

Strategy as Credit Assignment

Christina Fang

Department of Management
Stern School of Business
New York University
44 West 4th Street, 7-62
New York, NY 10012.
(212) 998-0241; cfang@stern.nyu.edu

March 28th, 2006

Preliminary Draft: Please Do Not Circulate

1. Introduction

The resource-based view has emerged as a dominant paradigm for research on strategy and competitive advantage. Although the initial formulations were mainly static, focusing on the economic rents to be derived from existing resource differentials, recent work has emphasized the dynamics of resources acquisition and development (e.g. Winter, 1995; Teece et. al., 1997). The strategic factor market argument (Barney, 1986) implies that unless resources can be acquired at a price below their future expected value, they cannot be a source of competitive advantage (Peteraf, 1993). Thus, according to this logic, the only systematic way for firms to acquire resources capable of generating economic rent is superior managerial insight into the future value of resources (Barney, 1986). In other words, if the strategic factor market was efficient, the only systematic source of inter-firm profitability differentials would be systematic differences in the valuation of resources. For a firm to come into possession of resources capable of generating economic rent, it has to develop systematically more accurate expectations about the future values of resources than other resource market participants (Makadok, 2001). These arguments have become the backbone of the Resource-Based View of the firm.

Despite the importance of resource valuation, there have been few attempts to model the development of differential valuations and in particular, the mechanisms by which firms create such heterogeneous expectations (Barney, 2002)¹ Most discussions either focus on the implications of an exogenously given random distribution of valuations (Makadok, 2000; Makadok and Barney, 2001) or implicitly assume an efficient strategic factor market in which resources, on average, are priced at their expected future value. However, the assumption of an efficient strategic factor market assumes away important and well-known problems of price formation and imputation in factor markets. In particular, except for the assumption of a fictional Walrasian auctioneer who has access to demand and supply functions of all economic agents,

¹ This comes from speech given by Jay Barney in AOM symposium 2002, Denver.

classical economic models of factor prices contain no model of how consumer demand come to be reflected in factor prices (Denrell, Fang and Winter, 2003). This process is non-trivial but is totally bypassed in general equilibrium models. In short, if resources are all assumed to have their respective true future value imputed correctly, then the only source of rent is chance, or random and exogenously determined distributions.

Yet, to have any kind of theory that can predict systematically sources of competitive advantage, we have to have a theory of why and how divergent input prices are formed. Equilibrium arguments regarding factor prices provide little insight into how these beliefs regarding valuations emerge. Furthermore, questions such as measuring these expectations, identifying the techniques used to form them, assessing the skill of managers at applying these techniques remain much less understood (Makadok, 2001).

More generally, the study of resource valuation involves a dialectic between information engineers on one hand and the students of behavioral decisions on the other. Information engineers hope to depict and design resource valuation with elements of intelligence and sensibilities, as evidenced by the equilibrium models. For the latter group, the problem is to understand the actual process of resource valuation. They focus on such things as the limits of rationality and biases inherent in human decisions (Tversky and Kahnman, 1974; Simon, 1978); the mechanisms individuals and organizations conduct employ in search and discovery activities (March and Simon, 1958; Cyert and March, 1963); and the models which approximate behaviors to some optimal benchmarks rather than deriving them (Sutton and Baro, 1998).

This paper studies how divergent resource valuations may emerge by reviewing the major elements of both traditions. We argue that resource valuation can be fundamentally seen as a process of gradual formation of expectations or beliefs about the value of resources. First, building on earlier work (notably Winter 1997, Kogut and Kulatilaka, 2001), we posit that the development of expectations about the valuation of resources lies at the heart of strategy research.

Accurate formation of such beliefs, however, is often hindered by the sequential nature of many strategic decisions. Since a coherent strategy is often carried out in stages, an earlier decision has to be made often without benefits of immediate feedback as to whether it contributes to the overall strategic goal. In Section 2, we show that the challenge of valuation in this context is formally equivalent to the challenge of credit assignment (Holland 1975). To further provide some analytical handle on the mechanism of valuation, we formulate sequential decisions in the language of optimal control theory in Section 3. This formulation makes clear that the correct valuation of actions, which may have long term consequences, is key to strategic decisions. However, while the tools of optimal control theory can serve as normative benchmark of sequential behaviors, their behavioral relevance is less clear. We briefly review the descriptive validity of this normative framework in Section 4 and discuss a characterization of the process by which accurate valuation of actions is attained in a behaviorally realistic way. We show how more accurate expectations about the future values of resources can be developed sequentially, incrementally and systematically. Lastly, we discuss both theoretical and empirical implications of viewing strategy as a process of valuation.

2. Canonical Formulation of Resource Valuation

Nearly every facet of life entails a sequence of decisions in an attempt to achieve certain goals. Driving to a destination requires a series of accelerations, braking, lane changes, and turns. Completing one's education means going through successive stages of schoolings. These diverse decisions share one common feature: earlier decisions all have to be made without knowing the eventual outcomes².

² There will be no accidents if the driver can predict perfectly how others will react to her driving.

A canonical example of such lack of immediate feedback is playing a game of chess. In a game of chess, payoff occurs only at the end of a game and the information provided then only tells whether the game has been won or lost. During the entire game, players have to make appropriate moves without being informed by direct feedback as to whether the moves chosen are good or bad (i.e., take the player closer to victory or defeat). Second, even with the ultimate feedback of victory or defeat, it is far from obvious which particular moves have resulted in a win or loss since a long sequence of actions has been played. Allen Newall, a winner of the National Medal of Science and a pioneer of computer science, once remarked:” it is extremely doubtful whether there is enough information in ‘win, lose, or draw’ when referred to the whole play of the game to permit any learning at all over available time scales’ (Minsky, 1961).

