b chooses alternative 1 with p_h in Region 6, and with p_l in Region 11, Δ_a does just the opposite. In particular: $\Lambda_a^6 = p_l(v_1 + h(x_a^6) + x_b^6) + (1 - p_l)(v_{ol})$, and $\Lambda_a^{11} = p_h(v_1 + h(x_a^{11}) + x_b^{11}) + (1 - p_h)(v_{oh})$. Thus, $\Lambda_a^6 + \Lambda_a^{11} - (\Phi_b^6 + \Phi_b^{11}) = (p_h - p_l)[(h(x_a^{11}) - x_b^6) - (h(x_a^6) - x_b^{11})] > 0$. The inequality results from $p_h > p_l$, and $h(x_a^{11}) = x_b^6 > h(x_a^6)$ combined with the stretching property.

For Regions 7 and 10, consider the pair of Region 7 attribute realizations $(h(x_a^{7.1}), x_b^{7.1})$, along with the equiprobable pair of realizations $(h(x_a^{7.2}), x_b^{7.2}) := (x_b^{7.1}, h(x_a^{7.1}))$, also from Region 7. To begin with, assume that $h(x_a^{7.1}) < x_b^{7.1}$. Now, the same argument used for the equiprobable pairs of attribute realizations in Regions 6 and 11 can be applied to the two equiprobable pairs in Region 7, resulting in $\Lambda_a^{7.1} + \Lambda_a^{7.2} - (\Phi_b^{7.1} + \Phi_b^{7.2}) > 0$. If $h(x_a^{7.1}) = x_b^{7.1}$ then $\Lambda_a^{7.1} + \Lambda_a^{7.2} - (\Phi_b^{7.1} + \Phi_b^{7.2}) = 0$. The same line of argument establishes the same results for Region 10.

For the remaining Regions 1,2,3,5,8,9,12,14,15, and 16, the same argument used for the equiprobable pairs of attribute realizations across Regions 6 and 11 can be used to show that $\Lambda_a^i + \Lambda_a^j - (\Phi_b^i + \Phi_b^j) > 0$ for (i, j) pairs (3,8), (2,12), (1,16), (5,15), and (9,14). In particular, for pairs (3,8)

and (2,12) $1 = p_h > p_l > 0$. For (5,15) and (9,14) $1 > p_h > p_l = 0$. For (1,16) and $1 = p_h > p_l = 0$.

Thus, for all pairs of regions considered other than 4 and 13, the expected payoffs associated with policy Δ_a are strictly larger than those associated with Ψ_b . On the other hand, in Regions 4 and 13, the expected payoffs are the same. Therefore, integrating across the joint probability density function $f(h(x_a), x_b)$ results in $\Lambda_a > \Phi_b$. Thus, $\Phi_a > \Phi_b$. ■

The following lemma makes explicit that any attribute can yield a realization that results in search terminating, and its alternative being chosen immediately.

Lemma 1 *Let there be $m \geq 2$ alternatives $I = \{1, \dots, m\}$, and at least one alternative with at least one searchable attribute, i.e. $\exists i \in I$ such that $|S_i| \geq 1$. Then for any alternative i with searchable attribute \tilde{x}_{ij} there exists a unique $N_{ij} \in \mathbb{R}$ such that under policy Ψ_{ij} (which searches \tilde{x}_{ij} first, then proceeds optimally thereafter) if $v_i + x_{ij} > N_{ij}$ then search stops immediately and alternative i is chosen.*

Proof. Call the current revealed value $E(V_i)$, v_i for short, and assume that $v_1 > v_2 > \dots > v_m$, without loss of generality. First consider the problem with just one searchable attribute \tilde{x}_{ij} among all of the alternatives. Clearly, there exists a unique $N_{ij} \in \mathbb{R}$ such that if $v_i + x_{ij} > N_{ij}$ then search ends (trivially) immediately after \tilde{x}_{ij} is searched, and alternative i is chosen. If $i = 1$ then $N_{ij} = v_2$. If $i > 1$ then $N_{ij} = v_1$.

Now consider the problem with at least two unsearched attributes among the m alternatives. There are two cases to consider. In the first case policy Ψ_{ij} begins by searching \tilde{x}_{ij} , and there is at least one other unsearched attribute \tilde{x}_{ik} , where $j \neq k$, in alternative i . If $i = 1$ then the probability that $v_1 + x_{1j}$ falls in the interval of $[v_2, v_2 + r_{22}]$ is greater than zero. Call $\Phi(x_{1j})$ the expected value of the policy that proceeds optimally after searching \tilde{x}_{1j} . If $v_1 + x_{1j}$ does fall in the interval $[v_2, v_2 + r_{22}]$, then $\Phi(x_{1j}) - v_2 > c > 0$, by definition of the reservation distance. However, as x_{1j} grows larger the expected benefit of continuing search continuously decreases in x_{1j} , and as x_{1j} gets arbitrarily large the expected benefit of searching an additional attribute converges to $-c < 0$. Thus, by the intermediate value theorem there exists a unique $N_{1j} \in \mathbb{R}$ such that if $v_1 + x_{1j} > N_{1j}$ then it is optimal to end search immediately after \tilde{x}_{1j} is searched, and choose alternative 1. If $i > 1$ then the same argument applies, with the only difference being that instead of considering the reservation interval associated with v_2 , one considers the positive probability that $v_i + \tilde{x}_{ij} \in [v_1, v_1 + r_{12}]$, assuring that $\exists x_{ij}$ such that $\Phi(x_{ij}) - v_1 > c > 0$.

In the second case there is just one unsearched attribute \tilde{x}_{ij} in alternative i , which is searched first under policy Ψ_{ij} . Consider the set of other alternatives, $j \neq i$ that contain at least one unsearched attribute, and call v_h the highest current revealed value among them. The probability that $v_{ij} + x_{ij}$ falls in the reservation interval $[v_h, v_h + r_{hk}]$, where r_{hk} is the reservation distance associated with \tilde{x}_{hk} , is greater than zero. Thus for values of x_{ij} that fall in this interval $\Phi(x_{ij}) - v_h > c > 0$. Following the same steps as above, $\exists N'_{ij}$ such that if $v_i + x_{ij} > N'_{ij}$ then search ends immediately after \tilde{x}_{ij} is searched, and alternative i is chosen. However, so far this argument has ignored the v_j for $j \in I$ corresponding to alternatives with no unsearched attributes. Call the maximum value among these v_a . If $v_a < N'_{ij}$ then the $N_{ij} = N'_{ij}$. If $v_a \geq N'_{ij}$ then $N_{ij} = v_a$. ■

The following corollary is sufficient to characterize the complete order of optimal sequential search when there are exactly two alternatives, each with any number of searchable attributes.

