Taming the antagonistic forces of exploration and exploitation in organizational search

Thorbjørn Knudsen

Strategic Organization Design Unit (SOD)

Dep. of Marketing and Management, University of Southern Denmark

Campusvej 55, DK-5230 Odense M, Denmark

tok@sam.sdu.dk; Tel +45 6550 1000

www.sdu.dk/sod

Nils Stieglitz

Strategic Organization Design Unit (SOD)

Dep. of Marketing and Management, University of Southern Denmark

Campusvej 55, DK-5230 Odense M, Denmark

nst@sam.sdu.dk; Tel +45 6550 1000

www.sdu.dk/sod

Draft of 11 September 2007

Abstract

The challenge of adjusting exploration and exploitation is usually portrayed as finding and maintaining a balance between two antagonistic forces. This challenge is rarely contrasted with the idea of sequencing exploration and exploitation as two distinctive stages of search. Building on prior work on the NK analysis in organizational studies, the present paper develops a modeling structure to address this important gap in our knowledge. We model exploration as a wider search among alternatives, and exploitation as the refinement of a chosen alternative. Our model allows consideration of both a set balance and a dynamic adjustment of exploration and exploitation within a single stage of search. The results reported here firmly establish that a two stage search is superior to a fixed balance of exploration and exploitation. This conclusion is quite robust. A dynamic adjustment version of our model puts fewer cognitive demands on decision makers, yet reaps significant gains by mimicking the two-stage search process. Overall, we suggest how the antagonistic forces of exploration and exploitation can be tamed.

Keywords: Exploration, exploitation, search, organizational adaptation, strategic organization design

© Knudsen & Stieglitz 2007

Introduction

Many decision problems related to strategy, organization design, and technology management are structured as a two-stage search process of exploration and exploitation (Greve, 2007).[1] This structuring has an intuitive appeal: Go find a promising technology and then proceed to refinement. Yet, remarkably little is known about the comparative effectiveness of sequencing exploration and exploitation in two distinct stages of organizational search.

Despite empirical evidence that points to the importance of two-stage sequencing of discovery and refinement of capabilities[2] and despite notions of a similar two-stage approach in relation to industry choice and to organizational change, most of our theories assume that exploration and exploitation must somehow strike a balance between the two (March, 1996; Katila & Ahuja, 2002; Rivkin & Siggelkow, 2007). Even though it is rarely made explicit, this idea assumes a single stage of organizational search. Balancing exploration and exploitation within a single stage of search is extremely challenging, however, because of the dysfunctional dynamics that tend to make them antagonistic forces (March, 1996).

Exploitation is associated with clear, immediate rewards that reinforce refinement search. It sacrifices the realization of long-term potential in the name of efficiency (March, 1991). As organizations tend to experience rewards for refinement, they are led into a dynamic that traps competence within a narrow domain – an observation supported in recent empirical research (Benner & Tushman, 2002).[3] Success in using a particular capability leads to more use, which improves the particular capability, which leads to more success, etc. (Levinthal & March, 1993; March & Sutton, 1997). Thus, agents become trapped by immediate positive feedback from competence within a rather narrow domain (Levinthal & March, 1993).

By contrast, exploration tends to lead agents into a failure trap that drives out exploitation. Failure provides negative feedback, which stimulates further exploration and change. As new ideas and capabilities fails, negative feedback

---

[1] Strategic management often portrays the path to competitive advantage as a two-stage search process. In stage one, firms are supposed to identify an attractive industry (from among many possible), and in stage two, they search for a superior competitive position within the industry (Porter, 1980; Rumelt, 1991; McGahan & Porter, 1997). Organizational change is frequently described in terms of a first stage of search among organizational configurations to be followed by a second stage of refinement of the configuration chosen in stage one (Miller & Friesen, 1980; Miller, 1996; Tushman & Romanelli, 1985).

[2] Two-stage sequencing of search is often used in biotechnology or medical products (Rosenkopf & Narr, 2001; Rotharmel, 2001; Rotharmel & Deeds, 2004). Furthermore, Burgelman (2002) portrayed Intel's strategy and its capability development in semiconductors as a temporal cycling between long periods of exploitation punctuated by short bursts of exploration. Gupta et al. (2006: 698) characterize this approach to balancing exploration and exploitation as "temporal sequencing".

[3] This dynamic is variously known as the competency trap (March, 1991; March & Sutton, 1997), the capability trap (Repenning &. Sterman, 2002) or the success trap (Levinthal & March, 1993).

reinforces the exploration for new ideas in an endless cycle of failure and unrewarding change (Levinthal & March, 1993). Thus, organizations become trapped by immediate negative feedback resulting from lack of competence (Levinthal & March, 1993). The difficulty in balancing exploration and exploitation lies in the antagonistic forces of immediate feedback from the two search modes. These antagonistic forces will induce a dynamics where exploration will tend to drive out exploitation, and *vice versa*.

We suggest here how the antagonistic forces of exploration and exploitation can be tamed. The proposed remedy is a two-stage sequencing in which a period of exploration is followed by exploitation. Despite the possible advantage of a two-stage sequencing of exploration and exploitation, it has not been systematically examined in the strategy and organization literatures. Also in neither our theories nor our formal models do we have systematic knowledge that allows comparison of a one-stage search process with a two-stage sequencing of exploration and exploitation.[4] The present paper develops a modeling structure to address this important gap in our knowledge. Our model also allows consideration of both a set balance and a dynamic adjustment of exploration and exploitation within a single stage of search.[5]

We build on and contribute to previous research in a number of ways. First, we model organizational search at two distinct levels. Agents search between task environments, and within a chosen task environment. We associate the former with exploration or discovery, the latter with exploitation or refinement. Second, we compare the efficacy of one-stage balancing with two-stage sequencing of exploration and exploitation. Third, we compare two-stage sequencing with a dynamic adjustment of exploration and exploitation. Fourth, we introduce sunk costs as a resource constraint on organizational search. Fifth, we analyze limited resource control and its impact on organizational search.

The paper proceeds as follows. In section two, we review prior research and elaborate on the role of cognitive constraints and resource control in search processes. Section three develops a modeling structure to address the open issues relating to a two-stage sequencing of exploration and exploitation. In section four, we report and discuss the simulation results. A fifth section summarizes our results and points to further avenues of research.

[4] A notable exception is Gavetti & Levinthal's (2000) model, which *can* be viewed as an example of a two-stage search within a particular task environment. In their model, agents first use a coarse-grained cognitive representation to identify a promising alternative and then proceed to a second stage of local search, which refines the features that were abstracted in the first stage of search. However, the emergence of cognitive representation is not an endogenous feature of their model.

[5] We interpret the dynamic adjustment model as the integration of exploration and exploitation, since the underlying equation defining the parameters and state variables does not change. However, dynamic adjustment may mimic two stages (or many stages) of search, depending on how the initial parameters and state variables for the equation are specified. The crucial aspect is that the balance between exploration and exploitation is dynamically adjusted based, among other things, on performance feedback.

Cognitive constraints and resource control

Our formal models portray agents with limited cognitive capacity as searching for new alternatives in a domain close to the status quo (Levinthal, 1997; Gavetti & Levinthal, 2000).[6] Agents typically engage in local search, favoring exploitation over exploration.

Local search often leads to impressive success in simple task environments where improvements can be achieved in piece-meal fashion (Stuart & Podolny, 1996; Fleming, 2001). More complex task environments, however, demand a broadening of search (Rivkin & Siggelkow, 2007). The ability to broaden search may be interpreted as a reflection of the cognitive capacity of agents (Gavetti and Levinthal, 2000; Fleming & Sørensen, 2004). Less myopic agents have the potential to engage in broader search and consider a wider range of alternatives. For example, scientific knowledge (Fleming & Sørensen, 2004), mental models (Siggelkow, 2001), or cognitive representations (Gavetti & Levinthal, 2000) may provide the necessary ability to benefit from sampling more distant alternatives.

Agents with better cognitive capacity can span a larger domain within a task environment in a way that leads to improvement over local search, a critical ability in complex domains. However, search may also extend beyond the traditional domain of the organization, into domains that are significantly different from the traditional ones. Nucor's search for a new capability to enter the flat-rolled steel segment represents such a case (Winter, 2000). Other well-known empirical examples that appear close to this characterization of organizational search include optoelectronics (Miyazaki, 1994) and pharmaceuticals (Rosenkopf & Narr, 2001; Rotharmel, 2001; Rotharmel & Deeds, 2004).

Theoretical research on organizational search spanning multiple domains has just begun. Gavetti, Levinthal & Rivkin (2005) invoke multiple performance landscapes to study the role of managerial analogies in new domains. Organizations gain experience which they apply to new domains. An important result is that a well-informed analogy is a particularly powerful conceptual guide in complex domains. Yet, prior research has not systematically examined how organizations apply analogies in their discovery and exploration of new domains. That is, our formal models have not considered choice among multiple domains as an endogenous feature of organizational search.

