On the value of efficient search*

Christopher J. Costello† and Michael B. Ward

October 14, 2004

Abstract

Costly search for rare items pervades society. We model R&D as a process of uncertain search and discovery. A collection of research leads can be searched at random or in a more efficient manner informed by additional information. We examine the value of information that facilitates efficient search. We find for a large class of problems that this value is very small. For example, when applied to the current controversy over biological prospecting, optimally ordering research leads improves the value of the collection only 3% above random search. Results contradict widely-held views on the value of the research process.

1 Introduction

What do a cure for cancer, the incandescent light bulb, and the proof of Fermat's Last Theorem have in common? These are just a few examples of research and development (R&D) applications that require costly search of a potentially large collection of research leads. Information acquired through, for example, preliminary testing, searching the literature, or developing a scientific framework, can improve search efficiency and enhance the value of the R&D project. However, acquiring this information may be costly. When determining whether to pursue additional information, the pivotal question, and the focus of this paper is: how large are the benefits of efficient search? And while a large theoretical literature addresses the question of how to conduct maximally efficient R&D (Granot and Zuckerman 1991; Gallini and Kotowitz 1985; Roberts and Weitzman 1981; Lucas 1971; Ross 1969; Charnes and Stedry 1966), the value of R&D efficiency has not been examined.

Across actual R&D applications, the amount of preliminary testing can vary significantly. Take for example two famous R&D projects by the American inventor, Thomas Edison. In the development of the incandescent light bulb, Edison's technicians laboriously waded, essentially at random, through thousands of different filament materials before finding that carbonized sewing thread met his durability and brightness criteria. In contrast to this largely unordered search, during the search for a domestic source of rubber, Edison carefully ranked hundreds of plant species according to their potential before extensively testing and cross-breeding varieties of the most promising species.

Why might some R&D applications call for a meticulously researched scientific framework to inform discovery, while others require no more than a random search of even large

*__DRAFT DATE October 14, 2004__. PLEASE DO NOT CITE

†Corresponding Author: Donald Bren School of Environmental Science & Management, 4410 Bren Hall, Santa Barbara, CA 93117, Costello@bren.ucsb.edu

collections of research leads? In this paper we examine theoretically and empirically the value of efficient R&D.

Intuitively, it would seem that efficient search must have high value. Indeed, Rausser and Small (2000, p. 174) argued "It is a powerfully general rule that no one ever searches for anything by examining large collections of objects in random order". Surprisingly, we find that for a large class of search problems, the value of information facilitating efficient search is trivially small. This result is shown by deriving an upper bound on the value of optimal search as a simple function of the pivotal characteristic parameters of R&D problems. Moreover, we determine what properties tend to make efficient search more, or less, valuable. Even in some seemingly favorable cases, the value is still small.

The paper is organized as follows. We develop in section 2 a model in which a collection of leads is ordered and searched sequentially. We derive the expected value of efficient search. The dependence of this value on features of the search problem are presented in section 3 as is a derivation of the theoretical upper bound on the value of efficient search for the most favorable non-parametric probability distribution.

In section 4 we apply our theoretical model empirically to the case of bioprospecting, where we find, contrary to previous studies, that the value of efficient search is only about 3%. We find that features of the bioprospecting problem (e.g. a very large number of highly diverse research leads) should lead to a higher, rather than lower, value of efficient search. Viewed in that light, we conclude that search problems with less extreme characteristics may benefit even less from efficient search. We continue in section 5 and consider the interplay between efficient search and an alternative approach to discovery in which a long list of leads is coarsely sorted into two piles - those worthy of future search and those that should be discarded. That section illustrates that the value of coarse sorting can be substantial implying that information facilitating efficient search can have significant value, but not for reasons that have thus far been explored. We conclude in section 6.

2 R&D as a search

In this section, we lay out a stylized model of R&D, with three distinct phases: identifying the collection of leads (Phase I), optionally refining information on quality of leads (Phase II), and searching the leads for an eventual success (Phase III).

2.1 Overview

At the beginning of a project, the researcher faces a large collection of leads of potentially varying quality (probability of success). In Phase I, the researcher coarsely sorts this collection into two piles: those worthy of further investigation, and those that should be discarded. Phase I sorting is conducted on the basis of crude prior knowledge of the leads' potentials. In the context of the light bulb example, one might think of Phase I as the choice to search among carbonized filament materials.[1] In the search for a domestic rubber source, Phase I might be thought of as the limitation of leads to endemic plant species with some known rubber content.

Once the collection of worthy leads has been identified, the researcher faces a choice that is central to our analysis: whether to engage in Phase II. In this step, the researcher

[1] In fact Edison purchased a patent on the use of carbonized filament from another inventor. The preferred embodiment listed in the patent, carbonized bamboo filament, was not sufficiently durable for commercial use.

can further refine his information about lead qualities in order to more efficiently search among them. Alternatively, the researcher can skip this phase, and simply search randomly among the collection of leads identified in Phase I as worthy.[2] In the search for a domestic source of rubber, Edison's team planted thousands of species for preliminary testing. After assessing the rubber content of these initial plantings, the team ranked the species in terms of promise for more thorough investigation with the top one being goldenrod. In contrast, in the light bulb search, testing of thousands of candidate filament materials proceeded in a fairly haphazard way. In our stylized view, we think of the light bulb search as skipping Phase II.

The actual search, Phase III, involves the full testing of leads for a success. In this phase, the researcher sequentially tests leads. These tests resolve all uncertainty. In the light bulb search, this involved checking whether a candidate filament in light bulbs met minimal durability and brightness criteria. In the case of rubber, this involved testing whether by cross-breeding of varieties a sufficiently low-cost, high-yield specimen could be developed.

This stylized view of R&D as a multi-stage process of lead identification, investigation, and uncertain discovery of patentable products is consistent with approaches taken in the literature (see, e.g. Gallini and Kotowitz (1985), Fudenberg et al. (1983), Roberts and Weitzman (1981), Charnes and Stedry (1966)).[3] For example, Fudenberg et al. (1983) model two stages of R&D: a preliminary invention phase followed by the development phase; our analyis maps these to Phase II and Phase III, respectively. Similarly Gallini and Kotowitz (1985) distinguish between the basic research and the development phases. As in Phase II here, the purpose of the basic research step is to reduce the number of research avenues so that in the development phase research can focus in on the few most promising projects.

As in Granot and Zuckerman (1991), Gallini and Kotowitz (1985), Lucas (1971) and others we will focus initially on an R&D project that is already under way. Here, the researchers have already identified in Phase I which leads from a potentially infinite initial set are worthy of further investigation. Most of this literature focuses on optimally engaging in Phase II, the second development stage of a project. While our focus is also on this phase, we will also consider the interplay between the basic research and the development phases of research.

