Adv. Appl. Prob. **45,** 51–85 (2013)
Printed in Northern Ireland

MONOTONE POLICIES AND INDEXABILITY FOR BIDIRECTIONAL RESTLESS BANDITS

K. D. GLAZEBROOK,* *Lancaster University*
D. J. HODGE,** *The University of Nottingham*
C. KIRKBRIDE,* *** *Lancaster University*

Abstract

Motivated by a wide range of applications, we consider a development of Whittle's restless bandit model in which project activation requires a state-dependent amount of a key resource, which is assumed to be available at a constant rate. As many projects may be activated at each decision epoch as resource availability allows. We seek a policy for project activation within resource constraints which minimises an aggregate cost rate for the system. Project indices derived from a Lagrangian relaxation of the original problem exist provided the structural requirement of indexability is met. Verification of this property and derivation of the related indices is greatly simplified when the solution of the Lagrangian relaxation has a state monotone structure for each constituent project. We demonstrate that this is indeed the case for a wide range of *bidirectional* projects in which the project state tends to move in a different direction when it is activated from that in which it moves when passive. This is natural in many application domains in which activation of a project ameliorates its condition, which otherwise tends to deteriorate or deplete. In some cases the state monotonicity required is related to the structure of state transitions, while in others it is also related to the nature of costs. Two numerical studies demonstrate the value of the ideas for the construction of policies for dynamic resource allocation, most especially in contexts which involve a large number of projects.

Keywords: Asset management; Gittins index; indexability; inventory management; Lagrangian relaxation; machine maintenance; monotone policy; stochastic dynamic programming; restless bandit; Whittle index

2010 Mathematics Subject Classification: Primary 90C40
Secondary 49L20; 90C39; 49M20

1. Introduction

Gittins [4], [5], in what is now regarded as a classical contribution, demonstrated the optimality of policies of index form for a version of the multiarmed bandit problem with a discounted reward criterion. Such problems present to the decision-maker a fixed range of alternatives (often called *projects*, *arms*, or *bandits*) which evolve stochastically when chosen and which earn state-dependent rewards as they do so. They represent one of the simplest interesting classes of models for *dynamic resource allocation*. Whittle [17] proposed an extension of this model class such that passive projects also evolve, though according to a different stochastic law. His setting called for m of n restless bandits (as he called them) to be active at any time, the goal being to devise a policy for project activation to maximise the time

Received 17 September 2010; revision received 28 May 2012.
* Postal address: Department of Management Science, Lancaster University, Lancaster, LA1 4YX, UK.
** Postal address: School of Mathematical Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK.
*** Email address: c.kirkbride@lancaster.ac.uk

average reward achieved over an infinite horizon. He proposed an index policy derived from a Lagrangian relaxation in which a subsidy is paid to passive projects, equivalently a charge is levied for project activation. While Whittle's index heuristic does not achieve optimality in general, Weber and Weiss [15], [16] have demonstrated a form of asymptotic optimality, under given conditions, in a limiting regime in which m and n grow in proportion. Whittle's model has proved extremely popular and has been proposed in many application contexts. These include the routeing of unmanned military aircraft (see, for example, [10]), opportunistic communication channel usage (see [11]), inventory routeing (see [2]), machine maintenance (see [18]), asset management (see [7]), and queueing control (see [1]).

Whittle's index is only defined for restless bandits which enjoy a structural property called *indexability*. This property requires that each bandit, when subject to a charge for its activation, should have an associated family of optimal policies in which the set of states in which activation is optimal is monotone decreasing in the charge. This property, while natural, is by no means universal. A complete understanding has certainly not been achieved. Niño-Mora [12], [13] has espoused an approach to indexability based on the achievable region approach to stochastic optimisation. See also [3] and [8].

In this paper we propose a simple direct approach to indexability in the context of families of more complex restless bandits than those envisaged by Whittle in that they require a state-dependent amount of some key resource (manpower, money, equipment) for their activation. A fixed amount of this resource is available in total at each decision epoch and as many of the projects may be activated as may be accommodated within the resource constraint. This class of models is introduced in Section 2 and has widespread application. For example, more depleted inventories require more resource for their replenishment, older machines may require more man-hours to be serviced, and more highly performing assets may need more investment to maintain their performance. Note that Glazebrook and Minty [6] developed a generalised Gittins index appropriate for a similar extension of the classical discounted reward multiarmed bandit problem. Furthermore, while the models of Jacko [9] feature state-dependent resource usage, no global resource constraint is imposed.

For models of the kind considered in this paper, indexability is studied by solving an average cost activation problem for each constituent restless bandit when a charge W is levied per unit of resource consumed. For indexability, we need to demonstrate the existence of activation policies which are monotone in W in the strong sense described above. We shall describe a simple direct approach which draws upon the fact that, for a large number of important cases, we can demonstrate the existence of optimal policies for the activation problems with resource charge which are state monotone, namely which have the character that the set of states in which activation of the project is optimal lie above (or below) some threshold. In Section 3 this is demonstrated for a class of so-called *bidirectional* projects in which the active and passive actions tend to move the project in opposite directions, and which are skip-free in the direction of passive evolution. This class includes the *spinning plates model* of Glazebrook *et al.* [7] and the *Ehrenfest project* of Whittle [17] as special cases, neither of which feature state-dependent resource consumption. Once state monotonicity of optimal activation policies has been established for this class, the subsequent analysis of indexability is straightforward. For example, should activation be optimal for large states (above an appropriate threshold) then indexability will be secured whenever the rate at which resource is consumed decreases as the threshold for activation rises. This kind of condition is often easy to check.

In Section 4 we seek to escape the skip-free assumption of passive evolution made in Section 3 in pursuit of the goal of broadening further the applicability of the material. A natural and

important domain of application is to inventory management where, under the passive action (no stock replenishment), an inventory will be depleted by the maximum of the remaining stock and one day's demand for it. If when managing a collection of such inventory-holding locations the key resource in the replenishment process is the availability of stock itself, then it is plain that the appropriate restless bandit model will indeed consume state-dependent amounts of this resource when activated, thus extending the scope of the models of Archibald *et al.* [2]. In Section 4 state monotonicity of optimal policies for the associated activation problem with resource charge is achieved by constraining the cost structure of the problem. This has the innovative consequence that indexability, namely the demonstration that optimal policies are monotone in the resource charge W, is initially demonstrated only over a finite W-range. It is also the case for these more complex models that indexability does not flow simply from the demonstrated state monotonicity of optimal policies for the activation problems as it did in Section 3. While applications to inventory management are highlighted in Section 4, the approach and the results have relevance to other application domains.

In the course of the paper, we use two numerical studies to illustrate how the theoretical results of the paper may be deployed to develop index heuristics for dynamic resource allocation and how strongly the resulting policies perform.

2. Indexability and index heuristics

We define a *restless bandit* model with general resource requirements $[\{\Omega_j, p_j^a, p_j^b, C_j^a, C_j^b, R_j\}; 1 \le j \le J]$ to be a Markov decision process (MDP) with average costs whose main features are as follows.

(i) The *state space* of the process is $\times_{j=1}^{J} \Omega_j$ with Ω_j the (finite or countable) state space of project (or bandit) j, denoted by P_j. For most of the examples discussed in the paper, Ω_j will be a finite subset of $\mathbb{N}$. The *state* of the process is observed at each time $t \in \mathbb{N}$. At time t, we write $\bar{X}(t) = \{X_1(t), X_2(t), \dots, X_J(t)\}$ for the process state with $X_j(t) \in \Omega_j$, the state space of bandit j.

