**Dynamic Choice Theory and Dynamic Programming**

David M. Kreps; Evan L. Porteus

*Econometrica*, Volume 47, Issue 1 (Jan., 1979), 91-100.

Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28197901%2947%3A1%3C91%3ADCTADP%3E2.0.CO%3B2-I

*Econometrica*, Vol. 47, No. 1 (January, 1979)

# DYNAMIC CHOICE THEORY AND DYNAMIC PROGRAMMING[1]

BY DAVID M. KREPS AND EVAN L. PORTEUS

Finite horizon sequential decision problems with a "temporal von Neumann-Morgenstern utility" criterion are analyzed. This criterion, as developed in [7], is a generalization of von Neumann-Morgenstern (expected) utility of the vector of rewards, wherein an individual's preferences concerning the timing of the resolution of uncertainty are taken into account. The preference theory underlying this criterion is reviewed and then extended in natural fashion to yield preferences for strategies in sequential decision problems. The main result is that value functions for sequential decision problems can be defined by a dynamic programming recursion using the functions which represent the original preferences, and these value functions represent the preferences defined on strategies. This permits citation of standard results from the dynamic programming literature, concerning the existence of (memoryless) strategies which are optimal with respect to the given preference relation.

## 1. INTRODUCTION AND SUMMARY

CONSIDER A FINITE HORIZON sequential decision problem, where at each time $t = 0, 1, \ldots, T$, an individual must choose an action $d_t$. The actions available are determined by the state at time $t$, $x_t$. Some random event takes place, determining an immediate payoff $z_t$ and the next state $x_{t+1}$. The probability distribution of the pair $(z_t, x_{t+1})$ is determined by the action $d_t$. The individual desires to select "best" actions contingent on states and other information that is relevant. In a conventional approach, "best" is defined through a cardinal utility function on the vector of payoffs: Each (measurable) strategy (i.e., contingent plan for selecting actions) induces a probability measure on the vector of payoffs, and strategies are ranked by the expected utility they thereby induce. An advantage of this criterion is that dynamic programming can be used to find "best" strategies, when they exist. One flaw with this sort of criterion, however, is that it ignores the temporal aspect of the resolution of uncertainty. That is, two strategies may induce the same probability distribution of payoffs, but one may cause the uncertainty to resolve at an earlier time than does the other. Insofar as there is value in earlier resolution of uncertainty, the individual may prefer the former strategy. In [7], we provided some motivation for the consideration of preference structures that account for this temporal resolution of uncertainty. (See also [**2**, **10**, **16**] for motivation in the context of consumption-investment budgeting.) Furthermore, we axiomatized, represented, and studied special cases of such preference structures. In this paper, we seek to show how dynamic programming can be used to analyze sequential decision problems when the individual has such a preference structure. We frequently cite results from [7], and we assume no further motivation for considering such preference structures is necessary. Thus, the reader may wish to refer to [7] before proceeding.

[1] This research was supported in part by a grant from the Atlantic Richfield Foundation to the Stanford Graduate School of Business.

In Section 2, we recount the relevant definitions and results from [7]. In that paper, two approaches to dynamic choice theory were given—one descriptive and the other normative. They were shown to be equivalent in a natural sense. We use only the normative approach here, as the resulting development closely parallels the conventional "expected utility" criterion development. The crucial notion introduced in Section 2 is a *temporal lottery*, which is a generalization of a lottery (or probability distribution) on the payoff vector. The generalization is that uncertainty is "dated" by the time of its resolution in a temporal lottery; in comparison, this "dating" is not encoded by a probability distribution on the payoff vector. This allows us to introduce preferences which can distinguish between temporal lotteries solely because the times at which their uncertainty resolves differ.

In Section 3, we define what we mean by a sequential decision problem and compare this with the conventional definition of such a problem and with what was termed a dynamic choice problem in [7]. After defining (measurable) strategies and policies, a crucial construction is given: Each strategy, payoff history, and state give rise to a corresponding temporal lottery, viz., the temporal lottery that the individual faces if, given the payoff history and state, he thereafter uses that strategy. (This parallels the usual construction of a probability measure on the payoff vector from a strategy, history, and state.)