2.2 Model

In this section we formalize the model of efficient search of those leads. Following Rausser and Small (2000), Simpson et al. (1996), Ross (1969), and others, N research leads identified in Phase I as worty of search are to be sequentially tested for a success. The search terminates upon the first success yielding gross payoff R. Each test involves search cost c, and lead i has success probability p_i, which is independent of the success of all other leads and, from the researcher's perspective, may be uncertain. As a consequence of Phase I research, leads are retained only if their expected benefits exceeded the cost of their search:

[2] Of course, it is likely that a researcher would have *some* priors on lead quality. Here, we take a stylized view that no prior information is available to rank lead quality. Then, if Phase II is omitted, search must proceed in random order. This stylized view is conservative in that we tend to understate the real information available to a researcher, and thus overstate the benefits of acquiring more information.

[3] We analyze the problem from the perspective of a single researcher or firm attempting to identify a success among a pool of research leads. This is distinct from the "patent race" literature in which multiple firms are competing to achieve the same success (see, e.g. Gallini and Kotowitz (1985) and Fudenberg et al. (1983)).

$p_iR > c$. The presence of sequential search with positive search cost is what distinguishes this model from that in Polasky and Solow (1995), who also consider the value of a collection of leads. Because they treat search cost as zero, search order is irrelevant, and therefore the type of information considered here could not have value.

The researcher may refine his beliefs about success probabilities by collecting additional information. The details of this updating process will depend on the R&D application, but may involve things like preliminary testing, developing a scientific framework, or incorporating indigenous knowledge. While this information will improve search efficiency, its acquisition may be costly. To calculate the value of the information, we take as a benchmark the case in which the researcher has an identical prior over all p_i's (that is, no way to distinguish between the quality of leads) in which case the collection of leads is examined in random order. Undertaking Phase II of the research process would provide further refining information that identifies the probabilities precisely and would thus facilitate a more efficient search. We use the phrase "value of efficient search" to denote the difference in the *ex ante* expected value of the collection of leads under fully optimal and random search; that is the difference between the expected value of the project with the information provided in Phase II and the expected value of the project when the researcher skips immediately from Phase I to Phase III. Because less precise information would confer lower value, our calculations represent an upper bound on the value of efficient search.

Let S be a collection of N research leads with an associated set of probabilities $\{p_i\}$. By the virtue of Phase I, each lead in this collection is of sufficiently high quality to justify its search; we relax this assumption in a later section. In this section, we calculate the value of optimally *ordering* those leads in the search queue.

Denote a specific ordering of the N research leads in set S by $S^\wedge$ with associated probabilities $\{p_1^\wedge, p_2^\wedge, ..., p_N^\wedge\}$, respectively. The level of certainty the researcher has about these probabilities depends on the information she has at the time of the search. In this arbitrary ordering, lead $S_i^\wedge$ is the i^{th} lead to be tested. The expected value of the ordered collection $S^\wedge$ is given by

$$V(S^\wedge) = R(1 - a_{N+1}(S^\wedge)) - c\sum_{n=1}^{N} a_n(S^\wedge) \tag{1}$$

where $a_n(S^\wedge) = \prod_{i=1}^{n-1}(1 - p_i^\wedge)$ is the probability that every preceding lead has been tested unsuccessfully. The formula (1) has an intuitive explanation. The first term on the right hand side is the expected revenue of the search: the probability that at least one lead contains a success times the revenue upon success. The total value of the ordered collection is that revenue less the expected cost – the expected number of trials until the first success multiplied by the cost of each trial.

2.3 Value of Search

We consider in particular two types of queues, each represented by a researcher who searches that queue. Consider first Researcher 1 who has completed Phase II and thus has a well-developed prior distribution over all of the probabilities of success; this information facilitates efficient search of the collection. In other words, she has a well-developed subjective belief about the success probabilities of her research leads, and can therefore order them optimally. The problem of efficiently ordering the search queue is then a special case of the "Pandora's box" problem analyzed by Weitzman (1979). In our case, an efficient search queue orders the leads in descending order of the probability of success[4]. We denote by S^* the optimally ordered queue. The *ex ante* expected value of the collection by Researcher 1 is then $V_1 = V(S^*)$.

How much would Researcher 1 be willing to pay, *ex ante*, for the queue S^* rather than some alternative ordering of the same leads, $S^\wedge$? This value is given by is given by:

$$V_1 - V(S^\wedge) = c\left(\sum_{n=1}^{N} a_n(S^\wedge) - \sum_{n=1}^{N} a_n(S^*)\right) \tag{2}$$

Note that expected revenue plays no role here because the probability of conducting a full search without any successes is independent of the search order: $a_{N+1}(S^*) = a_{N+1}(S^\wedge)$ for the ordering, $S^\wedge$. Equation 2 holds *for any* sequence $S^\wedge$.

At the other extreme is Researcher 2 who skips Phase II and thus lacks any information distinguishing one lead from another, and so searches the collection in random order. To be concrete we assume that Researcher 2 adopts the same distribution over the probabilities $\{p_i\}$ as Researcher 1 but does not know which lead is which, and so searches those leads at random. Denote by V_2 the *ex ante* expected value of the collection for Researcher 2. It is given by the sum of the values of all possible search queues divided by the number of such queues, as follows:

$$V_2 = E\left[V(S^\wedge)\right] = \frac{\sum_{S^\wedge} V(S^\wedge)}{N!}. \tag{3}$$

We denote the difference in value between Researcher 1 and Researcher 2 by $\Theta_{12} \equiv V_1 - V_2$, which gives the value of efficient search of the collection of leads.

An alternative measure of the value of efficient search that may be more informative is in percentage terms. We define this measure $\Pi_{12} = 100(V_1 - V_2)/V_1 = 100\Theta_{12}/V_1$ as the percent difference between the value of a collection when it is searched optimally and the value of the same collection when it is searched at random. This measure may be useful as it captures the value of efficient search relative to the scale of the optimal search problem. For example, Rausser and Small argue that efficient search dramatically changes conservation incentives. This is essentially a claim about the value of efficient search *relative* to random search. One natural measure of such relative impacts is the percentage difference captured by Π_{12}.

3 Analyzing the theoretical model

In this section we analyze the theoretical model of Section 2 to derive general properties of the value of efficient search, emphasizing the features of an R&D problem that would tend to make gathering information more, or less, valuable. Here, we assume the researcher has completed Phase I, and is faced with a collection of leads already known to be *individually* worthy of search, but whose precise probabilities may be unknown. In Section 4.3, we will address the value of coarsely sorting leads into those worthy and unworthy of search.

We begin by deriving in section 3.1 an upper bound on the value of efficient search that depends only on gross properties of the problem, rather than on the probabilities of all the leads. This result supports a simple rule of thumb to determine whether it might be worthwhile to gather more detailed information about lead quality prior to search. More

[4] A more general treatment would allow cost to vary across leads. Applying Weitzman's (1979) formula, it can be shown in that case that leads should be ordered in descending order of p_i/c_i.

generally, it illustrates that under a broad class of circumstances, the value of efficient search cannot be large. Second, we identify features of a search problem that tend to make efficient search more, or less valuable (section 3.2).