Corollary V2 *Let there be exactly two alternatives $I = \{1, 2\}$, with expected values $E(\tilde{V}_1) \geq E(\tilde{V}_2)$, w.l.o.g., and at least two attributes total across the alternatives, i.e. $|S_1| + |S_2| \geq 2$. If $F_{ia} \prec^{s.s.o} F_{jb}$ for $i, j \in I$ then $\Phi_{ia} > \Phi_{jb}$. If F_{ia} and F_{jb} are the same then $\Phi_{ia} = \Phi_{jb}$.*

Proof. Assume that alternatives 1 and 2 have n_1 and n_2 unsearched attributes respectively, i.e. $\tilde{x}_{11}, \tilde{x}_{12}, \dots, \tilde{x}_{1n_1}$ and $\tilde{x}_{21}, \tilde{x}_{22}, \dots, \dots, \tilde{x}_{2n_2}$. To ease notation refer to $E(V_1)$ as v_1 and $E(V_2)$ as v_2 . Start with the simplest case, in which the value of alternative 1 is $v_1 + \tilde{x}_{11}$, and that of alternative 2 is $v_2 + \tilde{x}_{21}$. Consider the expected value of the policy that first searches \tilde{x}_{11} , then proceeds optimally thereafter:

$$\begin{aligned} \Phi_{11} &= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} - 2c, v_2 + \tilde{x}_{21} - 2c\}]\}] \\ &= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} - \tilde{x}_{21} - 2c, v_2 - 2c\}]\}] \text{ (by translation)} \\ &= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 2c, v_2 - 2c\}]\}] \text{ (by symmetry)} \end{aligned}$$

But the final expression corresponds precisely to the expected value of the policy that first searches \tilde{x}_{11} , then proceeds optimally, for the search problem in which the (ex ante) value of alternative 1 is $v_1 + \tilde{x}_{11} + \tilde{x}_{21}$, and that of alternative 2 is v_2 .

More generally, in the search problem in which the value of alternative 1 is $v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1}$, and that of alternative 2 is $v_2 + \tilde{x}_{21} + \tilde{x}_{22} + \dots + \tilde{x}_{2n_2}$, the same two-step procedure of translation and symmetry can be applied repeatedly until the expected value of starting by searching any attribute and then proceeding in any order is the same as it would be if the search problem were

instead one in which the value of alternative 1 were $v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} + \tilde{x}_{21} + \tilde{x}_{22} + \dots + \tilde{x}_{2n_2}$, and that of alternative 2 were v_2 . Thus, choosing any order, for example, $x_{11}, x_{21}, x_{22}, x_{12}, \dots, x_{1n_1}$:

$$\begin{aligned}
\Phi_{11} &= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} - 2c, v_2 + \tilde{x}_{21} - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} - 3c, v_2 + \tilde{x}_{21} + \tilde{x}_{22} - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} - (n_1 + n_2)c, v_2 + \tilde{x}_{21} + \tilde{x}_{22} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c\}]\dots\}]]]]\}}] \\
&= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} - \tilde{x}_{21} - 2c, v_2 - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} - \tilde{x}_{21} - 3c, v_2 + \tilde{x}_{22} - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} - \tilde{x}_{21} - (n_1 + n_2)c, v_2 + \tilde{x}_{22} + \tilde{x}_{23} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c\}]\dots\}]]]]\}}] \text{ (by translation)} \\
&= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 2c, v_2 - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 3c, v_2 + \tilde{x}_{22} - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} + \tilde{x}_{21} - (n_1 + n_2)c, v_2 + \tilde{x}_{22} + \tilde{x}_{23} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c\}]\dots\}]]]]\}}] \text{ (by symmetry)} \\
&= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 2c, v_2 - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - \tilde{x}_{22} - 3c, v_2 - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} + \tilde{x}_{21} - \tilde{x}_{22} - (n_1 + n_2)c, v_2 + \tilde{x}_{23} + \tilde{x}_{24} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c\}]\dots\}]]]]\}}] \text{ (by translation)} \\
&= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 2c, v_2 - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} + \tilde{x}_{22} - 3c, v_2 - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} + \tilde{x}_{21} + \tilde{x}_{22} - (n_1 + n_2)c, v_2 + \tilde{x}_{23} + \tilde{x}_{24} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c\}]\dots\}]]]]\}}] \text{ (by symmetry)} \\
&= \dots \\
&= E_{\tilde{x}_{11}}[\max\{v_1 + \tilde{x}_{11} - c, v_2 - c, E_{\tilde{x}_{21}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} - 2c, v_2 - 2c, E_{\tilde{x}_{22}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{21} + \tilde{x}_{22} - 3c, v_2 - 3c, E_{\tilde{x}_{12}}[\max\{\dots, E_{\tilde{x}_{1n_1}}[\max\{v_1 + \tilde{x}_{11} + \tilde{x}_{12} + \dots + \tilde{x}_{1n_1} + \tilde{x}_{21} + \tilde{x}_{22} + \dots + \tilde{x}_{2n_2} - (n_1 + n_2)c, v_2 - (n_1 + n_2)c\}]\dots\}]]]]\}}]
\end{aligned}$$

Thus, the complete optimal order to search attributes in, given any number of attributes in each of the two alternatives, is characterized by Condition V.⁵² ■

A.2 Condition T

Result T.1 *Let there be three alternatives, the first two of which are ordered from highest to lowest expected value, w.l.o.g., i.e. $E(\tilde{V}_1) \geq E(\tilde{V}_2)$. Assume that alternatives 1 and 2 each have exactly one unsearched attribute, i.e. $|S_i| = 1$ for $i = 1, 2$, and alternative 3 has none, i.e. $S_3 = \{\emptyset\}$. If*

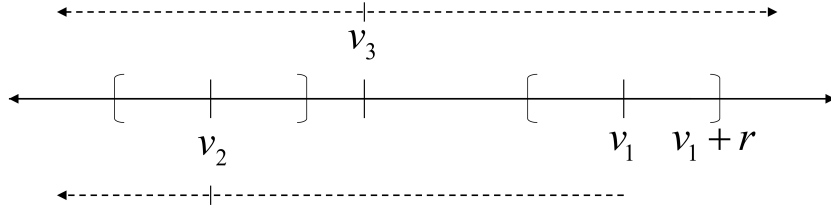
⁵²Thus, in the case of two attributes with the same distribution, it is clear (by symmetry) that the expected value of the policy that starts by searching either first, then proceeds optimally thereafter, is the same for both attributes.

$F_{1a} \preceq^{s.s.o} F_{2b}$, then $\Phi_{1a} \geq \Phi_{ib}$.

Proof. I first prove the result assuming identical attribute distributions, i.e. F_{12} and F_{22} , then extend to $F_{1a} \prec^{s.s.o} F_{2b}$.

Call the current revealed values of alternatives 1 and 2, that is $E(\tilde{V}_1)$ and $E(\tilde{V}_2)$, v_1 and v_2 for short. Because the cumulative distribution functions of alternatives 1 and 2 are identical, the reservation distances for both alternatives must be the same, i.e. $r_1 = r_2 = r > 0$. Notice also that by symmetry there exists an upper and lower reservation level for each alternative. If y falls within the reservation interval of v_j , i.e. $[v_j - r, v_j + r]$, where $j \neq i$, then the policy Ψ_{i2} will proceed optimally by searching alternative j . Otherwise, search will end immediately, with Ψ_{i2} choosing the alternative with the highest current revealed value, i.e. $k \in \arg \max\{v_i + x_{i2}, v_j, v_3\}$.