This construction allows us in Section 4 to extend in a natural fashion the (assumed) preference relation on temporal lotteries to a preference relation on strategies. Given a representation of the preferences on temporal lotteries (as described in Section 2), we give a dynamic programming recursion which defines value functions for strategies, and show that these value functions represent the preferences on strategies. Optimality of strategies is defined and optimality criteria are given. Finally, we adapt our results to the general operator approach to sequential decision problems as developed in [**12**], which permits us to cite further results and to investigate the optimality of strategies which have special structure. As examples, we give conditions under which there exists an optimal (measurable) strategy and under which there exists an optimal "memoryless" strategy. (These results are only sketched, as the mathematics employed are quite standard. Our objective in this paper is primarily to provide the connection between the dynamic choice theory of [7] and the standard theory of sequential decision problems/dynamic programming.)

Throughout, we deal only with finite horizon problems (that is, problems with only a finite number of times at which actions must be taken). A treatment of countable stage decision problems must await development of a theory of preferences for countable stage temporal lotteries. Also, we do not consider "almost-optimal" criteria, although the details of a treatment of such criteria will become apparent.

Mitten [**9**] and Sobel [**15**] have dealt with general ordinal criteria for sequential decision problems. In particular, Section 5 of Sobel formally subsumes the preference structures we deal with here. But by dealing only with the special preference structures axiomatized in [7], we are able to give sharper results. A

special case of the sequential decision problems treated in this paper was considered in [12]. It was given as an example of a problem that could not be modelled using a conventional affine operator or even extrema of affine operators. However, it was not axiomatized and there was no recognition of the phenomenon of temporal resolution of uncertainty. Finally, preferences which have identical mathematical representations as those considered here, but which are motivated somewhat differently, are developed by Selden [14].

## 2. DYNAMIC CHOICE THEORY

Let $T$ be a positive integer and, for each (time) $t = 0, 1, \ldots, T$, let $Z_t$ be a compact Polish (i.e., complete separable metric) space. The set $Z_t$ is the set of possible *payoffs* at time $t$, with generic element $z_t$. Define $Y_1 = Z_0$ and, recursively, $Y_t = Y_{t-1} \times Z_{t-1}$ for $t = 2, \ldots, T+1$. The set $Y_t$ is the set of payoff *histories* up to (but not including) the time $t$ payoff, with generic element $y_t = (z_0, \ldots, z_{t-1})$ Note that $Y_{T+1}$ is the set of complete payoff vectors. For notational convenience, we will sometimes write $Y_t$ or $y_t$ when $t = 0$; $Y_0$ can be taken to be any convenient singleton set.

The objects on which the individual is assumed to have primitive preferences are called *temporal lotteries*. There is uncertainty concerning the payoffs that will be received, and we assume that this uncertainty resolves at times $t = 0, \ldots, T$ and that all uncertainty concerning $z_t$ must resolve on or before time $t$. However, the uncertainty concerning $z_t$ may partially or completely resolve before time $t$, and the individual may prefer earlier or later resolution of this uncertainty. The concept of a temporal lottery is a formal way of "dating" uncertainty by the time of its resolution. This is accomplished by the following definitions.

Let $D_T^*$ be the space of Borel probability measures on $Z_T$, endowed with the Prohorov metric (i.e., the metric of weak convergence). Recursively, let $D_t^*$ be the space of all Borel probability measures on $Z_t \times D_{t+1}^*$. (This construction is possible because each $D_t^*$, metrized with the Prohorov metric, is compact and Polish, cf. [11].) Generic elements of $D_t^*$ are denoted by $d_t$. Note that each $D_t^*$ is a mixture space: If $d_t, d_t' \in D_t^*$ and $\alpha \in [0, 1]$, let $\alpha d_t + (1-\alpha)d_t'$ denote the element of $D_t^*$ which, for $A$ a Borel measurable subset of $Z_t \times D_{t+1}^*$, assigns measure $\alpha \cdot d_t(A) + (1-\alpha) \cdot d_t'(A)$ to $A$. Elements of $D_t^*$ can be depicted as probability (or event) trees with chance nodes for times $t, t+1, \ldots, T$, as described in [7]. (In that paper, degenerate "choice nodes" were included in the depiction of elements of $D_t^*$. Here, it is convenient to delete them.)