3.1 Upper bound on the value of efficient search

In general, the value of efficient search will depend in a complicated way on the probabilities of all the leads, as shown in equation (2). It seems intuitively plausible that for some favorable probability distributions the value of efficient search could be enormous. In this section we examine this question by deriving the non-parametric distribution of probabilities that gives the theoretically maximal value of efficient search.

What would the most favorable non-parametric probability distribution look like? Under optimal search, the worst lead must have probability of at least c/R to justify search. This is because the marginal value of the worst lead under optimal search is $p_i R - c$ and so if $p_i < c/R$ the lead would be dropped.[5] Holding constant the overall success probability, P, the probability distribution that yields the largest value of efficient search, therefore, would have $N-1$ leads of probability c/R. The remaining lead $\bar{p}$ is then set large enough that the overall success probability equals P, as required; this value is $\bar{p} = 1 - P/(1-\delta^{N-1})$. Under this probability distribution, Researcher 2 (who searches at random) must wade through a large number of marginal leads before eventually succeeding but Researcher 1 (who searches the best lead first) is most likely to find success on the first trial. Proposition 1 below calculates the value of efficient search under this probability distribution. Because this value would be smaller under any other probability distribution, we interpret it as an upper bound on the value of efficient search. Defining by $\delta = 1 - c/R$ the probability of failure when searching a marginal lead, we have:

Proposition 1 **(a)** *The nominal value of Phase II R&D,* Θ_{12}*, cannot exceed:*

$$\bar{\Theta}_{12} = c(\frac{\bar{p}}{1-\delta} - 1)(1 - \frac{1-\delta^N}{N(1-\delta)}). \tag{4}$$

(b) *The percentage value of Phase II R&D,* Π_{12}*, cannot exceed:*

$$\bar{\Pi}_{12} = 100\left[1 - \frac{1-\delta^N}{N(1-\delta)}\right] \tag{5}$$

A Proof is in the Appendix.

The calculations in Proposition 1 are important for two reasons. First, they will support our empirical investigation in the next section. Second, they provide rules of thumb that could help guide the direction of research. Because $\bar{\Theta}_{12}$ and $\bar{\Pi}_{12}$ do not require any knowledge of the probability distributions of leads, the bounding argument could be used in practice to provide an upper bound on the value of the research process even when very little is known about the nature of the leads themselves. $\bar{\Pi}_{12}$ is particularly useful in this

[5] Note that under random search the "cutoff" value could be slightly larger than c/R, so leads with probability $< c/R$ would not ever be searched. Calculating analytically this precise cutoff value is extremely complex. Simulations reveal that the cutoff is numerically extremely close to c/R (typically within less than 1%). Using c/R as the cutoff only overstates the value of efficient search. This is a second reason for calling the derived value an upper bound.

regard because it requires estimates only of the size of the collection of leads (N) and the ration of search cost to revenue ($c/R = 1 - \delta$). In particular, it is independent of the overall success probability, P so little information is required to assess whether further refining information would be useful in the R&D process. It is straightforward to show using the bounding argument that the attractiveness of efficient search increases in the pool of leads, N, and the relative size of c/R; $\frac{\partial \bar{\Pi}_{12}}{\partial N} > 0$ and $\frac{\partial \bar{\Pi}_{12}}{\partial \delta} < 0$. The empirical magnitude of $\bar{\Pi}_{12}$ remains an open question which we address in section 4.

3.2 What makes information valuable?

We have shown that the value of efficient R&D can be bounded above by the expressions in Proposition 1. However, the simplicity of the upper bounds comes by abstracting away from the potentially complicated exact distribution of lead qualities. While this allows simple insights into a difficult problem, it is also useful to understand what factors influence the value of search in a less extreme setting. In this subsection, we present comparative statics results for the exact value of efficient search. It is important to note that these comparative statics are marginal, in the sense that we assume we are already in Phase II; that is no leads are added or removed from the queue as parameters change.

The most straightforward parameters to analyze are cost c and revenue R. Assessing the effects of characteristics of changes in the distribution of leads is more challenging. Pertinent characteristics include the number of leads, the overall quality of leads, and the dispersion of leads which we call "spread". To do so a working definition of spread is needed. Suppose we replace two leads p_i, p_j in a pool with q_i, q_j such that $q_i > p_i > p_j > q_j$ and $(1 - p_i)(1 - p_j) = (1 - q_i)(1 - q_j)$. Then we say that the second pool has greater spread, as suggested by the values of p and q. The second condition is a normalization, which preserves the overall probability of success in the search.

The propositions below characterize the dependence of the expected value of efficient search on search payoff, search cost, and several properties of the distribution of leads. We present results for both the nominal (Θ_{12}, Proposition 2) and percentage (Π_{12}, Proposition 3) value of efficient search.

Proposition 2 *The nominal value of Phase II R&D,* Θ_{12}*:*

- **(a)** *is independent of the success payoff, R,*
- **(b)** *is linearly increasing in search cost, c,*
- **(c)** *decreases if all probabilities are increased by the same (small) amount,*
- **(d)** *increases as the spread of the lead probabilities increases, and*
- **(e)** *decreases (weakly) if the lowest probability lead is eliminated from the collection*

Proposition 3 *The percentage value of Phase II R&D,* Π_{12}*:*

- **(a)** *is decreasing in the success payoff, R,*
- **(b)** *is increasing in search cost, c,*
- **(c)** *decreases if all probabilities are increased by the same (small) amount.*
- **(d)** *increases as the spread of the lead probabilities increases.*

Proofs of these Propositions are included in the Appendix.

3.3 Discussion of comparative statics results

(a) dependence on revenue

Efficient search only improves search value insofar as it reduces search cost; it does not impact the probability of success of a search, and therefore, expected revenue is play no role in the nominal value of efficient search. However, the larger is R the smaller is a given percentage gain from efficient search.

(b) dependence on search cost

The fact that the value of efficient search is increasing in cost is not surprising: the higher the search cost, the higher the opportunity cost, and therefore, the greater is the value of forgoing searches of poor leads. In section 5 we examine a second effect of increasing c. As cost increases, the pool of acceptable leads may shrink, which decreases the value of efficient search. This tension is carefully analyzed in section 5.

(c) dependence on probability of success

We now explore the role of the overall probability of success of the search in determining the value of efficient search. To do so we need to be specific about how the overall probability of success is increased. One natural way to do so is to increase by a small amount all of the probabilities. Intuition may suggest that the less likely is an eventual success, the less important is information that facilitates optimal search. This intuition turns out to be precisely backwards. Careful analysis reveals that the savings in the expected number of searches as the probabilities are increased is smaller under the optimal ordering than under *any* alternative ordering. Therefore, the value of efficient search declines as the probabilities are increased.

(d) dependence on the spread of the probabilities

To begin, note that if all probabilities are equal, random search is optimal, so $\Theta_{12} = \Pi_{12} = 0$. In contrast, the more spread out is the probability distribution, the greater is the opportunity to exploit information to re-order the search. This is essentially the phenomenon that Rausser and Small (2000) intended their data to illustrate. Holding constant the overall probability of success the result is consistent with this intuition; higher spread increases the value of efficient search.

