Attributes*

Diego Klabjan, Wojciech Olszewski, and Asher Wolinsky

July 22, 2014

Abstract

A decision make (DM) considers the acquisition of a multi-attribute object with uncertain qualities which can be discovered at a cost. DM's problem is to decide how much to invest in the discovery and whether to adopt or discard based on partial information. We characterize the solution in some special cases and discuss the computability of the solution in more general cases.

1 Introduction

This paper considers a truly fundamental problem. A decision maker (DM) considers the acquisition of a multi-attribute object with uncertain qualities which can be discovered at a cost. DM's problem is to decide, on the basis of partial information, how much to invest in the discovery and whether to adopt or discard the object. Almost any decision one can think of (be it purchasing a good or deciding on a job offer) is of this sort.

The object in question has n attributes. Acceptance of the object gives DM utility $u(x_1, ..., x_n)$, where x_i is the level of attribute i. Rejection yields utility of V. At the outset DM does not know the x_i's. She knows that they are independent draws from distributions F_i and can discover the realization of each x_i at a cost c_i. We will assume throughout that $u(x_1, ..., x_n) = x_1 + ... + x_n$, but the analysis easily generalizes to weighted additive utility $u(x_1, ..., x_n) = \alpha_1 x_1 + ... + \alpha_n x_n$.

We study two scenarios of this decision problem. In the simultaneous scenario DM decides in advance on a set of attributes $S \subset \{1, ..., n\}$ to discover. After incurring the associated cost $\sum_{i \in S} c_i$

*We would like to thank Hector Chade and Rakesh Vohra for helpful comments and suggestions. Support from NSF grant SES-1123595 is gratefully acknowledged

and learning the realizations x_i, $i \in S$, she decides whether to accept or reject the object. In the *sequential scenario*, DM discovers attributes sequentially in the order that she chooses and can decide on acceptance or rejection at every point in the process. The discovery costs are incurred along the process accordingly.[1]

In the simultaneous scenario we show that, when two attributes are ordered by second-order stochastic dominance, and have the same cost of discovery, the dominated attribute will be more attractive for discovery (i.e., is included in the optimal set whenever the other is). When two attributes are ordered by a somewhat stronger criterion, we characterize the differences in the costs of discovery across attributes that can reverse the previous observation and make a dominating attribute more attractive for discovery. We also characterize the optimal set in some other special cases, in which it can be computed by a pseudo polynomial-time algorithm. We show, however, that polynomial-time algorithms may not exist[2].

In the sequential scenario, we analyze two special cases. For the case of simple binary F_i's and equal c_i's, we show that the optimal order of discovery is also in the reverse order of dominance. For the case of two independently and symmetrically distributed attributes with reservation utility equal to the expected value of the object, we completely characterize the optimal decision rule. The solution associates with each attribute an index that depends only on its distribution and cost; the discovery starts with the attribute with the higher index. This is thus reminiscent of Gittins' indices in the context of the multi-armed bandit model (see Gittins and Jones (1974) and Gittins (1989)) and of Weitzman's (1979) Pandora rule in the context of the search model.

Our model can be viewed as some sort of a search model, but it differs from conventional search models in a significant way. In such models the searcher examines separate objects rather than attributes of a single object. Consequently, the payoff from acceptance after examining k objects with values $x, ..., x_k$ depends only on them (in search with recall it is $\max\{x_1, ..., x_k\}$). In contrast, in our problem, the corresponding payoff $u(x_1,..., x_n)$ depends on all n attributes (including undiscovered ones) and typically u would not be the maximum function.

This difference from conventional search models might not seem so large, but it still makes the present problem more unwieldy in the sense that some "nice" results familiar from the search literature cannot be reproduced here. When $n > 2$, there is no simple analog in the sequential

[1] We do not consider other possible scenarios like deciding on the order of discovery in advance and then proceeding sequentially as might be the case in certain applications.

[2] It is an open question whether the optimal set can be computed by a pseudo polynomial-time algorithms in more general cases.

scenario to Weitzman (1979)'s elegant optimal strategy (*Pandora rule*) for the related sequential search model. In the simultaneous scenario the optimum cannot be found by the steepest ascent algorithm as is the case in Chade and Smith (2005)'s simultaneous search model.

Neeman (1995) studies the optimal search strategy of an agent who faces a sequence of i.i.d. multi-attribute objects, and can observe *only one* attribute of each object. His main result establishes the desirability of observing an attribute whose distribution is second-order stochastically dominated by all others and is thus reminiscent of our results on the precedence of second-order stochastically dominated attributes in the discovery.

Before proceeding to the analysis, it might be useful to point out a somewhat alternative interpretation. The x_i's can be viewed as realizations of signals about the value of the object, rather than as actual attributes. With this interpretation the model is still formally equivalent, though the specific additive payoff function that we assume might be more natural for the attribute interpretation.

2 Further assumptions and definitions

The model was already outlined in the introduction. It is further assumed[3] that

$$V = \int x_1 dF_1 = ... = \int x_n dF_n = 0.$$

Consider two random variables y and z with the mean zero[4] and cdfs G and H, respectively. Recall that y is *second-order stochastically dominated* by z, denoted $y \precsim^{s.o.} z$, if

$$\int_{-\infty}^{s} H(t)dt \leq \int_{-\infty}^{s} G(t)dt \text{ for all } s.$$

We will also use a stronger notion of stochastic dominance: y is *simply second-order stochastically dominated* by z, denoted $y \precsim^{s.s.o.} z$, if

$$G(t) \leq H(t) \qquad \text{for all } t \geq 0$$

and

$$H(t) \leq G(t) \qquad \text{for all } t \leq 0$$

[3]We will comment on this and other simplifying assumptions in Section 6.

[4]Since we assume that the means of all attributes are zero, we define the relevant concepts for distributions with zero mean.

This notion is a special case of Diamond and Stiglitz (1974)'s concept of *simple mean-preserving spread.*[5] In particular, *simple second-order stochastic dominance* implies *second-order stochastic dominance.* Proposition VI in the appendix contains a characterization of *simple second-order stochastic dominance* that is analogous to the familiar characterization of *second-order stochastic dominance.*

3 Simultaneous discovery of attributes

In this scenario DM selects a subset S of the attributes to be discovered simultaneously. After observing the attributes in S, it is optimal to accept the object iff $\sum_{i\in S} x_i \geq 0$.[6] Therefore, DM's payoff as a function of S is

$$U(S) = \int ... \int \max\left(0, \sum_{i\in S} x_i\right) d\prod_{i\in S} F_i(x_i) - \sum_{i\in S} c_i. \tag{1}$$

DM's problem: Choose S^* such that

$$S^* \in \arg\max_{S\subset\{1,...,n\}} U(S). \tag{2}$$

The first part of Proposition I says that, other things equal, an attribute would be selected "ahead" of an attribute that stochastically dominates it. The second part of Proposition I points out what would be a sufficiently large cost difference $c_i - c_j$ to offset the advantage of selecting an attribute i over attribute j that dominates it. We are able to prove this second result only for the stronger notion of *simple second-order stochastic dominance.*

Proposition I: (i) Suppose that $x_i \precsim^{s.o.} x_j$ and $c_i \leq c_j$. Given any S that does not contain i or j, $U(S \cup i) \geq U(S \cup j)$.

(ii) Suppose that $x_i \precsim^{s.s.o.} x_j$ and

$$c_i - c_j \geq \int_0^\infty x_i dF_i(x_i) - \int_0^\infty x_j dF_j(x_j). \tag{3}$$

Given any S that does not contain i or j, $U(S \cup i) \leq U(S \cup j)$.

Therefore, if $x_1 \precsim^{s.o.} ... \precsim^{s.o.} x_n$ and $c_1 \leq ... \leq c_n$, then there is a solution S^* of the form $S^* = \{1, ..., k\}$. In other words, if the attributes are ordered by second-order stochastic dominance, there is a solution S^* that consists of the k most dominated attributes.

[5] Their concept requires that H and G cross only once, and we use here at the special case in which the crossing point is the common mean.

[6] The decision in the case of $\sum_{i\in S} x_i = 0$ is inessential for the analysis.

If the attributes are ordered by *simple second-order stochastic dominance*, i.e. $x_1 \precsim^{s.s.o.} ... \precsim^{s.s.o.} x_n$ and (3) holds for $j = i + 1$ for all i, then there is a solution S^* of the form $S^* = \{k, ..., n\}$; that is, it consists of the k most dominant attributes for some $k \in \{1, ..., n\}$.