The space $D_0^*$ is the space of temporal lotteries. Contained within $D_0^*$ are temporal lotteries where no uncertainty resolves until time $t$ or after. Clearly, in these temporal lotteries the payoffs at times $0, \ldots, t-1$ must be deterministic, given by some $y_t \in Y_t$ and the "beyond time $t$" structure is given by some $d_t \in D_t^*$. So we can write $Y_t \times D_t^*$ (with generic element $(y_t, d_t)$) to denote the set of these special temporal lotteries. Note that in this notation, $Y_t \times D_t^* \cong Y_{t+1} \times D_{t+1}^*$.

A further piece of notation: For $d_t \in D_t^*$ and $f$ a bounded measurable function on $Z_t \times D_{t+1}^*$, the expectation of $f$ with respect to $d_t$ is denoted by $E_{d_t}[f]$.

We assume, as in [7], that the individual expresses a preference ordering on elements of $D_0^*$, denoted by $\succsim$, with $\sim$ and $\succ$ denoting indifference and strict preference, respectively, which satisfies the following axiom:

AXIOM N: *(a) The relation $\succsim$ is complete and transitive. (b) The relation $\succsim$ is continuous. (c) If $d_t, d_t' \in D_t^*$ and $y_t \in Y_t$ are such that $(y_t, d_t) \succ (y_t, d_t')$, then $(y_t, \alpha d_t + (1-\alpha)d_t'') \succ (y_t, \alpha d_t' + (1-\alpha)d_t'')$ for all $\alpha \in (0, 1]$ and $d_t'' \in D_t$.*

Parts (a) and (b) of the axiom should be clear. Part (c) is a "temporal substitution" property which allows us to apply the machinery of cardinal utility theory (cf. [4]) and obtain the following representation theorem.

THEOREM 1: *A relation $\succsim$ on $D_0^*$ satisfies Axiom N if and only if there exist continuous functions $u_t^*: Y_t \times Z_t \times R \to R$ for $t = 0, 1, \ldots, T-1$, and $u_T^*: Y_T \times Z_T \to R$ such that* (a) *for $t = 1, \ldots, T-1$, $u_t^*$ is strictly increasing in its third argument, and* (b) *if we define $U_T^*: Y_T \times D_T^* \to R$ by $U_T^*(y_T, d_T) = E_{d_T}[u_T^*(y_T, \tilde{z}_T)]$ and $U_t^*: Y_t \times D_t^* \to R$ $(t = 0, \ldots, T-1)$ by*

$$U_t^*(y_t, d_t) = E_{d_t}[u_t^*(y_t, \tilde{z}_t, U_{t+1}^*((y_t, \tilde{z}_t), \tilde{d}_{t+1})],$$

*then $(y_t, d_t) \succsim (y_t, d_t')$ if and only if $U_t^*(y_t, d_t) \geqslant U_t^*(y_t, d_t')$.*

(Here, $R$ denotes the real line. Tildes denote random variables.) The proof of this result may be found in [7]. The functions $\{u_t^*\}$ are called *basic utilities* representing $\succsim$. This preference structure is *not* representable by the expectation of a single cardinal utility function on the vector of payoffs. Here, the individual may distinguish between temporal lotteries when the only difference between them is in the timing of the resolution of uncertainty. Only when the individual is indifferent to the timing of resolution of uncertainty can his preferences be represented by a single utility function. (The formal conditions for this are Axiom N and $\alpha(y_t, d_t) + (1-\alpha)(y_t, d_t') \sim (y_t, \alpha d_t + (1-\alpha)d_t')$ for all $\alpha$, $y_t$, $d_t$, and $d_t'$.) The nature of the individual's preferences for earlier or later resolution of uncertainty are given by the shape of the $u_t^*$ as a function of their third argument. See [7] for further details.