(ii) At each time $t \in \mathbb{N}$ an action $\bar{a} = \{a_1, a_2, \dots, a_J\}$ is applied to the process, with $a_j \in \{a, b\}$ the action applied to bandit j. The action a is *active* and calls for the positive commitment of resource. The alternative action b is *passive*. Under action $\bar{a}$, the N projects evolve independently, each according to its own Markov law. We write

$$\begin{aligned}
&\mathbb{P}\{\bar{X}(t+1) = \bar{y} \mid \bar{X}(t) = \bar{x}, \bar{a}\} \\
&\quad = \prod_{j=1}^{J} \mathbb{P}\{X_j(t+1) = y_j \mid X_j(t) = x_j, a_j\} \\
&\quad = \prod_{j=1}^{J} \{p_j^a(x_j, y_j) I(a_j = a) + p_j^b(x_j, y_j) I(a_j = b)\}, \qquad (1)
\end{aligned}$$

where in (1) and elsewhere, I is an indicator.

(iii) For all j, C_j^a and C_j^b are *cost functions* which map the state space Ω_j to the positive reals $\mathbb{R}^+$. The expected cost incurred when action $\bar{a}$ is applied to the process in state $\bar{x}$

is written

$$C(\bar{a}, \bar{x}) = \sum_{j=1}^{J} \{C_j^a(x_j) I(a_j = a) + C_j^b(x_j) I(a_j = b)\},$$

and so costs are additive across projects.

(iv) For all j, $R_j : \Omega_j \to \mathbb{R}^+$ is a *resource consumption function*. The amount of resource consumed when action $\bar{a}$ is applied to the process in state $\bar{x}$ is written

$$R(\bar{a}, \bar{x}) = \sum_{j=1}^{J} R_j(x_j) I(a_j = a),$$

and is also additive across projects. The set of *admissible actions* in state $\bar{x}$ is given by

$$A(\bar{x}) = \{\bar{a}; R(\bar{a}, \bar{x}) \leq R\},$$

where R is the total resource available at each decision epoch. We shall suppose that $A(\bar{x}) \neq \varnothing$, $\bar{x} \in \times_{j=1}^{J} \Omega_j$, and, moreover, that there exists $\bar{x} \in \times_{j=1}^{J} \Omega_j$ such that $A(\bar{x}) \neq \{a, b\}^J$. Hence, resource availability is taken to strictly constrain the set of available actions.

(v) An admissible policy $\bar{\pi}$ is a rule for choosing an admissible action at each decision epoch. Such a rule can in principle depend upon the entire history (successive actions taken and states occupied) of the process to date. We shall seek an admissible policy to minimise the average cost per unit time incurred over an infinite horizon. See Puterman [14] for an account of the general theory relating to such problems. The standard approach to such problems utilises dynamic programming (DP). However, a pure DP approach is unlikely to yield insight and will be computationally intractable for problems of realistic size. Hence, the primary quest is for strongly performing *heuristic policies*. In what follows we shall restrict to the class $\bar{S}$ of admissible, stationary, deterministic, and Markov policies which are identified with functions $\bar{\pi} : \times_{j=1}^{J} \Omega_j \to \{a, b\}^J$ and which choose admissible actions without randomisation and on the basis of the current process state only.

Following the classical work of Gittins [4], [5] and Whittle [17] we shall seek heuristics in the form of *index policies*. Hence, we shall seek *calibrating functions* $W_j : \Omega_j \to \mathbb{R}$, $1 \leq j \leq J$, which will guide the construction of good actions in each process state. In order to develop such indices, we shall seek a decomposition of (a relaxation of) our optimisation problem into J individual problems, one for each bandit. It is these individual problems which, when suitably structured, will yield the calibrating functions required.

We proceed via a series of relaxations as follows: we state the optimisation problem of interest as

$$C^{\text{opt}}(\bar{x}) = \inf_{\bar{\pi} \in \bar{S}} \left\{ \sum_{j=1}^{J} C_j^{\bar{\pi}}(\bar{x}) \right\}. \tag{2}$$

In (2), $C_j^{\bar{\pi}}(\bar{x})$ is the average cost per unit time incurred by P_j under policy $\bar{\pi}$ from initial state $\bar{x}$. Note that we make no *a priori* assumption that this quantity is independent of $\bar{x}$. We now relax (2) by extending to a class of policies $\bar{S}(\bar{x})$ (some of which may be inadmissible) which are

stationary, deterministic, and Markov and for which the average resource consumed per unit time from initial state $\bar{x}$ does not exceed R. We express this latter condition by writing

$$\sum_{j=1}^{J} R_j^{\bar{\pi}}(\bar{x}) \leq R, \qquad \bar{\pi} \in \bar{S}(\bar{x}). \tag{3}$$

In (3), $R_j^{\bar{\pi}}(\bar{x})$ is the average resource consumed per unit time by P_j when policy $\bar{\pi}$ is applied from initial state $\bar{x}$. Plainly,

$$\bar{S}(\bar{x}) \supseteq \bar{S}, \qquad \bar{x} \in \bigtimes_{j=1}^{J} \Omega_j,$$

and we relax (2) to

$$\bar{C}^{\text{opt}}(\bar{x}) = \inf_{\bar{\pi} \in \bar{S}(\bar{x})} \left\{ \sum_{j=1}^{J} C_j^{\bar{\pi}}(\bar{x}) \right\}. \tag{4}$$

We now further relax the problem by (i) extending to the class of stationary, deterministic, and Markov policies in which all admissibility requirements are abandoned and (ii) incorporating the average resource consumption constraint (3) into the objective in a Lagrangian fashion. The class of policies described in (i) is denoted by $\bar{S}'$ and is identified with the set of all functions $\bar{\pi}\colon \bigtimes_{j=1}^{J} \Omega_j \to \{a, b\}^J$. We now write

$$C^{\text{opt}}(\bar{x}, W) = \inf_{\bar{\pi} \in \bar{S}'} \sum_{j=1}^{J} \{C_j^{\bar{\pi}}(\bar{x}) + W R_j^{\bar{\pi}}(\bar{x})\} - WR. \tag{5}$$

Plainly,

$$C^{\text{opt}}(\bar{x}, W) \leq \bar{C}^{\text{opt}}(\bar{x}) \leq C^{\text{opt}}(\bar{x}), \qquad \bar{x} \in \bigtimes_{j=1}^{J} \Omega_j,\ W \in \mathbb{R}^+.$$

It is clear that the Lagrangian relaxation (5) permits an additive decomposition by individual project. We write

$$C^{\text{opt}}(\bar{x}, W) = \sum_{j=1}^{J} C_j(x_j, W) - WR. \tag{6}$$

In (6), $C_j(x_j, W)$ is the value of a stochastic DP concerning P_j *only* in which we seek a stationary, deterministic, and Markov policy $\pi_j\colon \Omega_j \to \{a, b\}$ to minimise an aggregate of the rate of costs incurred by project activation and the rate of charges imposed for the resource consumed. We write, in a natural notation,

$$C_j(x_j, W) = \inf_{\pi_j \in S_j'} \{C_j^{\pi_j}(x_j) + W R_j^{\pi_j}(x_j)\}. \tag{7}$$

Note that the Lagrange multiplier W has an economic interpretation as a *charge imposed* per unit of resource consumed. Denote this optimisation problem for P_j by $P_j(x_j, W)$. From (6), an optimal policy for the Lagrangian relaxation in (5) runs optimal policies for the $P_j(x_j, W)$, $1 \leq j \leq J$, in parallel. We note at this point that we shall use $C_j^{\pi_j}$ and $R_j^{\pi_j}$ for the cost/consumption rates in (7) if these are known to be independent of x_j.