The intuition for the first part of Proposition I is that stochastically dominated distributions have thicker tails, and hence provide more significant information about the value of the object. Therefore, when their costs are not higher than those of discovering other attributes, they should be discovered first. This part follows immediately from the following lemma.

Lemma 1: Let G_0, G_1 and G_2 be the cdfs of random variables y_0, y_1 and y_2, and suppose that $Ey_0 = Ey_1 = Ey_2 = 0$ and $y_1 \precsim^{s.o.} y_2$. Then

$$\begin{aligned} &\int\int_{y_0+y_2\geq 0} (y_0 + y_2)\, dG_0(y_0) dG_2(y_2) \leq \\ \leq \quad &\int\int_{y_0+y_1\geq 0} (y_0 + y_1)\, dG_0(y_0) dG_1(y_1). \end{aligned}$$

This lemma says that the benefit from discovering an additional attribute is higher than it would be from discovering an attribute that second-order stochastically dominates it. When $y_0 \equiv 0$ this is immediately obvious from the definition of second order stochastic dominance. It takes another step to show that, as one would expect, this is the case also when y_0 is non-degenerate. The proof of Lemma 1 as well as those all subsequent results are relegated to the appendix.

The proof of the second part of Proposition I uses the following lemma and an additional argument presented in the appendix.

Lemma 2: Let G_0, G_1 and G_2 be the cdfs of random variables y_0, y_1 and y_2. Suppose that all three random variables y_0, y_1 and y_2 have mean 0, i.e., $Ey_0 = Ey_1 = Ey_2 = 0$. Suppose further that y_1 is *simply second-order stochastically dominated* by y_2, i.e., $y_1 \precsim^{s.s.o.} y_2$. Then

$$\begin{aligned} &\int\int_{y_0+y_1\geq 0} (y_0 + y_1)\, dG_0(y_0) dG_1(y_1) - \int_0^{+\infty} y_1 dG_1(y_1) \leq \\ \leq &\int\int_{y_0+y_2\geq 0} (y_0 + y_2)\, dG_0(y_0) dG_2(y_2) - \int_0^{+\infty} y_2 dG_2(y_2). \end{aligned}$$

Lemma 2 implies that, if $y_1 \precsim^{s.s.o.} y_2$, then the incremental benefit from the discovery of additional attribute y_0 (or attributes that sum up to y_0) over the benefit of discovering attribute y is lower when $y = y_1$ than when $y = y_2$.

The following example illustrates Lemmas 1 and 2.

Example 1: Suppose y_a with cdf G_a is distributed uniformly on the interval $[-a, a]$, and y_b with cdf G_b is distributed uniformly on the interval $[-b, b]$, where $b < a$. That is, y_a is *simply second-order stochastically dominated*, and so is *second-order stochastically dominated* by y_b. Then

$$\int_0^{+\infty} y_a dG_a(y_a) = \frac{a}{4}; \int_0^{+\infty} y_b dG_b(y_b) = \frac{b}{4},$$

and

$$\int\int_{y_a+y_b\geq 0} (y_a + y_b)\, dG_a(y_a)dG_b(y_b) = \frac{a}{4} + \frac{b^2}{12a}. \tag{4}$$

Suppose first that $y_0 = y_a$, and $y_i = y_{b_i}$ for some $b_2 < b_1 < a$. That is, $y_1 \precsim^{s.o.} y_2$. Lemma 1, in this case, says that (4) increases in b.

Suppose now that $y_0 = y_b$, and $y_i = y_{a_i}$ for some $b < a_2 < a_1$. That is, $y_1 \precsim^{s.s.o.} y_2$. Lemma 2, in this case, says that

$$\int\int_{y_a+y_b\geq 0} (y_a + y_b)\, dG_a(y_a)dG_b(y_b) - \int_0^{+\infty} y_a dG_a(y_a) = \frac{b^2}{12a}$$

decreases in a.

4 Sequential discovery of attributes

This section considers sequential discovery. DM decides on the order of discovery and at each point whether to stop with an acceptance or rejection decision. The analysis of this case is somewhat harder since the optimal decision at each point depends on the history. We therefore focus on two special cases. In the first case of binary, symmetric distributions we characterize the optimal order of discovery without fully characterizing the optimal rule. In the second case of two symmetrically distributed attributes we completely characterize the optimal rule. Both of these cases confirm the insight obtained in the simultaneous scenario that optimal discovery gives precedence to *second-order stochastically dominated* attributes.

Proposition II: Consider the binary symmetric case $\Pr(x_i = a_i) = \Pr(x_i = -a_i) = 1/2$, where $a_1 > a_2 > ... > a_n > 0$ and $c_i = c > 0$. In this case optimal discovery is in ascending order of the indices (i.e., descending order of the a_i's).

Obviously, in this case F_i *second-order stochastically dominates* F_j for all $j < i$. So, this result is consistent with the result of Proposition I. While, strictly speaking, this result is proved only for the binary case, the logic of the proof seems to apply to a broader class of cases in which the distributions of the attributes are ordered by second-order stochastic dominance. Sanjurjo (2013) has shown a version of this result for normally distributed attributes in his analysis of experimental data presented in Gabaix et. al. (2006).

The second result of this section is for the case of *two symmetrically*[7] and *continuously* distributed attributes. The symmetry together with the restriction to two attributes facilitate a fairly elegant result that does not seem to survive in a similar simple form under more general conditions. The continuity of the distributions is assumed for the sake of simplicity and is not essential.

Observe that if

$$c_i \geq \int_0^{+\infty} x_i dF_i(x_i). \tag{5}$$

it is optimal not to discover attribute i. If (5) holds with the reverse inequality, then the following equation has a unique solution $x_i^* > 0$

$$x_i^* = -c_i + \int_{-x_i^*}^{+\infty} (x_i^* + x_i) dF_i(x_i) \qquad i = 1, 2. \tag{6}$$

The meaning of x_i^* is that, if the total value of the attributes that have been discovered is x_i^*, DM is just indifferent between accepting the object right away or first discovering attribute i and then deciding optimally whether to accept or reject it.

The symmetry of F_i implies $\int_{-x_i^*}^{+x_i^*} x_i dF_i(x_i) = 0$ and

$$\int_{x_i^*}^{+\infty} (-x_i^*) dF_i(x_i) = -x_i^*[1 - F_i(x_i^*)] = -x_i^* F_i(-x_i^*) = \int_{-x_i^*}^{+\infty} x_i^* dF_i(x_i) - x_i^* \tag{7}$$

Therefore, Equation (6) is equivalent to the equation

$$0 = -c_i + \int_{x_i^*}^{+\infty} (-x_i^* + x_i) dF_i(x_i). \tag{8}$$

That is, if the total value of the attributes that have been discovered is $-x_i^*$, DM is indifferent between rejecting the object right away or first discovering attribute i and then deciding optimally whether to accept or reject it.

Notice finally that the value of x_i^* is determined only by attribute i (that is, by c_i and F_i), and is independent of attribute j.

Proposition III: Suppose that the object has only two symmetrically and continuously distributed attributes.

[7] A random variable y with mean zero and cdf G is *symmetric* if $G(-y) = 1 - G(y)$, for every $y \geq 0$.

According to every optimal strategy, if (5) holds, attribute i should not be discovered; otherwise, attribute i should be discovered first if $x_i^* > x_j^*$. After discovering attribute i, the object should be accepted immediately if $x_i > x_j^*$, rejected immediately if $x_i < -x_j^*$ and discovery should proceed to attribute j otherwise. In the last case, the final acceptance (rejection) will follow if $x_i + x_j$ is positive (negative).

The following corollary establishes the insights of Proposition I for the present case.

Corollary 1: (i) Suppose that $x_1 \precsim^{s.o.} x_2$ and $c_1 \leq c_2$. Then, attribute 1 should be discovered first.

(ii) Suppose that $x_1 \precsim^{s.s.o.} x_2$ and

$$c_1 - c_2 \geq \int_0^{+\infty} x_1 dF_1(x_1) - \int_0^{+\infty} x_2 dF_2(x_2),$$

then attribute 2 should be discovered first.

The analysis of this section can be replicated for discrete, and even for general distributions. If the distribution of x_i is discrete, then there exist ${}^-x_i^*$ and ${}^+x_i^*$ such that LHS(8) is no higher than the RHS at ${}^-x_i^*$, and LHS(8) is no lower that the right-hand side at ${}^+x_i^*$.

5 Computability

The above analysis focused on some specific cases. This section inquires about the more fundamental question regarding the possibility of computing solutions for a broader class of cases.