A special case of particular interest in the context of Markov decision problems is when the preference relation is *temporally separable*. By this we mean: If $(y_t, d_t) \succsim (y_t, d_t')$ for some $y_t \in Y_t$, then $(y_t', d_t) \succsim (y_t', d_t')$ for all $y_t' \in Y_t$. Verbally, preferences concerning what happens beyond time $t$ are independent of previous payoffs. If this assumption holds, the representation simplifies as follows. The function $u_T^*$ can be assumed to be a function of $z_T$ only and, for $t < T$, $u_t^*$ can be taken to be a function on $Z_t \times R$. (In the case where preferences can be represented by a single cardinal utility function on the vector of payoffs, this implies that the utility function is separable, a special case being the additive utility function which dominates the literature of Markov decision problems.)

One point concerning this dynamic choice theory should be carefully noted. We are implicitly assuming that the individual's preferences are temporally consistent in the following sense. His preferences for "time $t$" temporal lotteries (elements of $D_t^*$) at time $t$ depend on the payoff history $y_t$ and are given by $\succsim$ restricted to $\{y_t\} \times D_t^*$. (This is *not* an assumption of "stationarity" of preferences or what Sobel [**15**] calls "temporal persistence of preferences." Such an assumption would only make sense in our context in an infinite horizon model (ours has a finite horizon) and would entail requiring that preferences be stationary over time rather than that they be consistent.) This assumption does rule out consideration of "changing preferences" in the sense of Hammond [**5**] (among others).

## 3. SEQUENTIAL DECISION PROBLEMS, POLICIES, AND STRATEGIES

DEFINITION: A *sequential decision problem* over payoff spaces $\{Z_t; t = 0, \ldots, T\}$ is a collection $\{X_t, D_t, A_t(x_t); t = 0, \ldots, T\}$ of (a) *state spaces* $X_t$ with generic element $x_t$ which are Polish spaces, (b) *action spaces* $D_t$ with generic $d_t$, where $D_t$ is the space of Borel probability measures on $Z_t \times X_{t+1}$, metrized throughout by the Prohorov metric, and (c) for each $x_t \in X_t$, sets $A_t(x_t)$ of *feasible actions* from state $x_t$, which are subsets of $D_t$.

As in Section 2, we write $E_{d_t}[f]$ for the expectation of a bounded Borel measurable function $f$ on $Z_t \times X_{t+1}$ taken with respect to $d_t$. Also, $X_{T+1}$ is taken for notational convenience to be some singleton set. Note that $D_t$ as metrized is Polish and is compact if $X_{t+1}$ is.

This differences between our definition and the standard definitions of a sequential decision problem found in the literature (see Blackwell [**1**], Hinderer [**6**], Strauch [**17**], etc.) are as follows. The notion of a state is just as in the literature—the state is a statistic which tells the individual which actions are feasible. An action here, however, is *identified* as the probability distribution on pairs of immediate payoff and next state. In comparison, the literature typically has transition probabilities as functions of states and actions. This difference is purely semantic. In [**7**], we used the terminology dynamic choice problem rather than sequential decision problem. There, $x_t$ was used to denote the (necessarily closed) set of feasible actions itself, rather than a statistic indexing that set. That is, $A_t(x_t)$ would simply be $x_t$. The definition used here allows for greater generality—$A_t(\cdot)$ needn't be a continuous correspondence in the $X_t$ topology. Note that we do not (yet) wish to assume that $A_t(\cdot)$ has any structural properties, such as being a measurable correspondence.

DEFINITIONS: Given payoff spaces $\{Z_t\}$ and a sequential decision problem $\{X_t, D_t, A_t(x_t)\}$, an admissible *policy* for time $t$ is any Borel measurable function $\delta_t: Y_t \times X_t \to D_t$ such that $\delta_t(y_t, x_t) \in A_t(x_t)$ for all $y_t \in Y_t$ and $x_t \in X_t$. The set of (admissible) policies for time $t$ is denoted by $\Delta_t$. An (admissible) *strategy* is a vector of policies $\pi = (\delta_0, \delta_1, \ldots, \delta_T) \in \Delta_0 \times \ldots \times \Delta_T = \Pi$.