Figure 4



If $v_1 = v_2$ then clearly $\Phi_{12} = \Phi_{22}$, by symmetry. Otherwise, assume that $v_1 > v_2$, without loss of generality. Given this, in principal, the relative positioning of v_1, v_2 , and v_3 , along with the degree of overlapping with the reservation intervals of v_1 and v_2 , creates many different cases that must be considered separately (see Figure 4), as across cases expressions for the expected value(s) of one or both of the search policies change. For the purposes of the proof, it is sufficient to reduce these to the following four exclusive and exhaustive cases:

1. $v_3 \leq v_2 - r$
2. $v_2 - r < v_3 \leq v_1 - r$
3. $v_1 - r < v_3 \leq v_1 + r$
4. $v_1 + r < v_3$

For Case 1, consider Φ_{12} and Φ_{22} :

$$\begin{aligned}\Phi_{12} &= -c + \int_{-\infty}^{-(v_1-v_2)-r} v_2 dF_{12} + \int_{-(v_1-v_2)-r}^{-(v_1-v_2)+r} \int_{-\infty}^{\infty} (-c + \max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}\}) dF_{22} dF_{12} \\ &\quad + \int_{-(v_1-v_2)+r}^{\infty} (v_1 + \tilde{x}_{12}) dF_{12}\end{aligned}$$

$$\begin{aligned}\Phi_{22} &= -c + \int_{-\infty}^{(v_1-v_2)-r} v_1 dF_{22} + \int_{(v_1-v_2)-r}^{(v_1-v_2)+r} \int_{-\infty}^{\infty} (-c + \max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}\}) dF_{12} dF_{22} \\ &\quad + \int_{(v_1-v_2)+r}^{\infty} (v_2 + \tilde{x}_{22}) dF_{22}\end{aligned}$$

But these are precisely the expressions for Φ_{12} and Φ_{22} in the absence of v_3 . As a result, it follows immediately, by Corollary V2, that $\Phi_{12} = \Phi_{22}$. This is the unique case in which $\Phi_{12} \not> \Phi_{22}$, and corresponds to the somewhat uninteresting situation in which v_3 is so relatively low that there is no chance it will ever be chosen by either policy.

In Case 2 Φ_{22} remains the same as in Case 1, while v_3 leads to an increase in Φ_{12} that is positive (and monotonically increasing in v_3). As a result, $\Phi_{12} > \Phi_{22}$. To see this, first consider $v_3 = v_2 - r + \epsilon$, where $\epsilon > 0$ is sufficiently small so that v_3 falls in the reservation interval of v_2 . Under Ψ_{12} , in the case that $v_1 + x_{12} < v_2 - r + \epsilon$, search will proceed with \tilde{x}_{22} . This event occurs with positive probability, so, by construction of the reservation interval, leads to a gain in Φ_{12} relative to the counterfactual case in which v_3 is not available. Clearly, Φ_{12} is monotonically increasing in v_3 .

In Case 3 the expected value of search policy Ψ_{22} changes to the following:

$$\Phi_{22} = -c + \int_{-\infty}^{(v_1-v_2)+r} \int_{-\infty}^{\infty} (-c + \max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{12} dF_{22} + \int_{(v_1-v_2)+r}^{\infty} (v_2 + \tilde{x}_{22}) dF_{22}$$

Now, consider a sub-optimal search policy Δ_{12} , with expected value Λ_{12} , that searches \tilde{x}_{12} first, but otherwise mimics the decision rule of Ψ_{22} until it reaches \tilde{x}_{22} in search (if it does), at which point it behaves optimally thereafter. In particular, for realizations of \tilde{x}_{22} that lead Ψ_{22} to stop after searching the first attribute, and choose alternative 2, if Δ_{12} observes an x_{12} taking the same value, it stops search and chooses alternative 1 instead. The expected value of this policy is the

following:

$$\Lambda_{12} = -c + \int_{-\infty}^{(v_1-v_2)+r} \int_{-\infty}^{\infty} (-c + \max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{22} dF_{12} + \int_{(v_1-v_2)+r}^{\infty} (v_1 + \tilde{x}_{12}) dF_{12}$$

Taking the difference in expected values of the two policies yields:

$$\begin{aligned} \Lambda_{12} - \Phi_{22} &= \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{22} dF_{12} \\ &+ \int_{-\infty}^{(v_1-v_2)+r} \int_{-\infty}^{(v_1-v_2)+r} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{22} dF_{12} \\ &- \left(\int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{12} dF_{22} \right. \\ &+ \left. \int_{-\infty}^{(v_1-v_2)+r} \int_{-\infty}^{(v_1-v_2)+r} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{12} dF_{22} \right) + \int_{(v_1-v_2)+r}^{\infty} (v_1 - v_2) dF_{22} \\ &= \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{22} dF_{12} \\ &- \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{12} dF_{22} + \int_{(v_1-v_2)+r}^{\infty} (v_1 - v_2) dF_{22} \\ &= \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{2v_1 - v_2 + \tilde{x}_{12}, v_1 + \tilde{x}_{22}, v_1 - v_2 + v_3\}) dF_{22} dF_{12} \\ &- \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{v_1 + \tilde{x}_{12}, v_2 + \tilde{x}_{22}, v_3\}) dF_{12} dF_{22} \\ &+ \int_{(v_1-v_2)+r}^{\infty} \int_{(v_1-v_2)+r}^{\infty} (v_1 - v_2) dF_{22} dF_{12} \\ &= \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (\max\{2v_1 - v_2 + \tilde{x}_{12}, v_1 + \tilde{x}_{22}\}) dF_{22} dF_{12} \\ &- \int_{-\infty}^{(v_1-v_2)+r} \int_{(v_1-v_2)+r}^{\infty} (v_1 + \tilde{x}_{12}) dF_{12} dF_{22} + \int_{(v_1-v_2)+r}^{\infty} \int_{(v_1-v_2)+r}^{\infty} (v_1 - v_2) dF_{22} dF_{12} \\ &> 0 \end{aligned}$$

The first term in the last expression follows from plugging in the lower bound of the inner integral, which yields $2v_1 - v_2 = v_1 - v_2 + v_1 > v_1 - v_2 + v_3$. The second term follows from plugging in the upper and lower bounds of the outer and inner integrals, respectively, which yields $2v_1 - v_2 + r > v_1 + r > v_3$, meaning that $v_1 + \tilde{x}_{12}$ is the maximum over the entire region of integration. Clearly, the first term is larger than the second, and the third term is positive.

In Case 4:

$$\begin{aligned}
\Phi_{22} &= \int_{-\infty}^{v_3-v_2} v_3 dF_{22} + \int_{v_3-v_2}^{\infty} (v_2 + \tilde{x}_{22}) dF_{22} \\
&< \int_{-\infty}^{v_3-v_2} v_3 dF_{22} + \int_{v_3-v_2}^{\infty} (v_1 + \tilde{x}_{22}) dF_{22} \\
&< \int_{-\infty}^{v_3-v_1} v_3 dF_{22} + \int_{v_3-v_1}^{\infty} (v_1 + \tilde{x}_{22}) dF_{22} \\
&= \int_{-\infty}^{v_3-v_1} v_3 dF_{12} + \int_{v_3-v_1}^{\infty} (v_1 + \tilde{x}_{12}) dF_{12} \\
&= \Phi_{12}
\end{aligned}$$

This completes the proof under the assumption of identical attribute distributions. Now, for $F_{1a} \prec^{s.s.o} F_{2b}$ the arguments given for Cases 1-4 immediately extend. This is easiest to see by abusing notation to assume now that $F_{12} \prec^{s.s.o} F_{22}$. In Case 3, the positive final term in the last expression gets larger. In Case 4 Φ_{22} clearly falls while Φ_{12} remains unchanged, because it is not affected by alternative 2 or the distribution of \tilde{x}_{22} . ■

As discussed in Section 2, the special case in which $v_3 > v_1, v_2$ satisfies the requirements of Weitzman (1979)'s elegant reservation value rule, making that an alternative way of proving the result for Case 4 above, and half of Case 3.