Recall that the running time of an algorithm depends not only on the number of values $U(S)$ we have to compute, but also on the size of the parameter set. This includes the number of values each of the x_i's can take, and the sizes of these values as well as those of the c_i's, as measured by the number of digits in their binary expansions. Suppose the size of the parameters is commonly bounded by a number M. Then, an algorithm that is polynomial in M and n is called *pseudo polynomial.* To be truly polynomial, an algorithm must have running time that is either independent of M and polynomial in n (*strongly polynomial*), or polynomial in $\log M$, i.e., the number of bits to encode M on a computer, and n (*weakly polynomial*).[8]

[8] McCormick (2008) or Fujishige (2005) contain a more-detailed discussion. However, they devote much of their discussion of polynomial-time algorithms to minimization of submodular (or equivalently maximization of supermodular) functions. But our objective function is generally neither supermodular nor submodular.

Attention will be restricted to the simultaneous discovery case, since it is conceptually simpler. The question is whether the optimization problem (2) can be solved in polynomial, or pseudo-polynomial time.[9] Our first result suggests that in general we cannot expect the existence of polynomial-time algorithms.

Proposition IV: There is no polynomial-time algorithm for solving problem (2), even in the case in which each x_i has a binary distribution taking value a_i with probability $1/2$ and value $-a_i$ with probability $1/2$.

Since binary distributions taking values a_i and $-a_i$ with equal probabilities are ordered by *second-order stochastic dominance*, by Proposition I, there exist algorithms that deliver a solution S to problem (2) by computing $U(S)$ at a polynomial number of sets S of the form $S = \{1, ..., k\}$[10]. Thus, Proposition IV must reflect the fact that the computation of $U(S)$ for some sets S cannot be performed in polynomial time. In contrast, pseudo polynomial-time algorithms for solving problem (2) often exist.

Proposition V: If there exists an algorithm for solving problem (2) that requires computing $U(S)$ only for $P(n)$ sets S, where $P(n)$ is a polynomial function of n, then there also exists a pseudo polynomial-time algorithm for solving problem (2).

Of course, any pseudo polynomial-time algorithm requires computing $U(S)$ only for a polynomial number of sets S. Therefore, the questions regarding pseudo polynomial-time algorithms can be equivalently formulated in terms of the algorithms which require computing $U(S)$ only for a polynomial number of sets S.

6 Discussion and extensions

The results of this paper rely on fairly strong simplifying assumptions: (i) additive utility function; (ii) independence of the attributes; (iii) common means of the attributes equal to the reservation utility V; (iv) simple binary distributions or two symmetrically distributed attributes in the sequential discovery scenario.

[9] Of course, we must restrict attention here to distributions with finite support.

[10] However, we do not even know if there is a simple solution for all binary distributions. For example, whether for the case in which x_i takes values $a_i > 0$ and $b_i < 0$ and $c_i = c$ for all i there exists an algorithm for finding an optimal set of attributes S which requires computing $U(S)$ only at a polynomial number of sets S.

The additive utility combined with the independence of the attributes introduce sufficient separability into the optimization problems to make it possible to use marginal arguments with regard to the discovery.[11] In the absence of these assumptions, the discovery of an additional attribute would interact with the set of discovered attributes in more complex ways and would preclude the observations made in Proposition I.

Example 2: (a) Simultaneous search with three independently distributed attributes and

$$u(x_1, x_2, , x_3) = x_1 + x_2 + x_3 + x_1 \cdot x_2.$$

Suppose that $\Pr(x_i = 100) = \Pr(x_i = -100) = 1/2$, $i = 1, 2$; $\Pr(x_3 = 110) = \Pr(x_3 = -110) = 1/2$ and $c_i = 1000$ for all i. The optimal set of attributes to be discovered in this case is $\{x_1, x_2\}$, although each of them second-order stochastically dominates attribute x_3. Intuitively, the reason is that at the cost of order 1000, it only pays back to learn whether $x_1 \cdot x_2$ is equal to 10,000 or -10,000.

(b) Simultaneous search with three attributes but now with

$$u(x_1, x_2, , x_3) = x_1 + x_2 + x_3.$$

Suppose that the x_i's take on the same values as in part (a), but x_1 and x_2 are perfectly correlated while still independent of x_3; $c_i = 60$ for all i. The two optimal sets of attributes to be discovered are $\{x_1\}$ and $\{x_2\}$, not the stochastically dominated attribute x_3. The reason is that by learning the value of x_1 or x_2 DM learns the value of the other attribute. This enables DM to avoid expected loss of 100 $(= 200 \times \frac{1}{2}) > 60 = c_i$. In contrast, by checking x_3, DM avoids only expected loss of 55 $(= 110 \times \frac{1}{2}) < 60 = c_3$.

Obviously, the analysis of the general problem with other utility specifications or with correlated attributes would be useful in that it may open the way to other interesting applications. For example, the case in which different attributes are independent signals about the quality of an object is a very natural application but typically would involve a different utility function and perhaps some correlation. But, unfortunately, we have no significant result for these cases. Research on related topics (e.g., Gittins' indices, see Gittins and Jones (1974), Gittins (1989), or Weitzman's (1979) Pandora rule) suggests that the analysis would be rather intractable without the independence. Some small extensions are fairly immediate. For example, the first part of Proposition I generalizes

[11] The simple non-weighted additive form $u(x_1, ..., x_n) = \sum x_i$ is of course just a normalization and the analysis can be easily generalized to $u(x_1, ..., x_n) = \sum \alpha_i x_i$.

straightforwardly to $u(x_1, ..., x_n) = f(x_1 + ... + x_n)$, where f is an increasing function, but it is even unclear what would be a reasonable counterpart of the second part of Proposition I for this form of utility function.

The equality of the means to V is not crucial for the simultaneous scenario. It is not needed for the first part of Proposition I and a modified version of the second part of Proposition I can be proved for the case in which the means differ from V.

Analysis of the sequential discovery scenario is somewhat more challenging owing to the contingent nature of the decisions. Proposition II is stated and proved for the binary case. This assumption simplifies the argument since the absolute value of attribute j is always larger than that of $j+1$. It seems that a similar argument would hold for general symmetric distributions but it would be complicated by the absence of a deterministic ranking of the absolute values of the outcomes.

In the analysis of Proposition III (on sequential discovery), the symmetry, the restriction to two attributes and the equality of the mean to V all guarantee that the threshold between immediate acceptance and continued discovery, x_i^*, is just the negative of the threshold between rejection and continued discovery, $-x_i^*$. This property makes it possible to order the attributes using the x_i^*'s with the special feature being that x_i^* depends only the properties of attribute i. In the absence of any of these assumptions, these thresholds may not coincide and, in general, the ordering of the attributes would not captured by an index that is computed for each attribute independently of the other.

7 Appendix

Recall from Section 2 the notion of *simple second-order stochastic dominance* and the relation $\precsim^{s.s.o.}$ based on it. Let z and y be random variable with $E(z) = E(y) = 0$. Suppose that $y = z + x_z$ where the distribution of x_z depends on the realization of z. When the outcome of z is positive (negative) x_z takes on only nonnegative (non-positive) values. In such a case we will say that *y is obtained from z by moving mass away from the mean.*

Proposition VI: Consider random variables y and z with the mean zero and with cdfs G and H, respectively. Then the following statements are equivalent:

(i) $y \precsim^{s.s.o.} z$;

(ii) for any function $f : R \to R$ which is non-decreasing on $(-\infty, 0]$ and non-increasing on

$[0,+\infty)$,

$$\int_{-\infty}^{+\infty} f(y)dG(y) \leq \int_{-\infty}^{+\infty} f(z)dH(z); \tag{9}$$

(iii) y is obtained from z by moving mass away from the mean.

Proof. Proposition VI follows from analogous characterizations of first-order stochastic dominance. To see that (i) implies (ii), recall that a random variable z with cdf H first-order stochastically dominates a random variable y with cdf G (denoted $y \precsim^{f.o.} z$) iff

$$\int_{-\infty}^{+\infty} f(t)dG(t) \leq \int_{-\infty}^{+\infty} f(t)dH(t) \tag{10}$$

for any non-decreasing function f. We will denote the first order dominance relation by $y \precsim^{f.o.} z$ and also use $G \precsim^{f.o.} H$ in the same meaning.