A policy for time $t$ specifies a feasible action for each pair of payoff history and state. Note that we do not formally allow the selection of an action to depend on previous states and/or actions but only on previous payoffs. (Of course, one can technically incorporate the state history into the payoff history in applications (after compactifying $X_t$) and then use the machinery of Section 4 to show that the state history can be considered to be irrelevant if it is irrelevant to the individual's preferences.) Since nothing has been assumed about the measurability of $A_t(\cdot)$, the requirement that policies are measurable makes it possible that some $\Delta_t$ (and thus $\Pi$) is empty.

For each strategy $\pi$, time $t$ payoff history $y_t$ and state $x_t$, there exists a corresponding temporal lottery—the temporal lottery which arises if at time $t$ with history $y_t$ and state $x_t$ the individual uses strategy $\pi$. The formal definition of this temporal lottery follows.

Fix $\pi = (\delta_0, \delta_1, \ldots, \delta_T)$ and define $d_T^\pi: Y_T \times X_T \to D_T^*$ by $d_T^\pi(y_T, x_T) = \delta_T(y_T, x_T)$. For $t = T-1, \ldots, 0$, recursively define $d_t^\pi: Y_t \times X_t \to D_t^*$ by

$$(1) \qquad d_t^\pi(y_t, x_t) = \delta_t(y_t, x_t) \circ (1, d_{t+1}^\pi((y_t, \cdot), \cdot))^{-1},$$

where 1 here denotes the identity map on $Z_t$. Equation (1) can be interpreted as follows. Suppose (inductively) that $d_{t+1}^\pi(y_{t+1}, x_{t+1})$ is the $D_{t+1}^*$ "part" of the temporal lottery associated with $\pi$, $y_{t+1}$ and $x_{t+1}$. That is, $d_{t+1}(y_{t+1}, x_{t+1})$ is a measure on $Z_{t+1} \times D_{t+2}^*$. Fix $y_t$, $x_t$, policy $\delta_t$ (from strategy $\pi$), and a measurable subset $A$ of $Z_t \times D_{t+1}^*$. From $(y_t, x_t)$ and using $\delta_t$, the decision maker will be "in" $A$ if the ensuing $(z_t, x_{t+1})$ are such that $(z_t, d_{t+1}^\pi((y_t, z_t), x_{t+1})) \in A$. That is, he will be "in" $A$ if $(z_t, x_{t+1}) \in (1, d_{t+1}^\pi((y_t, \cdot), \cdot))^{-1} A$. The probability of this is of course $\delta_t(y_t, x_t) \circ (1, d_{t+1}^\pi((y_t, \cdot), \cdot))^{-1} A$. (An implication of this construction is that if $f: Y_{t+1} \times D_{t+1}^* \to R$ is bounded and measurable and $g: Y_{t+1} \times X_{t+1} \to R$ is defined by $g(y_{t+1}, x_{t+1}) = f(y_{t+1}, d_{t+1}^\pi(y_{t+1}, x_{t+1}))$, then $E_{d_t^\pi(y_t, x_t)}[f] = E_{\delta_t(y_t, x_t)}[g]$.) Of course, the temporal lottery corresponding to $\pi$, $y_t$, and $x_t$ is $(y_t, d_t^\pi(y_t, x_t))$.

## 4. DYNAMIC PROGRAMMING

Taken as primitive is a binary relation $\succsim$ on the set of temporal lotteries which satisfies Axiom N. Let $\{u_t^*\}$ be basic utilities which represent $\succsim$ in the sense of Theorem 1.

We begin by using $\succsim$ to induce preference relations on strategies. For each pair $(y_t, x_t)$, it is natural to say that strategy $\pi$ is $(y_t, x_t)$-*preferred* to $\pi'$ if the temporal lottery which arises from $\pi$, $y_t$, and $x_t$ is preferred to that arising from $\pi'$, $y_t$, and $x_t$. Formally,

$$\pi \succsim_{(y_t, x_t)} \pi' \quad \text{if} \quad (y_t, d_t^\pi(y_t, x_t)) \succsim (y_t, d_t^{\pi'}(y_t, x_t)).$$

Obviously, each $\succsim_{(y_t, x_t)}$ is complete and transitive. Furthermore, $\pi$ is said to be *optimal* if $\pi \succsim_{(y_t, x_t)} \pi'$ for all $y_t$, $x_t$, and $\pi'$.