In the same vein, investment in a particular course of strategic action has to be made without the benefit of knowing whether the action indeed helps in the attainment of the outcome. For instance, many strategic actions do not have any immediate or direct consequence, but instead set the stage for subsequent actions that bring the firms towards some actual payoff. The primary effect of R&D projects, advertising campaigns, enhancing (or diminishing) customer service is felt in the future not in the present. Consider a firm’s R&D decisions, which are among the primary determinants of successful adaptation over time. After an R&D strategy is carried out, payoffs are not known immediately since R&D decisions are typically located far away from the market. Their primary effect is felt in the future not in the present as a series of value chain activities such as manufacturing and marketing has to unfold sequentially. As such, it is difficult to take into account full consequences of R&D strategies when assessing alternatives at any given point in time. Yet a successful strategy requires the recognition of such ‘future’ repercussions. Many of the examples of strategic actions in Diedricxs and Cool (1989) depict firms accumulating resources to enable them to obtain a favorable future resource position. A key dimension of strategy formulation is the task of making appropriate choices about the strategic

expenditures such as advertising, R&D outlay with a view to accumulating required resources and skills such as brand loyalty or technological expertise (Diericks and Cool, 1989). Indeed, the very notion of strategic decisions in the strategy literature, in contrast to that of tactical decision-making, revolves around the fact that the former class of decisions entails longer-term consequences (Andrews, 1971).

While sequential decision processes exist in many diverse varieties, they share some common mathematical representations. Consider a simple prototypical example of a multi-stage production chain as seen in Figure 1. Suppose this firm specializes in the production of final output Y from an input resource X. It transforms X into Y through a series of 3 production stages, namely R&D, manufacturing, and testing correspondingly. The firm's problem is sequentially dependent in nature. Each department or stage is managed by a manager. At each stage, there are 9 available production choices (represented by directed arrows) and 3 types of corresponding intermediate resources (represented by nodes). For instance, resource #1 can be transformed into resource #4, #5 and #6 respectively by the R&D department in 3 different production processes. The R&D manager has to decide among 9 alternative productive processes which carry different costs as indicated by the numbers above each arrow. Similarly, each one of the resources in the manufacturing stage can be transformed into resource #7, #8 and #9. The latter intermediate resources can then be transformed in the testing stage into the final output Y. As a result of the sequential inter-dependency, however, decision makers in the earlier stages of the production such as the R&D manager have to make decisions as to which intermediate resource to target without knowing exactly how those intermediate resources may be used in downstream stages.

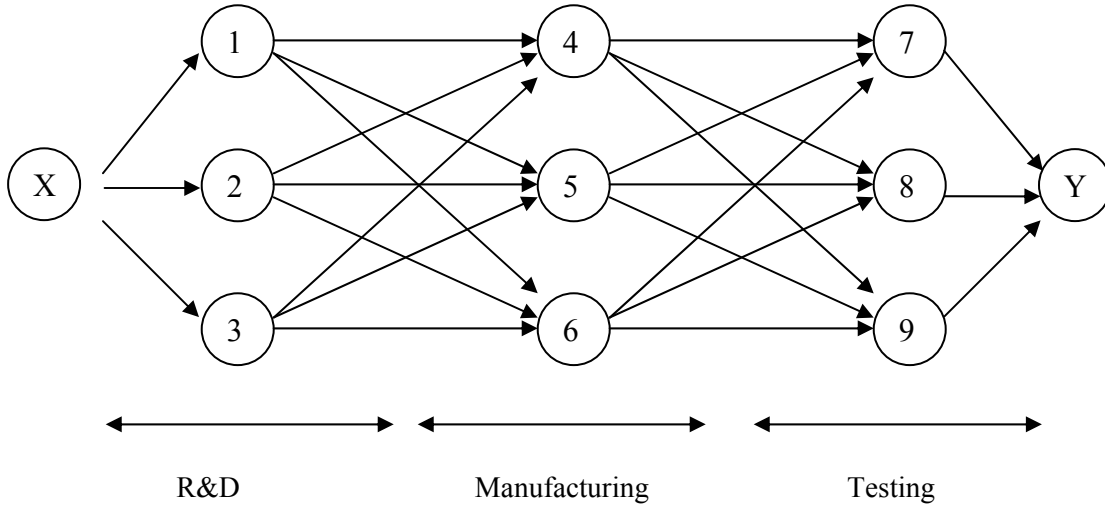


Figure 1: A Canonical Representation of Resource Valuation

The challenge in this simple example is – how do we come up with the right valuation for these heterogeneous resources? Alternatively, how should a firm decide which action or production process to undertake at each stage? It is clear that the answer to the first question simultaneously solves the second. They are the two sides of the same coin because once correct valuation of resources at each stage is obtained and optimal sequence of resources obtained, the appropriate flow decisions to accumulate those resources is also implicitly determined.