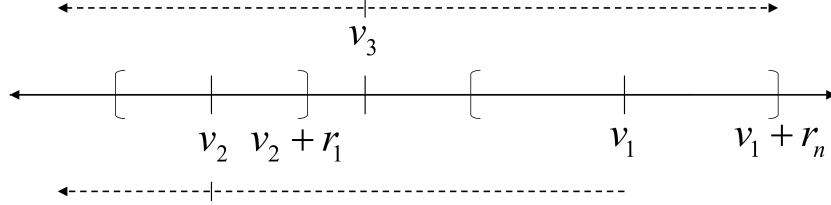
Result T.2 *Let there be three alternatives, the first two of which are ordered from highest to lowest expected value, w.l.o.g., i.e. $E(\tilde{V}_1) \geq E(\tilde{V}_2)$. Assume that alternative 1 has any number of unsearched attributes $S_1 \subseteq A$, alternative 2 has exactly one unsearched attribute, i.e. $|S_2| = 1$, and alternative 3 has none, i.e. $S_3 = \{\emptyset\}$. If $F_{1a} \preceq^{s.s.o} F_{ib}$ for all $b \in S_i$ and $i \in \{1, 2\}$, then $\Phi_{1a} \geq \Phi_{ib}$ for any b and i .*

Proof. As in the proof of Result T.1, begin with the assumption that the highest variance attribute of alternatives 1 and 2 is identically distributed, i.e. F_{12} and F_{22} are the same.

The first step is to confirm that the reservation distance of an alternative increases in its number of unsearched attributes. Start with the case of a single searchable attribute in alternative 1. As explained at the beginning of this section, there exists a unique reservation distance, call it r_1 (abusing notation), so long as the minimal condition $\int_0^\infty \tilde{x}_{12} dF_{12} > c$ is satisfied. In the case of two searchable attributes, each satisfying the minimal condition, call the reservation distance r_2

(abusing notation again). Clearly, $r_2 > r_1$, as, with positive probability, upon \tilde{x}_{12} being searched (it must be searched before \tilde{x}_{13} , by Condition V), it is optimal to then search \tilde{x}_{13} , meaning that the expected value of searching \tilde{x}_{12} to begin with is higher in the presence than absence of \tilde{x}_{13} . By induction it follows that $r_i > r_j$ for $i > j$.

Figure 5



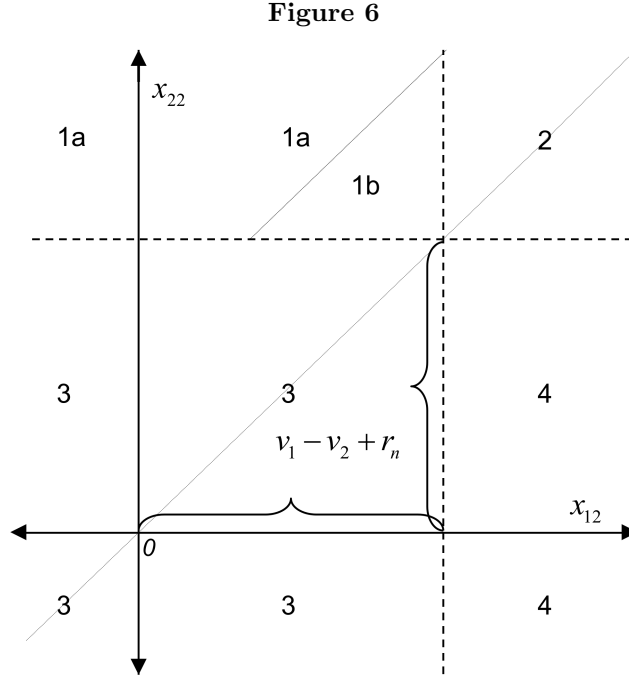
Now consider the case that alternative 1 has $n > 1$ unsearched attributes, with reservation distance r_n . Figure 5 shows how v_2 and v_3 can be rearranged to, in principle, require many different cases to be considered, as in Figure 4. As in the proof of Result T.1, the many possible rearrangements can be reduced to the following 4 cases:

1. $v_3 \leq v_2 - r_n$
2. $v_2 - r_n < v_3 \leq v_1 - r_n$
3. $v_1 - r_n < v_3 \leq v_1 + r_n$
4. $v_1 + r_n < v_3$

Figure 5 looks similar to the problem treated at the end of the proof of Result T.1, in which $F_{12} \prec^{s.s.o} F_{22}$. In particular, though in that problem there is only one unsearched attribute in each alternative, the reservation distance there is also greater for alternative 1 than it is for alternative 2. Consistent with this similarity, Cases 1,2, and 4 from the proof of Result T.1 extend immediately.

For Case 3, consider the search policy Ψ_{22} that begins by searching \tilde{x}_{22} , then proceeds optimally thereafter. Crucially, by Condition V, if a second attribute is searched, then it must be \tilde{x}_{12} . Now, consider a different search policy, Δ_{12} , that instead begins by searching \tilde{x}_{12} , then mimics the decisions made by Ψ_{22} until it reaches \tilde{x}_{22} in search (if it does), proceeding optimally thereafter. In particular, for values of x_{22} that lead Ψ_{22} to stop after the first search and choose alternative 2, for the same values of x_{12} policy Δ_{12} will stop search and choose alternative 1. On the other hand, for the same values of x_{22} that lead Ψ_{22} to continue on and search \tilde{x}_{12} , Δ_{12} continues on and searches \tilde{x}_{22} . Clearly, Δ_{12} is a sub-optimal search policy.

Given that in Case 3 v_3 lies in the reservation interval of v_1 , i.e. $v_3 \in (v_1 - r_n, v_1 + r_n)$, the only way that Ψ_{22} will immediately stop after searching \tilde{x}_{22} is if $x_{22} > v_1 - v_2 + r_n$. Otherwise, the policy proceeds to search \tilde{x}_{12} . Similarly, Δ_{12} continues searching unless $x_{12} > v_1 - v_2 + r_n$. Both of these thresholds are illustrated in the (x_{12}, x_{22}) plane of Figure 6, creating Regions 1-4. A dotted line corresponding to $x_{22} = x_{11}$ reflects symmetric equiprobable regions and sub-regions on either side. Crucially, for any pair of realizations (x_{12}, x_{22}) lying in Region 3, both attributes will be searched, with both policies proceeding identically thereafter, thus the same alternative will be chosen (eventually) under both Ψ_{22} and Δ_{12} , yielding identical payoffs. Thus, the only way that the expected payoffs of the two policies can differ is if at least one of them ends search immediately after the first attribute is searched, which occurs for the pairs of attribute realizations (x_{12}, x_{22}) contained in Regions 1,2, and 4 in Figure 6.



In Region 2 both Δ_{12} and Ψ_{22} choose the alternative that they begin searching immediately after the first attribute is searched. For any pair of realizations (x_{12}, x_{22}) there is a symmetric equiprobable pair (x_{22}, x_{12}) lying in the same region. Across these two equiprobable pairs the difference in (expected) payoffs favors policy Δ_{12} over Ψ_{22} , by the following amount: $(v_1 + x_{12} + v_1 + x_{22}) - (v_2 + x_{22} + v_2 + x_{12}) = 2(v_1 - v_2) > 0$.

In Region 4 $v_1 + x_{12}$ is chosen after the first attribute is searched under Δ_{12} , and after the second attribute is searched under Ψ_{22} . Thus, for each pair of realizations (x_{12}, x_{22}) the (expected)

payoff under Δ_{12} is higher than that under Ψ_{22} , by $c > 0$.

In Region 1 there are two sub-regions, 1a and 1b. Policy Ψ_{22} chooses alternative $v_2 + x_{22}$ after the first attribute is searched in both Regions 1a and 1b. Policy Δ_{12} always searches two attributes before choosing an alternative, and chooses $v_2 + x_{22}$ so long as $x_{22} \geq v_1 - v_2 + x_{12}$, which corresponds to Region 1a. In this case policy Δ_{12} earns $-c < 0$ relative to policy Ψ_{22} . On the other hand, if $x_{22} < v_1 - v_2 + x_{12}$ then Δ_{12} instead chooses $v_1 + x_{12}$, resulting in a relatively smaller difference in payoffs between the two policies of $v_1 + x_{12} - c - (v_2 + x_{22}) = v_1 - v_2 + (x_{12} - x_{22}) - c > -c$.