This problem is beginning to attract interest. Levinthal & Posen (2007) analyze organizational adaptation and the efficacy of selection processes. Two interrelated domains represent the task environments of the departments within a single firm. The overall performance of a firm depends on organizational choices in both task environments, with the interactions between task environments substantially complicating the efficacy of organizational search processes. An important finding is that the myopia of selection processes leads to the long-run

[6] When search capacity is constrained by the cognitive capabilities of agents (Simon, 1955) or by scarce resources (Stigler, 1961), the balancing of exploration and exploitation is far from trivial.

survival of firms with inferior capabilities. None of these contributions, however, have explicitly addressed the problem of searching among multiple domains (or task environments). This appears to be a major limitation in our theoretical understanding.

In contrast to previous models in the strategy and organization literature, we consider the possibility of organizational search on two levels.[7] Our model portrays exploration as search between domains or alternative task environments (e.g. different capabilities) and exploitation as search within a particular domain (refinement of a particular capability). Both the exploration and the exploitation phases of search are characterized by the cognitive constraints of the agent.[8] In the exploration phase, less myopic agents evaluate a wider range of task environments before they single one out for refinement. And in the exploitation phase, less myopic agents consider a wider range of improvements from the alternative that has been singled out. Thus, agents may search very different (or similar) task environments and they may consider very radical (or incremental) adjustments within a chosen task environment.

Our model systematically examines the implications of variation in cognitive capacity and constraints (see Table 1). As cognitive constraints limit the capabilities of agents, they tend to limit search among task environments, and they tend to engage in local search within task environments. At the (lower cognitive) limit, agents are confined to a randomly allocated task environment within which they engage in local search. Most of our formal models have examined this special case even though empirical and theoretical research has commonly pointed to organizational search among multiple task environments.

In addition to cognitive constraints, search is limited by the allocation of organizational resources. For example, the development of new capabilities often involves substantial sunk cost investments (Winter, 2000). In previous contributions, search was limited by the time constraint implicit in the length of simulations (e.g. Levinthal, 1997), or by variable search costs (Kauffmann et al., 2000). To better capture the dynamics of exploration among capabilities, we consider the constraints imposed by entry or switching costs. Agents have to pay a non-recoverable entry fee to search a task environment. This represents a sunk cost (Baumol, Panzar, Willig, 1982). The entry fee may create lock-in effects and, since agents also have to bear the cost when they exit a task environment, a lock-out effect (Ghemawat, 1991).

[7] Our conceptualization of exploration and exploitation is grounded in decision theory and the adaptive systems approach, which often consider the N-armed bandit problem as a representation of the exploration and exploitation trade-off (Robins, 1959). A gambler with a finite budget faces the problem of maximizing her total payoffs. When pulled, each lever of the bandit provides a reward drawn from a distribution associated to that specific lever. Initially, the gambler has no knowledge about the levers, but through repeated trials (exploration), she can identify and focus on the most rewarding levers (exploitation).

[8] The notions of broad and local search are commonly used to characterize the scope of search within a particular task environment (e.g. Levinthal & Warglien, 1999; Kauffmann et al., 2000). We consider local and broad search within a single task environment (refinement) and as characterizations of search among landscapes (discovery).

Virtually all formal models of organizational adaptation and search treat the task environment as a stable set of choice attributes under the full control of the agents (Knudsen & Winter, 2007 is a notable exception). In essence, agents have power over all relevant input factors or resources for a chosen configuration. Variations in performance are due to the choices made by an agent and not the result of external events. This portrayal of organizational search implies that quality changes in the upstream or downstream resources would not distort and frustrate organizational search. Compared to many real world situations, this seems to be highly artificial. In the development of capabilities, Winter (2000: 986) observes that "assessment of the consequences of specific adjustments may be difficult or slow even ex post if there is substantial random variation in trial outcomes". Pinpointing the cause of performance variations may be a challenge, because they may result from events beyond the control of the organization. Furthermore, Knudsen & Winter (2007) emphasize limited control over policy attributes as a major obstacle to the expansion of a business in a heterogeneous physical space. More importantly, assuming full control over input factors could suppress critical aspects of organizational search.

Two contrasting aspects of less than full control might be relevant. First, having no control over some resources or choice attributes may hamper search as agents continually have to accept some degree of randomness in their performance outcomes. Search may be frustrated as agents fail to stabilize and refine a choice configuration. Second, less than full control over input factors could facilitate search.[9] Having more control makes the decision problem more complex, with more possible configurations to be searched. Relinquishing control simplifies search by helping the agents focus on a limited set of choice attributes, by accepting a mean outcome for the attributes beyond their control. In this way, agents may rapidly improve performance by focusing search on only a few attributes.

The outlined features relating to cognitive constraints and resource control suggests several basic scenarios, as presented in table 1.

[9] This idea is elaborated in Knudsen & Winter (2007).

Resource control Landscapes	Unlimited	Limited
One	(1a) Local search (1b) Broad search	(3a) Local search (3b) Broad search
Multiple	(2a) Local search (2b) Broad search	(4a) Local search (4b) Broad search

Table 1

Scenario (1) represents the traditional case of search within one task environment (e.g. Levinthal, 1997). It includes two possibilities relating to cognitive constraints on search: (1a) represents the well-known possibility of local, myopic search, and (1b) represents the case of more powerful cognition admitting broader search within a task environment. These possibilities are also present in the three remaining scenarios.

Since the properties of scenario (1) are well known, we do not analyze it further. Instead, we focus on the three remaining scenarios, which have not been studied in the literature. In Scenario (2), agents face the additional problem of identifying an attractive performance landscape. They can search both among and within task environments. In Scenario (3), we add the problem of limited resource control. Scenario (4) analyzes all three dimensions of the decision-problem.

For scenarios (2) to (4), we examine the efficacy of the three search strategies which take center stage in the present work: A) one-stage balancing of exploration and exploitation, B) two-stage sequencing of exploration and exploitation, C) dynamic adjustment of exploration and exploitation. The section below elaborates on the modeling structure that we have developed to address these issues.

The model

Kauffman's (1993) NK model has been widely used in the study of organizational search. We use a variant of this model and extend it to study organizational search in multiple task environments to represent distinct capabilities. Our model structure has three major building blocks. The first establishes the nature of the task environments; the second specifies the cognitive and resource constraints of organizational search. The third building block establishes the alternative search strategies for exploration and exploitation.

Multiple performance landscapes

We generate a set of performance landscapes that represent separate task environments. Task environments may differ in size, complexity, and degree of resource control.[10] In each task environment, a choice set consists of N binary attributes. These represent organizational routines as the building blocks of capabilities. The attributes may relate to sourcing, production, sales, support function, etc. Each attribute can take on two states, so there are 2^N different organizational configurations in each landscape. The performance landscape created by the model is a mapping of the set of attributes onto performance values. The performance values of each of the N attributes are determined by random draws from a uniform distribution over the unit interval.[11] The overall performance of a configuration in one landscape is the average of the values assigned to each of the N attributes.

The attributes of an organizational configuration may be more or less interdependent. Attributes are interdependent if the value of each of the N individual attributes depend on both the state of that attribute itself and the states of K other attributes. If K = 0, attributes are independent. As K increases, more and more attributes of a configuration become interdependent, with K = N – 1 being the case of interdependence among all attributes. The number of interdependencies given by K determines the surface of the performance landscape. With higher values of K, there are more local peaks, and performance differences among neighboring configurations differing only in a single attribute become relatively more pronounced.

[10] The specification of more than one task environment therefore allows for a more penetrating analysis of exploration and exploitation as landscapes may differ in the global optimum. Even if agents have reached the global optimum in one landscape, they may still be outperformed by agents in higher performing landscapes. This essential property of exploration and exploitation cannot be captured by studying just one performance landscape.

[11] Since our studies include rather large performance landscapes (N= 50), we report raw values from the NK runs rather than normalized results.

Resource control

Resource control captures the idea that performance is shaped by an agent's ability to consciously manipulate the task environment and to pinpoint the causes of performance variations. We model the absence of resource control as random fluctuation in the state of a policy attribute. While attributes that are controlled do not vary over time, uncontrolled attributes fluctuate at random between their possible states (0, 1). The controlled attributes in a configuration, denoted by N_A, may be changed by the agent. The remaining attributes N_I of a configuration are beyond the control of agents and fluctuate randomly. They represent events outside the agent's control, for example, the costs and quality input factors, the lack of control over distribution channels, etc. Therefore, lack of control is equated with variability in outcomes. The task environments may differ in the number of uncontrolled attributes N_I and in the rate of random variations.

Constraints on organizational search

We look at two factors that constrain organizational search: cognitive limits and a fixed financial budget. Limited cognition or bounded rationality constrains organizational search by restricting alternative generation to the immediate proximity or neighborhood of current behavior. Broad search, on the other hand, reflects the greater cognitive capabilities of agents.

Local and broad search

Since the properties of local search are well known, we use it as our baseline for search within task environments. With local search, agents revise policy attributes by flipping a randomly chosen single bit and examining the outcome. If the result is improvement, the proposed revision is implemented; if not, the proposed revision is rejected.[12] Notably, agents can only experiment with the controlled attributes N_A in a configuration.