The connection with dynamic programming is that we shall use the basic utilities which represent $\succsim$ to represent the relations $\succsim_{(y_t, x_t)}$. For $\pi = (\delta_0, \delta_1, \ldots, \delta_T)$, define *value functions for strategy* $\pi$ by the following recursion. Let

$v_T(\pi; y_T, x_T) = E_{\delta_T(y_T, x_T)}[u_T^*(y_T, \tilde{z}_T)$ and, for $t = T-1, \ldots, 0$, define $v_t(\pi; \cdot, \cdot)$: $Y_t \times X_t \to R$ by

$$(2) \qquad v_t(\pi; y_t, x_t) = E_{\delta_t(y_t, x_t)}[u_t^*(y_t, \tilde{z}_t, v_{t+1}(\pi; (y_t, \tilde{z}_t), \tilde{x}_{t+1}))].$$

Also, define *optimal value functions* by

$$f_t(y_t, x_t) = \sup_{\pi \in \Pi} v_t(\pi; y_t, x_t),$$

where the suprema are taken pointwise. (The terminology used will be justified shortly.)

LEMMA: *For $\pi \in \Pi$, $y_t \in Y_t$, and $x_t \in X_t$, $v_t(\pi; y_t, x_t) = U_t^*(y_t, d_t^\pi(y_t, x_t))$.*

PROOF: We proceed by backward induction in $t$. For $t = T$, $v_T(\pi; y_T, x_T) = E_{\delta_T(y_T, x_T)}[u_T^*] = U_T^*(y_T, \delta_T(y_T, x_T)) = U_T^*(y_T, d_T^\pi(y_T, x_T))$. Assume the result is true for $t+1$. Then

$$\begin{aligned} U_t^*(y_t, d_t^\pi(y_t, x_t)) &= E_{d_t^\pi(y_t, x_t)}[u_t^*(y_t, \tilde{z}_t, U_{t+1}^*((y_t, \tilde{z}_t), \tilde{d}_{t+1}))] \\ &= E_{\delta_t(y_t, x_t)}[u_t^*(y_t, \tilde{z}_t, U_{t+1}^*((y_t, \tilde{z}_t), d_{t+1}^\pi((y_t, \tilde{z}_t), \tilde{x}_{t+1}))] \\ &\quad \text{(by (1))} \\ &= E_{\delta_t(y_t, x_t)}[u_t^*(y_t, \tilde{z}_t, v_{t+1}(\pi; (y_t, \tilde{z}_t), \tilde{x}_{t+1}))] \text{ (by induction)} \\ &= v_t(\pi; y_t, x_t) \quad \text{(by definition).} \end{aligned}$$

*Q.E.D.*

Immediate from the lemma and the definitions is the following.

THEOREM 2: *For all $\pi$, $\pi'$, $y_t$, and $x_t$, $\pi \succeq_{(y_t, x_t)} \pi'$ if and only if $v_t(\pi; y_t, x_t) \geq v_t(\pi'; y_t, x_t)$. Strategy $\pi$ is optimal if and only if $v_t(\pi; y_t, x_t) = f_t(y_t, x_t)$ for all $y_t$ and $x_t$.*

With this result and the recursive definition of $\{v_t\}$ given by equation (2), we can call upon the extant theories of finite horizon dynamic programming to investigate issues such as the existence and possible structure of optimal strategies. The general operator approach of [**12**] can be utilized as follows.

Let $V_t$ denote the set of bounded Borel measurable real valued functions defined on $Y_t \times X_t$ for $t = 0, \ldots, T+1$. For $\delta \in \Delta_t$, $v \in V_{t+1}$, and $t < T$, define the operator $H_{t\delta}$ on $V_{t+1}$ by $[H_{t\delta}v](y_t, x_t) = E_{\delta(y_t, x_t)}[u_t^*(y_t, \tilde{z}_t, v((y_t, \tilde{z}_t), \tilde{x}_{t+1}))]$. For $t = T$, define $H_{T\delta}$ on $V_{T+1}$ by $[H_{T\delta}v](y_T, x_T) = E_{\delta(y_T, x_T)}[u_T^*(y_T, \tilde{z}_T)]$. The continuity of $u_t^*$ insures that $H_{t\delta}: V_{t+1} \to V_t$. Also, for $v \in V_{t+1}$, define $A_t v = \sup_{\delta \in \Delta} H_{t\delta} v$, where the supremum is taken pointwise. Standard examples (see, for example [**17**]) show that $A_t v$ need not be in $V_t$.