(e) dependence on the size of the pool of leads

Proposition 2e states that Θ_{12} decreases if the poorest quality lead is eliminated from the collection. While the formal proof (see appendix) is somewhat complex, it makes intuitive sense. Eliminating the poorest quality lead from the collection reduces the advantage Researcher 1 (who orders the collection) has over Researcher 2 (who searches at random). One could ask more generally what is the expected effect on Θ_{12} of eliminating a randomly selected lead from the collection (rather than the least promising lead)? From numerous numerical calculations, it appears that doing so always decrease Θ_{12} in expectation, though we cannot prove this formally.

In comparing the nominal value of efficient search, Θ_{12}, to the percentage value of efficient search, Π_{12}, it is noteworthy that the comparative statics all work in the same direction (with the exception of the comparative statics with respect to R which does not affect Θ_{12}).

The above results show that efficient search is most valuable when search cost is high, the number of leads is large, the distribution of success probabilities is highly spread out, and the overall probability of success is small.

4 Bioprospecting as a search problem

Simpson, Sedjo, and Reid (1996) consider a bioprospecting model in which the leads are endemic plant species present in biodiversity hotspots around the world. In that model, the researcher has no prior information on how to order the search, and so all leads are treated as having the same probability of success, $\tilde{p}$, and they are searched at random. The search terminates upon the identification of a natural compound with high pharmaceutical value. They find that the marginal value of a species is likely to be small regardless of $\tilde{p}$. When $\tilde{p}$ is small, the marginal value of a species is small because any one species is unlikely to produce a success. When $\tilde{p}$ is large, the marginal value is small because species are close substitutes (any one species is likely to be redundant). Using a species area curve relationship, they translate the value per species into a value per hectare. The maximum value per hectare is about \$21 and is in Western Ecuador, the most biodiverse region on earth.

In sharp contrast is the result derived by Rausser and Small (2000) in which leads are kilo-hectares of biologically diverse land (they consider the same 18 biodiversity hotspots considered by Simpson et al.). In that model, success probabilities are heterogeneous, and the researcher benefits from ordering her search to improve efficiency. Rausser and Small (2000) conclude that the most biodiverse hectare of land on earth (again in Western Ecuador) has a bioprospecting value of \$9,177 given an efficient search queue. Rausser and Small attribute this dramatic increase (from \$21/hectare to \$9,177/hectare) to efficient search. The marginal values of land in all 18 biodiversity hotspots are reproduced in Table 1 below.

4.1 An experiment on the bioprospecting data

We now apply the value of efficient search model derived in section 2 to bioprospecting as a search problem. While the Simpson et al. and Rausser and Small models did differ slightly, the central message of Rausser and Small (2000) is that information that facilitates optimal search significantly improves value, and therefore raises the likelihood of private sector conservation. They refer to the random search assumption in Simpson et al. (1996) as a "nearly cost maximizing approach to discovery" (p. 175). In that debate, search efficiency appears to have an important effect; under efficient search the marginal values increase 440-fold (from \$21/hectare to \$9,177/hectare). Here we carefully examine that result, and segue into an examination of the generality of that claim.

As a point of departure, we conduct the following experiment. We employ the setup and parameter values from Rausser and Small (2000) to determine the marginal values and the value of the collection of 18 biodiversity hotspots under three different assumptions about the information available to the researcher (and therefore about the search order). The first case is analogous to our Researcher 1, who has already conducted Phase II of R&D and therefore searches the pool of leads in the most efficient order. The second case is analogous to our Researcher 2, who searches the leads at random. Finally, for illustrative purposes, we present a third case in which the leads are searched in the maximally inefficient order - in ascending order of the likelihood of success.

	Incremental Value ($/hectare)		
Biodiversity Hotspot	Optimal	Random	Backwards
Western Ecuador	9,177	8,836	8,455
Southwest Sri Lanka	7,463	7,190	6,882
New Caledonia	5,473	5,277	5,056
Madagascar	2,961	2,863	2,751
W. Ghats of India	2,026	1,963	1,890
Phillippines	1,973	1,912	1,841
Atlantic Coast Brazil	1,867	1,809	1,744
W. Amazonia	1,043	1,012	977
Tanzania	811	787	760
C. Floristic S. Africa	632	614	593
Peninsular Malaysia	539	522	503
SW Australia	435	420	402
Ivory Coast	394	379	362
North Borneo	332	317	301
Eastern Himalaya	332	317	301
Colombian Choco	231	215	198
Central Chile	231	215	198
CA Floristic Province	0	-20	-43
Value of collection	$V_1 = 42.1\text{x}10^9$	$V_2 = 41.1\text{x}10^9$	$40.1\text{x}10^9$

The first column of Table 1 lists the 18 biodiversity hotspots examined by Simpson et al. (1996) and Rausser and Small (2000), which were originally described by Myers (1988) and Myers (1990). The second through fourth columns provide the marginal values associated with each hotspot assuming: perfect information which facilitates fully efficient ordering of the collection (column 2), no information which means leads must be searched in random order (column 3), and to explore the true cost maximizing approach to discovery, we examine the case in which leads are searched in the least efficient order – backwards (column 4)[6]. The details of these calculations are provided in the Appendix. Rather than increasing marginal values by two or more orders of magnitude, as is suggested by the debate in the literature, even perfect information about lead successes turns out to increase marginal values by only about 3%. The conclusion for the total value of the collection is commensurate: $42.1 billion under optimally-ordered search vs. $41.1 billion under random-ordered search[7].

4.2 Features of the bioprospecting problem

Above we demonstrated that even perfect information facilitating efficient search has relatively low value in the bioprospecting example. Here, we examine to what extent is this

[6]Note that the worst lead has a negative marginal value under both random and backwards search. This occurs because the probability of the worst lead is exactly c/R, and as noted in section 3.1 the marginal value of that lead would be negative under alternative search orders; thus that lead should be eliminated from the collection under random or backwards search. Doing so would increase the value of the collection slightly (by about 1%), bolstering our result that efficient search has low value.

[7]In another paper (Costello and Ward, 2004) we examine in detail the root causes of the difference between the marginal values obtained by Simpson et al. (1996) and Rausser and Small (2000). Most of the difference can be attributed simply to differences in parameters.

result an artifact of specific features of the bioprospecting problem. We employ the results from the theoretical model in section 2 to gain insight.

The bioprospecting data have several salient features. There is a very large pool of leads (74,600 kilo-hectares). Although success probabilities vary dramatically among the leads, the probabilities are all quite small. The greatest probability is 0.000105, and the smallest is about two orders of magnitude smaller. Given the empirically small value of efficient search in the bioprospecting problem, it is tempting to conclude that at least some of these features are responsible for the low value of efficient search. In fact we find the opposite; each of these factors tends to increase the value efficient search. We explore the impact of these features by calculating the empirical comparative statics parallelling Proposition 2.