In addition to pure local search, we analyze a combination of local and broad (distant) search that has not been examined in prior research. Initially, agents engage in local search by flipping a single, randomly chosen policy attribute. If an agent's performance does not increase by local search (since it has been trapped on a local peak), the agent broadens the scope of search by incrementally increasing the number of controlled attributes that are changed (cf. Cyert and March, 1963; Nelson & Winter, 1982).[13] A new configuration is only implemented if its performance is

[12] We limit our study to perfect evaluation and do not consider the elaborations introduced by Knudsen & Levinthal (2007) relating to imperfect evaluation.

[13] Broad search is initially very coarse-grained and then becomes more fine-grained. The scope of search is increased. In the first pass of broad search, the probability that an attribute is mutated is initially set at 0.1 (the threshold for useful mutations when N=10). It is then increased by increments of 0.1 until it reaches the limit of 1. In the next pass, the probability that an attribute is mutated is initially set at 0.05. It is then increased by increments of 0.05 until it reaches the limit of 1. In the third pass, the incremental mutation probability is set to 0.03, etc. (with 4 passes, this procedure has reached the threshold of

higher than the existing level. This combination of local and broader search appears to have some support in empirical research (Fleming and Sørensen, 2004).

Fixed budget

Furthermore, organizational search is limited by fixed time period and budget *B*. The financial budget constrains search among performance landscapes. Agents have to invest a non-recoverable fee every time they enter a task environment. The fee represents a sunk cost investment and makes entry more risky as agents can get trapped in a performance landscape. Obviously, the higher the budget and the longer the time period, the more explorative search agents are able to engage in before they proceed to exploit a particular performance landscape.

Organizational search among and within performance landscapes

In the setting described above, agents have to allocate scarce resources to the exploration of task environments and exploitation within a task environment. They need to identify the task environment with the highest expected performance and refine the configuration within that landscape.

We characterize agents as boundedly rational, yet capable of some level of forward looking deliberation. They are able to form expectations or beliefs about the relative performance of a number of task environments.[14] An agent forms beliefs on the basis of knowledge generated through prior search processes. To model this relationship between past search efforts and the formation of beliefs, we turn to research on reinforcement learning. The Softmax algorithm, attributed to Luce (1959), provides a straight-forward way to model the formation of beliefs of agents (Sutton & Barto, 1998; Vermorel & Mohri, 2004). Recent findings in neuroscience suggest that the choices of human agents in uncertain environments correspond to this algorithm (Daw et al., 2006). The Softmax algorithm makes the choice of a performance landscape at time step *t* dependent on the observed mean performance $\bar{x}_i$ of the task environments *d* and on the proclivity for exploration τ:

(1) $$p(t) = e^{\bar{x}_i/\tau} / \sum_{i=1}^{d} e^{\bar{x}_i/\tau}$$

0.02 for useful mutations when N=50). Numerous simulations show that our model of broad search, apart from being realistic, outperforms the notion of distant search where all attributes in a bit-string are mutated.

[14] The formation of expectations or beliefs about performance appears to be an essential component in the analysis of exploration and exploitation. Agents form more estimates during exploration that are useful in guiding further exploitation. This link between exploration and exploitation is not considered in the common use of the local and broad search dichotomy.

The parameter τ, called the temperature, influences the degree to which an agent adheres to prior beliefs. A lower value of τ increases the probability that an agent remains within the performance landscape chosen before, as the most attractive one. Hence, it places a higher emphasis on exploitation. A higher τ downplays the role of past search processes and increases the probability that an agent samples and explores a different landscape. Note that agents continually update the estimates of mean performances. The more an agent explores landscapes, the better the estimates of mean performances.

In the Softmax algorithm, the temperature τ encapsulates the balance between exploration among landscapes and exploitation within landscapes. The critical question then is how agents set τ. We specify three alternative approaches to the determination of τ, each representing a different balancing of exploration and exploitation. The first approach aims to achieve a fixed balance of exploration and exploitation with a single stage of search. The second approach considers a more sophisticated two-stage sequence of exploration and exploitation, with the length of the exploration period depending on the number of task environments. The third approach takes its point of departure in models of adaptive organizational search (e.g. Levinthal & March, 1981; Nelson & Winter, 1982; March, 1988; Greve 2003) and allows agents to continually adapt the balance between exploration and exploitation.

One-stage fixed balancing of exploration and exploitation

With one-stage balancing of exploration and exploitation, decision makers set a parameter (temperature τ eq. 1) that determines a fixed balance between exploration and exploitation. The higher this parameter, the more the balance is pushed towards exploration – search among landscapes dominates search within landscapes. The lower the temperature, the more the balance is pushed towards exploitation – search within landscapes dominates search between landscapes.

Two-stage sequencing of exploration and exploitation

With two-stage sequencing of exploration and exploitation, the temperature τ (eq. 1) determines the mode of search. The decision maker takes a number of samples in exploration mode and then firmly shifts to exploitation mode. In exploration mode, the temperature τ is set so high that search among landscapes dominates search within landscapes. By contrast, in exploitation mode, the temperature is set so low that search within landscapes dominates search among landscapes.

Thus, the definition of exploration and exploitation admits a continuum of grades between two extremes with one extreme implying an incessant exploration among performance landscapes and with the other extreme implying that the landscape initially assigned at random is exploited forever. Our proposal of two-stage sequencing relates to a definition of exploration and exploitation that admits a gradation of these search strategies.

The effective sequencing of exploration and exploitation depends largely on the number of alternatives to be searched. The more alternatives an agent is confronted with, the more samples need to be taken before the estimates of the munificence of task environments become reliable. The length of the exploration period is determined by the optimal sample size. We assume that each agent samples one landscape in each time step. When an agent returns from an excursion to another landscape, local search is resumed from the configuration that was used before departure from this landscape. If agents have not visited a landscape before, the starting configuration is randomly determined. Following Even-Dar et al. (2002), the α-optimal sample size S with probability $1 - \delta$ is given by

(2) $$S = d/\alpha^2 \log (d/\delta)$$

with d being the number of performance landscapes. The parameter α signifies the tolerance of the agent, and δ the confidence level. If a more precise and reliable estimate is desired, the agent must spend more time on exploration. The sample size S determines the time step at which an agent switches from the exploration stage (high τ) to the exploitation stage (low τ). In essence, agents build up a broad stock of knowledge about the task environments (exploration with high τ) and then proceed to refine this knowledge (exploitation with low τ).

How exactly is the sample size S determined? Decision makers choose a tolerance level α and confidence level δ. The tolerance level captures the deviation from the best possible alternative. With a tolerance of α=0.20, decision makers will accept an alternative that has 20% less performance than the best possible. The confidence level δ captures the risk of accepting a bad alternative. With a confidence level of δ =0.95, decision makers will accept a bad alternative with probability 0.05. That is, the alternatives they accept will be worse than the limit set by the tolerance level. With a tolerance of α= 0.20, 5% of the accepted alternatives will have performance that deviates more than 20% from the best possible. The two sensitivity parameters (α and δ) and the number of performance landscapes to be searched (d) jointly determine the total number of samples S that the decision maker will allocate to exploration (see eq. 2). The lower the tolerance level α, the higher the confidence level δ, and the more landscapes d, the bigger the sample size.

How are samples in stage one (exploration mode) allocated among landscapes? Decision makers use eq. 2 to determine the total number of samples S, but the actual number of samples allocated to the examination of each particular landscape is endogenously determined. As decision makers gain experience with the performance landscapes, their perceptions of their quality are reflected in the Softmax algorithm. When a decision maker stumbles on good alternatives within a particular landscape, it is increasingly likely that other landscapes will be disregarded. Effort is directed to the most promising landscape. In other words, effort is directed towards land rich in oil, not to land that appears barren.

Dynamic adjustment of exploration and exploitation

Alternatively, firms may dynamically adjust the balance of exploration and exploitation within a one stage search process. In actual practice, there are a number of reasons why firms may prefer such a strategy. The optimal sample size may be unrealistic from a practical perspective or the number of alternatives may be unknown. In addition, agents may gain by continually updating the balance instead of sticking to a fixed sequence of exploration and exploitation.

In our model, we consider a simple mechanism that agents use to adjust the temperature τ in the Softmax algorithm. They decrease exploration if they have been successful in the past (cumulated performance W) and increase exploration if they have a large financial budget [15] B:

(3) $$\tau = \beta\ (B/W).$$

The parameter β is exogenously given and reflects how much reliance the agent puts on prior knowledge. Ceteris paribus, a low β tones down the effect of the temperature τ, indicating a higher reliance on prior existing knowledge. Agents tend to prefer a higher level of exploration (higher τ) the higher is their budget B, the smaller is the cumulated performance W, and the less they rely on prior knowledge. We assume that agents commit to a fixed budget B set aside for exploration. When that budget is depleted, exploration ceases.

Generally, exploration tends to decrease over time as the budget is depleted and cumulated performance increases, resulting in a decreasing temperature τ. This adaptive balance of exploration and exploitation roughly corresponds to empirical findings on new product development that highlight the feedback between research and development (e.g. Iansiti, 1998). In the development of technologically complex products, firms synchronize exploration (research) and exploitation (development) through an adaptive process of technology integration. Exploration does not end with the concept freeze (allowing firms to settle on a concept earlier) and might even be intensified if the expected product performance falls below a certain aspiration level.