Now, when computing $\Lambda_{12} - \Phi_{22}$ by integrating across all 4 regions, Region 2 is positive, Regions 1 and 4 are positive, on balance (by symmetry), and Region 3 is zero. Thus, $\Lambda_{12} > \Phi_{22}$. Then, necessarily, $\Phi_{12} > \Phi_{22}$.

This completes the proof under the assumption that the highest variance attribute in alternatives 1 and 2 are identically distributed. Now, for $F_{1a} \prec^{s.s.o} F_{2b}$, where a is the highest variance attribute in alternative 1, and b is the highest variance attribute in alternative 2, it is easiest to see how the result extends by abusing notation to now simply say that $F_{12} \prec^{s.s.o} F_{22}$. With this, the arguments given for Cases 1,2, and 4 immediately extend. For Case 3, it is straightforward to extend the argument from above, using an approach similar to that in the proof of Condition V. In particular, consider a search policy Δ'_{12} , with expected payoffs Λ'_{12} , that begins by searching the higher variance attribute \tilde{x}_{12} , rather than \tilde{x}_{22} . For each realization x_{22} of \tilde{x}_{22} there is a unique realization x_{12} of \tilde{x}_{12} such that $F_{22} = F_{12}$. The correspondence x_{12} to x_{22} can be represented by the function $h : \mathbb{R} \rightarrow \mathbb{R}$, and because h is invertible, $h^{-1}(x_{22}) = x_{12}$. If $x_{22} > 0$ then $h^{-1}(x_{22}) = x_{12} > 0$ and $h^{-1}(x_{22}) - x_{22} > 0$ (by definition of h). Likewise, if $x_{22} < 0$ then $h^{-1}(x_{22}) = x_{12} < 0$ and $h^{-1}(x_{22}) - x_{22} < 0$. These differences in payoffs are symmetric around $x_{22} = 0$, so $h^{-1}(x_{22}) - x_{22} = -(h^{-1}(-x_{22}) - (-x_{22}))$.

Δ'_{12} strictly follows the decision rules of Ψ_{22} , but uses $h(x_{12})$ in place of x_{22} , until it reaches \tilde{x}_{22} in search (if it does). In the case that \tilde{x}_{22} is reached, Δ'_{12} switches to perfectly mimicking the decision rule of Ψ_{22} for the remainder of search. Notice that in principle Δ'_{12} is not an optimal policy, i.e. $\Lambda'_{12} \leq \Phi_{12}$.

Now consider Figure 6 with just one change: replace x_{12} , on the horizontal axis, with $h(x_{12})$. That is, the figure now represents an $(h(x_{12}), x_{22})$ plane. For any pair of attribute realizations (x_{12}, x_{22}) payoff differences between the two search policies Δ'_{12} and Φ_{22} remain the same as when F_{12} and F_{22} were identically distributed in Regions 3, because both policies choose the same alternative, and in 4, where both policies choose the first alternative, which has (expected) payoffs of

$v_1 + x_{12}$. By contrast, in Region 2 Δ'_{12} chooses the first alternative, with payoffs $v_1 + x_{12}$, while Φ_{22} chooses the second, with payoffs $v_2 + x_{22}$. Thus, because $x_{12} - h(x_{12}) > 0$ the difference in payoffs between Δ'_{12} and Φ_{22} increases. Finally, in Region 1 when $x_{22} \geq v_1 - v_2 + x_{12}$ both policies choose the second alternative, so there is no change in the difference of payoffs. However, for $x_{22} < v_1 - v_2 + x_{12}$ policy Δ'_{12} instead chooses alternative one, with payoffs $v_1 + x_{12}$, so the difference between policies increases by $x_{12} - h(x_{12}) > 0$. Therefore, because for all possible pairs of attribute realizations (x_{12}, x_{22}) differences in payoffs under Δ_{12} and Φ_{22} either increase or remain the same relative to the case in which F_{12} and F_{22} are identically distributed, it immediately follows that $\Delta'_{12} > \Phi_{22}$. Thus $\Phi_{12} > \Phi_{22}$. ■

Result T.3 *Let there be $m \geq 3$ alternatives, the first $m - 1$ of which are ordered from highest to lowest expected value, w.l.o.g., i.e. $E(\tilde{V}_i) \geq E(\tilde{V}_j)$ if $1 \leq i < j \leq m - 1$. Assume that the first alternative has any number of unsearched attributes, i.e. $S_1 \subseteq A$, while alternatives 2 through $m - 1$ each have exactly one unsearched attribute, i.e. $|S_i| = 1$ for $2 \leq i \leq m - 1$, and alternative m has none, i.e. $S_m = \{\emptyset\}$. If $F_{1a} \preceq^{s.s.o} F_{ib}$ for all $b \in S_i$ and $i \in \{1, 2, \dots, m - 1\}$, then $\Phi_{1a} \geq \Phi_{ib}$ for any b and i . Further, if $F_{jc} \preceq^{s.s.o} F_{ib}$ for $2 \leq j < i$ then $\Phi_{jc} \geq \Phi_{ib}$.*

Proof. The proof proceeds in the following order: (1) the case in which there are just three alternatives, one unsearched attribute in each, (2) add an alternative with no unsearched attributes, (3) extend to any number of attributes in alternative 1, (4) extend to any number of alternatives with an unsearched attribute.

(1) Consider three alternatives $I = \{1, 2, 3\}$, one unsearched attribute in each, and to begin with, identical attribute distributions, i.e. $S_i = \{2\}$ and F_{i2} independent of i .

First consider the search policy Ψ_{22} that starts by searching \tilde{x}_{22} , then proceeds optimally thereafter. By Lemma A.1 there exists a unique $N_{22} \in \mathbb{R}$ such that if $v_2 + x_{22} \geq N_{22}$ then search ends immediately, and alternative 2 is chosen. Crucially, by Result T.1, immediately after \tilde{x}_{22} is searched, if another attribute is to be searched, then it must be \tilde{x}_{12} . Now if $v_2 + x_{22} < N_{22}$ then there are two cases to consider. In Case 1 there exists $L_{22} \in \mathbb{R}$ such that if $v_2 + x_{22} < L_{22}$ then search ends immediately, and alternative 1 is chosen. In this case, if $L_{22} < v_2 + x_{22} < N_{22}$ then search proceeds with \tilde{x}_{12} . In Case 2, no such L_{22} exists, so search proceeds with x_{12} .

In Case 1, by construction, \tilde{x}_{32} will never be searched, and alternative v_3 will never be chosen, so it is as if the third alternative does not exist. On the other hand, in this case policy Ψ_{12} may

or may not choose v_3 with positive probability (depending on the proximity of v_2 to v_3). Thus, because $\Phi_{12} = \Phi_{22}$ in the absence of a third alternative by Corollary V2, Φ_{12} is at least as large as Φ_{22} . An analogous argument can be applied to demonstrate that Φ_{12} and Φ_{22} are at least as large as Φ_{32} .

Case 2 here is analogous to Case 3 in the proof of Result T.2, so I refrain from repeating the lengthy argument. As in Case 1, an analogous argument can be applied to demonstrate that $\Phi_{12}, \Phi_{22} > \Phi_{32}$.