Dependence on cost, c

Proposition 2 shows that the value of efficient search, Θ_{12} is linearly increasing in search cost, c. In comparing random-ordered search to optimally-ordered search in the numerical example above, we find that the slope is about $2.1 million: a $1 increase in cost leads to a $2.1 million increase in the value of efficient search, Θ_{12}.

It is tempting but incorrect to attribute the low value of efficient search to the low search cost, c. This is due to a subtlety that involves eliminating unacceptably poor leads from the collection as c increases. We analyze these dynamics in section 5.

Dependence on magnitudes of the probabilities

We showed above that increasing the probabilities decreased Θ_{12}. Consider for example, increasing each probability in the empirical example by $\tilde{p} = .000012$ (the Simpson et al. (1996) probability). In that case, the value of moving from random ordered search to fully efficient search, Θ_{12}, decreases by about $284 million (28%).

Dependence on spread

Above we made both intuitive and theoretical arguments about why increasing the spread would increase the value of efficient search. For example, consider increasing the spread of the empirical distribution of lead probabilities in the bioprospecting example. If the highest probability lead is increased by 1% (to 1.06E-4) and the lowest probability lead is decreased accordingly to leave the overall success probability unchanged (to 1.07E-6), Θ_{12} increases by $1.7 million.

Dependence on size of the pool

For illustrative purposes, consider the Rausser and Small (2000) data which consist of a large pool of 74,600 leads. It is perhaps tempting, but incorrect, to believe that this large pool of leads contributes to the relatively small value of efficient search. Adding a randomly selected lead to the optimal queue has a value of about $453 compared to a value of $430 when a randomly selected lead is added to the random queue, meaning that Θ_{12} increases by about $23 with the addition of a randomly selected lead. This result seems to hold more generally. Suppose for example the list of leads is doubled. Then Θ_{12} increases from about $1 billion to about $3.1 billion.

Dependence on the distribution of probabilities

The empirical comparative statics above assess how marginal changes in any particular parameter affect the value of efficient search. But the probabilities of success in the bioprospecting data set are somewhat *ad hoc* and are meant by the authors to be only illustrative. We would like to determine empirically how large the value of efficient search might become under more favorable probability distributions. To do so, we invoke Proposition 1, which derives an upper bound on the value of efficient search for *any* non-parametric probability distribution. Applying the proposition to the parameters from the Rausser and Small search problem ($c = 485$, $R = 4.5e8$, and $N = 74,640$) underscores the low value of efficient search. It turns out that even when the most favorable probability distribution is used, the percentage value of efficient search ($\bar{\Pi}_{12}$) is still a mere 3.9%.

In this light, we can revisit the empirical example of bioprospecting where the parameter values should lead to a large, rather than small, value of efficient search. This suggests that for search problems with empirical characteristics that are less extreme than those in the bioprospecting example (i.e. for problems with lower search cost, less diverse leads, or higher probability of success), efficient search will have an even lower value than the one derived here.

4.3 Coarse sorting

To illustrate the power of Phase I R&D, we provide the following example. Suppose the collection of biodiversity hotspots had contained a less promising biodiversity region. To make the example concrete, suppose the biodiversity-rich country of Argentina was included in the original collection. Argentina contains about 1100 endemic plant species over 273,669 kilo-hectares for a success probability of p_A =4.8E-8.[8] Since $p_A R < c$, Argentina would be eliminated from the queue in Phase I. But before Phase I the researcher contains insufficient information to eliminate Argentina and all leads (including Argentina) are searched in random order, the *ex ante* expected value of the collection is only \$11.7 billion. This can be compared with a value of \$41.1 billion when Argentina is removed but the remaining leads are searched in random order (Phase I only), and a value of \$42.1 billion when Argentina is removed and the remaining leads are searched in fully efficient order (Phases I and II). In this example, about 73% of the value of the collection is lost from keeping Argentina in the queue, while only 2.3% of the value is lost by not searching the "good" leads (i.e. the original leads) efficiently.

Even more striking is the case in which the collection has a sufficiently large number of low probability leads to render the expected value of random search negative. In the absence of further refining information, such collections would never be searched. For example, suppose in addition to the California Floristic Province (hotspot # 18), the remainder of the United States was included in the original collection of biodiversity hotspots. This region contains 1900 endemic plant species over 891,296 kilo-hectares for a probability of success of p_{US} =2.5E-8. If no information distinguishing lead quality is available, adding the US renders the total collection of leads valueless (searching this entire collection at random would entail an expected *loss* of \$56 billion), so the collection would never be searched. In such cases, a coarse-grained sorting of leads can eliminate the low quality leads from the collection thereby saving search costs and substantially increasing the value

[8]We assume here the Rausser and Small (2000) formula for the probability of success, which is $\tilde{p} * 1100/\text{kilo-hectares}$.

of the remaining collection. Even when the remaining collection of leads is searched in random order the value of the collection is vastly improved.

5 A comparison of information types

We use the phrase "optimal ordering" to refer to Phase II R&D - further refinement of the search sequence over the promising leads. We compare these two approaches for efficient search by considering another hypothetical investigator, Researcher 3, to supplement our existing two. Recall that Researcher 1 completes Phases I and II of the project, and so benefits both from coarse sorting and from optimal ordering. Researcher 2 completes Phase I, but skips Phase II and so benefits only from coarse sorting. Researcher 3 is entirely naive, skipping both Phases I and II, and so inspects the full collection of leads including those for which $p_i > c/R$. Clearly, the value of any collection of leads (recall this is denoted V_i for Researcher i) will vary between the researchers, with $V_1 \geq V_2 \geq V_3$.

In accordance with previous notation, define by $\Theta_{ij} \equiv V_i - V_j$ the difference in the value of the collection for Researcher i and the value of the collection for Researcher j. Then Θ_{12} describes the value of optimally ordering a collection of leads that are "search worthy" and Θ_{13} describes the value of coarse sorting and optimally ordering the collection. In percentage terms, these differences are defined as follows:

$$\Pi_{ij} \equiv 100 * \frac{V_i - V_j}{V_i} \tag{6}$$

Using this definition, Π_{12} is the percentage loss (relative to optimally ordered search) from random search of the good leads and Π_{13} is the combined percentage loss from neither coarsely sorting nor optimally ordering.

5.1 Increasing c and decreasing R

In that case, information allows the researcher to fine-tune her search sequence to economize on search cost (Phase II), though we have shown that this value may be negligible. More generally, however, information may lead to another type of efficiency gain. In particular, after updating her priors on the basis of the acquired information, the researcher may decide that certain leads contained in the original list are no longer worthy of search. This determination depends on the interplay between the probabilities $\{p_i\}$, the search cost c, and the revenue on success R. As discussed above, leads for which the expected payoff is less than the search cost (i.e. those for which $p_i R < c$) are not worthy of search and should be eliminated from the collection.