In order to develop ideas further we shall need to explore issues relating to the structure of solutions to the problems $P_j(x_j, W)$, $1 \leq j \leq J$. We shall say that a policy is *optimal*

for $P_j(W)$ if it minimises $C_j(x_j, W)$ for all $x_j \in \Omega_j$. Hence, the solution to $P_j(W)$, should one exist, is a policy which is optimal for $P_j(x_j, W)$ uniformly over initial states x_j. Let $\pi_j : \Omega_j \to \{a, b\}$ be some policy for $P_j(\cdot, W)$; we write $b_j(\pi) := \{x_j \in \Omega_j; \pi_j(x_j) = b\}$ for its *passive set*, namely the set of states in which π_j takes the passive action. The natural requirement that the passive set of an optimal policy for $P_j(W)$, should one exist, be increasing in the resource charge W is expressed in Definition 1 below along with other related and stronger notions. Since in this definition we are describing properties of *individual projects*, we shall drop the project identifier j from the notation. We shall do this for much of the rest of the paper.

Definition 1. Project P is indexable if there exists a collection of policies $\{\pi(W),\ W \in \mathbb{R}\}$ such that

(i) $\pi(W)$ is optimal for $P(W)$ for all $W \in \mathbb{R}$; and

(ii) the passive set $b(\pi(W))$ is increasing in W.

The related index function $W : \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is given by

$$W(x) = \begin{cases} -\infty & \text{if } x \in b(\pi(W)) \text{ for all } W \in \mathbb{R}, \\ \infty & \text{if } x \notin b(\pi(W)) \text{ for all } W \in \mathbb{R}, \\ \inf\{W; x \in b(\pi(W))\} & \text{otherwise.} \end{cases}$$

An indexable project is strictly indexable whenever there is an index function which is one-to-one and whose range is contained in $\mathbb{R}$.

Remark 1. In the models discussed in the paper, it may be that there are isolated W-values at which there is more than one choice of optimal policy $\pi(W)$. Even when this is the case, the corresponding related index functions are all equal. Hence, in practice, there may be some arbitrariness in relation to the specification of optimal policies for $P(W)$ but none in relation to index functions.

Remark 2. From the above definitions, there exist optimal policies for the Lagrangian relaxation in (5) of index form when all J constituent projects P_j are indexable. In state $\bar{x}$ there exists an optimal policy $\bar{\pi}(W)$ which will apply the active action to all projects P_j for which $W_j(x_j) \geq W$ and the passive action to all projects for which $W_j(x_j) < W$. Furthermore, should there exist a positive W-value, $W(R)$ say, for which the policy $\bar{\pi}(W(R))$ achieves equality in (3) for all initial states then it must be optimal for the relaxation in (4). These facts motivate the design of heuristic policies in which the choice of action is based on current values of the related index functions $W_j,\ 1 \leq j \leq J$.

Remark 3. For most practical purposes, $\mathbb{R}$ can be replaced by $\mathbb{R}^+$ throughout Definition 1. This modification replaces Definition 1(ii) in requiring that only $b(\pi(W))$ be increasing in W over $\mathbb{R}^+$ and modifies the related index function by setting $W(x) = 0$ if $x \in b(\pi(W)),\ W \in \mathbb{R}^+$. We shall say that the project is (*strictly*) *indexable* over $\mathbb{R}^+$. This is the approach taken, for example, in Glazebrook and Minty [6]. To see why such a formulation may be adequate, indeed natural, for the development of index heuristics, note that in some applications (including, for example, the inventory problems of Section 4) it may be best in some states *not* to attempt to use as much of the available resource as possible. Hence, imagine that the model $[\{\Omega_j, p_j^a, p_j^b, C_j^a, C_j^b, R_j\}; 1 \leq j \leq J]$ is enhanced to include a number of additional 'do nothing' projects, each of which has a single state $*$, incurs no costs, and consumes resource at a constant rate. Such projects are trivially indexable, with $W(*) = 0$. Suppose that these

'do nothing' projects collectively exhaust all of the available resource. If all the projects P_j are indexable over $\mathbb{R}^+$ then there is an optimal solution to the Lagrangian relaxation which is as follows. When $W = 0$, in state $\bar{x}$ activate all the 'do nothing' projects and all conventional projects P_j for which $W_j(x_j) > 0$. When $W > 0$, activate all conventional projects P_j for which $W_j(x_j) > W$. In this solution, the 'do nothing' projects are always preferred to any conventional project whose index over $\mathbb{R}^+$ is zero. It is also true that, for any $\bar{x}$, the number of activated projects is decreasing in W. It is then natural to develop a policy for the original problem by taking this solution to the Lagrangian relaxation and, in every state $\bar{x}$, increasing W from zero until enough projects are rendered passive that an admissible action is achieved. These observations motivate the description of index heuristics given in Remark 4 below. For the simpler models in Section 3, we are able to analyse (strict) indexability as in Definition 1, while for the more complex models of Section 4, we deal with (strict) indexability over $\mathbb{R}^+$.

Remark 4. Following Remarks 2 and 3, a natural policy proposal for the optimisation problem (2) when all projects P_j are indexable (whether over $\mathbb{R}$ or $\mathbb{R}^+$) constructs an action $\bar{a}$ in the process state $\bar{x}$ as follows. List the projects in descending order of the index values $W_j(x_j)$, with largest first. Proceed down the list and choose the active action for projects until *either* the resource constraint is violated by any further activation of projects *or* the next index on the list is nonpositive. Repeat for all states. This is the *greedy index heuristic* (GI heuristic). It is possible for the GI heuristic to perform poorly because it makes inadequate use of the available resource. One proposed fix for this is to replace the index $W_j(x_j)$ by the product $W_j(x_j)R_j(x_j)$ in the above prescription. We thus derive the *weighted greedy index heuristic* (WGI heuristic). An alternative approach is to choose action $\bar{a}$ to maximise the sum

$$\sum_{j=1}^{J} W_j(x_j) I(a_j = a)$$

within the resource constraint. This is the *knapsack index heuristic* (KI heuristic). We develop a *weighted knapsack index heuristic* (WKI heuristic) in the obvious way.

Over the next two sections of the paper, we elucidate a project structure which is both natural and which also yields indexability. In doing so we are then able to develop strong heuristic policies for applications which include asset management, machine maintenance, and inventory routeing.

3. Bidirectional bandits, monotone structures, and indexability

Throughout the theoretical development of this section and the next, we shall focus on structural properties of individual projects and drop the project identifier j from the notation. Hence, the focus of our study will be project $P = \{\Omega, p^a, p^b, C^a, C^b, R\}$ and the associated optimisation problem $P(W)$. Recall that in the latter the challenge is to determine some policy $\pi : \Omega \to \{a, b\}$ to minimise an average cost rate for the project which aggregates costs directly incurred through project activation with charges levied for resource consumption.

In this section we focus on projects which arise in many applications in which the active and passive actions tend to move the project state in opposite directions. In a *descendant project*, the active action consumes resource to move the state downwards to states in which costs are incurred at a lower rate. Under the passive action, a descendant project tends to deteriorate via transitions upwards to states in which costs are incurred at a higher rate. In *ascendant projects* these directions are reversed. In a descendant (ascendant) project, good policies will

call for the active action in high (low) states and we should expect index values to be larger for these. In Definition 2 below and for the remainder of this section up to the comments following Theorem 2, we suppose that project P has the finite state space $\Omega = \{0, 1, \dots, N\}$.