By conditioning search on the relation between a fixed budget for search and resulting cumulative gains, the dynamics will tend to resemble a two-stage search process. The critical factor is the exogenously given parameter β. With a high β, in initial time periods exploration dominates and in later time periods, exploitation takes over. Thus, a high β will induce a dynamic that mimics a two-stage sequencing of exploration and exploitation. With lower values of β, the initial stage of exploration will endure longer, but be less pronounced. As β approaches zero at the limit, agents will become trapped in the task environment to which they were initially assigned.

[15] The financial budget relevant in organizational search typically concerns R&D activities.

Results

In our model, decision makers face the challenge of searching among and within performance landscapes. Within this context, we compare one stage balancing with two-stage sequencing of exploration and exploitation. Essentially, decision makers try to locate the best performance landscape and then proceed to locate the best alternative within this landscape.

We begin by comparing one-stage fixed balancing with two-stage sequencing of exploration and exploitation, and then comparing the two-stage sequencing procedure with the dynamic adjustment of exploration and exploitation. We have restricted our analysis and exposition to d= 4 task environments with the same N and different Ks.[16] We compare four rather small landscapes of N= 10 and K= 0, 1, 5, and 9, with larger landscapes of N= 50 and K= 0, 1, 25, 49.[17] Each landscape is randomly seeded. Our results reported here show the average of 15 simulation runs for each parameter with 1000 time steps. With N= 50, the decision maker faces the daunting task of considering some 4.5E15 alternatives distributed across four task environments of varying complexity. With N=10, the task is somewhat lighter with 4096 alternatives distributed across the four task environments.

One-stage and two-stage search compared

We examine two representative one-stage search strategies with a fixed balance between exploration and exploitation. The first is pushed towards the extreme pole of exploitation with a very low temperature set to $\tau = 0.005$. The second is set to a level of $\tau = 1$ where exploration tends to dominate exploitation. Additional robustness checks indicate that our results are invariant to different values of temperature in the interval between 0.5 and 1.

In the two-stage search, the temperature is again set to 1 during the exploration stage and then, after a total of S samples, the decision maker shifts to exploitation mode by reducing the temperature to $\tau = 0.002$. The values for α and δ were initially set to 0.2 and 0.05 respectively. According to eq. 2 this gives a total of 438 samples in exploration mode. Moreover, entry costs were set to 0.1, with a fixed budget of 50. To better understand the dynamics of the two-stage search, we analyzed its properties in more detail by varying entry costs, temperature τ, α and δ.

---

[16] For models with a fixed exploration-exploitation balance, our results can be extrapolated to a higher number of landscapes by adjusting for sample size as shown in eq. 2. For models with dynamic adjustment, the effect of a higher number of landscapes can be gauged by reference to eq. 3.

[17] Landscapes with N= 10 are commonly used in the organization and management literature. We add the large N= 50 landscapes in order to capture the common instance where the possibilities vastly dominate the time constraints on search.

Since the properties of local search within the context of one-stage search are well known, we do not analyze it further here. Instead, we focus on a number of scenarios relating to multiple performance landscapes and limited resource control (see table 1), which have not been studied in the literature. We analyze both local and broad search within the context of multiple performance landscapes. Both of these search modes are viewed here as particular instances of exploitation, or refinement search, which are used to further improve performance within a task environment. Compared to local search, a combination of both local and more distant search – representing fewer limits on cognitive abilities – dramatically increases performance. Note that the issue of exploration is altogether different from what is here labeled "local" and "broad" search. It concerns the shifts between performance landscapes that are guided by the Softmax algorithm.

Figure 1 compares the average performance of one-stage and two-stage search for small landscapes (N=10) and full control over attributes. Two-stage search clearly outperforms one-stage search strategies, whether they tend towards extreme exploitation or favor more exploration. According to Figure 1, one-stage search in the exploitation mode is characterized by huge initial performance increases. As firms use one-stage exploitation, they tend to become stuck in a random landscape and therefore single-mindedly concentrate on refinement of their randomly chosen configuration. As a result, performance rapidly improves. However, after this initial period of rapid improvement, gains from single-minded exploitation level off. In contrast, one-stage search in the exploration mode tends to have a longer period of performance gains. Performance continues to increase slightly over the entire 1000 time steps because of the unyielding exploration between landscapes. Overall, one-stage exploitation shows better performance than one-stage search, favoring more exploration.[18] Thus, the model reflects a bias towards exploitation as the initial performance gains are much higher (the well-known signature of competency traps).

[18] Over the ultra-long run, one-stage search in the exploration mode would often outperform one-stage search in the exploitation mode.

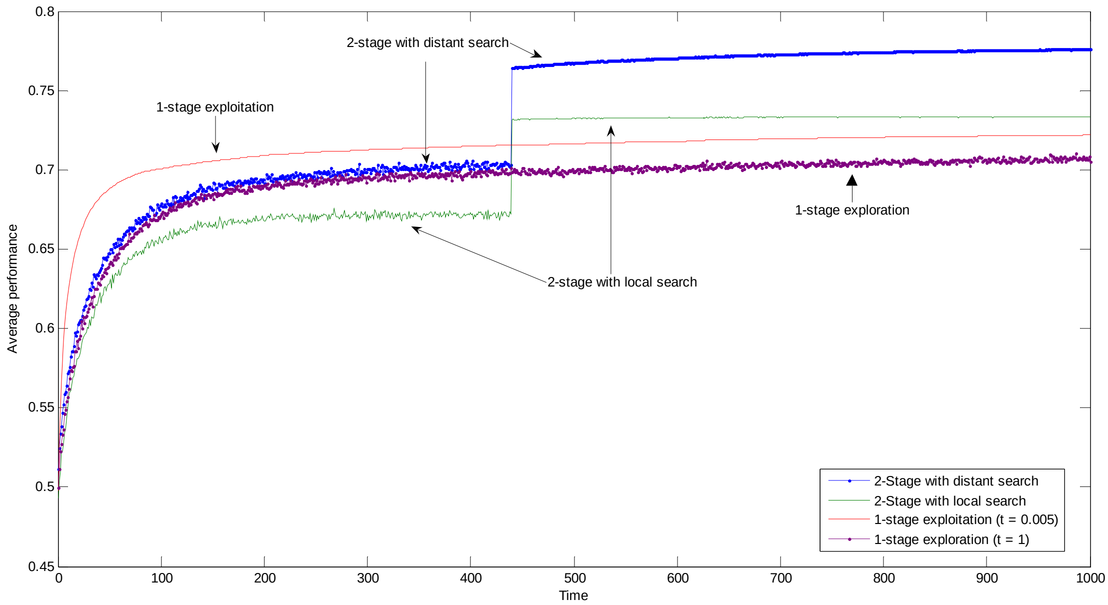

Figure 1: Two-stage search compared to one-stage search for N=10 and full control (entry cost = 0.1, α = 0.2, δ = 0.05, τ= 1)

The two-stage search strategy produces a much smaller initial improvement than one-stage search in the exploitation mode. However, after the exploration stage has been concluded (*S*= 438 time steps), the switch from exploration to exploitation is associated with a huge increase in performance. This dramatic gain is achieved because of the knowledge sampled in the exploration stage. In stage two, agents use their knowledge about performance landscapes to settle on the task environment perceived to be the most attractive. Within this performance landscape, they start from the configuration they have so far identified as the best.

When refinement in the two-stage strategy is achieved by local search, much of the gains are already realized in the exploration stage, while performance increases continue if a combination of local and distant search is used in the refinement stage. What is striking is that the two-stage model, even if it is limited to local search in the refinement stage, clearly outperforms the one-stage search strategies based on the powerful combination of local and distant search. More generally, Figure 1 demonstrates that two-stage sequencing of exploration and exploitation clearly dominates a one-stage fixed balancing of these search modes in terms of achieved performance (and this is also the case for cumulated performance).[19]

How does loss of control over attributes alter the relative performance characteristics of the search strategies? Figure 2 reports average performance when

[19] Although some agents outperform all others with one-stage search in the exploitation mode, it is the most risky strategy. In this case, performance critically depends on the initial random assignment to a performance landscape. This is supported by a comparison of the standard deviation of accumulated performance. The mean of the standard deviations for one-stage search in the exploitation mode is 51, while it is only 13 for the two-stage strategy with distant search.

agents lose control of 25 percent of the attributes. Overall, the performance of all search strategies decreases vis-à-vis the full control case. Moreover, the relative superiority of two-stage search increases further. With an even higher loss of control (0.75), the one-stage strategies suffer even more.[20] Balancing search between and within landscapes becomes even more difficult with less than full control. As uncontrolled attributes change state at random, they are in a sense unimportant. However, in order to realize this, a large number of samples is needed. Otherwise, firms are caught in a failure trap. They continue chasing phantoms as attractive stochastic variations appear and then vanish. As shown in Figure 2, a sequencing of exploration and exploitation is a powerful antidote to this failure trap.