Suppose one can define a distinguished subset of $V_t$, denoted $V_t^*$ and called the set of *structured value functions*, and distinguished subsets of each $\Delta_t$, denoted $\Delta_t^*$ and called the sets of *structured policies*, such that the following condition is met.

CONDITION PA (Preservation and Attainment): *If* $v \in V_{t+1}^*$ *(for* $t \leq T$*), then* $A_t v \in V_t^*$ *and there exists* $\delta \in \Delta_t^*$ *such that* $H_{t\delta} v = A_t v$.

Then the standard procedure of dynamic programming can be utilized.

THEOREM 3: *If condition PA holds, then* (a) $f_t \in V_t^*$ *for* $t = 0, \ldots, T$, (b) $f_t = A_t f_{t+1}$ *for* $t = 0, \ldots, T$, (c) $\pi = (\delta_0, \ldots, \delta_T)$ *is optimal if and only if* $H_{t\delta_t} f_{t+1} = A_t f_{t+1}$ *for each* $t$, *and* (d) *there exists a strategy* $\pi \in \Delta_0^* \times \Delta_1^* \times \ldots \times \Delta_T^*$ *which is optimal.*

The proof of Theorem 1 in [12] can easily be modified to give this result. (Note that $H_{t\delta}$ is isotone because $u_t'$ is nondecreasing in its third argument. Also, the proof in [12] nominally requires that the $H_{t\delta}$ be Lipschitzian, which may not be the case here, but is only used in [12] for $\varepsilon$-optimality. In fact, the Lipschitzian condition would not be necessary even if our objective was $\varepsilon$-optimality—the compactness of the $Z_t$ allows us to replace the Lipschitzian condition with one of uniform continuity.)

We conclude with a pair of lemmas and corollaries to Theorem 3. The lemmas give conditions under which Condition PA holds (i) for $\Delta_t^*$ being the set of all Borel measurable policies and (ii) for $\Delta_t^*$ being the set of "memoryless" policies. The corollaries then harvest the results of Theorem 3 for these cases. We make use of the selection theorem of Schäl [13, Corollary 4] which slightly generalizes that of Dubins and Savage [3].

LEMMA: *If* $A_t(x_t)$ *is compact in* $D_t$ *and* $A_t(\cdot)$ *is an upper semi-continuous correspondence for all* $t$, *then condition PA holds for* $\Delta_t^* = \Delta_t$ *and* $V_t^*$ *the set of bounded u.s.c. functions on* $Y_t \times X_{t+1}$.

A special case of the lemma is the case in which all the $Z_t$ and $X_t$ are discrete and each $A_t(x_t)$ is compact. Then $A_t$ is automatically u.s.c.

PROOF: If $v$ is a bounded u.s.c. function on $Y_{t+1} \times X_{t+1}$, then $u_t^*(y_t, z_t, v((y_t, z_t), x_{t+1}))$ is as well, because $u_t^*$ is continuous and nondecreasing in its third argument. Then (cf. Maitra [8, Lemma 4]) $w((y_t, x_t), d_t) = E_{d_t}[u_t^*(y_t, \tilde{z}_t, v((y_t, \tilde{z}_t), \tilde{x}_{t+1}))]$ is bounded and u.s.c. on $Y_t \times X_t \times D_t$. That $w$ and $A_t(\cdot)$ are u.s.c. imply that $\sup_{d_t \in A_t(x_t)} w((y_t, x_t), d_t)$ is bounded and u.s.c. Thus $A_t v \in V_t^*$. Finally, [13, Theorem 2] yields the existence of $\delta \in \Delta_t^*$ such that $A_t v = H_{t\delta} v$. *Q.E.D.*

COROLLARY: *If* $A_t(x_t)$ *is compact in* $D_t$ *and* $A_t(\cdot)$ *is an u.s.c. correspondence, then there exists an optimal (measurable) strategy.*

DEFINITIONS: A policy $\delta \in \Delta_t$ is *memoryless* (it could be called Markov) if $\delta_t$ is $x_t$-measurable; that is, if $\delta(y_t, x_t) = \delta(y_t', x_t)$ for all $x_t$, $y_t$, and $y_t'$. A strategy $\pi$ is memoryless if it is composed of memoryless policies.