Definition 2. A project P is descendant if

(D1) $p^a(x, y) = 0$, $y > x$, and $p^a(x, x) < 1$, $x \in \Omega \setminus \{0\}$;

(D2) $p^a(0, 0) = 1$;

(D3) $p^b(x, y) = 0$, $y > x + 1$, and $p^b(x, x + 1) > 0$, $x \in \Omega \setminus \{N\}$.

A project P is ascendant if

(A1) $p^a(x, y) = 0$, $y < x$, and $p^a(x, x) < 1$, $x \in \Omega \setminus \{N\}$;

(A2) $p^a(N, N) = 1$;

(A3) $p^b(x, y) = 0$, $y < x - 1$, and $p^b(x, x - 1) > 0$, $x \in \Omega \setminus \{0\}$.

It will also usually be the case that costs $C^a(x)$ and $C^b(x)$ are nondecreasing (respectively, nonincreasing) in x for descendant (respectively, ascendant) projects and this is borne out in the examples discussed in Section 3.3. However, the indexability theory for bidirectional restless bandits developed in Section 3.1 and Section 3.2 depends upon project structure through only p^a and p^b. To make that clear, we prefer *not* to include conditions on project cost structure in Definition 2. The nature of the resource consumption function $R(x)$ varies considerably between applications.

Note from (D3) that evolution under the passive action is *right skip-free*, reflecting the notion that project deterioration tends to be gradual. In (D3) a possible leftward jump toward a lower state under the passive action may model, for example, some catastrophic failure of a machine, say, with enforced replacement/renewal at high cost. See Example 2 below for an instance of this. Similar comments apply to ascendant projects.

Please note that a descendant (ascendant) project becomes ascendant (descendant) under a redesignation of states under the transform $x \to N - x$. Hence, in these finite-state cases, we need only develop theory for the descendant case. We shall always adopt a choice of state variable (and, hence, of project type) which seems natural for the example concerned and/or which is consistent with conventions adopted in related literature. Extensions of the theory to projects with countable state space are discussed later.

3.1. Monotone policies are optimal for $P(W)$

The study of the indexability of descendant and ascendant projects is considerably simplified by the fact that there always exists a *monotone policy* which is optimal for $P(W)$. Denote the monotone policy π such that

$$\pi(y) = b \quad \Longleftrightarrow \quad 0 \le y \le x - 1$$

by x^{D}. Note that 0^{D} takes the action a in all states while $(N + 1)^{\mathrm{D}}$ takes the action b in all states. Denote the monotone policy π such that

$$\pi(y) = a \quad \Longleftrightarrow \quad 0 \le y \le x$$

by x^{A}. Note that $(-1)^{\mathrm{A}}$ takes the action b in all states while N^{A} takes the action a in all states.

Proposition 1. (i) *If project P is descendant, there exists some policy x^{D} which is optimal for $P(W)$.*

(ii) *If project P is ascendant, there exists some policy x^{A} which is optimal for $P(W)$.*

The correspondence between descendant and ascendant projects noted above means that we need only prove Proposition 1(i). Before we proceed to the proof, we need some preparatory ideas and notation. Assume henceforth that P is descendant. Consider the state process $\{X(t),\ t \in \mathbb{N}\}$ evolving under policy x^{D}. This is a Markov chain with one-step transition matrix $p^{x^{\mathrm{D}}}$, where

$$p^{x^{\mathrm{D}}}(y,\cdot) = p^b(y,\cdot)I(0 \le y \le x-1) + p^a(y,\cdot)I(x \le y \le N), \qquad y \in \Omega.$$

We call this Markov chain the x^{D}-chain. Write $D(x) \subseteq \Omega$ for the *communicating class* of state x for the x^{D}-chain. Similarly, write $D'(N)$ for the communicating class of state N for the $(N+1)^{\mathrm{D}}$-chain. The following result is preparatory to the proof of the above proposition.

Lemma 1. (i) $D(0) = \{0\}$ *and* 0 *is an absorbing state for the* 0^{D}*-chain.*

(ii) *If $x \in \Omega \setminus \{0\}$ then $D(x)$ has the form $\{y_x, y_x+1, \dots, x\}$, where $y_x < x$ and is closed for the x^{D}-chain.*

(iii) *$D'(N)$ has the form $\{y'_N, y'_N+1, \dots, N\}$, where $y'_N \le N$ and is closed for the $(N+1)^{\mathrm{D}}$-chain.*

Proof. Part (i) is a trivial consequence of conditions (D1) and (D2). We now prove part (ii). Consider the x^{D}-chain with initial state x. It is clear from conditions (D1) and (D3) that if $X(0) = x$ then the chain can never enter the subset $\{y;\ y \ge x+1\}$ and, hence, there does not exist $y > x$ with $y \in D(x)$. Plainly, condition (D1) together with the right skip-free evolution of (D3) imply that there exists $y < x$ for which $y \in D(x)$. Use y_x to denote the smallest such and consider any y in the range $y_x < y < x$. Condition (D3) implies that evolution of the chain from y_x to y in $y - y_x$ time units has positive probability as does evolution from y to x in $x - y$ time units. This, together with the fact that $y_x \in D(x)$, plainly implies that $y \in D(x)$. It then follows that $D(x)$ must have the form claimed in the lemma.

Suppose now that $D(x)$ is *not* closed and, hence, that there exist $y \in D(x)$ and $z \notin D(x)$ with $p^{x^{\mathrm{D}}}(y,z) > 0$. From the above argument, it follows that $z < y_x$. However, under condition (D3), evolution from z to y in $y - z$ time units has positive probability, from which it follows that $z \in D(x)$, which is the required contradiction. This completes the proof of part (ii). Part (iii) is proved similarly.

We are now in a position to prove Proposition 1 for descendant projects. Before we do so, the reader should recall the notation $C^{\pi}(y)$ and $R^{\pi}(y)$ for the average cost incurred and the resource consumed, respectively, by the project per unit time when operating under policy π from initial state y. We thus have cost rate $C^{x^{\mathrm{D}}}(y)$ and resource consumption rate $R^{x^{\mathrm{D}}}(y)$ when monotone policy x^{D} is applied. However, it is clear that, under policy x^{D}, any descendant project will become trapped in $D(x)$ in finite time and, hence, that $C^{x^{\mathrm{D}}}(y)$ and $R^{x^{\mathrm{D}}}(y)$ do not depend upon y. We thus use the notation $C^{x^{\mathrm{D}}}$ and $R^{x^{\mathrm{D}}}$ for descendant projects (and $C^{x^{\mathrm{A}}}$ and $R^{x^{\mathrm{A}}}$ for ascendant projects) without ambiguity.

Proof of Proposition 1. Consider policy $\pi \in S'$ for descendant project P. We suppose initially that $\pi(0) = b$, and define the integers r^{π} and x^{π}_r, y^{π}_r, $1 \le r \le r^{\pi}$, as follows.

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

If $\pi \neq (N+1)^{\mathrm{D}}$ then set

$$x_1^\pi = \min\{z; \pi(z) = a\}.$$

Otherwise, set $x_1^\pi = y_1^\pi = N+1$. If $x_1^\pi \leq N$ and $\pi(z) = b$ for some z, $x_1^\pi + 1 \leq z \leq N$, then set

$$y_1^\pi = \min\{z; z \geq x_1^\pi + 1 \text{ and } \pi(z) = b\}.$$

Otherwise, set $y_1^\pi = N+1$. If $y_1^\pi = N+1$ then stop and set $r^\pi = 1$.