This completes the proof for step (1), when F_{i2} is independent of i . On the other hand, when F_{1a} is the highest variance attribute of alternative i , and we perform a symmetric mean preserving spread on either F_{22} , F_{32} , or both, then an analogous argument to that used in the last part of the proof of Result T.1 can be applied here to prove that $\Phi_{12} > \Phi_{22}, \Phi_{32}$, where for expositional simplicity we abuse notation by calling F_{1a} simply F_{12} (allowing F_{22} and/or F_{32} now to s.s.o dominate F_{12}).

(2) Now add $v_4 \in \mathbb{R}$, a fixed value (with no associated unsearched attribute), to the problem with alternatives 1,2, and 3, from step (1). Once again, begin by assuming that F_{i2} is independent of i .

Reconsider the two cases from step (1) with the introduction of v_4 . In Case 1, if $v_4 < L_2$ then v_4 has no effect on policy Ψ_{22} , and is never chosen. Therefore, it follows immediately from step (1) that $\Phi_{12} > \Phi_{22}$. If $L_2 \leq v_4 < N_2$ then alternative 3 and \tilde{x}_{32} have no effect on Φ_{22} , so it is as if alternative 3 does not exist. Then, by Result T.1, it follows immediately that $\Phi_{12} > \Phi_{22}$. Finally, if $v_4 \geq N_2$ then by an argument analogous to that provided in Case 4 of Result T.1 $\Phi_{12} > \Phi_{22}$.

In Case 2 v_4 is never chosen when either policy Ψ_{22} or Δ_{12} ends search immediately after the first attribute is searched. Therefore, when it is chosen, it is chosen by both policies. As a result, in the Regions of the (x_{12}, x_{22}) plane over which the payoffs of Φ_{22} and Δ_{12} can differ, the differences remain unchanged from those considered in Case 2 of Step (1). Thus, it immediately follows that $\Phi_{12} > \Phi_{22}$.

Throughout step (2), analogous arguments can be used to prove that $\Phi_{12}, \Phi_{22} > \Phi_{32}$.

(3) Now, consider any number of n attributes in alternative 1, and continue with the assumption that F_{i2} is independent of i . By Condition V, it is impossible that any s.s.o-dominated attribute within an alternative is searched first in an optimal policy. Thus, in the optimal policy, either \tilde{x}_{12} , \tilde{x}_{22} , or \tilde{x}_{32} , must be searched first. Consider the policy Ψ_{22} that searches \tilde{x}_{22} first, then proceeds optimally thereafter. Crucially, by Result T.2, if it continues search then it must next search

\tilde{x}_{12} . But then the argument provided in step (2) immediately extends, as the comparison between Φ_{22} and Λ_{12} can be made solely on the basis of (up to) the first two attributes searched by each policy, which are the \tilde{x}_{12} and \tilde{x}_{22} in both cases. An analogous argument can be used to show that $\Phi_{12}, \Phi_{22} > \Phi_{32}$.

(4) An induction argument will be used to show that the result from step (3) extends when any number of alternatives, each with at most one searchable attribute, are added to the problem. Consider the search problem in which any number of alternatives $m - 4$, where $m > 4$ are added to the problem in (3), and that each of these alternatives has exactly one unsearched attribute with identical distribution, i.e. $S_i = \{2\}$ for $1 < i < m$ and F_{i2} is independent of i . In addition, any number of alternatives, each without a single unsearched attributes, can be added to the problem as well, but only the maximum value between these and v_4 need be considered. This maximum value will be represented by v_{max} . Now, step (3) shows that the result holds for the case of $m = 4$. Next, assume that the result holds for $m = b$, where $b > 4$ and show that it also holds for the case of $m = b + 1$. Given that the result holds for $m = b$, in the case that $m = b + 1$, consider the policy Ψ_{i2} that starts by searching the attribute of any alternative $i \in \{2, 3, \dots, m - 1\}$, then proceeds optimally thereafter. By the inductive assumption the second attribute searched, if one is to be searched, must be \tilde{x}_{12} . Thus, an argument analogous to that used in steps (2) and (3) can be applied, as the comparison of Φ_{i2} and Λ_{12} will only depend on search of (up to) the first two attributes, \tilde{x}_{12} and \tilde{x}_{22} . An analogous argument can be used to show that $\Phi_{i2} > \Phi_{j2}$, where $i \in \{1, 2, \dots, m - 2\}$, $j \in \{2, \dots, m - 1\}$, and $i > j$.

Finally, consider the case that $F_{i2} \prec^{s.s.o} F_{j2}$ for any $i, j \in I$, and $i > j$. Because the relative expected values of Φ_{i2} and Δ_{j2} can be reduced to a comparison of the expected payoffs over the first two attributes searched, x_{i2} and x_{j2} , it immediately follows from an argument analogous to that used at the end of Result T.2 that $\Phi_{i2} > \Phi_{j2}$.

Finally, consider the case that $F_{ia} \prec^{s.s.o} F_{jb}$ with the convenient abuse of notation $F_{i2} \prec^{s.s.o} F_{j2}$ for $i, j \in I$, and $i > j$. Because the relative expected values of Φ_{i2} and Δ_{j2} can be reduced to a comparison of the expected payoffs over the first two attributes searched, x_{i2} and x_{j2} , it immediately follows from an argument analogous to that used at the end of Result T.2 that $\Phi_{i2} > \Phi_{j2}$. ■

B Numerical Computations

In order to confirm that Condition T holds more generally than in the search problems considered in Results T1-T3 of Section 2, I run numerical computations for search problems with up to multiple searchable attributes (two) in each of multiple alternatives (three). These dimensions allows the searcher, after searching an attribute, to (1) search another attribute in the same alternative, (2) switch to search an attribute in one of multiple other alternatives, or (3) immediately end search and choose the highest valued alternative. Notice that these are the same qualitative features present in versions of the problem with arbitrarily large dimensions.

The sequential search problems that I consider each have multiple alternatives/attributes, full recall, no order restrictions, a constant cost of search per attribute, and normally distributed attribute distributions with a full second-order stochastic dominance ordering in each alternative. The normal distribution is used in order to match this feature in the GLMW experimental design. Because of limits to numerical tractability, bounded discrete, rather than unbounded continuous distributions will be used, as described below. In order to test Condition T on dimensionally larger versions of the search problems, that match the exact dimensions of GLMW’s design, I run simulations in Section C.

In the explicit computations performed for each search problem reported below, the separation between the current revealed values of each alternative, the variance of the attributes, and the cost of searching are determined at random, ex ante. Current revealed values for the first through third alternatives are assigned values $x_{11} = 10$, $\tilde{x}_{21} = 10 - \tilde{y}$, and $\tilde{x}_{31} = x_{21} - \tilde{z}$, respectively, where \tilde{y} and \tilde{z} are uniformly distributed i.i.d. random variables with support $[0, 10]$.⁵³ All searchable attributes’ values are drawn from a discretized normal distribution yielding integer values. The standard deviation of the left-most searchable attribute is drawn from a continuous uniform distribution with support $[0, 30]$. As in GLMW, there is a decaying variance structure from left to right; here a searchable attribute one column to the right has 0.81 the variance of the attribute immediately to its left. The search cost is drawn from a uniform distribution with support $[0, 1]$. Due to the induced computational burden by these Reimann sums with many layers of nesting it is not possible to give infinite supports to the searchable attributes. For the results reported here the

⁵³There is no loss in generality in fixing x_{11} because what matters, when determining which is the optimal attribute to search first, are the differences $x_{11} - x_{21}$ and $x_{21} - x_{31}$. Likewise, notice that the values of attributes x_{11}, x_{21} , and x_{31} are realized “before” the search problem starts. The uniform distribution is used here to increase the robustness of results, by generating a more even distribution of relative distances between alternatives than what would result if the distribution were instead normal.