LEMMA: *If $A_t(x_t)$ is compact in $D_t$, $A_t(\cdot)$ is an u.s.c. correspondence, and preferences are temporally separable (cf. Section 2), then condition PA holds with $V_t^*$ the set of bounded u.s.c. $x_t$-measurable functions and $\Delta_t^*$ the set of memoryless (time $t$) policies.*

The proof is apparent, following the sketch of the proof of the previous lemma.

COROLLARY: *If $A_t(x_t)$ is compact in $D_t$, $A_t(\cdot)$ is an u.s.c. correspondence, and preferences are temporally separable, then there exists an optimal memoryless strategy.*

Note that in this case, the recursions analogous to "Bellman's equations" are

$$f_t(x_t) = \sup_{d \in A_t(x_t)} E_d[u_t^*(\tilde{z}_t, f_{t+1}(\tilde{x}_{t+1}))],$$

instead of the standard equations for separable von Neumann-Morgenstern utility:

$$f_t(x_t) = \sup_{d \in A_t(x_t)} E_d[\alpha_t(\tilde{z}_t) + \beta_t(\tilde{z}_t) \cdot f_{t+1}(\tilde{x}_{t+1})]$$

(for $\beta_t > 0$).

*Stanford University*

*Manuscript received July, 1977; revision received November, 1977.*

REFERENCES


[1] BLACKWELL, D.: "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36 (1965), 226–235.

[2] DRÈZE, J., AND F. MODIGLIANI: "Consumption Decisions under Uncertainty," *Journal of Economic Theory*, 5 (1972), 308–335.

[3] DUBINS, L., AND L. SAVAGE: *How To Gamble If You Must*. New York: McGraw-Hill, 1965.

[4] FISHBURN, P.: *Utility Theory for Decision Making*. New York: John Wiley and Sons, 1970.

[5] HAMMOND, P.: "Changing Tastes and Coherent Dynamic Choice," *Review of Economic Studies*, 21 (1976), 159–173.

[6] HINDERER, K.: *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*. New York: Springer-Verlag, 1970.

[7] KREPS, D., AND E. PORTEUS: "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46 (1978), 185–200.

[8] MAITRA, A.: "Discounted Dynamic Programming on Compact Metric Spaces," *Sankhya*, 30A (1968), 211–216.

[9] MITTEN, L.: "Preference Order Dynamic Programming," *Management Science*, 21 (1974), 43–46.

[10] MOSSIN, J.: "A Note on Uncertainty and Preferences in a Temporal Context," *American Economic Review*, 59 (1969), 172–174.

[11] PARTHASARATHY, K.: *Probability Measures on Metric Spaces*. New York: Academic Press, 1970.

[12] PORTEUS, E.: "On the Optimality of Structured Policies in Countable Stage Decision Problems," *Management Science*, 22 (1975), 148–157.

[13] SCHÄL, M.: "A Selection Theorem for Optimization Problems," *Archiv der Mathematik*, 25 (1974), 219–224.
[14] SELDEN, L.: "A New Representation of Preferences over 'Certain × Uncertain' Consumption Pairs: The 'Ordinal Certainty Equivalent' Hypothesis," *Econometrica*, 46 (1978), 1045–1060.
[15] SOBEL, M.: "Ordinal Dynamic Programming," *Management Science*, 21 (1975), 967–975.
[16] SPENCE, M., AND R. ZECKHAUSER: "The Effect of Timing of Consumption Decisions and the Resolution of Lotteries on the Choice of Lotteries," *Econometrica*, 40 (1972), 401–403.
[17] STRAUCH, R.: "Negative Dynamic Programming," *Annals of Mathematical Statistics*, 37 (1966), 871–890.