Suppose that condition (i) of Proposition VI is satisfied for random variables y and z with cdfs G and H. Then, $G(0) = H(0)$. Denote this number by a and consider four random variables with cdfs G_-, H_-, G_+ and H_+:

(a) G_- coincides on $(-\infty, 0)$ with G, and $G_-(0) = 1$ (i.e., G_- has an atom of mass $1-a$ at 0);

(b) H_- coincides on $(-\infty, 0)$ with H, and $H_-(0) = 1$;

(c) G_+ coincides on $(0,+\infty)$ with G, and $G_+(0) = a$ (i.e., G_+ has an atom of mass a at 0);

(d) H_+ coincides on $(0,+\infty)$ with H, and $H_+(0) = a$.

Then, $G_- \precsim^{f.o.} H_-$ and $H_+ \precsim^{f.o.} G_+$. This implies that (10) holds with G replaced with G_- and H replaced with H_-, and also with G replaced with H_+ and H replaced with G_+.

To show that condition (ii) of Proposition VI is satisfied, consider a function $f : R \to R$ which is non-decreasing on $(-\infty, 0)$ and non-increasing on $(0,+\infty)$, and define two functions f_- and f_+ such that

(e) f_- coincides on $(-\infty, 0]$ with f, and is constant on $[0,+\infty)$;

(f) f_+ coincides on $[0,+\infty)$ with $-f$, and is constant on $(-\infty, 0]$.

Since f_- and f_+ are non-decreasing,

$$\int_{-\infty}^{+\infty} f_-(y)dG_-(y) \leq \int_{-\infty}^{+\infty} f_-(z)dH_-(z) \text{ and } \int_{-\infty}^{+\infty} f_+(z)dH_+(z) \leq \int_{-\infty}^{+\infty} f_+(y)dG_+(y).$$

These inequalities imply respectively

$$\int_{-\infty}^{0} f(y)dG(y) \leq \int_{-\infty}^{0} f(z)dH(z) \text{ and } -\int_{0}^{+\infty} f(z)dH(z) \leq -\int_{0}^{+\infty} f(y)dG(y).$$

which together yield (9).

To see that (iii) implies (i), recall that $y \precsim^{f.o.} z$ iff

$$y = z + x_z, \tag{11}$$

where the random variable x_z takes non-positive values and whose distribution may depend on the realization of z. Equivalently, $y \precsim^{f.o.} z$ iff $z = y + x_y$, where x_y is a random variable with nonnegative values.

Suppose that condition (iii) of Proposition VI is satisfied for random variables y and z. Consider two pairs of random variables: y_- and z_-, and y_+ and z_+ such that

(g) $y_- = y$ if $y \leq 0$ and $y_- = 0$ if $y > 0$, and $z_- = z$ if $z \leq 0$ and $z_- = 0$ if $z > 0$;

(h) $y_+ = y$ if $y \geq 0$ and $y_- = 0$ if $y < 0$, and $z_- = z$ if $z \geq 0$ and $z_- = 0$ if $z < 0$.

Then, (11) is satisfied for $z = z_-$ and $y = y_-$, and for $z = y_+$ and $y = z_+$. Thus, z_- stochastically dominates y_-, and y_+ stochastically dominates z_+. This yields

$$H_-(t) \leq G_-(t) \text{ and } H_+(t) \geq G_+(t)$$

for all t, where G_- and H_- are the cdfs of y_- and z_-, and G_+ and H_+ are the cdfs of y_+ and z_+, respectively. And this yields $G(t) \geq H(t)$ for $t \leq 0$, and $G(t) \leq H(t)$ for $t \geq 0$.

Finally, to see that (ii) implies (iii), suppose that (ii) is satisfied. By applying it to functions f that are non-decreasing on $(-\infty, 0]$, non-negative at 0, and equal to 0 on $(0, +\infty)$, we get that y_- and z_- satisfy (10) and hence $y_- \precsim^{f.o.} z_-$. It suffices to restrict attention to functions f with these properties, since one can add to any function f, without affecting (10), a constant no smaller than $f(0)$; and then, since variables y_- and z_- take positive values with probability 0, one can change, also without affecting the condition, the values of the so obtained function on $(-\infty, 0]$ to 0.

Similarly, z_+ and y_+ are shown to satisfy condition (10) and hence $z_+ \precsim^{f.o.} y_+$ by applying condition (ii) to functions $-f$ such that f is equal to 0 on $(-\infty, 0)$, non-negative at 0, and non-increasing on $[0, +\infty)$. Thus, y_- and z_- satisfy (11) and so do z_+ and y_+, which is exactly the condition defining moving mass away from the mean.■

7.1 Proofs of Lemmas 1 and 2, and the second part of Proposition I

Proof of Lemma 1: We need to show that the integral

$$\int \int_{y_0 + y_i \geq 0} (y_0 + y_i)\, dG_0(y_0) dG_i(y_i) \tag{12}$$

becomes higher when $y_i = y_2$ gets replaced with $y_i = y_1$, where $y_1 \precsim^{s.o.} y_2$.

Recall that a second-order stochastically dominated variable, such as y_1, can be represented as a mean-preserving spread of a variable that dominates it, such as y_2, that is, as the compound lottery such that in the first stage, we have a lottery y_2; and in the second stage, we randomize each possible outcome of y_2 further so the final outcome is $y_2 + x$, where the distribution of x has mean zero, and may depend on the outcome of y_2.

Consider a pair of realizations of variables y_0 and y_2 such that $y_0 + y_2 \geq 0$ and variable x such that $y_1 = y_2 + x$ contingent on this particular realization of y_2. For any such pair y_0 and y_2, we have that

$$y_0 + y_2 = \int (y_0 + y_2 + x) dG(x) \leq \int \max\{0, y_0 + y_2 + x\} dG(x) = \int_{y_0+y_2+x\geq 0} (y_0 + y_2 + x) dG(x).$$

This yields

$$\begin{aligned}
\int\int_{y_0+y_2\geq 0} (y_0 + y_2)\, dG_0(y_0) dG_2(y_2) \quad &\leq \quad \int\int_{y_0+y_2\geq 0} \left(\int_{y_0+y_2+x\geq 0} (y_0 + y_2 + x) dG(x) \right) dG_0(y_0) dG_2(y_2) \\
&\leq \quad \int\int_{y_0+y_1\geq 0} (y_0 + y_1)\, dG_0(y_0) dG_2(y_1).
\end{aligned}$$

The proof is illustrated in Figure 1(a). For any pair of outcomes y_0 and $y_i = y_2$ such that $y_0 + y_i + x \geq 0$ for all outcomes of x, the addition of x to y_i does not affect expression (12), because the expected value of x is equal to 0 (an example of such a case is depicted in dashed arrows), then. And for any pair of outcomes y_0 and $y_i = y_2$ such that $y_0 + y_i + x < 0$ for some outcomes of x (an example of such a case is depicted in solid arrows), the addition of x to y_i makes expression (12) higher, because the outcomes such that $y_0 + y_i + x < 0$ are located out of the region of integration.

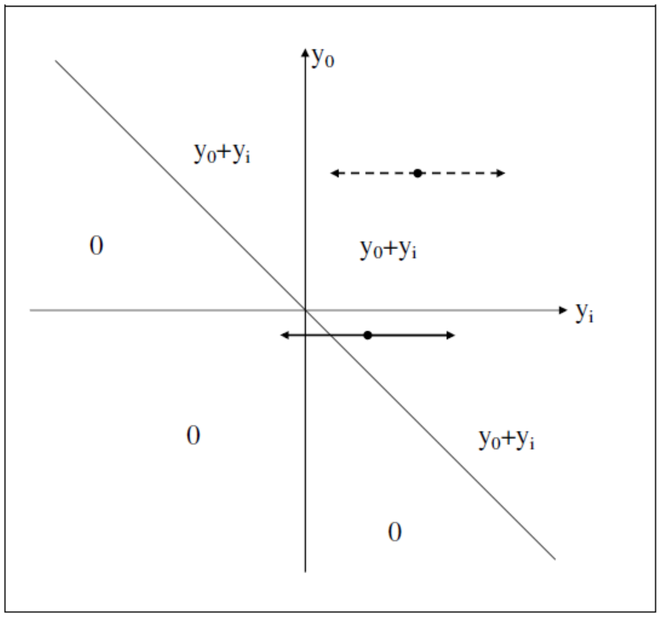

Figure 1 (a)