We have argued that there can be significant value derived from eliminating inferior leads from the collection. In the Rausser and Small bioprospecting model, c and R were chosen specifically so the marginal value of the worst lead would be zero (see p. 192). In that example if either c were higher (than \$485) or R were lower (than $4.5E8$), an efficient search would entail eliminating some leads from the queue. This coarse sorting of the collection introduces a tension vis-a-vis earlier results in this paper. Proposition 2 shows that as c increases (or R decreases) we would expect Π_{12} to increase. But Proposition 2(e) shows that dropping the worst lead decreases the value of efficient search. If the worst lead happens to be marginally worth searching in the first place (as is the case when it is dropped as a result of increasing c or decreasing R), the percentage value of efficient search must also decrease. Indeed, simple experimentation on these numbers reveals that as the size of

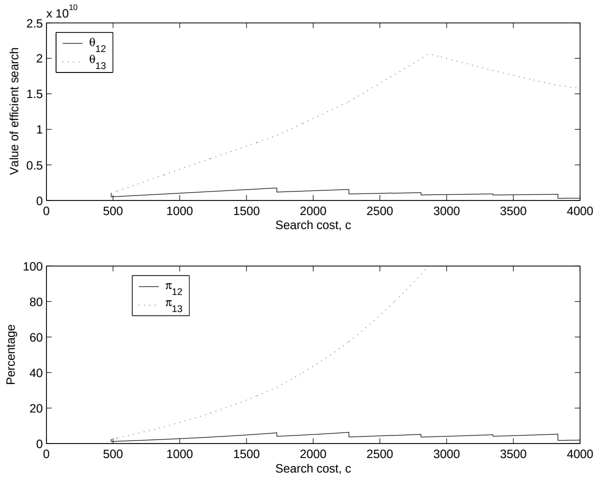

Figure 1: Value of efficient search (top) and percentage value relative to V_1 (bottom) as search cost, c increases. Dashed lines depict the difference between Researcher 1 (full information) and Researcher 3 (no information). Solid lines depict the difference between Researcher 1 (full information) and Researcher 2 (coarse sorting).

the collection diminishes (as coarse sorting eliminates inferior leads from the collection), so does Π_{12}. Which effect dominates is an empirical question which we answer in the following experiment.

How well will the researchers fare as c rises (or R falls)? The relative performance of the three researchers can be conveniently summarized by plotting our information measures Θ and Π against rising c (see Figure 1) or falling R (see Figure 2).

In the top panel of Figure 1 the dotted line measures Θ_{13} as c varies on the horizontal axis. The solid line illustrates Θ_{12}. Similarly the bottom panel of Figure 1 shows Π_{13} (dotted) and Π_{12} (solid) as c varies on the horizontal axis. The kinks mark the values of c at which leads are eliminated from the optimal queue. The salient feature of both plots is that, by using only coarse sorting, Researcher 2 does almost as well as Researcher 1 (the solid line is relatively low). Inspection of Figure 1 reveals that for the bioprospecting data, the percentage loss from inefficient search *never exceeds 6.5%* regardless of the search cost. In contrast, Researcher 3 with no information does very poorly indeed, losing 100% of the value for $c >$\$2,867.

The two panels of Figure 2 perform the parallel experiment for decreasing R rather than increasing c. Qualitatively, the results are the same. Coarse sorting is almost as efficient as full information search. By contrast, wholly random search is ineffective. The maximum loss from inefficient search again never exceeds 6.5%.

These experiments reveal that information may have quite large value but for a reason that has not been fully explored. In the example developed here, almost all the value is accrued by knowing what to search, not when to search it. For a given collection of leads worthy of search, Phase II R&D adds relatively little additional value.

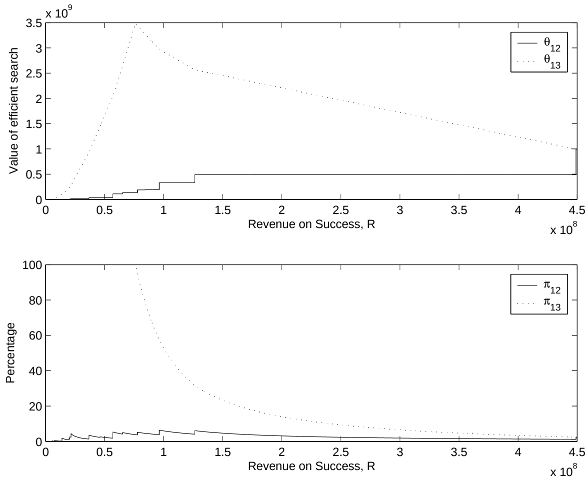

Figure 2: Value of efficient search (top) and percentage value relative to V_1 (bottom) as revenue upon success, R, decreases (right to left). Dashed lines depict the difference between Researcher 1 (full information) and Researcher 3 (no information). Solid lines depict the difference between Researcher 1 (full information) and Researcher 2 (coarse sorting).

6 Discussion

One purpose of research is to resolve uncertainty and thus inform better decisions. Whether searching for WMD, the cure for cancer, or the meaning of life, efficient search of a collection of leads must have higher value than random investigation of the same collection. Previous work suggested that dramatic dividends typically accrue from the use of an organizing scientific framework to better sequence search (Rausser and Small 2000). However, because such organizing information is often costly to acquire, the pivotal question is: how large is the value of information that facilitates efficient search?

We derived an expression for the value of efficient search, and calculated its theoretical upper bound as a function of the key parameters of any R&D problem. In the bioprospecting problem, we found this upper bound to be exceedingly small; about 4% of the total value of the collection. We argued that features of the bioprospecting problem were particularly well positioned to generate a high value of efficient search. For search problems with less extreme characteristics (e.g. smaller N or less heterogeneous probabilities) efficient search would be even less valuable.

Does that imply that basic scientific research and development has only negligible value? No. The result that even under favorable conditions the value of efficient search is often small applies only to a collection of leads that has already been identified as worthy of search. But leads for which $pR < c$ should never be searched. Information that allows the researcher to eliminate very low quality leads from the collection may have tremendous value, even when the good leads are subsequently searched at random. In many cases, information enabling coarse sorting can provide incentive to search a collection that was previously valueless. Again, in that case, the additional value of optimally ordering the good leads may be negligible.

The principles of this kind of search problem apply much more broadly to any research inquiry. In all search applications of this type, the researcher faces the conceptually separate questions of which leads to search and in what order to search them. Based on our analysis, the former question is much more crucial than the latter.

7 Appendix

Proof of Proposition 1.