integer support for every discretized normal distribution is $\{-10,-9,-8,\dots,0,\dots,8,9,10\}$, resulting in 21 mass points.⁵⁴

Given the random variation in (1) observable x_{i1} values, for $i \in \{1, 2, 3\}$, (2) the variance of searchable attributes' values, and (3) search costs, the analysis of each considered search problem is fairly robust—search can be virtually costless relative to the variance of searchable attributes, in which case all attributes should be searched almost always, while on the other extreme, search can be arbitrarily expensive relative to the variance of searchable attributes, in which case no attributes should be searched. Also, randomly drawn searchable attribute variances can on one extreme (high variance) be such that the distribution is essentially uniform over the truncated support, and on the other extreme (low variance) roughly all of the normal distribution's mass is contained within the support $\{-10,\dots, 10\}$. For versions of the search problem smaller than the three alternative, three attribute problem, it is relatively easier to stretch the supports, which results in larger separation between the expected values of different search policies. On the other hand, the three by three search problem took 17 days of computation time with support $\{-10,\dots, 10\}$, to do 100 computations, so the number of computations presented below are 250 for the three by three search problem, but 1000 for all of the others.

Table 8 presents results from the computations. Row labels indicate the search problem being analyzed. Table 8's first column header indicates how many computations were run (N) for each type of search problem. The remaining columns, from left to right, separately report the average (across computations) expected values of starting sequential search with attribute \tilde{x}_{12} , \tilde{x}_{22} , or \tilde{x}_{32} , respectively. Not shown in the table is that for every computation it is confirmed that $\Phi_{i2} \geq \Phi_{j2}$, when $i > j$, as predicted by Condition T.

While differences in the average expected payoffs of different search policies, when averaging across many iterations, presented in Table 8 may seem small, these average differences are attenuated here by a number of factors. For one, v_3 can take relatively very low values, ensuring that in the problems presented in rows one and two, for example, $\Phi_{12} = \Phi_{22}$, by Results T.1 and T.2. To illustrate further, in the simplest search problem, presented in row 1 of Table 8, if v_3 is simply drawn from $v_1 + \tilde{\epsilon}$, where $\tilde{\epsilon}$ is uniformly distributed on support $[-5, 5]$ (rather than from $v_2 - 5 + \tilde{\epsilon}$), then the average expected payoffs become $\Phi_{12} = 12.49$ and $\Phi_{22} = 12.24$, increasing the separation

⁵⁴In order to increase the tractability of computations, I prune expressions when going from smaller nested versions of the problem to larger ones, by assuming that Conditions V and T hold in the smaller versions once numerical results there corroborate this. When I repeat the analysis for each problem, without assuming anything, or pruning, I get similar results, but can only confirm (15 days of computation time) with up to 7 mass points, rather than 21.

Table 8: *Results from testing necessary Condition T*

Search Problem	N	Φ_{12}	Φ_{22}	Φ_{32}
$x_{11} + \tilde{x}_{12}$ $x_{21} + \tilde{x}_{22}$ x_{31}	1000	10.69	10.67	-
$x_{11} + \tilde{x}_{12} + \tilde{x}_{13}$ $x_{21} + \tilde{x}_{22}$ x_{31}	1000	10.96	10.94	-
$x_{11} + \tilde{x}_{12} + \tilde{x}_{13}$ $x_{21} + \tilde{x}_{22} + \tilde{x}_{23}$ x_{31}	1000	11.25	11.23	-
$x_{11} + \tilde{x}_{12} + \tilde{x}_{13}$ $x_{21} + \tilde{x}_{22}$ $x_{31} + \tilde{x}_{32}$	1000	11.16	11.10	10.88
$x_{11} + \tilde{x}_{12} + \tilde{x}_{13}$ $x_{21} + \tilde{x}_{22} + \tilde{x}_{23}$ $x_{31} + \tilde{x}_{32}$	1000	11.32	11.28	11.02
$x_{11} + \tilde{x}_{12} + \tilde{x}_{13}$ $x_{21} + \tilde{x}_{22} + \tilde{x}_{23}$ $x_{31} + \tilde{x}_{32} + \tilde{x}_{33}$	250	11.46	11.41	11.22

considerably. A second reason the the separations are attenuated are the bounded supports of attribute distributions. A third reason is that, given the great variation allowed in parameterizing each search problem, the ratio of variance to search costs can be made very high, or very low. The consequence of either is that it does not really matter where search starts, either because it will certainly continue (the former case) or certainly end (latter case) once the first attribute is searched. In order to get an idea of the costliness of violating Condition T in GLMW’s rich search experiment, I run simulations in Section C that show the costs are considerable.

C Econometric Analysis of Alternative Switching

The GLMW dataset is sufficiently rich to allow regressions to be run on the individual subject level. As a result, I run a separate conditional logit regression for each subject and sequential alternative switch pairing, for the first three alternative switches performed by each subjects across their search problems. The results of these regressions suggest that “irrational” spatial preferences and rational value preferences are roughly equally strong in subjects’ search behavior.

When a subject is deciding which of the eight alternatives to search first there are two main dimensions on which the alternatives differ. The first dimension, which contains the only information of importance to the rational searcher, is the always observable attribute 1 value of each alternative.⁵⁵ The second dimension on which alternatives differ is the spatial location of each alternative, which is irrelevant to the rational searcher. Thus the two independent variables in the conditional logit model are the value of attribute 1, drawn from a mean-zero normal distribution (integer-rounded), and the height of the alternative, which can take any value between 1(top) and 8(bottom). This simple model is used in order to count the number of subjects for which the effect of either independent variable is statistically distinguishable from zero. Rational subjects would only consider the value of attribute 1 for each alternative, so only its corresponding estimated coefficient would be statistically distinguishable from zero.⁵⁶

First Alternative Searched

Results from the first alternative searched regression in Table 9 report that the effect of attribute 1 value is statistically distinguishable from zero for 290 of the 390 experimental subjects. However, the spatial, or height effect, is also distinguishable from zero for 275 of the 390 subjects.

Aside from the 373 subjects for whom the regression coefficients are estimated successfully, there are 17 subjects with perfectly identified alternative switches for one of the two independent variables. 16 of these subjects always choose the spatially highest alternative (top), while one always searches the alternative with highest attribute 1 value. These respective subjects are added into Table 9 in order to report the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified. Also reported in the table are the number of subjects with both independent variable coefficients statistically distinguishable from zero, and the number of subjects with strictly one coefficient statistically distinguishable from zero, along with their respective average coefficient values.⁵⁷

⁵⁵This statement would not generally be true if searching one alternative rather than another conferred significant time gains, but it seems fair to assume that time gains of this type are relatively small enough to ignore.

⁵⁶Please see Appendix C for formal specifications of the systems of equations used to estimate conditional logit coefficients, here and later in this section.

⁵⁷The interpretation of estimated coefficients in the conditional logit model is not entirely straightforward. Perhaps the easiest way of interpreting the coefficients is by using an equivalence easily derived from the original likelihood expressions for each alternative: $\log(\frac{P_m}{P_s}) = \beta_V(V_{tm} - V_{ts}) + \beta_H(H_{tm} - H_{ts})$ where m and s are different alternatives, e.g. $m, s \in \{1, 2, \dots, 8\}$, V is for value and H for height, and t is the index identifying the search problem. All else equal, a change in the log ratio of the relative probabilities of searching any pair of given alternatives corresponds to an equal change in the product of either of the coefficients and the relative difference in its corresponding independent variable across the two alternatives. Thus, a natural way of comparing coefficients is to determine how much of a change in one independent variable is necessary to offset the effect of a one unit increase in the other. For example, the results in Table 9 suggest that a subject for whom both value and height are significant would search one alternative

By simply counting the number of subjects for which each coefficient is statistically distinguishable from zero the results suggest that alternatives' attribute 1 values and spatial locations are of roughly equal importance to subjects.