■

Proof of Lemma 2: We need to show that the integral

$$\int\int_{y_0+y_i\geq 0} (y_0+y_i)\, dG_0(y_0)dG_i(y_i) - \int_0^{+\infty} y_i dG_i(y_i) \tag{13}$$

becomes lower when $y_i = y_2$ gets replaced with $y_i = y_1$, where $y_1 \precsim^{s.s.o} y_2$. To show this it will be convenient to transform expression (13) as follows

$$\int\int_{y_0+y_i\geq 0} (y_0+y_i)\, dG_0(y_0)dG_i(y_i) - \int_0^{+\infty} y_i dG_i(y_i) =$$

$$= \int\int_{y_0+y_i\geq 0} y_i dG_0(y_0)dG_i(y_i) - \int\int_{y_i\geq 0, -\infty<y_0<+\infty} y_i dG_i(y_i)dG_0(y_0) + \int\int_{y_0+y_i\geq 0} y_0 dG_0(y_0)dG_i(y_i) =$$

$$= \int\int_{y_0+y_i\geq 0, y_i\leq 0} y_i dG_i(y_i)dG_0(y_0) - \int\int_{y_0+y_i\leq 0, y_i\geq 0} y_i dG_i(y_i)dG_0(y_0) +$$

$$+ \int\int_{y_0+y_i\geq 0} y_0 dG_0(y_0)dG_i(y_i) =$$

$$= \int\int_{y_0+y_i\geq 0, y_i\leq 0} (y_0+y_i) dG_i(y_i)dG_0(y_0) + \int\int_{y_0+y_i\leq 0, y_i\geq 0} (-y_0-y_i) dG_i(y_i)dG_0(y_0).$$

The sequence of transformations is easier to understand with reference to Figure 1(b). The first equality is obvious. To see the second equality notice that the left-hand side of that equality has the integral of y_i over the region $y_0 + y_i \geq 0$ minus the integral of y_i over the region $y_i \geq 0$, which is exactly the integral over the region above line $y_0 + y_i = 0$ and to the left of line $y_i = 0$ minus the integral of y_i over the region below line $y_0 + y_i = 0$ and to the right of line $y_i = 0$. The third inequality follows from the same argument applied to y_0 instead of y_i. Notice here that

$$\int\int_{y_0+y_i\geq 0} y_0 dG_0(y_0)dG_i(y_i) = \int\int_{y_0+y_i\geq 0} y_0 dG_0(y_0)dG_i(y_i) - \int\int_{y_i\geq 0,-\infty<y_0<+\infty} y_0 dG_i(y_i)dG_0(y_0)$$

because the very last integral is equal to the expected value of y_0, which is 0.

Consider the integral

$$\int\int_{y_0+y_i\geq 0, y_i\leq 0} (y_0 + y_i)dG_i(y_i)dG_0(y_0).$$

Consider also a pair of realizations of variables y_0 and y_2 such that $y_2 > 0$, and a variable x that is added to this particular realization of y_2 to obtain variable y_1 by moving mass away from the mean (see Proposition VI). Then, we must have that $x \geq 0$, and no pair of realizations y_0, y_1 so obtained can belong to the region of integration for $y_i = y_1$. Thus, a pair y_0, y_1 can belong to the region of integration for $y_i = y_1$ only when $y_2 \leq 0$, but then we also have that $x \leq 0$, and that the pair of realizations y_0 and y_2 must belong to the region of integration for $y_i = y_2$.

Therefore,

$$\begin{aligned}\int\int_{y_0+y_1\geq 0, y_1\leq 0} (y_0 + y_1)dG_1(y_1)dG_0(y_0) &= \int\int_{y_0+y_2+x\geq 0, y_2\leq 0} (y_0 + y_2 + x)dG(x)dG_2(y_2)dG_0(y_0) \\ &\leq \int\int_{y_0+y_2\geq 0, y_2\leq 0} (y_0 + y_2)dG_2(y_2)dG_0(y_0).\end{aligned}$$

A similar argument applies to the integral

$$\int\int_{y_0+y_i\leq 0, y_i\geq 0} (-y_0 - y_i)dG_i(y_i)dG_0(y_0).$$

Indeed, consider also a pair of realizations of variables y_0 and y_2 such that $y_2 < 0$, and a variable x that is added to this particular realization of y_2 to obtain variable y_1 by moving mass away from the mean. Then, we must have that $x \leq 0$, and no pair of realizations y_0, y_1 so obtained can belong to the region of integration for $y_i = y_1$. Thus, a pair y_0, y_1 can belong to the region of integration for $y_i = y_1$ only when $y_2 \geq 0$, but then we also have that $x \geq 0$, and that the pair of realizations y_0 and y_2 must belong to the region of integration for $y_i = y_2$.

Therefore,

$$\int\int_{y_0+y_1\le 0, y_1\ge 0}(-y_0-y_1)dG_1(y_1)dG_0(y_0) = \int\int_{y_0+y_2+x\le 0, y_2\ge 0}(-y_0-y_2-x)dG(x)dG_2(y_2)dG_0(y_0)$$
$$\le \int\int_{y_0+y_2\le 0, y_2\ge 0}(-y_0-y_2)dG_2(y_2)dG_0(y_0).$$

The arguments are illustrated in Figure 1(b). For example, consider the integral over the region above line $y_0 + y_i = 0$ and to the left of line $y_i = 0$. If a pair of realizations y_0, y_2 lies in this region for $y_i = y_2$, then a pair of realizations y_0, y_1 such that y_1 was obtained by adding variable x from Proposition VI (iii) to this particular realization y_2 may or may not belong to the region above line $y_0 + y_1 = 0$ and to the left of line $y_1 = 0$, as illustrated by a dashed arrow and a solid arrow, respectively. The value of $y_0 + y_1$ is in each case lower than the value of $y_0 + y_2$. In addition, if a pair of realizations y_0, y_2 lies outside the region above line $y_0 + y_2 = 0$ and to the left of line $y_2 = 0$, then a pair of realizations y_0, y_1 such that y_1 was obtained by adding variable x from Proposition VI (iii) to this particular realization y_2 also lies outside the region above line $y_0 + y_1 = 0$ and to the left of line $y_1 = 0$.

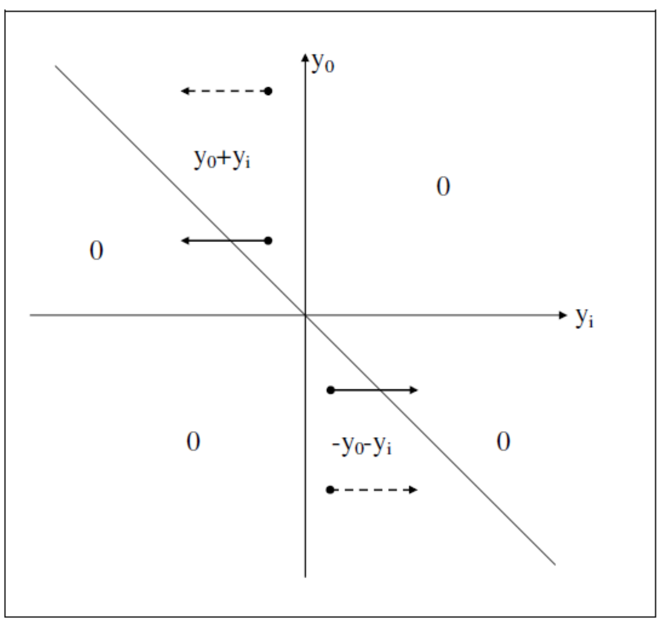

Figure 1 (b)

■

Proof of the second part of Proposition I: Applying Lemma 2 to

$$y_0 = \sum_{k \in S} x_k; \;\; y_1 = x_i; \;\; y_2 = x_j.$$

we have that

$$\int ... \int_{x_k, k \in S \cup j} \max\left(0, \sum_{k \in S \cup j} x_k\right) - \int ... \int_{x_k, k \in S \cup i} \max\left(0, \sum_{k \in S \cup i} x_k\right) \geq$$

$$\geq \int_0^{+\infty} x_j dF_j(x_j) - \int_0^{+\infty} x_i dF_i(x_i).$$

By assumption,

$$\int_0^{+\infty} x_j dF_j(x_j) - \int_0^{+\infty} x_i dF_i(x_i) \geq c_j - c_i.$$

Since $\sum_{k \in S \cup j} c_k - \sum_{k \in S \cup i} c_k = c_j - c_i$, it follows from (1) that $U(S \cup j) \geq U(S \cup i)$.■

7.2 Proof of Proposition II

The proof is by induction. Suppose that this result is true when m attributes are left and suppose that $m+1$ are left. Let a_j be maximal in this set and suppose that it is optimal to start with k such that $a_k < a_j$. By the inductive assumption, if the process continues after discovering k, it continues with j. The optimality of discovering k implies that the decision after the realization a_k is different than after the realization $-a_k$, since otherwise it would be beneficial to postpone the discovery of attribute k to a point in which it might be relevant. Thus, the possible decisions after discovering k are

Table 1

a_k	accept	continue	accept
$-a_k$	reject	reject	continue

Consider next the alternative of starting with attribute j, making the same decisions after a_j and $-a_j$ as prescribed above for a_k and $-a_k$, then continuing with k if the decision is to continue and thereafter following the optimal rule. The expected discovery cost is the same with both rules. Their outcomes coincide when $(x_k, x_j) = (a_k, a_j)$ or $(-a_k, -a_j)$. Their outcomes may differ only when $(x_k, x_j) = (a_k, -a_j)$ or $(-a_k, a_j)$. The following three tables describe all the ways in which the decisions may differ in these cases.