With loss of control, the advantage of broader search within landscapes also declines dramatically. The randomly determined attributes prevent agents from getting stuck on a local peak too early in the search process, enhancing the efficacy of local search. In addition, many agents never realize the advantage of more distant search as their performance continues to fluctuate, even if the configuration does not change. Agents fail to broaden their search and continue to search locally, not realizing that performance differences are triggered by the uncontrolled part of the configurations.

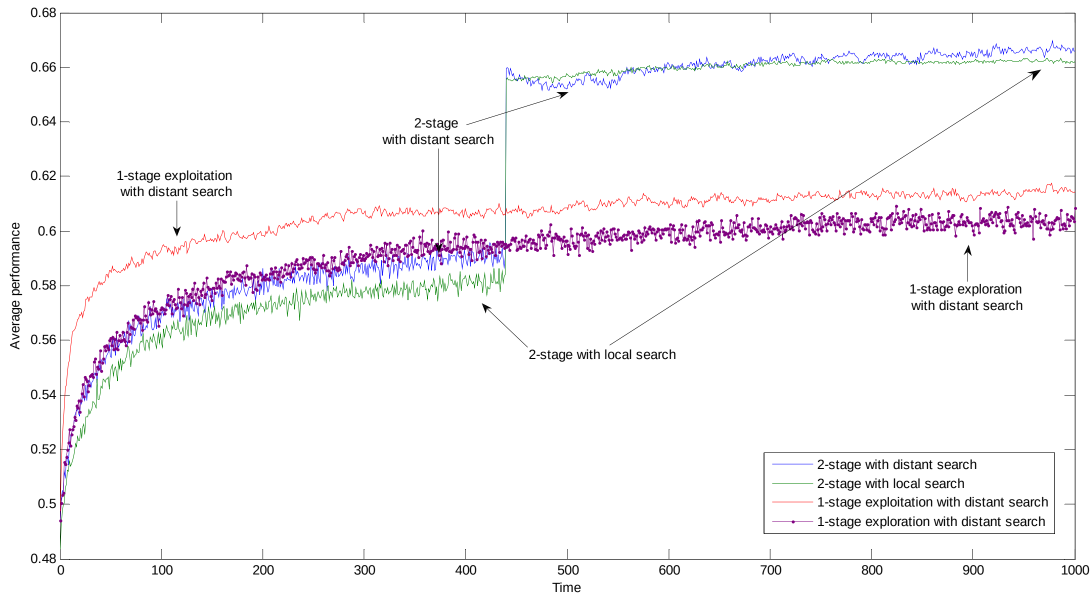

Figure 2: Two-stage search compared to one-stage search for N=10 and 25% loss of control (entry cost = 0.1, α = 0.2, δ = 0.05, τ = 1)

With larger landscapes (N=50), the relative advantages of two-stage search become even more pronounced (figure 3). It is perhaps surprising that two-stage search with a refinement mode comprising both local and distant search actually decreases average (and cumulated) performance compared to refinement mode

[20] With a 75% loss of control, the average performance of both one-stage strategies hovers around 0.53, while the two-stage strategies reach an average performance of about 0.59. These results are not shown here, but can be obtained upon request.

limited to local search *only*. The reason is that decision makers are now negotiating huge spaces of possibilities. In the refinement stage, most agents have located in the simple landscape with K=0. In that case, simple gradient search outperforms distant search which misleads agents.

With N=50, a 25% loss of control does not change the overall picture of the relative advantage of two-stage search. Yet, a huge loss of control (75%) gives one-stage search strategies a relative gain as they are applied in larger landscapes.[21] The reason for this relative gain is that a huge loss of control simplifies the task at hand because the many uncontrolled attributes "average out", in which case agents are less prone to chase stochastic fluctuations. Ironically, agents will tend to escape the failure trap because their behavior is shot through with random fluctuation; an interesting variant of Ashby's law of requisite variety.

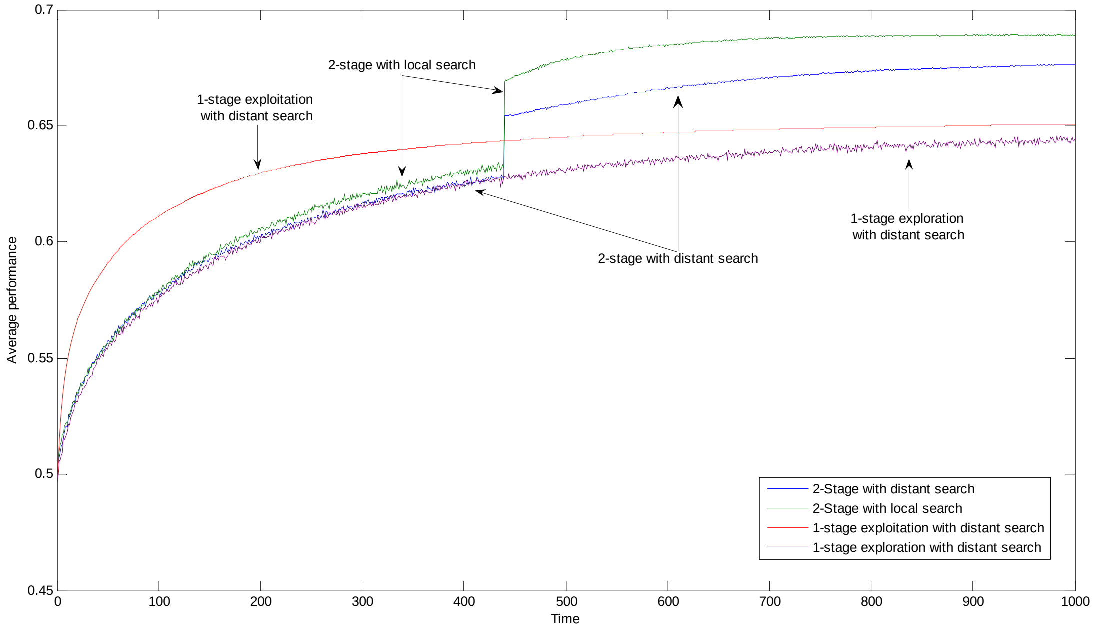

Figure 3: Two-stage search compared to one-stage search for N=50 and full control (entry cost = 0.1, α = 0.2, δ = 0.05, τ = 1)

Not surprisingly, overall performance of all search strategies declines with loss of resource control as the granularity of search becomes coarser. The lack of control simplifies the agents' task by substantially lowering the number of configurations that need to be searched. However, the random element within the configuration frustrates the search efforts of agents and lowers the overall effectiveness, especially of broader search. With a coarser granularity of search (loss of control), the relative superiority of two-stage search becomes even more pronounced.

In general, the findings provide powerful support for the conjecture that two-stage sequencing of exploration and exploitation dominates one stage-balancing of the two forces. One-stage search in the exploitation mode is a high-risk search

[21] These results are not shown here, but can be obtained upon request.

strategy that may benefit a few agents, with the majority remaining stuck in an inferior landscape. One-stage search in the exploration mode, on the other hand, foregoes the potential performance benefits of refining a configuration. A two-stage search strategy combines the benefits of these two approaches. Agents form expectations about the relative performance of the four landscapes based on an optimal sample size and then proceed to refine the highest performing configuration in the most attractive landscape.

Two-stage search: Lock-in and adjusting the duration of the exploration stage

The effective balancing of exploration and exploitation in a two-stage search process vitally depends on aligning the budget with the intended sample size. If the budget does not support a sufficiently long period of exploration, agents get locked-in prematurely. They get stuck in a landscape too early and may regret their pick after the completion of the exploration stage. Still, the performance impact of a lock-in is not as pronounced as one might expect. If agents are capable of broader search within landscapes, an extended period of exploitation compensates for the lack of extended exploration. Even with local search only, a longer exploitation stage provides amble opportunities for improving performance in simple landscapes. In more complex landscapes, however, performance decreases more sharply by a lock-in, since agents quickly reach a local peak. Overall, these findings confirm our conjecture that one-stage search in exploitation mode is not a substitute for a two-stage search process.

To prevent a lock-in, agents may reduce the sample size in the exploration stage by adjusting the confidence or the tolerance level. Reducing the confidence level to 80% (α= 0.2) decreases the length of the exploration stage markedly. However, agents are still able to differentiate between landscapes, and achieve large performance gains after completion of the first explorative stage (see figure 4).