This challenge of resource valuation is a variant of a more general challenge known as the credit assignment problem in artificial intelligence (Samuel, 1959; Minsky, 1961; Holland, Holyoak, Nisbett, and Thagard, 1986; Levinthal 2000) – how should one assign the credit arising from the overall sequence of actions to each of the antecedent actions? When action and payoffs are separated across time, it is non trivial to consider assigning appropriate credit to individual moves that lead to good positions and blame to moves that lead to poor positions.

Even in this simple example, it is not immediately obvious which constitute answers to these valuation questions. Clearly, the value of an existing resource depends on the levels of other resources that it might be transformed into. Values of resources at earlier stages hinge upon the subsequent transformations downstream. For instance, to the extent that the new product and process development find their origin in customer requests (von Hippel 1978), it may be harder to

develop technological know how for firms who do not have an extensive service network (Dierickx and Cool, 1989). Similarly in our simple example, the value of resource #3 depends on the value of downstream resources such as resources #4, #5, #6, #7, #8 and #9. Given our objective to minimize the total production costs for the firm, it is important to recognize and properly account for the sequential nature of dependency inherent in this set up. For instance, even if it is cost effective to transfer resource #1 into resource #4, the latter may not represent the most cost effective intermediate resource to reach the final output Y. The value of resource #4 in turn depends on the values of all subsequent resources such as resources #7, #8 and #9.

In this sense, decisions need to take into account any downstream or long term costs. A correct valuation of an action requires that not only the immediate payoffs (i.e. costs or benefits), but also the downstream or subsequent payoffs that can be obtained from the resource, be taken into account. Rational choice cannot be based on myopic evaluation alone; but instead requires the computation of values of all resources that constitutes a particular resource combination or configuration.

How does the firm choose the optimal resource configuration that will minimize the total costs of production? One straightforward solution approach is simply direct search – go through all possible configurations and choose the optimal one. In the example above, if N and M denote the number of stages and resources per stage, then the total number of possible resource configurations is M^N or 27 (i.e. 3^3). While this seems easy enough, direct search quickly loses its allure as M and N grows in magnitude. For instance, imagine a firm with 10 stages of production processes, each with 90 different resources possible, this generates a state space of 10^{90} , a number that is greater than the total number of atoms in the entire observable universe. This implies that it is close to impossible to compute all the resources valuations by iterating over the entire state space. In the section below, we explore how a formal formulation of the credit assignment problem can help illuminate other more and plausible solutions.

3. Strategy as Optimal Control

It is well known that a broad spectrum of processes, whether encountered by engineers, economists or management consultants, may be subsumed under the abstract formulation of optimal control theory (Bellman, 1961: 3). For instance, cruise control system in a car works by carefully controlling the inputs (e.g. air fuel mixture being fed into the engine) in order to maintain the speed of a car at a constant level. Just as engineers often need to alter the behavior of a system (e.g. to reduce wasteful heat loss from an electrical system), an economic system needs to be 'controlled' to achieve certain desirable goals (e.g. to minimize unemployment). In optimal control framework, the decision maker and his immediate environment are considered to constitute a system that may be in one of a number of states. It distinguishes *state* variables which describe the states of the environments, from *control* variables, over which decision makers have control. State variables are therefore those aspects of the system that cannot be changed easily in the short run. Such a distinction is central to any attempt at descriptive theorizing, since the motion of an organization through time cannot be adequately traced or described otherwise. In particular, decision making is conceived as a process of achieving continuous control over a system in order to produce a desired outcome, rather than a resolution of a discrete choice dilemma (Brehmer, 1990).

Winter (1987) was the first to point out that many strategic initiatives can be recast in the language and form of optimal control without too much sacrifice of content and often with the benefits of revealing gaps, limitations or vagueness in the particular perspective. Indeed, most of the approaches to strategic problems in the literature do not emphasize the possibility of translating the analysis into the language, if not the formalism, of control theory (Winter, 1987). With the tools of optimal control now available, it is immediately clear that a requisite first step in strategic thinking is the identification of the attributes of the organization that are considered

subject to directed change. This echoes the distinction between stocks and flows in strategic assets (Dierickx and Cool, 1989). Just as water level in a bathtub cannot be easily adjusted, resources take a consistent pattern of resource flows to accumulate and therefore may not be subject to change. For instance, advertising expenditures can be viewed as investments with a view to accumulate brand loyalty, while R&D expenditures can be viewed as similar investment to accumulate technological expertise. The essence of these organizational resources and capabilities is their cumulateness and stock nature (Dierickx and Cool, 1989). Such organizational resources and capabilities constitute important state variables in the optimal control framework (Winter, 1987).

Optimal control theories aim to derive the optimal behavior of a system, through the use of a set of mathematical tools such as dynamic programming. It follows that a firm's choice of appropriate time path for flows (such as annual R&D expenditures) may be re-formulated in a dynamic programming framework. A sequence of actions leading to some ultimate resource accumulation may emerge as the outcome of such an analysis. For instance, a firm can be seen as choosing a capability level that maximizes the value of the projects over time (Kogut and Kulatilaka, 2001). In particular, they model the tangible benefits as well as switching costs associated with choosing a different capability level. Accordingly, core competence can be defined as the capability set (i.e. the combination of organization and technology elements) that permits the firm to dynamically choose the optimal strategy for a given price realization of the strategic factor.