Table 2

i	$(a_k, -a_j)$	$(-a_k, a_j)$
k first	accept	reject
j first	reject	accept

ii	$(a_k, -a_j)$	$(-a_k, a_j)$
k first	continue	reject
j first	reject	continue

iii	$(a_k, -a_j)$	$(-a_k, a_j)$
k first	accept	continue
j first	continue	accept

For tables 2(i) and 2(ii), it is obvious that discovering j first is superior, since the difference between the expected benefit of either accepting or continuing over rejection is larger for the case $-a_k, a_j$ than for the case $a_k, -a_j$ by at least $a_j - a_k > 0$.

It remains to evaluate table 2(iii). Given the set S of undiscovered attributes and the sum X of the values of the already discovered attributes, define $\Phi(S|X)$ to be the expected payoff of continuing with an optimal policy, not counting the costs that have already been incurred. It is defined inductively by

$$\Phi(\{\ell\}|X) = \max\{0, X, \frac{1}{2}\max\{X - a_\ell, 0\} + \frac{1}{2}\max\{X + a_\ell, 0\} - c\}$$

and

$$\Phi(S|X) = \max_{\ell \in S}\max\{0, X, \frac{1}{2}\Phi(S\backslash\ell|X - a_\ell) + \frac{1}{2}\Phi(\{S\backslash\ell|X + a_\ell) - c\}$$

Consider now the situation described by table (iii). Discovering j first yields higher payoff than discovering k first iff the difference between the expected continuation value and the value of immediate acceptance is larger after $a_k, -a_j$ than after $-a_k, a_j$. That is, if $\Phi(S|Y + a_k - a_j) - (Y + a_k - a_j) > \Phi(S|Y + a_j - a_k) - (Y + a_j - a_k)$, where S is the set of undiscovered attributes (excluding j and k) and Y is the sum of the values of attributes discovered before j and k were reached. Now, this follows from the fact that $\Phi(S|X) - X$ is decreasing in X, if $|S| \leq m$ (where m is given by the inductive assumption in the beginning of the proof). This fact is established by induction on $|S|$.

For $S = \{\ell\}$, $\Phi(\{\ell\}|X) - X = \max\{-X, 0, \frac{1}{2}\max\{-a_\ell, -X\} + \frac{1}{2}\max\{a_\ell, -X\} - c\}$ which is decreasing in X. Suppose now that this is true for all sets S such that $|S| \leq t < m$ and consider an S such that $|S| = t + 1$. Since $|S| \leq m$, the identity of the first attribute to be discovered out of S is independent of X (by the inductive assumption made the beginning of the proof). Let ℓ be that attribute and observe that

$$\Phi(S|X) - X = \max\{-X, 0, \frac{1}{2}[\Phi(S\backslash\ell|X - a_\ell) - (X - a_\ell)] + \frac{1}{2}[\Phi(S\backslash\ell|X + a_\ell) - (X + a_\ell)] - c\}$$

is decreasing in X, since by the inductive assumption $\Phi(S\backslash\ell|X - a_\ell) - (X - a_\ell)$ and $\Phi(S\backslash\ell|X + a_\ell) - (X + a_\ell)$ are decreasing in X. ■

7.3 Proofs of Proposition III and Corollary 1

Proof of Proposition III: Suppose that $x_1^* < x_2^*$. We will show that attribute 1 should be discovered first. Other assertions of Proposition III are straightforward. The payoffs contingent on any possible pair of realizations of the two attributes are exhibited in Figure 2(a). The top row in each area is the payoff from discovering the realization of attribute 1 first, and playing the optimal continuation strategy contingent on any realization of this attribute. According to this strategy, DM should accept the object whenever $x_2^* < x_1$; she should reject the object whenever $x_1 < -x_2^*$; and when $-x_2^* < x_1 < x_2^*$, she should discover the realization x_2 of attribute 2, accept the object when $x_1 + x_2 > 0$ and reject the object when $x_1 + x_2 < 0$.

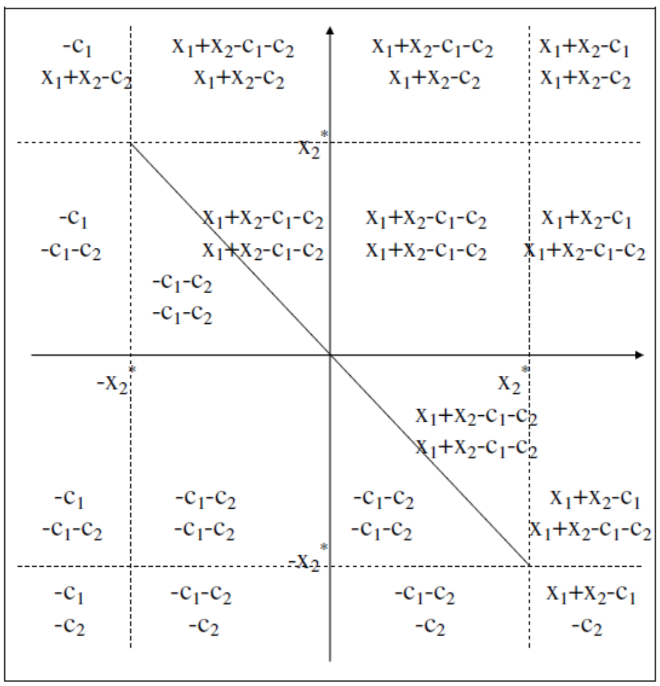

Figure 2 (a)

The bottom row is the payoff from discovering the realization of attribute 2 first, but then playing the (possibly suboptimal) continuation strategy, according to which DM should accept the object whenever $x_2^* < x_2$; she should reject the object whenever $x_2 < -x_2^*$; and when $-x_2^* < x_2 < x_2^*$, she should discover the realization x_1 of attribute 1, accept the object when $x_1 + x_2 > 0$ and reject the object when $x_1 + x_2 < 0$. (The optimal continuation strategy would have threshold x_1^* not x_2^*.)

We obtain Figure 2(b) from Figure 2(a) by deleting the common components of the corresponding top and bottom payoffs. Finally, we obtain Figure 2(c) from Figure 2(b) by means of (6) and (8). Notice that for any given value of x_1, the component $-c_2$ appears in the bottom row of Figure 2(b) either for every single value of x_2 or for no value of x_2. The component $-c_2$ appears for every single value of x_2 when $x_1 > x_2^*$ or when $x_1 < -x_2^*$. In the former case, we can replace $-c_2$ with $x_2^* + x_2$ for $x_2 < -x_2^*$ and with 0 for $x_2 > -x_2^*$. In the latter case, we can replace $-c_2$ with $x_2^* - x_2$ for $x_2 > x_2^*$ and with 0 for $x_2 < x_2^*$. These changes do not affect the integral of the bottom row payoff across all pairs of realizations of the two attributes. To see this, observe first that (6) and (8) imply

$$x_i^* F_i(-x_i^*) = -c_i + \int_{-x_i^*}^{+\infty} x_i dF_i(x_i) \tag{14}$$

and

$$x_i^*[1 - F_i(x_i^*)] = -c_i + \int_{x_i^*}^{+\infty} x_i dF_i(x_i) \tag{15}$$

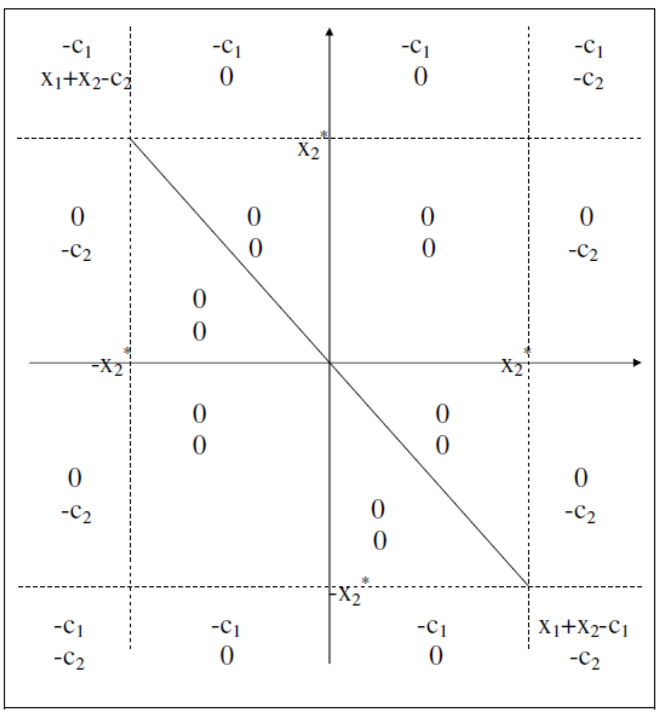

Figure 2 (b)