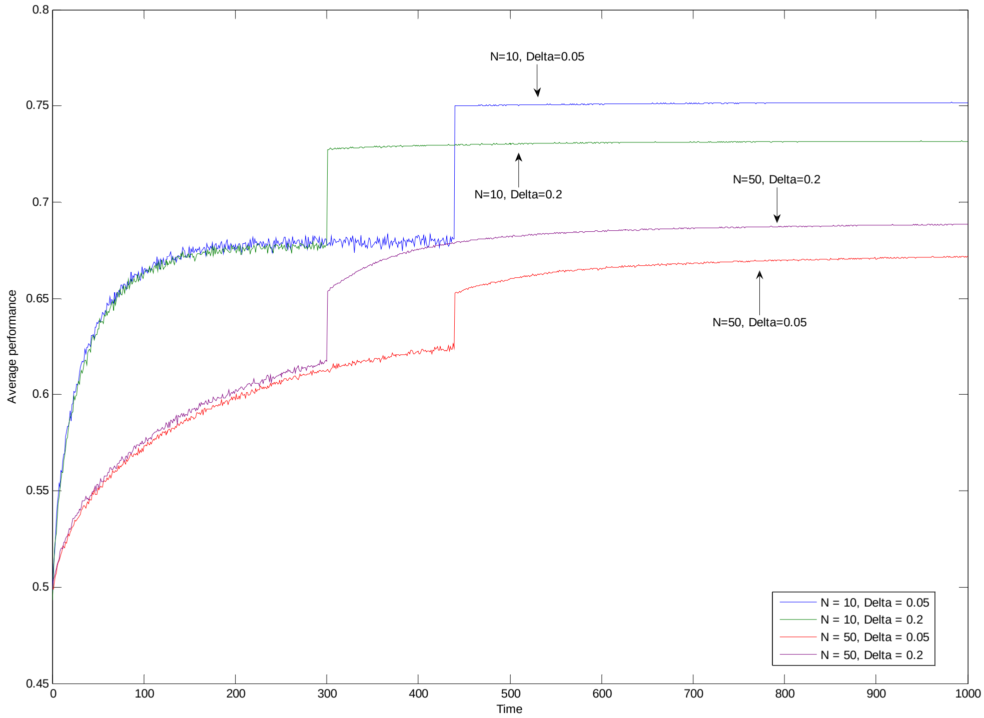

Figure 4: Two-stage search with different confidence levels and task environments (N=10/50, full control, entry costs = 0.1; α = 0.2, δ = 0.05, 0.2, τ = 1)

With a shorter time for exploration, agents compensate with more exploitation in the second stage of search. The overall effect depends on the performance differences between landscapes and the relative contribution of exploitation. In small task environments search tends to be exhaustive, a fact that favors a more extended exploration stage (lower confidence, δ, or higher tolerance, α) and a second stage with few opportunities for refinement. In contrast, large task environments exhibit more refinement opportunities and therefore leave more room for exploitation, a fact that may justify a shorter exploration stage (this effect becomes more pronounced with lack of control). Indeed, it is remarkable that a shorter period of search leads to better performance in the N=50 case. These landscapes are so huge that it is apparently better to briefly take aim and then go for the most promising alternative. Another way to adjust the duration of exploration stage is to increase the tolerance level α. Again, a shorter exploration stage particularly hurts agents with full control, leading to a drop in overall performance.[22] With lack of resource control, on the other hand, average performance remains roughly the same even with a shortened exploration stage, since agents compensate with more exploitation in the long run. The lack of control actually helps these agents, as it forces them from the local peaks and stimulates continued gains from search in the exploitation stage.

Overall, two-stage search seems to be fairly sensitive to variations in α and δ. Agents with full control benefit from a longer exploration stage as it helps them form more reliable estimates of the relative performance of landscapes. If they are only

[22] The results are available from the authors upon request.

capable of local search, a longer exploitation period does not seem to increase performance, since agents often get stuck on local peaks fairly early in the exploration stage. Therefore, a shortening of the exploration stage has no corresponding benefit in more effective exploitation. This is not true for situations characterized by lack of control over attributes. In these situations, agents may compensate for a shortened exploration stage by more effective exploitation as random variations in performance attributes prevent agents from getting stuck. The inherent experimentation from lack of control over attributes thus compensates for a shorter exploration stage.

Two-stage sequencing compared to dynamic adjustment

In the following, we compare the two-stage sequencing of search with the dynamic adjustment of the balance between exploration and exploitation. The two-stage sequencing effectively constrains agents to a second stage of exploitation. In contrast, dynamic adjustment gradually reduces the temperature based on performance feedbacks and thereby "keeps the door open" to some level of exploration even in later time periods.

Figure 5 compares the performance characteristics in small (N=10, Fig. 5a) and in large landscapes (N=50, Fig. 5b) and full resource control. It is evident that the two strategies have very different characteristics. In general, average performance rises faster with dynamic adjustment than with the two-stage sequencing of search, but the latter catches on with the dramatic gain in performance that follows the transition to the exploitation stage. The effectiveness of the two search strategies critically depends on the size of landscape and on the cognitive abilities of the agents. The size of the landscape shapes the potential gains of focused exploitation, while the cognitive abilities influence whether agents take advantage of it. The ability to search broadly offers clear advantages in small landscapes (Fig. 5a). However, distant search appears to be a mixed blessing in large landscapes Fig. 5b). The reasons for this are two-fold. First, agents are not firmly focused on exploiting immediate gains, but stray away from local search as soon as a sampled configuration entails a decrease in performance. Second, the probability of finding superior configurations with distant search decreases sharply in large landscapes. Hence, distant search clearly outperforms agents capable of local search only in small landscapes, but this finding is reversed in the large landscapes.

In small landscapes (Fig. 5a), a firm commitment to an extended period of exploitation, while beneficial with distant search, hurts agents that are capable of local search only. They get trapped on a local peak soon after switching to exploitation and thereby lack the ability to take advantage of an extended period of exploitation. In this case, a modicum of exploration, as provided by the dynamic adjustment model, actually helps agents to continue the search for performance improvements between landscapes. Then again, in large landscapes (Fig. 5a), in which finding a local peak takes much longer, focused exploitation combined with local search outperforms all other search strategies. "Keeping the door open" to exploration here distracts agents from realizing incremental gains in a huge space of possibilities.

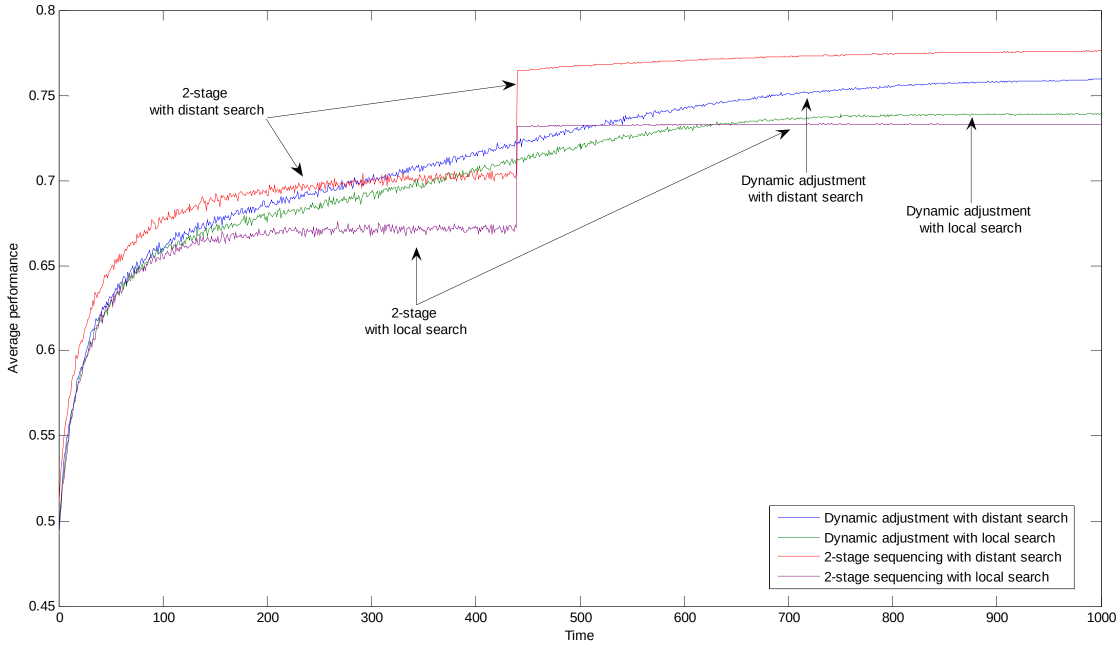

Figure 5a: Two-stage sequencing compared to dynamic adjustment in simple landscapes (N=10, full control, entry costs = 0.1; α = 0.2, δ = 0.05, Beta = 1, temperature = 1)