We propose here a more general formulation in the context of resource valuation in order to determine the appropriate time paths of flows relevant for building required assets stocks. At a point in time t , a value function captures present value of all future benefits given optimal future behavior. Given the distinction between state and control variables, the utility of choosing an action a in some state s is the expected immediate reward from taking that action plus the

discounted sum of long term ‘delayed’ rewards over the rest of the agent’s lifetime. Formally, we denote the utility by a value function $V(s, a)$, which gives the expected return for taking action a in a state s (and thereafter following an optimal policy) for all possible state-action pairs (s, a) . It is defined as below:

$$V(s_t, a) = \max_a \{ \text{Reward}(s_t, a) + \gamma \text{Exp}[V(s_{t+1}, a^*)] \}, \quad (1)$$

where Reward is the immediate reward of choosing action a in t , s_{t+1} is the state that will be reached after a is chosen, $E[V(s_{t+1}, a)]$ is the expected maximum utility from the next period through the end of the planning horizon given that a is chosen in t ; and γ is a discount factor. This equation indicates that in each period, a firm decides which course of actions to undertake (i.e. flows), and which resource position to attain (i.e. stocks) by evaluating the total expected rewards (which is the sum of current reward and some discounted payoff from future resource positions). If it chooses action a , it realizes some immediate benefits. However, such an action may make a valuable future resource position s_{t+1} possible. Since this future benefits come at a later time, the discounted expected value should also be taken into account.

Defining such a utility function for all possible states and actions, we have a system of Bellman equations. If there is M stages and N possible resource positions for each stage, we have $N*M$ number of possible nodes. This system of equation has $N*M$ number of unknowns, to which the true utility associated with each resource position is the solution. With this formulation, solutions can be explicitly obtained by many standard methods such as dynamic programming.³

In contrast to direct search which enumerate through all possible resource configurations, dynamic programming finds the optimal combination in more efficient ways. Instead of treating a configuration as a unit of analysis, dynamic programming recognizes the sequential nature of the problem by moving systematically from one side to the other, building the best solution as it goes.

³ Solving directly is akin to an exhaustive search, looking ahead at all possibilities, computing their probabilities of occurrence and their desirabilities in terms of expected rewards.

Many strategy problems are recursive in nature in that optimal strategy from any arbitrary time t on depends only on the state of the system at that time t and does not depend on the paths that the choice variable have taken up to that point (Bellman, 1957). Dynamic programming exploits this feature by following the *principle of recursive optimality* (Bellman, 1957), reusing the optimal strategy for a state at time t in determining optimal strategies for any state earlier than t . In this way, dynamic programming breaks down the problem into sub-problems and exploits the recursive structure inherent in the problem by reusing solutions to sub-problems. For instance, an optimal policy at time $T+1$ is only computed once and treated the ‘optimal value to go’ from that time onwards. It can then be used to determined optimal strategy for time T , $T-1$, $T-2$ etc. In contrast to direct search, dynamic programming does not compute all possible solutions based on paths or configurations. Instead, it employs a ‘divide and conquer’ strategy and reuses computed solutions to the sub-problems repeatedly to reduce the number of computations. Mathematically, instead of computing M^N times, it is guaranteed to find the optimal solutions in less than some *polynomial* function of M and N (Sutton and Barto, 1998).

Given the formulation in (1) above, normative solutions are obtained by dynamic programming techniques based on *backward induction*. That is, the decision maker looks ahead to all future periods and trace out the entire path of all of future actions and payoffs till the terminal period T . One then evaluates the optimal action for that terminal period first, before solving for optimal action for the earlier period $T-1$, given the action in T , *and so on*. It is a process of first looking to the end⁴. In this way, the decision maker can determine the optimal

⁴ Alternatively, the recursive procedure can be based on *forward induction*, where the initial stage is solved first before moving forward one stage at a time, until all stages are included. While both forward and backward induction should give the same results, only backward induction is possible in cases involving uncertainty. This is because future states are not independent of the uncertain evolutions of the current states. In stochastic problems, it may be impossible to guarantee (with probability 1) that any given state can be reached. Thus, the ‘optimal value to arrive’ does not make sense. Rather, it is much more reasonable to compute ‘optimal value to go’ under backward induction. Furthermore, backward induction follows the order in which decisions are made sequentially as outcomes become revealed.

path of actions by recursively unfolding history backwards⁵. Detailed procedures involved are summarized in the Appendix.

In this way, valuation of any particular resource emerges as the optimal outcome, and should have accounted for precisely both short run returns as well as long run ones. In other words, an optimal amount of credit is assigned to the relevant individual actions according to how much each has contributed to the overall outcome.

4. Strategy as Credit Assignment

While optimal control provides a precise and internally consistent language to study resource valuation, this stylized account seems unlikely to provide a realistic description of resources valuations actually arise from human decision making behavior. Given the complexity of computing optimal decision thresholds, it seems implausible that even expert decision makers would literally follow this approach. For instance, optimal control framework assumes that the task can be fully specified by a sequence of states and actions. Yet, prior specification of full decision tree is almost impossible in reality due to a lack of full information. Real life decisions are often characterized not only by risk but also by uncertainty in the Knightian sense (Knight, 1921). Firms typically deploy their firm-specific capabilities whose rents depend partly on unfathomable futures. Failures in a certain course of action may shed light on possibilities that are not envisioned at the conception of the strategy (Adner and Levinthal, 2004). Furthermore, even if the resource valuation challenge can be reduced to a dynamic programming problem, the solution can still be hard to reach. As the number of stages grows, the number of nodes expands exponentially. Even numerical methods cannot compute a reasonable approximation, let alone analytical ones. Bellman refers to this limitation as the ‘curse of dimensionality’. While the

⁵ This technique has been extensively applied to game theory. Backward induction finds Nash equilibrium of each sub game of the original game.

formulation of strategy as optimal control problems is both intuitive and insightful, the behavioral relevance of the optimal control framework is less clear.