Now, the desired result for the first case follows from (14) and from

$$\int_{-x_2^*}^{+\infty} x_2 dF_2(x_2) = \int_{-x_2^*}^{+x_2^*} x_2 dF_2(x_2) + \int_{+x_2^*}^{+\infty} x_2 dF_2(x_2) = -\int_{-\infty}^{-x_2^*} x_2 dF_2(x_2),$$

which in turn follows from the symmetry of x_2. Similarly, for the second case the desired result follows from (15).

We replace $-c_1$ in the top row of Figure 2(b) in a similar fashion. More precisely, the component $-c_1$ appears for every single value of x_1 when $x_2 > x_2^*$ or when $x_2 < -x_2^*$. In the former case, we replace $-c_1$ with $x_2^* + x_1$ for $x_1 < -x_2^*$ and with 0 for $x_1 > -x_2^*$. And in the latter case, we replace $-c_1$ with $x_2^* - x_1$ for $x_1 < -x_2^*$ and with 0 for $x_1 > -x_2^*$. This will increase the integral of the top row across all pairs of realizations of the two attributes by (14) and (15). Indeed, we would not affect the integral by replacing $-c_1$ with $x_1^* + x_1$ for $x_1 < -x_1^*$ and with 0 for $x_1 > -x_1^*$ in the former case, and replacing $-c_1$ with $x_1^* - x_1$ for $x_1 < -x_1^*$ and with 0 for $x_1 > -x_1^*$ in the latter case. However, $x_1^* < x_2^*$, so the left-hand sides of (14) and (15) exceed the right-hand sides of the two equations, respectively.

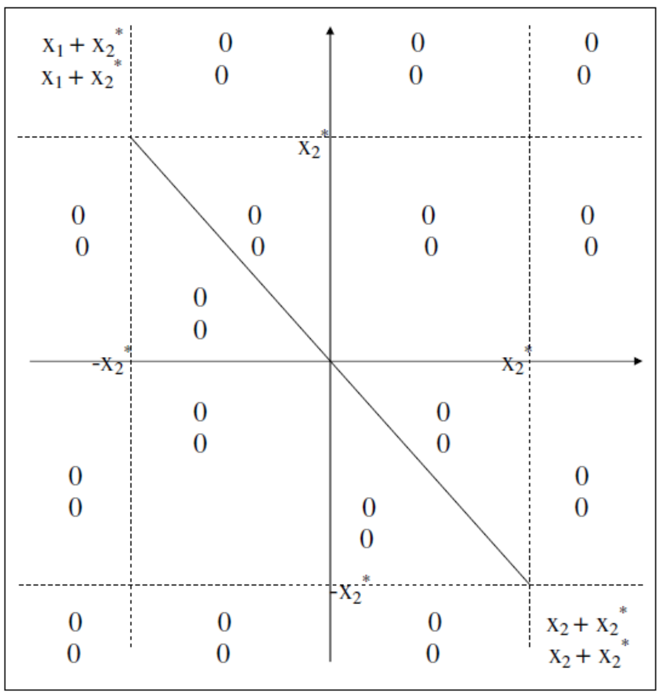

Figure 2 (c)

The entries in the top and bottom rows of Figure 2(c) coincide. This implies that the expected payoff from discovering attribute 2 first, followed by playing a suboptimal continuation strategy is no lower than the expected payoff from discovering attribute 1 first, followed by playing the optimal continuation strategy.■

Proof of the first part of Corollary 1: We first write formula (8) as

$$c_i = \int_{-\infty}^{+\infty} \max\{0, (-x_i^* + x_i)\} dF_i(x_i).$$

Notice that the function which is integrated on the right-hand side of this equation is convex. Thus, $x_1 \precsim^{s.o.} x_2$ implies that for any given value $x^* = x_1^* = x_2^*$, the right-hand side is no lower for $i = 1$ than for $i = 2$. Since $c_1 \leq c_2$, we have that $x_1^* \geq x_2^*$.■

Proof of the second part of Corollary 1: Add $c_i - \int_0^{+\infty} x_i dF_i(x_i)$ to both sides of (8) to get

$$c_i - \int_0^{+\infty} x_i dF_i(x_i) = \int_0^{+\infty} \max\{-x_i, -x_i^*\} dF_i(x_i).$$

By Proposition VI, $x_1 \precsim^{s.s.o.} x_2$ implies that for any given value $x^* = x_1^* = x_2^*$, the right-hand side is no lower for $i = 2$ than for $i = 1$. Since, by assumption, the left-hand side is no higher for $i = 2$ than for $i = 1$, we have that $x_1^* \leq x_2^*$.■

7.4 Proofs of Propositions IV and V

In the following we use the notation $[h] = \{1, \ldots, h\}$ for any integer h. Let

$$g(n) = E\left[\max\left(0, \sum_{i=1}^{n} x_i\right)\right].$$

Then it holds $U([n]) = g(n)$ and thus working with g is equivalent to dealing with U. From this it follows that if it is NP-complete to compute g, then it is also NP-complete to evaluate $U(S)$. (It is already hard for a specific $S = \{1, 2, \ldots, n\}$.)

7.4.1 Proof of Propositions IV

Suppose that $a_1, ..., a_n$ are positive natural numbers. Let $a = (a_1, ..., a_n)$, $\overline{a} = (a_1, ..., a_n, a_{n+1})$, where $a_{n+1} = 1$,

$$M := \frac{1}{2} a([n]),$$

and for any $\varnothing \neq S \subset [n]$,

$$a(S) := \sum_{i \in S} a_i.$$

Define $\overline{M}$ and $\overline{a}(\overline{S})$, where $\varnothing \neq \overline{S} \subset [n+1]$, for $\overline{a}$ in a manner similar to M and $a(S)$ for a.

Let $u(a)$ stand for the number of sets S such that $a(S) = M$. The idea is to show that if we were able to compute expression $g(n)$ in polynomial time, then we would also be able to compute $u(a)$ in polynomial time. And then, by checking whether $u(a) = 0$, we would be able to solve the problem of partitioning, which is known to be NP-complete (see, e.g., Garey and Johnson (1979)).

Assume, without loss of generality, that M is an integer. (Indeed, if M is not an integer, there obviously exists no S such that $a(S) = M$.)

If S denotes set $\{i : x_i = -a_i\}$ and T denotes set $\{i : x_i = a_i\}$, then

$$\sum_{i=1}^{n} x_i = a([t]) - a(S) = (a([n]) - a(S)) - a(S) = -2a(S) + a([n]),$$

and so

$$\sum_{i=1}^{n} x_i \geq 0 \;\Leftrightarrow\; a(S) \leq \frac{1}{2} a([n]).$$

Thus,

$$g(n) = \frac{1}{2^n} \left\{ a([n]) + \sum \left[(-2a(S) + a([n])) : \varnothing \neq S \subset [n],\, a(S) \leq \frac{1}{2} a([n]) \right] \right\} =$$

$$= \frac{1}{2^{n-1}} \left\{ \frac{1}{2} a([n]) + \sum \left[\left(\frac{1}{2} a([n]) - a(S) \right) : \varnothing \neq S \subset [n],\, a(S) \leq \frac{1}{2} a([n]) \right] \right\}.$$

Further, define

$$h(a) := 2^{n-1} \cdot g(n) - \frac{1}{2} a([n]) = \sum_{\varnothing \neq S \subset [n] : a(S) \leq M} (M - a(S)),$$

$$f(a) := |\varnothing \neq S \subset [n] : a(S) \leq M|,$$

$$h(a, -1) = \sum_{\varnothing \neq S \subset [n] : a(S) \leq M-1} (M - 1 - a(S))$$

and

$$f(a, -1) := |\varnothing \neq S \subset [n] : a(S) \leq M - 1|.$$

If we were able to compute $g(n)$ in polynomial time, then we would be able to compute $h(a)$ and $h(\overline{a})$. We will now derive a non-singular system of five linear equation with variables $f(a)$, $f(\overline{a})$, $f(a, -1)$, $h(a, -1)$, $u(a)$ and with constant terms including $h(a)$ and $h(\overline{a})$. Solving this system, we will find the value of $u(a)$.