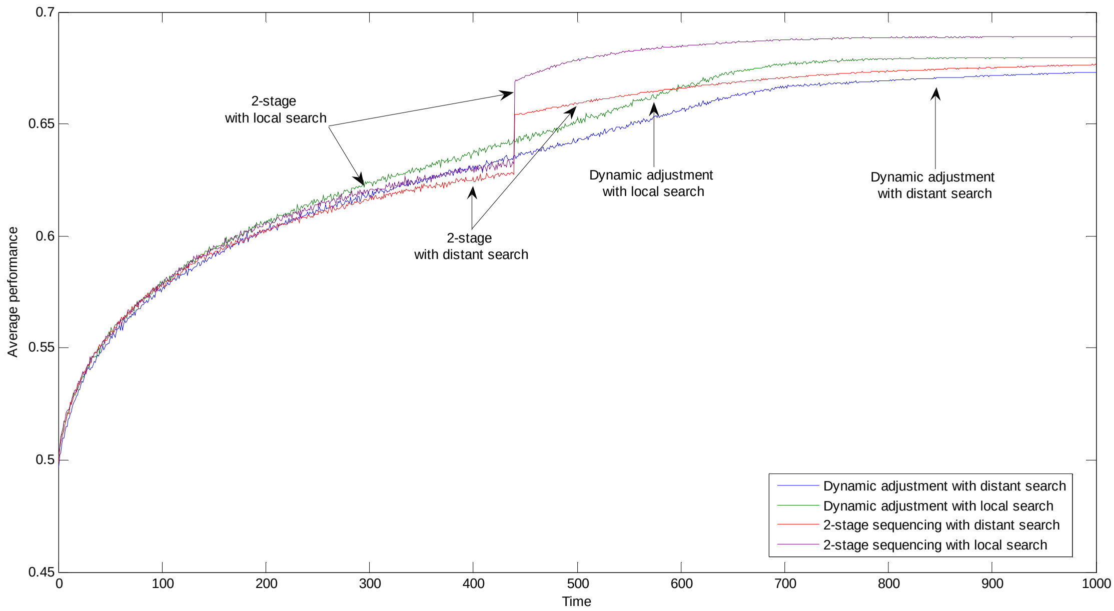

Figure 5b: Two-stage sequencing compared to dynamic adjustment in large landscapes (N=50, full control, entry costs = 0.1; α = 0.2, δ = 0.05, Beta = 1, τ = 1)

The success of dynamic adjustment critically depends on a firm commitment to a fixed budget that is not replenished during search.[23] Otherwise, agents may engage in too much exploration. To study the effects of a flexible budget, we lifted the budget constraint present in versions of the dynamic adjustment model reported up until this point. We did this by setting entry costs to zero. The balance between exploration and exploitation is still adjusted by performance feedbacks, but the adjustment is more gradual as exploration is not lowered by a declining budget. The impact on the effectiveness of search is quite pronounced (figure 6). When the budget constraint is lifted from the dynamic adjustment model, there is a dramatic decline in performance as agents are never making notable commitments to a particular task environment.

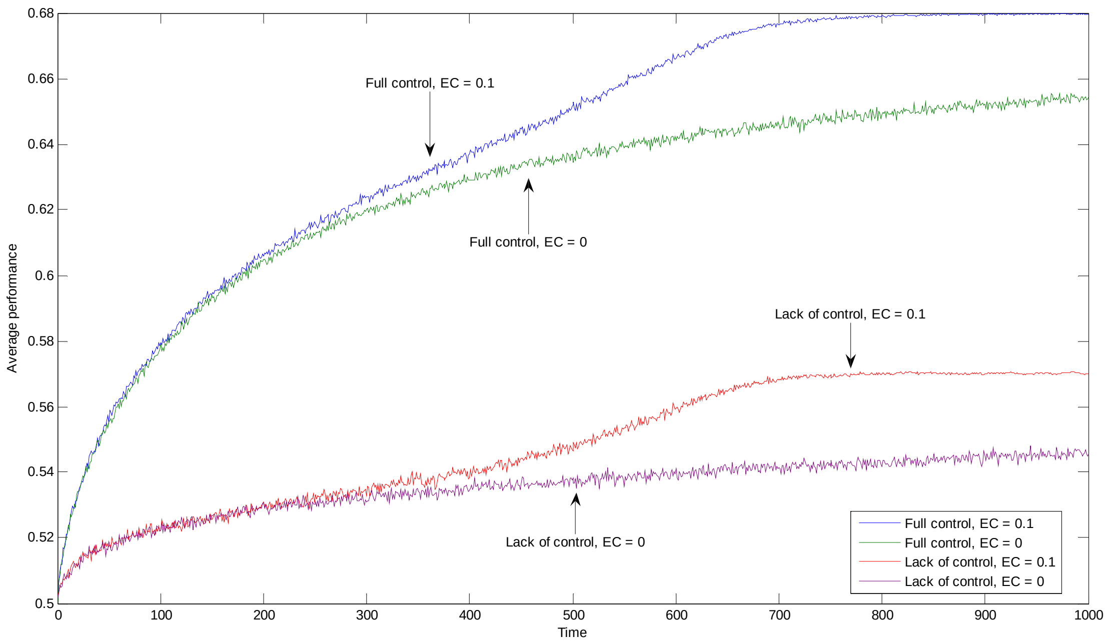

Figure 6: Dynamic adjustment and budget decline (N=50, full control/lack of control 0.75, entry costs = 0, 0.1, Beta = 1)

[23] The results for the dynamic adjustment model were quite robust to variations of parameter β, which determines the prior tendency for exploration. With a high value β, firms begin the dynamic adjustment period with much exploration across landscapes. As the value of β is lowered this tendency reduces. The results for the dynamic adjustment model reported here assume a value of β set to 1. We tested the robustness of these results by setting β to 0.1, 0.5 and 5. With 0.1, agents do not spend enough time searching between landscapes, leading to a decline in overall performance. No performance differences were observed when β was varied between 0.5 and 1. Even with a very high β (= 5) performance did not decline in the full control case for N = 10 and N = 50, and only declined slightly with less control over attributes. In this case, dynamic adjustment closely mimics the two-stage sequencing with a stable exploration stage pattern followed by exploitation.

Overall, the performance characteristics of the dynamic adjustment model were quite remarkable when compared to those for the two-stage search sequence. Effective two-stage sequencing puts a much higher demand on the decision-maker. For effective sampling, she must set two parameters (α and δ) to reasonable levels and find the right temperature τ for the two stages. Even though the results are fairly robust to variations in the temperature, changes in α and δ have significant performance implications. In addition, the decision maker must align exploration with the budget constraint to prevent the adverse effects of unintended lock-in during the exploration stage. The requirements for effective search with the dynamic adjustment model are less strict. The decision-maker needs only to set the initial temperature and specify a hard budget constraint. In addition, since dynamic adjustment lowers exploration with a declining budget, decision-makers are able to avoid lock-in to an inferior landscape.

A final issue pertains to resource control in the dynamic adjustment model. Here it affects performance in subtle ways. With loss of control, dynamic adjustment with local search consistently outperforms all other search strategies in simple landscapes. It gradually tones down exploration (from a relatively modest level) without eliminating it even after many time periods. Therefore, dynamic adjustment is superior to other search strategies in differentiating among small landscapes and selecting the most attractive one. However, in large landscapes with lack of resource control, the performance characteristics between strategies converge.

Conclusions

Many decisions relating to strategy, organization design, and technology management are structured as problems involving the discovery of a broad set of alternatives with potential gains from refinement. Agents have to strike a balance between the exploration among alternatives and the exploitation of one alternative. There are well known pathologies associated with attempts to balance exploration and exploitation within a one-stage search process with a fixed balance, i.e. the competency trap and the failure trap. One advantage of two-stage sequencing of search – an initial stage of exploration followed by a subsequent stage of exploitation – is that it can eliminate these pathologies. Another advantage is that a two-stage search strategy captures the stylized features of many problems in organization design, strategy, and technology management. Empirical evidence broadly supports the idea that agents first choose from a broader set of competing alternatives (e.g. technologies) with uncertain potential and then proceed to refine a particularly promising alternative. We find broad empirical support in the literature on technology research and development projects.

Despite the possible advantages of a two-stage sequencing of exploration and exploitation, this possibility has not been systematically examined in the strategy and organization literatures. Building on prior work on the NK analysis in organizational studies, we have developed a modeling structure that allows detailed study of these issues. Our approach extends prior research on organizational search and adaptation in a number of ways. First, we modeled organizational search as a process of exploration among task environments to be followed by exploitation within a chosen task environment. Second, we compared the efficacy of one-stage balancing with a two-stage sequencing of exploration and exploitation. Third, we compared two-stage sequencing with dynamic adjustment of exploration and exploitation. Fourth, we introduced sunk costs as a resource constraint on organizational search. Fifth, we analyzed limited resource control and its impact on organizational search.

The results reported here firmly establish that two stages of search with a set balance of exploration and exploitation are superior to one stage, at least when decision makers face the problem of choosing from multiple alternatives that hold potential gains from further development. This conclusion is quite robust to variations in the size of performance landscapes, loss of control over policy attributes, duration of the two stages of search and duration of exploration in the first stage of search. The essential and critical feature of a two stage search strategy is commitment. The two stage search model assumes commitment of a fixed period of time to exploration and then a firm shift to a second stage of exploitation. If this constraint is lifted, the gains from two-stage search evaporate.

We also examined a dynamic adjustment version of our model inspired by prior models of adapting aspirations. Our dynamic adjustment model is attractive because it appears to capture the stylized features of R&D reported in much of the recent empirical research on new product development. This model puts fewer cognitive demands on decision makers, yet reaps significant gains by mimicking the two-stage search process. Also in this model, commitment is a critical feature. Essentially, the decision-maker is required to specify a hard budget constraint. In addition, the decision maker must stimulate a high initial level of exploration (if in doubt, exaggerate the initial level of exploration).