(a) Let P be overall probability of success. Let δ be $1 - c/R$. By Proposition 2d the maximal spread distribution will give an upper bound for Θ_{12}. Given N leads, the maximal spread distribution satisfying P has $N-1$ leads at $1-\delta$. The high-value lead has probability $\bar{p}$ satisfying

$$1 - \delta^{N-1}(1-\bar{p}) = P.$$
$$\bar{p} = 1 - (1-P)/\delta^{N-1}.$$

Given this distribution, the expected number of searches can be calculated by averaging over all N distinct positions for $\bar{p}$. If we have T initial low leads, the expected number of searches in that queue is (for example)

$$1 + \delta + \delta^2 + \dots \delta^T + \delta^T(1-\bar{p})(1 + \delta + \delta^2 + \dots \delta^{N-T-2}).$$
$$= (1-\delta^{T+1})/(1-\delta) + \delta^T(1-\bar{p})(1-\delta^{N-T-1})/(1-\delta)$$
$$= (1-\delta^{T+1})/(1-\delta) + (1-\bar{p})(\delta^T - \delta^{N-1})/(1-\delta)$$

Averaging over all positions and reducing the sums yields the expected number of searches in random order (Researcher 2):

$$ES_2 = (1-\delta)^{-1}(1 - (1-\bar{p})\delta^{N-1} + \frac{1-\delta^N}{N(1-\delta)}(1-\bar{p}-\delta)).$$

Now the optimal search queue orders $\bar{p}$ first. In that case, the expected number of searches in optimal order (Researcher 1) is

$$ES_1 = 1 + (1-\bar{p})(1-\delta^{N-1})/(1-\delta)$$

So the maximal Θ_{12} is

$$\bar{\Theta}_{12} = c(ES_2 - ES_1)$$

Collecting terms in the expected search expressions and simplifying yields

$$\bar{\Theta}_{12} = c(\frac{\bar{p}}{1-\delta} - 1)(1 - \frac{1-\delta^N}{N(1-\delta)}).$$

(b) By Proposition 3d, the distribution which maximizes spread also maximizes Π_{12}. This distibution has $N-1$ leads of probability c/R and the remaining lead with probability 1. The bound on percentage loss from random search, $\bar{\Pi}_{12}$, is simply $100\frac{\bar{\Theta}_{12}}{R-c}$. The numerator of this expression is derived in part (a) above. The denominator is V_1, the value of the collection under fully efficient search. Dividing the numerator and denominator by R and collecting terms yields $\bar{\Pi}_{12} = 100\left[1 - \frac{1-\delta^N}{N(1-\delta)}\right]$, which is the desired expression.

■

Lemma 1 *Increased spread decreases the value of random search and increases the value of ordered search.*

Proof of Lemma 1. The expected value of a random search can be evaluated by taking the average value of all search order permutations. Consider any specific permutation $\{..., i, ..., j, ...\}$ and pair it with the corresponding permutation which reverses the position of i and j. For example, let these permutations be $\{1, i, 2, j, 3, 4\}$ and $\{1, j, 2, i, 3, 4\}$. Let $\{p_i\}$ be the set of lead probabilities, and let $P_k = 1 - p_k$. The expected number of searches under the first permutation is $1+P_1+P_1P_i+P_1P_iP_2+P_1P_iP_2P_j+P_1P_iP_2P_jP_3$. The average search duration over both permutations is thus $1+P_1+(P_1+P_1P_2)(P_i+P_j)/2+P_1P_iP_2P_j+P_1P_iP_2P_jP_3$. Now consider an increase in spread that increases p_i to q_i and decreases p_j to q_j such that $P_iP_j = Q_iQ_j$, where $Q_k = 1 - q_k$. For the new set of probabilities, the only terms that change in the average search expression are those with the element $(P_i + P_j)$. The leading terms do not depend on i or j. The trailing terms are unchanged because of the normalization assumption. So, the increase in search duration attributed to the spread is $(P_1 + P_1P_2)(Q_i + Q_j - P_i - P_j)/2$. This term is positive since $P_iP_j = Q_iQ_j$ and $Q_j > P_j > P_i > Q_i$. This same approach applies to *any* pair of permutations which switches the positions of lead i and j. Averaging over all such pairs gives the expected value of a random search. ■

Proof of Proposition 2. Consider an alternative researcher A who searches the collection in order $S^\wedge$. Then let $\Theta_{1A} = V(S^*) - V(S^\wedge)$.

(a) Θ_{1A} is independent of R because the probability of eventual success is independent of success order. Since Θ_{12} represents a random ordering of Θ_{1A} over all collections $S^\wedge$, Θ_{12} is also independent of R.

(b) Θ_{1A} is linear in c, so Θ_{12} is also linear in c.

(c) The optimally ordered sequence is S^*; the alternative sequence is $S^\wedge$. Letting $P_i^* \equiv (1 - p_i^*)$ and $P_i^\wedge \equiv (1 - p_i^\wedge)$, the value of ordered search is:

$$\Theta_{1A} = c\left[(1 + P_1^\wedge + P_1^\wedge P_2^\wedge + ...) - (1 + P_1^* + P_1^* P_2^* + ...)\right]$$

. To prove our result, consider increasing each p_i $(\forall i)$ by some amount, k. We now have

$$\begin{aligned}\frac{\Theta_{1A}}{c} &= \left((P_1^\wedge - k) + (P_1^\wedge - k)(P_2^\wedge - k) + (P_1^\wedge - k)(P_2^\wedge - k)(P_3^\wedge - k) + ...\right) \\ &\quad - ((P_1^* - k) + (P_1^* - k)(P_2^* - k) + (P_1^* - k)(P_2^* - k)(P_3^* - k) + ...)\end{aligned} \tag{7}$$

Taking the derivative, and evaluating at $k = 0$ gives:

$$\begin{aligned}\frac{1}{c}\frac{d\Theta_{1A}}{dk}|_{k=0} &= \left[P_1^*\left(\frac{1}{P_1^*}\right) + P_1^*P_2^*\left(\frac{1}{P_1^*} + \frac{1}{P_2^*}\right) + P_1^*P_2^*P_3^*\left(\frac{1}{P_1^*} + \frac{1}{P_2^*} + \frac{1}{P_3^*}\right) + ...\right] \\ &\quad - \left[P_1^\wedge\left(\frac{1}{P_1^\wedge}\right) + P_1^\wedge P_2^\wedge\left(\frac{1}{P_1^\wedge} + \frac{1}{P_2^\wedge}\right) + P_1^\wedge P_2^\wedge P_3^\wedge\left(\frac{1}{P_1^\wedge} + \frac{1}{P_2^\wedge} + \frac{1}{P_3^\wedge}\right) + ...\right]\end{aligned} \tag{8}$$

which is term-by-term non-positive, meaning that $\frac{d\Theta_{1A}}{dk} \leq 0$ with equality only if $S^{\wedge} = S^*$. Since this result holds for any sequence $S^{\wedge}$, it holds for the randomized sequence.

(d) The result is a direct consequence of Lemma 1.

(e) As before, let S^* denote the optimally ordered collection of leads and let $S^{\wedge}$ be an alternative ordering. We require notation for these ordered collections for the case in which the worst lead is omitted from the collection. We denote these by $\tilde{S}^*$ and $\tilde{S}^{\wedge}$, respectively.

Let $p_t^{\wedge} \equiv p_N^*$ be the value of the lowest lead.