Since all numbers $a(S)$ are integers, $M - a(S) = 0$ in the expression for $h(a)$ unless $a(S) \leq M - 1$. This yields that

$$h(a) = h(a, -1) + f(a, -1). \tag{16}$$

Further, observe that

$$-1 = f(a) + f(a, -1) - f(\overline{a}). \tag{17}$$

Indeed, $f(\overline{a})$ is the number of non-empty sets $\overline{S} \subset [n+1]$ such that $\overline{a}(\overline{S}) \leq \overline{M}$. There are three types of such sets $\overline{S}$: (i) If $n+1 \notin \overline{S} \subset [n+1]$, then $\overline{a}(\overline{S}) = a(\overline{S}) \leq \overline{M} = M + 1/2 \Leftrightarrow a(\overline{S}) \leq M$; (ii) If $n+1 \in \overline{S} \neq \{n+1\}$, then for $S := \overline{S} - \{n+1\}$, we have $\overline{a}(\overline{S}) = a(S) + 1 \leq \overline{M} = M + 1/2 \Leftrightarrow a(S) \leq M - 1$; (iii) If $\overline{S} = \{n+1\}$, then $\overline{a}(\overline{S}) = 1$. This yields that $f(\overline{a}) = f(a) + f(a, -1) + 1$.

Similarly,

$$h(\overline{a}) = \sum_{\varnothing \neq \overline{S} \subset [n+1] : \overline{a}(\overline{S}) \leq \overline{M}} (\overline{M} - \overline{a}(\overline{S})) =$$

$$= \sum_{\varnothing \neq \overline{S} \subset [n] : \overline{a}(\overline{S}) \leq \overline{M}} (\overline{M} - \overline{a}(\overline{S})) + \sum_{\{n+1\} \in \overline{S} \subset [n+1] : \overline{a}(\overline{S}) \leq \overline{M},\ \overline{S} \neq \{n+1\}} (\overline{M} - \overline{a}(\overline{S})) + \left(\overline{M} - 1\right) =$$

$$= \sum_{\varnothing \neq S \subset [n] : a(S) \leq M} (M + 1/2 - a(S)) + \sum_{\varnothing \neq S \subset [n] : a(S) \leq M-1} (M - 1/2 - a(S)) + (M - 1/2) =$$

$$= h(a) + \frac{1}{2} f(a) + h(a, -1) + \frac{1}{2} f(a, -1) + M - 1/2.$$

This is equivalent to

$$2\left(h(\overline{a}) - h(a) + 1/2 - M\right) = f(a) + f(a, -1) + 2h(a, -1). \tag{18}$$

Clearly,

$$u(a) = |\varnothing \neq S \subset [n] : a(S) = M| = |\varnothing \neq S \subset [n] : a(S) \leq M| - |\varnothing \neq S \subset [n] : a(S) \leq M - 1|$$

$$= f(a) - f(a, -1);$$

that is,

$$0 = -u(a) + f(a) - f(a, -1). \tag{19}$$

Finally,

$$|\varnothing \neq S \subset [n] : a(S) \leq M - 1| + |\varnothing \neq S \subset [n] : a(S) = M| + |\varnothing \neq S \subset [n] : a(S) \geq M + 1| = 2^n - 1.$$

Since $a(S) \leq M - 1 \Leftrightarrow a([t]) \geq M + 1$, where T stands for the complement of S,

$$2|\varnothing \neq S \subset [n] : a(S) \leq M - 1| + |\varnothing \neq S \subset [n] : a(S) = M| = 2^n - 1;$$

that is,

$$2^n - 1 = 2f(a, -1) + u(a). \tag{20}$$

We can now solve the system of equation (16)-(20), and compute $f(a)$, $f(\overline{a})$, $f(a,-1)$, $h(a,-1)$, $u(a)$ as a function of $h(a)$ and $h(\overline{a})$. Indeed, the matrix

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 2 & 0 \\ 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 2 & 0 & 1 \end{bmatrix}$$

is non-singular. Therefore, if we were able to compute $h(a)$ and $h(\overline{a})$ in polynomial time, then we would also be able to compute $u(a)$ in polynomial time. ■

7.4.2 Proof of Proposition V

The proof follows immediately from the following lemma.

Lemma 3: There exists a pseudo polynomial-time algorithm for computing $g(n)$.

Proof of Lemma 3: Suppose that x_i, $i = 1, ..., n$, takes values a_i in a finite set S_i. For $t = 1, ..., n$ and $a^t = (a_1, ..., a_t)$, define

$$p_t(a^t) := \prod_{i=1}^{t} \Pr\{x_i = a_i\},$$

and for any $t = 1, ..., n$, define

$$V_t(b) = \sum \left\{p_t(a^t) \cdot (a([t])) : a([t]) \geq b\right\}.$$

We will develop a recursive definition of $V_t(b)$:

$$V_t(b) = \sum \left\{p_t(a^t) \cdot (a_t + a([t-1])) : a([t]) \geq b\right\} =$$

$$= \sum \left\{p_t(a^t) \cdot a_t + \Pr\{x_t = a_t\} p_{t-1}(a^{t-1}) \cdot (a([t-1])) : a([t]) \geq b\right\} =$$

$$= \sum \left\{p_t(a^t) \cdot a_t : a([t]) \geq b\right\} + \sum_{a_t} \Pr\{x_t = a_t\} V_{t-1}(b - a_t).$$

Let

$$U_t(b) := \sum \left\{p_t(a^t) \cdot a_t : a([t]) \geq b\right\}.$$

Then

$$U_t(b) = \sum_{a_t} a_t \Pr\{x_t = a_t\} \cdot \left(\sum \left\{p_{t-1}(a^{t-1}) : a([t-1]) \geq b - a_t\right\}\right).$$

Let

$$z_t(b) := \sum \left\{ p_t(a^t) : a([t]) \geq b \right\}.$$

Then

$$z_t(b) = \sum_{a_t} \Pr\{x_t = a_t\} \cdot \left(\sum \left\{ p_{t-1}(a^{t-1}) : a([t-1]) \geq b - a_t \right\} \right) =$$

$$= \sum_{a_t} \Pr\{x_t = a_t\} \cdot z_{t-1}(b - a_t).$$

Therefore, $z_t(b)$ can be computed recursively, and so $U_t(b)$ and $V_t(b)$ can be computed recursively.

The recursive definition of $V_t(b)$ provides an algorithm for computing $g(n) = V_n(0)$ whose running time is polynomial in the following three parameters: n,

$$s := \max_{i=1,\dots,n} |S_i|,$$

and the common bound L on the size of $a_i \in S_i$, $i = 1, ..., n$. Note that for each $t = 1, ..., n$, functions z_t, U_t and V_t need to be computed for $b = 0, 1, ..., L$. ■

It is straightforward to observe that the presented algorithm is valid for computing $U(S)$ for an S by relabeling the elements in S.

8 References

Chade H. and L. Smith (2005): "Simultaneous Search," *Econometrica* **74**, 1293–1307.

Diamond, P. A. and J. E. Stiglitz (1974): "Increases in Risk and in Risk Aversion," *Journal of Economic Theory*, **8**, 337-360.

Fujishige S. (2005). *Submodular Functions and Optimization.* Elsevier, Amsterdam.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006): "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, **96**, 1043-1068.

Garey M.R. and D.S. Johnson (1979): *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freeman & Co., New York.

Gittins, J.C. and D.M. Jones (1974): "A dynamic allocation index for the sequential design of experiments," in *Progress in Statistics (J. Gani et al., eds.)*, 241–266, North-Holland, Amsterdam.

Gittins, J.C. (1989). *Multi-armed bandit allocation indices.* John Wiley & Sons, Ltd., New York.

McCormick S.T. (2008): "Submodular Function Optimization," Mimeo.

Neeman, Z. (1995): "On Determining the Importance of Attributes with a Stopping Problem," *Mathematical Social Sciences*, **29**, 195-212.

Sanjurjo, A. (2013): "Search with Multiple Attributes: Theory and Empirics," Mimeo.

Weitzman, M.L. (1979): "Optimal Search for the Best Alternative," *Econometrica*, **47**, 641-654.