If resource valuation does not proceed optimally, what can be a minimal set of behavioral assumptions regarding the mechanism of resource valuation? Several intuitive principles emerge as valuable insights on which a behaviorally realistic model of sequential decisions can be built.

First, while the behavioral relevance of dynamic programming may be questioned, the basic tenet that it pays to exploit the sequential structure of the problem would seem integral of any human intuition. As uncertainty is gradually resolved with time, it makes sense to start from the ultimate objective, and deduce appropriate actions backwards. For instance, Shively, Woodward and Stanley (1999) provide some recommendation about how to approach the academic job market. They start with Year 5, when one begins a job. Starting from that end, they move backward in time, making suggestions about how a student should put all pieces together. For instance, given a job interview in Year 5, it makes sense to do interviews at professional meetings at Year 4. In order to do so, one should submit a paper to a professional conference in Year 3. Submission at Year 3 could not have taken place without first passing comprehensives at Year 2 and taking courses in Year 1. Finally, one should make sure that the end outcomes in one's phd application are as clear and strong as possible to increase the chance of admission. Decision makers first recognize the value of actions that immediately precede the final outcome. Once the value of such immediate actions is recognized, a new causal link is established. Part of the value of a given action can in turn be attributed to its own preceding action, and the 'credit' from the outcome is distributed further backwards to earlier antecedent actions. In this way, 'credit' cascades backwards down the sequential path of action, a mechanism known as bucket brigade (Holland, 1975). As such, even though the mathematics of dynamic programming is not realistic descriptively, its formal calculus is promising for a psychological interpretation (Kleiter, 1973; Brehmer 1990, 1995; Gibson, Fichman and Plaut, 1997).

Second, a viable way to solve the problem should not assume full knowledge. In reality, decision makers do not know all the possible immediate, and long term consequences of their actions. Information about immediate costs or benefits can only be obtained once the decision makers take action or try something out. An action that has not been experienced either personally or vicariously (through observations of rivals) cannot be assessed with the same certainty and confidence. Learning therefore is critically conditioned by actual past experience. Solution proceeds incrementally, gradually and approximately by largely repeating what has worked in the past. In the absence of full knowledge, the only viable basis of action is performance payoffs, or feedback from the external environment. This is consistent with the law of effect (Thorndike, 1911), which states that actions that produce positive outcomes are reinforced, and the strength of the reinforcement is proportional to the strength of the outcomes. Feedback models have been widely used to explain a wide range of human and motor behaviors.

Third, while it makes sense to start with the end in mind, there exist considerable evidence that individual decision makers are not capable of seeing far ahead into the future. When anticipating future consequences of current actions, decision makers can only look one or two periods into the future and consider limited future outcomes (Simon, 1955; 1959). As reviewed by Hutchinson and Meyer (1994), psychological experiments have demonstrated convincingly that people are not capable of looking ahead to all future periods and often do not fully utilize past information to inform future decisions. For instance, Cripps and Meyer (1994) found little evidence of forward planning in the timing of durable replacements. Individuals do not appear to think about end positions, preferring instead to engage in forward induction over an extremely limited horizon (Anderson, Pirolli and Farrell, 1988). However, it is important to stress that the very basis of forward looking is derived from backward looking learning (March, 1994). Inferences from historical experiences are folded back into the actions that create subsequent history and the past is seen as imposing itself on the present through development of routines

(Nelson and Winter, 1982). While this way of learning is sub-optimal (as compared to Bayesian learning), it is clearly more realistic and requires minimum assumptions of any insight.

These intuitive principles underlie some recent development in artificial intelligence. A simple family of learning algorithms known as temporal differencing (Kaelbling, 1993; Watkins, 1996; Sutton & Barto, 1998; Denrell, et al., 2004) has emerged in the field of machine learning as a means to achieve solutions to optimal control problems (Sutton, 1998; Sutton and Barto, 1998). This family of models explicitly combines elements from both dynamic programming and feedback-based learning models. In short, dynamic programming equations such as (1) are converted into simple updating rules, and strategies that produce success over time are reinforced more than those which do not produce success over time. In this sense, the temporal differencing approach offers a stark contrast to direct optimization methods such as genetic algorithms, which attempts to search the policy space directly. Decision makers or agents are modeled as learning optimal behavior through trial-and-error interactions with an external environment. In stark contrast to the dynamic programming approach which requires a complete knowledge of the environment, this approach requires only on line experience and feedback from the environment. No complete knowledge of the environment (including transitional probabilities etc) is required. More importantly, this particular algorithm exploits the sequential nature of tasks in the same way as dynamic programming. It has been shown effective in determining the optimal decision in a similar manner as the dynamic programming approach and the solution reached has been shown to converge to optimal as the number of iterations approach infinity.

$$V(s_t, a) \leftarrow (1 - \alpha) V(s_t, a) + \alpha * \{ \text{Reward}(s_t, a) + \gamma \max[V(s_{t+1}, a^*)] \} \quad (2)$$

As compared to (1), the only additional parameter is α , which captures the extent to which our existing valuation is revised or updated based on new information.