Our approach leaves many unexplored avenues of inquiry relating to cognitive constraints, search among multiple performance landscapes, and resource control. One natural extension of the model is consideration of a wider range of search strategies. For example, agents could continuously cycle through stages of exploration and exploitation, or even start with a stage of exploitation followed by exploration. Or agents might use different approaches to search, e.g. a choice of landscapes based on the geometric mean (Kelly criterion) rather than the arithmetic mean. The dynamic adjustment model could be extended by allowing for a replenishing budget based on past performance (making the commitment to exploration less firm). In addition, the amount of exploration might depend on the average accumulated performance of other agents (cf. Knudsen, 2007). Underperforming agents in an unattractive task environment could then be forced to reignite exploration after an extended period of exploitation.

A limitation of the current model is that agents may locate only in a single task environment. However, when it comes to industries, firms may spread their investments across more than one industry through corporate diversification. The Softmax algorithm could be used to model the resource allocation process for the business units of the firm, while the relatedness between industries might be represented by making the performance of one industry dependent on presence in related industries (positive or negative association). Thus, the model can be extended to study the evolution of corporate coherence.

Another avenue of research might consider the dynamic emergence of new landscapes or introduce relative performance shifts. This could, in a stylized way, capture the essence of technological development and industry dynamics, since landscapes would evolve through a period of growth and stagnation. An interesting extension would be to introduce the competitive dynamics between agents (cf. Lenox, Rockart & Lewin, 2006).

Lastly, the aspect of limited resource control needs more attention in formal modeling (Knudsen & Winter, 2007). Limited resource control introduces noise and the granularity of search in a convincing manner (cf. Levinthal, 1997; Knudsen & Levinthal, 2007) since agents may not exactly pinpoint the cause of a performance increase or decline. This appears to be a reasonable representation of many actual search problems, since the outcome of an action may also depend on (possibly unobserved) outside events, with no clear-cut way to disentangle the compound impact. This makes feedback-based learning more difficult. A possibly interesting extension could be to make resource control an endogenous feature of the model. By investing in resource control, agents could determine the optimal amount of resource control that would balance two of its conflicting properties: simplification of the search on the one hand, and random fluctuations that frustrate search, especially in small and in complex landscapes, on the other.

We hope this contribution will benefit both research and practice by directing attention to these issues, providing a modeling structure with which they can be examined in a systematic way, and providing a first set of robust results that point to the advantages of two-stage sequencing and the dynamic adjustment of exploration and exploitation.

References

Baumol, W.J., J.C. Panzar, and R.D. Willig 1982, *Contestable markets and the theory of industry structure*, New York: Harcourt Brace Jovanovich.

Benner, M.J. and M.L. Tushman. (2003), Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited, *Academy of Management Review* 28, 238-256 .

Cyert, R.M. and J.G. March (1963), *A behavioral theory of the firm*, Englewood Cliffs: Prentice-Hall.

Daw, N.D., J.P. Doherty, P. Dayan, B. Seymour, R.J. Dolan (2006), Cortical substrates for exploratory decisions in humans, *Nature* 441, 876-879.

Dosi, G. (1982), Technological paradigms and technological trajectories, *Research Policy* 11, 147-162.

Even-Dar, E., S. Mannor and Y. Mansour (2002), PAC Bounds for Multi-Armed Bandit and Markov Decision Processes, *Fifteenth Annual Conference on Computational Learning Theory* (COLT), 255-270.

Fleming, L. (2001), Recombinant uncertainty in technological search, *Management Science* 47, 117-132.

Fleming, L. and O. Sørensen (2004), Science as a map in technological search, *Strategic Management Journal* 25, 909-928.

Gavetti, G. and D. Levinthal (2000), Looking forward and looking backward: Cognitive and experiential search, *Administrative Science Quarterly* 45, 113-137.

Gavetti, G., D. A. Levinthal and J.W. Rivkin, Strategy making in novel and complex worlds: the power of analogy, *Strategic Management Journal* 26, 691-712.

Ghemawat, P. (1991), *Commitment*, New York: Free Press.

Greve, H.R. (2003), *Organizational Learning from Performance Feedback*: A Behavioral Perspective on Innovation and Change, Cambridge, UK: Cambridge University Press.

Greve, H.R. (2007), Exploration and exploitation in product innovation, *Industrial and Corporate Change* (forthcoming).

Gupta, A.K., K.G. Smith, and C.E. Shalley (2006), The interplay between exploration and exploitation, *Academy of Management Journal* 49, 693-706.

Hansen, MT. and B. Lovas 2004, How do multinationals companies leverage technological competencies?, *Strategic Management Journal*, Special Issue, 25, 801-822.

Hodgson, G.M. and T. Knudsen (2006), Balancing inertia, innovation, and imitation in complex environments, *Journal of Economic Issues* 60, 287-295.

Iansiti, M. (1998), *Technology integration*, Boston: Harvard Business School Press.

Katila, R. and G. Ahuja (2002),. Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introductions, *Academy of Management Journal* 45, 1183-1194.

Kauffman, S.A. (1993), *The origins of order*, New York: Oxford University Press

Kauffman, S.A., J. Lobo and R. Macready (2000), Optimal search on a technology landscape, *Journal of Economic Behavior and Organization* 43, 141-65.

Knudsen, T. (2007), Reference Groups and Variable Risk Strategies, *Journal of Economic Behavior & Organization* (forthcoming).

Knudsen, T. and D. A. Levinthal (2007), Two faces of search: Alternative generation and alternative evaluation, *Organization Science* 18, 39-54.

Knudsen, T. and Winter, S.G. (2007), *An Evolutionary Model of Spatial Competition*, Draft prepared for the DRUID conference 2007.

Lenox, M.J., S.F. Rockart and A.Y. Lewin (2006), Interdependency, competition, and the distribution of firm and industry profits, *Management Science* 52, 757-772.

Levinthal, D. A. (1997), Adaptation on rugged landscapes, *Management Science* 43, 934-951.

Levinthal, D.A. and J.G. March (1981), A model of adaptive organizational search, *Journal of Economic Behavior and Organization* 2, 307-333.

Levinthal, D. A. and J.G. March (1993), The myopia of learning, *Strategic Management Journal* 14, 95-112.

Lippman, S.A and K.F. McCardle, K.F. (1991), Uncertain search: A model of search among technologies with uncertain values, *Management Science* 37, 1474-1490.

Luce, R.D. (1959), *Individual Choice Behaviour*. New York: Wiley.

March, J.G. (1988), Variable risk preferences and adaptive aspirations, *Journal of Economic Behavior and Organization* 9, 5-24.

March, J.G. (1991), Exploration and exploitation in organizational learning, *Organization Science* 10, 71-87.

March, J.G. (1996), Continuity and change in theories of organizational action, *Administrative Science Quarterly*, 41, 278-287.

March, J.G. and R.I. Sutton (1997), Organizational performance as a dependent variable, *Organization Science* 8, 698-706.

March, J.G. and H.A. Simon (1958), *Organizations*, New York: Wiley.

McGahan, A.M. and M.E. Porter (1997), How much does industry matter, really?, *Strategic Management Journal* 18, 15-30.

Miller, D. and P. H. Friesen (1980), Momentum and revolution in organizational adaptation, *Academy of Management Journal* 23, 591-614.

Miller, D. (1996): Configurations revisited, *Strategic Management Journal* 17, 505-512.

Miyazaki, K. (1994): Search, learning and accumulation of technological competencies: The case of optoelectronics, *Industry and Corporate Change* 3, 631-654.

Nelson, R.R. and S.G. Winter (1982*), An evolutionary theory of the economic change*, Cambridge, Ma.: Harvard University Press.

Porter, M.E. (1980), *Competitive Strategy*, New York: Free Press.

Robbins, H. (1952), Some Aspects of the Sequential Design of Experiments, *Bulletin of the American Mathematical Society* 55, 527–535.

Rothaermel, F.T. and D.L. Deeds (2004), Exploration and exploitation alliances in biotechnology: A system of new product development, *Strategic Management Journal* 25, 201-221.

Rivkin, J.W. and N. Siggelkow (2001), Patterned interactions in complex systems: Implications for exploration, *Management Science*, 1068-1085

Rumelt, R.P. (1991), How much does industry matter?, *Strategic Management Journal* 12, 167-185.

Siggelkow, N. (2001), Change in the presence of fit: The rise, the fall and the renaissance of Liz Claiborne, *Academy of Management Journal* 44, 838-857.

Simon, H. A. (1955), A behavioral model of rational choice, *Quarterly Journal of Economics* 69, 99-118.

Stigler, G.S. (1961), The economics of information, *Journal of Political Economy* 69, 213-225.

Stuart, T.E. and J.M. Podolny (1996), Local search and the evolution of technological capabilities, *Strategic Management Journal* 17, 21-38.

Sutton, R.S. and A.G. Barto (1998), *Reinforcement learning*, Cambridge, Ma.: MIT Press.

Tushman, M. L. and E. Romanelli (1985), Organizational metamorphosis: A punctuated equilibrium model of organizational evolution, *Research in Organizational Behavior* 7, 171-222.

Vermorel, J. and M. Mohri (2004), Multi-Armed Bandit Algorithms and Empirical Evaluation, Paper presented at *European Conference of Machine Learning*, 2004.

Winter, S.G. (2000): The Satisficing Principle in Capability Learning, Strategic Management Journal 21, 981-996.