We now proceed inductively. Suppose that we have x_r^π, y_r^π, $1 \leq r \leq r'$. If $y_{r'-1}^\pi < N+1$ and $y_{r'}^\pi = N+1$, then stop and set $r^\pi = r'$. If $y_{r'}^\pi < N+1$, we continue as follows. If $\pi(z) = a$ for some z, $y_{r'}^\pi + 1 \leq z \leq N$, then set

$$x_{r'+1}^\pi = \min\{z; z \geq y_{r'}^\pi + 1 \text{ and } \pi(z) = a\}.$$

Otherwise, set $x_{r'+1}^\pi = y_{r'+1}^\pi = N+1$. If $x_{r'+1}^\pi \leq N$ and $\pi(z) = b$ for some z, $x_{r'+1}^\pi + 1 \leq z \leq N$, then set

$$y_{r'+1}^\pi = \min\{z; z \geq x_{r'+1}^\pi + 1 \text{ and } \pi(z) = b\}.$$

Otherwise, set $y_{r'+1}^\pi = N+1$, and then stop and set $r^\pi = r'+1$.

We have

$$0 < x_1^\pi < y_1^\pi < \cdots < x_{r^\pi}^\pi \leq y_{r^\pi}^\pi = N+1.$$

We now define the subsets Ω_r^π, $1 \leq r \leq r^\pi$, by

$$\Omega_r^\pi := \{y_{r-1}^\pi, y_{r-1}^\pi + 1, \ldots, y_r^\pi - 1\}, \qquad 1 \leq r \leq r^\pi, \tag{8}$$

where we set $y_0^\pi = 0$. Note that $x_r^\pi \in \Omega_r^\pi$, $1 \leq r \leq r^\pi - 1$. The actions taken by policy π when applied to the states in any subset Ω_r^π, $1 \leq r \leq r^\pi - 1$, have the form $\{bb\ldots baa\ldots a\}$. In the case of $\Omega_{r^\pi}^\pi$ this remains valid when $x_{r^\pi}^\pi < y_{r^\pi}^\pi$. When $x_{r^\pi}^\pi = y_{r^\pi}^\pi$, then $\pi(x) = b$, $x \in \Omega_{r^\pi}^\pi$. We now consider these cases in turn.

Case 1: $x_{r^\pi}^\pi < y_{r^\pi}^\pi = N+1$. Consider project P evolving under policy π. We say that π has a *cycle* at $x_r^\pi \in \Omega_r^\pi$ if the communicating class $D(x_r^\pi) \subseteq \Omega_r^\pi$. Note that it is trivial to deduce from Lemma 1 that π certainly has a cycle in set Ω_1^π. It follows from the fact that the set $D(x)$ is closed for the x^{D}-chain that, should the state process $\{X(t),\ t \in \mathbb{N}\}$ evolving under π ever enter one of the $D(x_r^\pi)$ at which π has a cycle, then it will remain there. Suppose that $X(0) \in \Omega_r^\pi$, $r > 1$, and that π has a cycle at x_r^π. It is clear from the nature of the one-step transition matrix for the state process evolving under π, denoted by p^π, that in finite time the system will *either* evolve to $\bigcup_{v \leq r-1} \Omega_v^\pi$ *or* it will be trapped in $D(x_r^\pi)$. If $X(0) \in \Omega_r^\pi$, $r > 1$, and π has no cycle at x_r^π, evolution of the system state to $\bigcup_{v \leq r-1} \Omega_v^\pi$ is guaranteed in finite time. It now follows simply that, from any initial state, the system evolving under π will in finite time be trapped in one of the $D(x_r^\pi)$ at which π has a cycle almost surely.

Let $X(0) = x$ and write $p^\pi(r \mid x)$ for the probability that the system state, evolving from x under policy π, becomes trapped in $D(x_r^\pi)$ at which π has a cycle. Conditional upon this entrapment, the cost rate incurred in the problem $P(W)$ under π over an infinite horizon is the same as that incurred from any initial state under the monotone policy $x_r^{\pi^{\mathrm{D}}}$. This latter assertion follows from the fact that, under conditions (D1)–(D3), the system state evolving from any state under monotone policy y^{D} becomes trapped in the closed set $D(y)$ in finite time, almost surely. We thus have an expression for the cost rate under π from initial state x of the form

$$C^\pi(x) + WR^\pi(x) = \sum_r p^\pi(r \mid x)\{C^{x_r^{\pi^{\mathrm{D}}}} + WR^{x_r^{\pi^{\mathrm{D}}}}\} \geq \min_{0 \leq y \leq N+1} (C^{y^{\mathrm{D}}} + WR^{y^{\mathrm{D}}}), \tag{9}$$

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

where the summation is over those r for which π has a cycle at x_r^π. See (7) and note that the average cost rates and consumption rates for the monotone policies featuring in (9) are independent of the initial state x. From (9) we infer the existence of a monotone policy y^{D} with cost rate no higher than π, uniformly over initial states.

Case 2: $x_{r^\pi}^\pi = y_{r^\pi}^\pi = N+1$. If $X(0) \notin \Omega_{r^\pi}^\pi$ then the analysis and conclusions of case 1 apply. Suppose now that $X(0) = x \in \Omega_{r^\pi}^\pi$. Over the range $1 \le r \le r^\pi - 1$ we continue to say that π has a cycle at x_r^π if $D(x_r^\pi) \subseteq \Omega_r^\pi$. However, the changed structure of $\Omega_{r^\pi}^\pi$ in this case means that we now say that π has a cycle at N if $D'(N) \subseteq \Omega_{r^\pi}^\pi$. Suppose that π has a cycle at N. It will then follow that, from any initial state $x \in \Omega_{r^\pi}^\pi$, the system state will, in finite time, *either* evolve to $\bigcup_{v \le r^\pi - 1} \Omega_v^\pi$ *or* will be trapped in $D'(N)$. Should π not have a cycle at N then evolution to $\bigcup_{v \le r^\pi - 1} \Omega_v^\pi$ in finite time is guaranteed. From this point, the argument which yields the conclusion that there exists a monotone policy x^{D} with cost rate for $P(W)$ no higher than that under π, uniformly over initial states, is a minor perturbation of that which concludes the consideration of case 1. This now concludes the consideration of policies $\pi \in S$ for descendant project P for which $\pi(0) = b$.

When $\pi(0) = a$, we obtain the values x_r^π, y_r^π, $1 \le r \le r^\pi$, by setting $x_1^\pi = 0$ and thereafter proceeding according to the algorithm above (8). The sets Ω_r^π, $1 \le r \le r^\pi$, are as defined there. This case differs from the preceding setup in that now all actions taken by π in Ω_1^π are a. However, this means that, should the system state ever enter Ω_1^π then absorption in state 0 is guaranteed, as would be the case from any state under monotone policy 0^{D}. With this observation, a modest adjustment to the preceding discussion also yields the conclusion that there exists a monotone policy x^{D} with cost rate no higher than π uniformly over initial system states. This concludes the proof of Proposition 1(i).