This particular algorithm has gained recent recognition in psychology (Berthier, Rosenstein and Barto, 2005; Sutton and Barto, 1981) and has produced results consistent with the robust property of the ‘law of effect’ (Thorndike, 1911). Striking experimental support is also found in neural science (Daw and Dayan, 2004; McClure, Daw and Montague, 2003, O’Doherty et al, 2004) following a discovery in the primate ventral tegmental area of neurons whose firing closely resembles the predicted patterns (Schultz, et al, 1997; Montague et al, 1996).

An illustration of how a behavioral model of resource valuation is helpful. Given the valuation challenge faced by our hypothetical firm in Section 2, a plausible solution to this combines elements of dynamic programming with well-known human limitations and biases in learning. It may proceed as follows. Initially the firm has little idea about either the immediate transformation costs between any two pairs of resources, or the best resource configuration. Suppose the firm happens to choose the resource configuration of #1, #6, #9 in the first trial. From this experience, the firm learns about the immediate costs of transformation among these three resources respectively. More importantly, it learns something about the potential cost effectiveness of these resources in combination. The firm assigns some credit to each of these resources, recognizing their potential contribution to the overall outcome. In the second trial, the firm has now some vague idea of what has worked in the past, so it is likely to continue some exploration by choosing somewhat randomly the resource configuration. Suppose this second trial yields a resource combination of #2, #5 and #9. This particular resource configuration is now reinforced. In particular, since resource #9 has been chosen twice, its value is likely to be updated more positively than similar resources at the same stage. This implies that if the firm has to decide which resource to accumulate at the final stage, it is likely to choose resource #9 given its higher perceived value. Suppose that resource #9 is recognized as the most valuable to develop or acquire, the firm’s next task is to find out what partial resource configuration will get to resource #9 in the most cost effective way. Instead of solving the resource valuation problem for

3 stages (i.e. R&D, manufacturing and testing), the firm now solves only a 2-stage problem. In this backward induction, as experience accumulates, the firm develops more detailed valuation of individual resources at stages further removed from the terminal stage. Over time, this enables the firm to choose the best resource at respective stages, yielding a sequence that increasingly resembles what is optimal.

5. Implications

These general ideas have some obvious conceptual, theoretical and empirical implications. A behavioral model of resource valuations provides a vehicle for addressing several important questions in strategy.

5.1 Fundamental Distinction between RBV and Dynamic Capabilities

Recent strategic thoughts have devoted considerable attention to the distinction between two rent generating mechanisms: those prescribed by the RBV and those prescribed by dynamic capabilities perspective. For instance, Makadok (2001) distinguishes between resource-picking and capability building as two alternative rent generating mechanisms. It is argued that RBV implies that everything occurs before the acquisition of a resource whereas dynamic capabilities refer to the ability to deploy those resources once owned. Thus the distinction between the two perspectives is whether we are looking at a point in time that is either before or after an acquisition. While these conceptual distinctions are instrumental in clarifying our basic understanding, they also tend to under-emphasize commonalities and shared underpinnings.

Viewing strategy as a process of valuation offers a different illumination of this distinction. Both RBV and Dynamic capabilities are fundamentally about valuation whether done internally or externally via the market. The strategic concern is the same whether inside or outside the firm. The issue is the development of correct belief structure and expectations about the

various components of the firm's strategy. It is about how expectations are formed. Accurate valuation of resources is key not only to an external market but also useful in internal markets as well (in e.g. resource allocations to competing projects). The issue in the context of internal decision-making can be seen in the classical resource allocation processes of firm where investments need to be made on different courses of action. Furthermore, how one deploys the resource depends critically on how one values it in the first place. Decision makers prioritize their time everyday and allocate their energy to things that they value most first. The valuation issue is not only germane to acquisition in an external market; but also to allocation internally as well. Viewing strategy as valuation highlights the common strategic concern shared by a myriad of strategic decisions, regardless of whether they fall outside firm boundaries.

5.2 The Existence of Strategic Opportunity

Framing resource valuation as an optimal control problem, we implicitly embrace a long familiar heuristic principle in economic thinking -- the principle of full imputation. This principle states that a proper economic valuation of a collection of resources is one that precisely accounts for the returns these resources make possible (Bellman, 1957). It implies that the maximized present value attainable from the system should be attributed to the initial state together with features of the environment and the laws of the change (Winter, 1987). As such, if value is indeed fully imputed to various components of a system, there will be little room for strategy as any net return can only be due to blind chance. Full imputation however is unrealistic due to reasons such as lack of complete markets (Denrell, Fang and Winter, 2003) among other reasons highlighted above.

Our credit assignment approach, in contrast, can be used to formalize Winter's (1987) conceptualization of strategy as heuristics. Different heuristic frames provide a different list of things that influence profitability in the long run and of how these relate to things that are

controllable in the short run (Winter, 1987). As such, strategizing can be seen as the application of imperfect heuristics to problem solving and implementation (Kogut and Kulatilaka, 2001). While the model above captures a plausible and behaviorally grounded way of solving the sequential decision problem, it does not dictate that all applications of the same decision making process will result in the same valuation outcome. As alluded before, deviations from the optimal may be systematically linked to factors such as incentives, exploration and so forth. Different heuristics used by different firms will generate differential resource valuations, and there is little guarantee that they will converge toward a theoretical optimal. Imputation is therefore less than perfect. Since strategic analysis is necessarily imperfect in the real world, there is always room for improvement. Viewing strategy as a potentially imperfect process of resource valuation liberates strategy from the straitjacket of equilibrium reasoning. This implies that economic rent is expected to exist systematically as a function of a continual process of situation analysis, decision-making, action taking and evaluation of results.