Table 9: *First alternative searched: results*

	# of significant subjects + (PI)	β_V	β_H
β_V sig.	290+(1)=291	0.112	-0.378
β_H sig.	275+(16)=291	0.061	-0.795
β_V & β_H sig.	202	0.079	-0.516
β_V sig., β_H not	88+(1)=89	0.187	-0.063
β_H sig., β_V not	73+(16)=89	0.010	-1.568

Second Alternative Searched

When the searcher finishes search in the first alternative explored, and considers which alternative to search next, she can now consider two additional dimensions on which remaining alternatives differ. The first is the absolute distance (in rows) of each of the remaining 7 alternatives to the first alternative searched. The second is whether the considered alternative is above or below the previously searched one. Thus the model for second alternative switch uses as its independent variables, for each of the remaining 7 unsearched alternatives, attribute 1 value (V), height of the alternative (H), absolute distance of the candidate alternative from the first alternative searched (D), and an indicator variable indicating whether the candidate alternative is below the first alternative searched (I).

As in the case of the regression for the first alternative searched, I report the number of subjects for which each variable yields a coefficient estimate statistically distinguishable from zero, along with average coefficient values for these subjects (Table 10). The effect of attribute 1 value is statistically distinguishable from zero for 287 of the 390 subjects. Distance is statistically distinguishable from zero for 247 subjects, while now height is for 127, and the down dummy for 43. However, in this regression there are subjects who are perfectly identified in the down indicator. Many subjects, for example, always search alternative 1 first and alternative 2 second, thus they are always perfect height types, perfect down types, and perfect distance types. Correspondingly, these coefficients are either dropped or estimated incorrectly in the regressions. There are 99 subjects perfectly identified spatially lower only if its attribute 1 value were at least 7 (cents) higher.

in Height, Distance, or Down.⁵⁸ Of the 99 cases, 83 always switch downward, 51 always switch to an adjacent alternative, and 25 always switch to the (spatially) highest remaining alternative. Table 10 reports the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified.

Table 10: *Second alternative searched: results*

	# of significant subjects + (PI)	β_V	β_D	β_H	β_I
β_V sig.	287	0.173	-1.423	0.672	-0.017
β_D sig.	247+(51)=298	0.074	-2.679	1.384	-1.471
β_H sig.	127+(25)=152	0.226	-4.903	2.125	-0.488
β_I sig.	43+(83)=126	0.068	-1.268	0.216	0.760
β_V & β_D sig.	204	0.087	-1.941	1.087	-1.096
β_V sig., β_D not	83	0.386	-0.150	-0.348	2.636
β_D sig., β_V not	43+(16)=59	0.013	-6.182	2.794	-3.253

Third Alternative Searched

The specification of the regression for the third alternative switch is identical to that used for the second. The results are also similar. The effect of attribute 1 values are statistically distinguishable from zero for 250 subjects, as is distance for 238, height for 96, and the down dummy for 44. For the third alternative switch there are 80 subjects with regression coefficients dropped or inestimable. Of these 80 subjects 52 always switch downward, 52 always switch to an adjacent alternative, and 30 always choose the (spatially) highest remaining alternative. I add these perfectly identified types into Table 11 to report the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified.

Despite rationality dictating that subjects only consider attribute 1 values when deciding which unsearched alternative to switch search to, results of the conditional logit analysis indicate that subjects systematically incorporate specific non-optimal considerations into their alternative switching behavior. The first alternative switch analysis suggests that the spatial location (non-optimal consideration) of the alternative and its attribute 1 value (optimal consideration) are of roughly equal importance to subjects. This finding corresponds to a general proclivity of subjects in the GLMW data to start with the spatially highest alternatives, then switch downward. 78% of all alternative switches, in fact, are downward. Regressions for the second and third alternative switches reveal

⁵⁸As long as at least 90% of transitions comply with perfectly identified behavior the subject is considered perfectly identified

Table 11: *Third alternative searched: results*

	# of significant subjects + (PI)	β_V	β_D	β_H	β_I
β_V sig.	250	0.100	-0.895	0.140	0.967
β_D sig.	238+(52)=290	0.053	-2.175	0.792	0.338
β_H sig.	95+(30)=125	0.034	-6.504	3.236	1.122
β_I sig.	44+(52)=96	0.050	-0.741	-0.561	1.808
β_V & β_D sig.	165	0.070	-1.251	0.223	1.070
β_V sig., β_D not	85	0.160	-0.203	-0.022	0.769
β_D sig., β_V not	73+(52)=125	0.015	-4.261	2.079	-1.315

that attribute 1 value and adjacency (non-optimal consideration) are the most prevalent switching considerations, and are roughly equally represented in subjects' search. The regressions also suggest that consideration of alternative height, as well as tendency of switching downward, are less prominent, but also present. I provide an interpretation of these results in Section 7.

D Econometric Specification

I report the systems of equations used to estimate conditional logit coefficients in Section 5. Three different sets of coefficients are estimated for each subject: one for each of the first three alternative switches. For regressions, by individual subject, on the first alternative searched:

$$\ln\left(\frac{P_{it1}}{P_{it8}}\right) = \beta_{Vi}V_{it1} + \beta_{Hi}H_{it1} + \epsilon_{it1},$$

...

$$\ln\left(\frac{P_{it7}}{P_{it8}}\right) = \beta_{Vi}V_{it7} + \beta_{Hi}H_{it7} + \epsilon_{it7}$$

where $i \in \{1, 2, 3, \dots, 390\}$ is an individual subject, $t \in \{1, 2, 3, \dots\}$ is the task,

$P \equiv 1$ if alternative is searched, 0 otherwise

$V \equiv$ Value of attribute 1 for given alternative, $V \sim N(0, \sigma^2)$

$H \equiv$ Spatial height of Alternative, $H \in \{1, 2, 3, \dots, 8\}$ where 1 is top and 8 is bottom.

The same specified system of equations is used for regressions, by individual subject, for switches from the first alternative searched to the second, and for switches from the second to the third.

$$\ln\left(\frac{P_{it1}}{P_{it8}}\right) = \beta_{Vi}V_{it1} + \beta_{Hi}H_{it1} + \beta_{Di}D_{it1} + \beta_{Ii}I_{it1} + \epsilon_{it1},$$

...

$$\ln\left(\frac{P_{it7}}{P_{it8}}\right) = \beta_{Vi}V_{it7} + \beta_{Hi}H_{it7} + \beta_{Di}D_{it7} + \beta_{Ii}I_{it7} + \epsilon_{it7}$$

where $i \in \{1, 2, 3, \dots, 390\}$ is an individual subject, $t \in \{1, 2, 3, \dots\}$ is the task,

$P \equiv 1$ if alternative is searched, 0 otherwise

$V \equiv$ Value of attribute 1 for given alternative, $V \sim N(0, \sigma^2)$

$H \equiv$ Spatial height of Alternative, $H \in \{1, 2, 3, \dots, 8\}$

where 1 is top and 8 is bottom.

$D \equiv$ Absolute distance of searched alternative from just searched alternative, $D \in \{1, 2, 3, \dots, 7\}$

$I \equiv 1$ if alternative is below just searched alternative, 0 otherwise