3.2. Sufficient conditions for the (strict) indexability of P and the related indices

In light of Proposition 1, the minimisation in (7) which defines problem $P(W)$ can now be re-expressed as

$$C(W) = \min_{0 \le x \le N+1} (C^{x^{\mathrm{D}}} + W R^{x^{\mathrm{D}}}) \tag{10}$$

in the descendant case and by

$$C(W) = \min_{-1 \le x \le N} (C^{x^{\mathrm{A}}} + W R^{x^{\mathrm{A}}})$$

in the ascendant case. We continue to focus on the descendant case. As is plain from the above discussion, the system state evolving under policy x^{D} becomes trapped in finite time in the set $D(x)$ when $x \in \{0, 1, \dots, N\}$ and in the set $D'(N)$ when $x = N+1$. Consider some fixed x in the range $\{1, \dots, N\}$. Suppose that $X(0) = x$ and that policy x^{D} is applied. Use τ_x^{D} to denote the time of the first return of the system state to x, namely,

$$\tau_x^{\mathrm{D}} := \min\{t; t \in \mathbb{Z}^+,\ X(t) = x,\ \text{and there exists } s < t \text{ with } X(s) \ne x\}.$$

During $[0, \tau_x^{\mathrm{D}})$ there is an initial sojourn in state x of expected duration $(1 - p^a(x,x))^{-1}$, during which resource is consumed at rate $R(x)$ as the active action is applied and costs are incurred at rate $C^a(x)$. This is followed by departure from x and sojourns in states $y \in D(x) \setminus \{x\}$ in which x^{D} takes the passive action and no resource is consumed. It will be convenient to write $m^{\mathrm{D}}(x) = \mathbb{E}\{\tau_x^{\mathrm{D}}\}$ and $c^{\mathrm{D}}(x)$ for the total expected cost incurred by P (excluding charges levied for resource consumption) during $[0, \tau_x^{\mathrm{D}})$. It follows from the above discussion that

$$m^{\mathrm{D}}(x) \ge 1 + (1 - p^a(x,x))^{-1}$$

and $c^{\mathrm{D}}(x) \geq C^a(x)(1 - p^a(x,x))^{-1}$. Standard theory yields

$$C^{x^{\mathrm{D}}} = c^{\mathrm{D}}(x)\{m^{\mathrm{D}}(x)\}^{-1}, \quad R^{x^{\mathrm{D}}} = R(x)\{m^{\mathrm{D}}(x)(1 - p^a(x,x))\}^{-1}, \quad 1 \leq x \leq N. \tag{11}$$

The special features of the system evolving under 0^{D} and $(N+1)^{\mathrm{D}}$ yield the expressions

$$C^{0^{\mathrm{D}}} = C^a(0), \qquad R^{0^{\mathrm{D}}} = R(0),$$

and

$$C^{(N+1)^{\mathrm{D}}} = c^{\mathrm{D}}(N+1)\{m^{\mathrm{D}}(N+1)\}^{-1}, \qquad R^{(N+1)^{\mathrm{D}}} = 0; \tag{12}$$

the new quantities in (12) having their natural meanings. We are now ready to state our main result regarding indexability.

Theorem 1. (i) *Descendant project P is indexable if $R^{x^{\mathrm{D}}}$ is decreasing in x over the range $0 \leq x \leq N+1$.*

(ii) *Descendant project P is strictly indexable if $R^{x^{\mathrm{D}}}$ is decreasing in x over the range $0 \leq x \leq N+1$ and*

$$W^{\mathrm{D}}(x) := \frac{C^{(x+1)^{\mathrm{D}}} - C^{x^{\mathrm{D}}}}{R^{x^{\mathrm{D}}} - R^{(x+1)^{\mathrm{D}}}} \tag{13}$$

is finite and strictly increasing in x over the range $0 \leq x \leq N$. Under these conditions, $W^{\mathrm{D}}\colon \Omega \to \mathbb{R}$ is the index function.

Proof. Write x_W for the largest x-value achieving the minimum in (10). Plainly, $C(W)$ is increasing, continuous, piecewise linear, and concave in W with right W-gradient equal to $R^{x_W^{\mathrm{D}}}$ everywhere. Concavity implies that $R^{x_W^{\mathrm{D}}}$ is decreasing in W. The hypothesis of part (i) then implies that x_W must be increasing in W. Hence, the collection $\{x_W^{\mathrm{D}},\ W \in \mathbb{R}\}$ constitutes a family of optimal policies whose passive sets $b(x_W^{\mathrm{D}}) = \{0, 1, \ldots, x_W - 1\}$ are increasing in W. This concludes the proof of part (i). Under the hypotheses of part (ii), it is easy to see that the minimum in (10) is achieved by x if and only if $W^{\mathrm{D}}(x-1) \leq W \leq W^{\mathrm{D}}(x)$, $0 \leq x \leq N+1$, where we take $W^{\mathrm{D}}(-1) = -\infty$ and $W^{\mathrm{D}}(N+1) = \infty$. Hence, if we take

$$\pi(W) = x^{\mathrm{D}} \quad \Longleftrightarrow \quad W^{\mathrm{D}}(x-1) < W \leq W^{\mathrm{D}}(x), \qquad 0 \leq x \leq N+1,$$

then $\{\pi(W),\ W \in \mathbb{R}\}$ is a collection of optimal policies for $P(W)$ whose related index function is W^{D}. The latter is strictly increasing and, hence, one-to-one. Hence, P is strictly indexable. This concludes the proof.

In the strictly indexable cases of Theorem 1(ii), W^{D} is the index function. The index in (13) may be thought of as a marginal rate at which additional costs are incurred by being passive rather than active in state x per unit of resource saved thereby. Niño-Mora [13] referred to such quantities as *marginal productivity indices*. In the more general indexable cases of part (i), index values coincide with the W-values at which points of discontinuity in the right-gradient of $C(W)$ occur. These are easily obtained from the following algorithm.

Algorithm 1. *Step 1.* Calculate

$$W_1 = \min_{0 < x \leq N+1} \frac{C^{x^{\mathrm{D}}} - C^{0^{\mathrm{D}}}}{R^{0^{\mathrm{D}}} - R^{x^{\mathrm{D}}}}, \tag{14}$$

and write x_1 for the largest minimiser. If $x_1 = N+1$, stop.

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

Step k. From the first $k-1$ steps we have W-values $W_1 < W_2 < \cdots < W_{k-1}$ and minimising x-values $x_1 < x_2 < \cdots < x_{k-1}$. If $x_{k-1} = N+1$, stop. Otherwise, calculate

$$W_k = \min_{x_{k-1} < x \le N+1} \frac{C^{x^{\mathrm{D}}} - C^{x^{\mathrm{D}}_{k-1}}}{R^{x^{\mathrm{D}}_{k-1}} - R^{x^{\mathrm{D}}}} \tag{15}$$

and write x_k for the largest minimiser. Proceed to step $k+1$ and continue. Define $L \in \mathbb{Z}^+$ from the relation $x_L = N+1$.

It is straightforward to verify that the optimal policies $\{x^{\mathrm{D}}_W,\ W \in \mathbb{R}\}$ yield the related index function

$$W(x) = \sum_{k=1}^{L} W_k I(x_{k-1} \le x < x_k), \tag{16}$$

where we set $x_0 = 0$.

It is trivial to modify Theorem 1 and the content following to the case of ascendant assets. For completeness, we now give the version of the main result which applies.

Theorem 2. (i) *Ascendant project P is indexable if $R^{x^{\mathrm{A}}}$ is increasing in x over the range $-1 \le x \le N$.*

(ii) *Ascendant project P is strictly indexable if $R^{x^{\mathrm{A}}}$ is increasing in x over the range $-1 \le x \le N$ and*

$$W^{\mathrm{A}}(x) := \frac{C^{(x-1)^{\mathrm{A}}} - C^{x^{\mathrm{A}}}}{R^{x^{\mathrm{A}}} - R^{(x-1)^{\mathrm{A}}}}$$

is finite and strictly decreasing in x over the range $0 \le x \le N$. Under these conditions, $W^{\mathrm{A}}\colon \Omega \to \mathbb{R}$ is the index function.