A good illustration of this is the drive to build a fully integrated accounting system, such as the balanced scorecard (Kaplan and Norton, 1992). When evaluating internal performance, a balance scorecard system takes into account of a fuller array of drivers, in addition to available financial numbers. Correct valuation of different activities critically depends on an ability to recognize good actions to which due credit should be assigned. More importantly, different balance scorecard systems yield differential assignment of credit to the same activity, and as a result, produce differential performance dynamics.

5.3 The Discovery of Strategic Opportunities

In essence, this behavioral model of resource valuation is premised on the idea of credit assignment. It captures how resource valuation can proceed based on repeated past experience with minimum assumptions of knowledge and foresight. However, it also highlights an array of

potential factors that may affect the efficacy of the process of discovery an ‘optimal’ resource configuration. For instance, to what extent can we assume that organizations or firms are unitary actors? If decision making occurs at the individual stage or departmental level rather than at the overall firm level, the process of resource valuation is further complicated. In the absence of efficacious centralized coordination, resources at different stages of the production chain may not be correctly valued. For instance, if the R&D manager refuses to explore new nodes at her department, other downstream managers would be severely constrained in their searches. Whether an ‘optimal’ configuration is obtained depends on factors such as the exploration rates, incentives to share information, and other idiosyncratic influences at the sub unit level. Another important factor is the timing of uncertainty resolution. For instance, if novel resources constantly arrive at specific stages of the production chain, uncertainty expands according around those stages. This creates a dependency in resource valuation in other stages. Our model is critically contingent upon the assumption that uncertainty is gradually reduced as times goes by. In a production context where the only common uncertainty comes from the market, time becomes a proxy for distance to the market. However, in other contexts such as innovations, the direction of uncertainty resolution is less clear.

It will be interesting to explore systematically how such factors that influence the discovery of strategic opportunities are not homogeneously relevant across organizations, decision situations and time. There is reason to expect that there may exist systematic variation in the information behavior of organization and the individuals in them. For instance, the model can be used to explore *when* differential valuations are most likely to develop. For example, by varying the complexity of the production chain and the number of local optima, it is possible to test Schoemaker’s (1990) conjecture that the scope for economic rents, as a result of different valuations, is largest in situations of intermediary complexity.

To conclude, this paper takes a first step in characterizing some general properties of strategy as a process of sequential decision making. We build on recent work on the application of optimal control framework to strategy, and point out the behavioral limitations of this work. We find that several intuitive principles emerge as valuable insights on which a behaviorally realistic model of sequential decisions can be built. These efforts represent an attempt at systematically analyzing the intricacies of strategic decision making in an internally consistent and behaviorally relevant manner. Such a perspective of viewing strategy as a process of valuation will benefit from future work that outlines more explicitly both the organizational forces that drives valuation and the empirical implications.

References

Argote, L. (1999). *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Boston, Kluwer Academic Publishers.

Bellman, R. (1957). *Dynamic Programming*. New Jersey, Princeton University Press.

Bellman, R. E. (1961). *Adaptive Control Processes: A Guided Tour*, Princeton University Press, NJ.

Cyert, R. and J. March (1963). *A Behavioral Theory of The Firm*. Englewood Cliffs, New Jersey, Prentice Hall.

Denrell, J., C. Fang, and S. Winter (2003). "The Economics of Strategic Opportunities" *Strategic Management Journal*, 24(10): 977-990, 2003.

Denrell, J., C. Fang, and D. Levinthal (2004). "From T-mazes to Labyrinths: Learning from Model-based Feedback", *Management Science*, 50(10), 1366-1378, 2004

Dierickx, I. and K. Cool (1989). "Asset Stock Accumulation and Sustainability of Competitive Advantage." *Management Science* 35(12): 1504-1511.

Eisenhardt, K. M. and J. A. Martin (2000.). "Dynamic Capabilities: What Are They?" *Strategic Management Journal*(21): 1105-1121.

Holland, J. (1992). *Adaptation in Natural and Artificial Systems*. Cambridge MA, MIT Press.

Holland, J., K. Holyoak, et al. (1986). *Induction: Processes of Inference, Learning and Discovery*. Cambridge, MIT Press.

Holland, J., K. Holyoak, et al. (1986). *Induction: Processes of Inference, Learning and Discovery*. Cambridge MA, MIT Press.

Kaelbling, L. (1993). *Learning in Embedded Systems*. Cambridge, MIT Press.

Kogut, B. and N. Kulatilaka (2001). "Capabilities as Real Options." *Organization Science* 12(6): 744-758.

Levinthal, D. (1997). "Adaptation on Rugged Landscapes." *Management Science*(43): 934-950.

Levitt, B. and J. March (1988). "Organization Learning." *Annual Review of Sociology*(14): 319-340.

Makadok, R. (2001). "Toward a Synthesis of the Resource Based and Dynamic Capability Views of Rent Generation." *Strategic Management Journal* 22: 387-401.

Makadok, R. and J. B. Barney (2001). "Strategic Factor Market Intelligence: An Application of Information Economics to Strategy Formulation and Competitor Intelligence." *Management Science* 47(12): 1621-1638.