Let $a_{i,j} = \prod_{k=i}^{j-1}(1 - p_k^*)$. For $i \geq j, a_{i,j} = 1$

Let $b_{i,j} = \prod_{k=i}^{j-1}(1 - p_k^{\wedge})$. For $i \geq j, b_{i,j} = 1$

Then,

$$\Delta\Theta = (V(S^*) - V(S^{\wedge})) - (V(\tilde{S}^*) - V(\tilde{S}^{\wedge})).$$

The expected revenue for the pairs $V(S^*), V(S^{\wedge})$ and $V(\tilde{S}^*), V(\tilde{S}^{\wedge})$ are the same, because the same set of leads is being searched (in different order). So, Θ is determined only by search costs.

Let EC be the expected search cost for each case. Then rearranging,

$$\Delta\Theta = (EC(S^{\wedge}) - EC(\tilde{S}^{\wedge})) - (EC(S^*)) - EC(\tilde{S}^*)).$$

Note that the marginal value of lead $p_t^{\wedge}$ in the alternative queue is:

$$(p_t^{\wedge} R - c)b_{1,t} - \frac{p_t^{\wedge}}{1 - p_t^{\wedge}} \sum_{i=t+1}^{N} (p_i^{\wedge} R - c)b_{1,i}.$$

So, $EC(S^{\wedge}) - EC(\tilde{S}^{\wedge}) = c\ (b_{1,t} - p_t^{\wedge} \sum_{i=t+1}^{N} b_{1,i})$. Similarly, $EC(S^*) - EC(\tilde{S}^*) = c\, a_{1,N}$.

Substituting in these expressions yields

$$\Delta\Theta = c\ (b_{1,t} - \frac{p_t^{\wedge}}{1 - p_t^{\wedge}} \sum_{i=t+1}^{N} b_{1,i} - a_{1,N}).$$

Since the worst lead goes last in the optimal queue, $a_{1,N} = b_{1,t}\ b_{t+1,N+1}$. Using this fact, then factoring out $b_{1,t}$ yields

$$\Delta\Theta = c\ b_{1,t}(1 - \frac{p_t^{\wedge}}{1 - p_t^{\wedge}} \sum_{i=t+1}^{N} b_{t,i} - b_{t+1,N+1}).$$

We now bound the summation term. Since $p_t^{\wedge}$ is the worst lead, $b_{t,i} \leq (1 - p_t^{\wedge})^{i-t}$. So,

$$\sum_{i=t+1}^{N} b_{t,i} \leq \sum_{i=t+1}^{N} (1 - p_t^{\wedge})^{i-t} = (1 - (1 - p_t^{\wedge})^{N-t})(1 - p_t^{\wedge})/p_t^{\wedge}.$$

Substituting in this bound,

$$\Delta\Theta \geq c\, b_{1,t}((1 - p_t^{\wedge})^{N-t} - b_{t+1,N+1}).$$

By the same argument, $b_{t+1,N+1} \leq (1 - p_t^{\wedge})^{N-t}$. Substituting in this bound,

$$\Delta\Theta \geq 0.$$

Since this is true for any alternative queue, it must also be true in expectation for a random search queue; dropping the worst lead decreases (weakly) Θ_{12}.

■

Proof of Proposition 3.

(a) Note that $\Pi_{12} = 100\Theta_{12}/V_1$. Π_{12} is decreasing in R because Θ_{12} is independent of R and V_1 is increasing in R.

(b) Π_{12} is increasing in c because Θ_{12} is increasing in c, and V_1 is decreasing in c.

(c) Additively increasing all leads by the same small amount decreases Θ_{12} and increases V_1, thus it decreases Π_{12}.

(d) V_1 increases with spread. So, π is maximized when spread is maximized.

■

Details of calculations for Table 1

- Column 2 of Table 1 gives the marginal values and total value under optimally-ordered search. These calculations follow precisely the calculations in Rausser and Small (2000). Briefly, the total value of the collection is found by backwards induction by using the Bellman equation:

$$J_n = p_n R + (1 - p_n) J_{n+1} - c \tag{9}$$

  where J_n is the expected value of continuing the search when$n - 1$ leads have been tested unsuccessfully. This is solved backwards starting with $J_{N+1} = 0$, and using the parameters $c = 485$, $R = 4.5e8$, and the p_n values given in Table 1 on page 193 in Rausser and Small (2000). The marginal values are given by:

$$v_n = a_n[p_n(R - J_{n+1}) - c]. \tag{10}$$

  Rausser and Small assume that 26.435 independent searches are conducted each year, and that the discount rate is 0.10, so each resulting marginal value is multiplied by the factor: 290.79=26.435(1.1)/.1. The result is the marginal value per lead, where a lead is a kilo-hectare. To obtain the marginal value of a hectare, Rausser and Small divide this number by 1000. This procedure gives the values in Column 2 of Table 1 of this paper.

- Column 3 gives the marginal values and total value under random search. We calculate these values by the following procedure:

- Randomize the search sequence.
- Calculate the marginal value of each lead according to the procedure above.
- Store the marginal value of each lead under that particular search sequence.
- Repeat this exercise over 10,000 randomized search sequences.
- Take the average of the marginal value for each lead.

- Column 4 gives the marginal values and total value under backwards search. We calculate these values by ordering the leads from lowest to highest probability of success. We then perform exactly the same procedure as in the first item above to calculate the marginal values under backwards search.

References

Charnes, A. and A. Stedry (1966). A chance-constrained model for real-time control in research and development management. *Management Science 12*(8), B353–B361.

Costello, C. and M. Ward (2004). Search, bioprospecting, and biodiversity conservation: Comment. *UCSB Working Paper*.

Fudenberg, D., R. Gilbert, J. Stiglitz, and J. Tirole (1983). Preemption, leapfrogging and competition in patent races. *European Economic Review 22*, 3–31.

Gallini, N. and Y. Kotowitz (1985). Optimal R and D processes and competition. *Economica 52*, 321–334.

Granot, D. and D. Zuckerman (1991). Optimal sequencing and resource allocation in research and development projects. *Management Science 37*(2), 140–156.

Lucas, R. (1971). Optimal management of a research and development project. *Management Science 17*(11), 679–697.

Myers, N. (1988). Threatened biotas: hot spots in tropical forests. *Environmentalist 8*(3), 187–208.

Myers, N. (1990). The biodiversity challenge: expanded hot-spots analysis. *Environmentalist 10*(4), 243–256.

Polasky, S. and A. Solow (1995). On the value of a collection of species. *Journal of Environmental Economics and Management 29*(3), 298–303.

Rausser, G. and A. Small (2000). Valuing research leads:bioprospecting and the conservation of genetic resources. *Journal of Political Economy 108*(1), 173–206.

Roberts, K. and M. Weitzman (1981). Funding criteria for research, development, and exploration projects. *Econometrica 49*(5), 1261–1288.

Ross, S. (1969). A problem in optimal search and stop. *Operations Research 17*, 984–992.

Simpson, D., R. Sedjo, and J. Reid (1996). Valuing biodiversity for use in pharmaceutical research. *Journal of Political Economy 104*(1), 163–185.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica 47*(3), 641–654.