The sufficient condition for indexability in Theorem 1(i) has a natural interpretation. As x increases, so the policy x^{D} takes the active (remedial) action in fewer states. As this happens, the condition requires that the rate at which the resource is consumed under the policy decreases. While this condition seems intuitive, it is certainly not universal. Indeed, we know that indexability is not guaranteed, even for Whittle's restless bandits in which resource consumption is constant over projects and states. See, for example, Whittle [17]. In the case of our more complex projects, recall from (11) that $R^{x^{\mathrm{D}}} = R(x)\{m^{\mathrm{D}}(x)(1 - p^a(x,x))\}^{-1}$ and, hence, the requirement that $R^{x^{\mathrm{D}}}$ be decreasing in x plainly makes demands on both the project's resource consumption function R and its stochastic structure p^a, p^b. In Section 3.3 below we give examples of indexable descendant projects with state space $\mathbb{N}$ in which it is natural to assume that the resource consumption function R is nondecreasing. In Example 1, it is enough for indexability that R also be concave while in Example 2, the requirement is that the increasing function R should meet another condition (19) which also involves the project's stochastic structure. The index $W^{\mathrm{D}}(x)$ of Theorem 1(ii) may be understood as a *fair charge* (per unit of time and per unit of resource) for the application of the resource (deployment of the active action) in state x. When $W = W^{\mathrm{D}}(x)$, both actions a and b are optimal for $P(W)$ in state x.

Extension of the above material to descendant projects with countable state space $\mathbb{N}$ is usually easily achieved. Obvious modifications are made to Definitions 1 and 2. Consider any stationary, deterministic, and Markov policy π for a descendant project P with state space $\mathbb{N}$ and its active set

$$a(\pi) := \{x; \pi(x) = a\}.$$

If $a(\pi)$ is unbounded then, for any initial state x, there exists a smallest member of $a(\pi)$, x_π say, which exceeds x. The state process for project P evolving from x under π can never leave the finite set $\{0, 1, \ldots, x_\pi\}$. We argue as in the proof of Proposition 1 that there must exist some monotone policy y^{D}, say, whose cost rate from x is no greater than that of π. Should $a(\pi)$ be bounded, write N_π for its largest member. Evolution of the state process of P under π from any initial state in the finite set $\{0, 1, \ldots, N_\pi\}$ is covered by the same argument. Should P's evolution from an initial state $x \geq N_\pi + 1$ under policy π be such that entry into the set $\{0, 1, \ldots, N_\pi\}$ in finite time is *either* almost sure *or* has probability zero, then the conclusion that there exists a monotone policy y^{D} with no larger cost rate from x persists. In the latter case, this monotone policy could be ∞^{D} which takes the passive action in all states. Consideration of indexability/index functions then involves obvious modification to Theorem 1 and the material following its statement and proof. Note that, in applying an appropriate version of the algorithm around (14)–(16) to a project with state space $\mathbb{N}$, there will be *either* a finite number of distinct index values ($L < \infty$) *or* a countably infinite number.

3.3. Examples of indexable projects

Example 1. (*Machine maintenance.*) Consider descendant project P with countable state $\mathbb{N}$ and stochastic dynamics given by

$$p^a(x, 0) = 1 - p = 1 - p^a(x, 1), \qquad x \in \mathbb{N},$$

and

$$p^b(x, x) = 1 - p = 1 - p^b(x, x + 1), \qquad x \in \mathbb{N}.$$

In words, a machine is subject to gradual deterioration in performance at constant rate. Larger states correspond to higher degrees of wear. In the absence of any maintenance intervention (i.e. under the passive action) the machine spends an expected time p^{-1} in each state x before moving to state $x + 1$. Under the active action, the machine state is changed instantaneously to 0 (the pristine state) after which a single passive deterioration step takes place. Note that this setup is covered by the above comments on countable state projects. We take $C^a \equiv 0$ and, hence, the only costs incurred in taking the active action concern the charge for maintenance resource. We assume that C^b is increasing with $C^b(0) = 0$.

Plainly, $R^{x^{\mathrm{D}}} = px^{-1}R(x)$ and, hence, from (a modest extension to) Theorem 1, P is indexable when the resource function R is increasing and concave. Furthermore,

$$C^{x^{\mathrm{D}}} = x^{-1} \sum_{y=0}^{x-1} C^b(y),$$

from which we infer that we have strict indexability when the index function

$$W^{\mathrm{D}}(x) = \frac{C^{(x+1)^{\mathrm{D}}} - C^{x^{\mathrm{D}}}}{R^{x^{\mathrm{D}}} - R^{(x+1)^{\mathrm{D}}}} = \frac{x\{\sum_{y=0}^{x} C^b(y)\} - (x+1)\{\sum_{y=0}^{x-1} C^b(y)\}}{p\{(x+1)R(x) - xR(x+1)\}}, \qquad x \in \mathbb{Z}^+, \quad (17)$$

is strictly increasing. Observe that the active action is suboptimal in state 0 and that the implied index value is $W^{\mathrm{D}}(0) = -\infty$. In the affine resource function case $R(x) = R_1 + R_2 x$, $R_1 > 0$ and $R_2 > 0$, W^{D} is strictly increasing if C^b is. More generally, if we extend C^b, R to smooth functions from $\mathbb{R}^+$ to itself then it is straightforward to show that W^{D} in (17) will be strictly increasing if there exists such an extension for which

$$\frac{C^{b'}(x)}{R'(x-1) - R'(x)} \qquad (18)$$

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

is strictly increasing over $x \geq 1$, where a prime denotes differentiation. Strictly indexable examples with index given by (17) include

(RC1) $R(x) = R_1 + R_2 \ln(x+1)$; C^b increasing and convex;

(RC2) $R(x) = R_1 + R_2 \ln(x+1)$; $C^b(x) = C_1 x^\alpha$, $\alpha \in [0, 1)$;

(RC3) $R(x) = (R_1 + R_2) - R_2 \exp(-\theta x)$; C^b increasing and convex;

(RC4) $R(x) = (R_1 + R_2) - R_2 \exp(-\theta x)$; $C^b(x) = C_1\{1 - \exp(-\eta x)\}$, $\eta < \theta$;

(RC5) $R(x) = R_1 x^\alpha$, $\alpha \in (0, 1)$; $C^b(x) = C_1 x^\beta$, $\alpha < \beta$.

In the above examples, R_1, R_2, and C_1 are positive constants. The numerical study in Section 3.4 below considers index policies for problems in which there are between 10 and 140 machines to be maintained. In the study, the specification of each machine is as above, with the resource and cost functions as in (RC3).

Example 2. (*Machine maintenance.*) We develop a machine maintenance model in which costs are incurred at (catastrophic) machine breakdowns. Any such breakdown is deemed to be followed by renewal of the machine at some high cost (K). Breakdowns can be preempted by maintenance actions in the form of a planned renewal of the machine which consumes resource $R(x)$ when undertaken in machine state x. As in the previous example, large machine states correspond to more advanced age/wear. Stochastic dynamics for descendant project P are given by

$$p^a(x, 0) = 1, \qquad x \in \mathbb{N},$$

and

$$p^b(x, 0) = p(x) = 1 - p^b(x, x+1), \qquad x \in \mathbb{N},$$

where the sequence $\{p(x),\ x \in \mathbb{N}\}$ of breakdown probabilities is assumed nondecreasing. Direct computation yields

$$m^{\mathrm{D}}(x) = \left(\sum_{y=0}^{x}\left[\prod_{z=0}^{y-1}\{1 - p(z)\}\right]\right)\left[\prod_{y=0}^{x-1}\{1 - p(y)\}\right]^{-1}, \qquad x \in \mathbb{N}.$$