March, J. (1991). "Exploration and Exploitation in Organization Learning." *Organization Science*(2): 71-87.

Nelson, R. and S. Winter (1982). *An Evolutionary Theory of Economic Change*. Cambridge, Harvard University Press.

Samuel, A. (1959). "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development* (31): 211-229.

Samuel, A. (1967). "Some Studies in Machine Learning Using the Game of Checkers II - Recent Progress." *IBM Journal of Research and Development* (11): 601-617.

Shively, G., Woodward, R., & Stanley, D. (1999). Strategy and etiquette for graduate students entering the academic job market. *Review of Agricultural Economics*, 21, 513-526.

Sutton, R. (1988). "Learning to Predict by the Method of Temporal Differences." *Machine Learning*(3): 9-44.

Sutton, R. and A. Barto (1998). *Reinforcement Learning: an Introduction*. Cambridge, MIT Press.

Watkins, C. (1989). *Learning From Delayed Rewards*. Cambridge, Kings' College.

Winter, S. (1987). Knowledge and Competence as Strategic Assets. *The Competitive Challenge*. D. Teece, Ballinger: 159-185.

Appendix: Solving the Resource Valuation Problem Using Dynamic Programming

Suppose that the costs of transforming resource i into resource j are listed in the table below. For instance, to transform resource 2 into resource 6, the stipulated cost is 100.

| | X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Y |
|---|---|----|----|----|-----|-----|------------|-----|-----|-----|-----|
| X | | 10 | 10 | 10 | | | | | | | |
| 1 | | | | | 50 | 75 | 130 | | | | |
| 2 | | | | | 125 | 100 | 100 | | | | |
| 3 | | | | | 80 | 40 | 30 | | | | |
| 4 | | | | | | | | 160 | 60 | 50 | |
| 5 | | | | | | | | 90 | 150 | 80 | |
| 6 | | | | | | | | 180 | 150 | 120 | |
| 7 | | | | | | | | | | | 175 |
| 8 | | | | | | | | | | | 160 |
| 9 | | | | | | | | | | | 150 |
| Y | | | | | | | | | | | |

To obtain the dynamic programming solution to this problem, we first define:

V_i = total minimum costs downstream from resource i ;

C_{ij} = cost of transforming resource i into resource j .

Given that the last stage T involves the transformation from resources 7, 8, or 9 into the final product of Y , we start from the very end and calculate the total costs starting from any one of the resources 7, 8 or 9.

(T-1) Stage: $V(9) = 150$; $V(8) = 160$; $V(7) = 175$;

These calculations are then used to derive the V_i s in the (T-2) Stage:

$$\begin{aligned} V(6) &= \min \{ C(6, 7) + \underline{V(7)} ; C(6, 8) + \underline{V(8)} ; C(6, 9) + \underline{V(9)} \} \\ &= \min \{ 180 + \underline{175} ; 150 + \underline{160} ; 120 + \underline{150} \} \\ &= \min \{ 355, 310, 270 \} \\ &= 270; \end{aligned}$$

$$\begin{aligned} V(5) &= \min \{ C(5, 7) + \underline{V(7)} ; C(5, 8) + \underline{V(8)} ; C(5, 9) + \underline{V(9)} \} \\ &= \min \{ 90 + \underline{175} ; 150 + \underline{160} ; 80 + \underline{150} \} \\ &= \min \{ 265, 310, 230 \} \\ &= 230; \end{aligned}$$

$$V(4) = \min \{ C(4, 7) + \underline{V(7)} ; C(4, 8) + \underline{V(8)} ; C(4, 9) + \underline{V(9)} \}$$

$$\begin{aligned}
&= \min \{160+175; 60+160; 50+150\} \\
&= \min \{335; 220; 200\} \\
&= 200;
\end{aligned}$$

Similarly, we can obtain the valuations of resources at the (T-3) Stage:

$$\begin{aligned}
V(3) &= \min \{ C(3, 4) + \underline{V(4)}; C(3, 5) + \underline{V(5)}; C(3, 6) + \underline{V(6)} \} \\
&= \min \{ 80+200; 40+230; 30+270 \} \\
&= \min \{ 300; 270; 300 \} \\
&= 300;
\end{aligned}$$

$$\begin{aligned}
V(2) &= \min \{ C(2, 4) + \underline{V(4)}; C(2, 5) + \underline{V(5)}; C(2, 6) + \underline{V(6)} \} \\
&= \min \{ 120+200; 100+230; 100+270 \} \\
&= \min \{ 320; 330; 370 \} \\
&= 320;
\end{aligned}$$

$$\begin{aligned}
V(1) &= \min \{ C(1, 4) + \underline{V(4)}; C(1, 5) + \underline{V(5)}; C(1, 6) + \underline{V(6)} \} \\
&= \min \{ 50+200; 75+230; 130+270 \} \\
&= \min \{ 250; 305; 400 \} \\
&= 250;
\end{aligned}$$

(T-4) Stage:

$$\begin{aligned}
V(X) &= \max \{ C(X, 1) + \underline{V(1)}; C(X, 2) + \underline{V(2)}; C(X, 3) + \underline{V(3)} \} \\
&= \max \{ 10+250; 10+320; 10+300 \} \\
&= \max \{ 260; 330; 310 \} \\
&= 260;
\end{aligned}$$

Given this list of calculations, we can see the least costly resource configuration between X and Y is 1 → 4 → 9.