Since we have

$$R^{x^{\mathrm{D}}} = R(x)\{m^{\mathrm{D}}(x)\}^{-1},$$

it is straightforward to establish that the sequence $\{R^{x^{\mathrm{D}}},\ x \in \mathbb{N}\}$ will be decreasing and P will be indexable whenever $\{R(x),\ x \in \mathbb{N}\}$ is increasing such that

$$R(x)\{1 - p(x)\}^{-1} \geq R(x+1), \qquad x \in \mathbb{N}. \tag{19}$$

However, (19) is achieved if and only if there exists a sequence $\{\delta(x),\ x \in \mathbb{N}\}$ with $\delta(x) \in [0, 1]$ for all x and

$$R(x+1) = \{1 - \delta(x)p(x)\}\{1 - p(x)\}^{-1}R(x), \qquad x \in \mathbb{N}. \tag{20}$$

Furthermore, we have

$$C^{x^{\mathrm{D}}} = K\left(\sum_{y=0}^{x-1} p(y)\left[\prod_{z=0}^{y-1}\{1 - p(z)\}\right]\right)\left(\sum_{y=0}^{x-1}\left[\prod_{z=0}^{y-1}\{1 - p(z)\}\right]\right)^{-1}, \qquad x \in \mathbb{Z}^+.$$

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

If we suppose that the active action is disallowed in state 0 then strict indexability will be achieved when

$$W^{\mathrm{D}}(x) := \frac{C^{(x+1)^{\mathrm{D}}} - C^{x^{\mathrm{D}}}}{R^{x^{\mathrm{D}}} - R^{(x+1)^{\mathrm{D}}}}$$

is strictly increasing over $\mathbb{Z}^+$. Straightforward calculation yields the conclusion that this will be achieved when the δ sequence above satisfies

$$\delta(x)p(x)\{1 - p(x)\}\{1 - \delta(x)p(x)\}^{-1} \geq \delta(x+1)p(x+1), \qquad x \in \mathbb{N}. \tag{21}$$

If we specialise to the case $p(x) = p,\ x \in \mathbb{N}$, then we see that any sequence $\{\delta(x),\ x \in \mathbb{N}\}$ of values in [0, 1] yields indexability in (20). However, (21) implies that *either* $\delta(x) = 1,\ x \in \mathbb{N}$, or that $\delta(x^*) < 1$ for some $x^* \in \mathbb{N}$ and that, for $x \geq x^*$, $\delta(x)$ decreases to zero faster than the log-concave sequence obtained by enforcing equality in (21) over this range.

Example 3. (*Asset management.*) Consider ascendant project P with finite state space $\{0, 1, \ldots, N\}$ which models a reward earning/cost incurring asset whose reward/cost performance is strongest for high states. The asset deteriorates under the passive action, but this deterioration can be arrested by an active intervention. We shall suppose that, under the active action a applied in any state x, resource of fixed magnitude R leads to an upward movement in the asset state of random size X. All such upward transitions are truncated at the maximal state N and so strong improvements to the project are less easily won for high states. We suppose that X has probability mass function (PMF) $\{p_y,\ y \in \mathbb{Z}^+\}$. Active stochastic dynamics are given by

$$p^a(x, x+y) = p_y I(1 \leq y < N - x) + \bar{F}_{N-x} I(y = N - x), \tag{22}$$

where $\bar{F}_n = \sum_{y \geq n} p_y$. Passive dynamics model gradual deterioration of the asset at constant rate and are given by

$$p^b(x, x) = 1 - p = 1 - p^b(x, x-1), \qquad 1 \leq x \leq N,$$

and

$$p^b(0, 0) = 1.$$

It is straightforward that $R^{(-1)^{\mathrm{A}}} = 0$ and

$$R^{x^{\mathrm{A}}} = pR\{\bar{F}(x)\}^{-1}, \tag{23}$$

where

$$\bar{F}(x) = \sum_{y=1}^{N-x} \bar{F}_y.$$

It is straightforward that $R^{x^{\mathrm{A}}}$ is increasing in x over the range $-1 \leq x \leq N$ and, hence, from Theorem 2, that P is indexable. We take $C^a \equiv 0$, and assume that C^b is decreasing (equivalently, returns are increasing in the state). We then have $C^{(-1)^{\mathrm{A}}} = C^b(0)$ and

$$C^{x^{\mathrm{A}}} = \left\{\sum_{y=1}^{N-x} C^b(x+y)\bar{F}_y\right\}\{\bar{F}(x)\}^{-1}, \qquad 0 \leq x \leq N. \tag{24}$$

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

From (22)–(24) we conclude that

$$W^{\mathrm{A}}(0) = \left[\sum_{y=1}^{N}\{C^b(0) - C^b(y)\}\bar{F}_y\right](pR)^{-1}$$

and

$$W^{\mathrm{A}}(x) = \left[\sum_{y=1}^{N-x}\{C^b(x) - C^b(x+y)\}\{p_y\bar{F}(x) + \bar{F}_y\bar{F}_{N-x+1}\}\right] \times (pR\bar{F}_{N-x+1})^{-1}, \qquad 1 \le x \le N. \tag{25}$$

An additional assumption that C^b is strictly convex guarantees that $W^{\mathrm{A}}(x)$ is strictly decreasing over the range $1 \le x \le N$. This then yields from Theorem 2 an appropriate form of strict indexability if the active action is mandated for the project in state 0. Project indices are given by (25). To extend strict indexability to the range $0 \le x \le N$ requires an additional stipulation that $C^b(0) > \bar{C}$, where $\bar{C}$ is chosen to enforce $W^{\mathrm{A}}(0) = W^{\mathrm{A}}(1)$.

Further generalisation and simplification is possible when we consider the limit $N \to \infty$ in the above such that active stochastic dynamics become

$$p^a(x, x+y) = p_y, \qquad y \in \mathbb{N},$$

while passive dynamics are unchanged. We consider a general resource function $R\colon \mathbb{N} \to \mathbb{R}^+$, and now have $R^{(-1)^{\mathrm{A}}} = 0$ and

$$R^{x^{\mathrm{A}}} = pR(x)(\mathbb{E}\{X\})^{-1}, \qquad x \in \mathbb{N}.$$

Indexability is now guaranteed for any increasing R. Direct computation yields

$$W^{\mathrm{A}}(x) = [C^b(x) - \mathbb{E}\{C^b(x+X)\}][p\{R(x) - R(x-1)\}]^{-1}, \qquad x \in \mathbb{Z}^+,$$

which will be strictly increasing whenever C^b is strictly decreasing and strictly convex and R is strictly increasing and strictly convex. Under these conditions, we infer strict indexability when the active action is mandated in state 0. As before, to extend strict indexability to $\mathbb{N}$, we need to impose a lower bound on $C^b(0)$.

3.4. Numerical study

We now briefly present some numerical evidence in illustration of the power and efficacy of indices as building blocks for the construction of strongly performing heuristics. We shall take the machine maintenance problems given above as Example 1 and consider scenarios which involve the maintenance of collections of machines which number between $J = 10$ and $J = 140$. All such examples are comfortably beyond the realm where the calculation of optimal policies by standard DP methods is feasible. Each of our machines will share a common deterioration rate p, and will have resource consumption function and cost function as given in (RC3). To model the resource required to maintain machine j, we make the parameter choices $R_{1j} = 1 - \exp(-z_j/10)$, $R_{2j} = 1 - R_{1j}$, and $\theta_j = 0.1$, while the cost rate under the passive action in state x is given by $C_j^b(x) = y_j x^2$, $1 \le j \le J$, where the machine specific constants y_j and z_j are now specified.

We shall make the choice $z_j = 20j/J$, and so, as j increases, the machines become increasingly resource intensive to maintain. We choose the cost parameters y_j to reflect

the likely corresponding importance of the more demanding machines in cost terms and set $y_j = 2+Yj/J$ for some constant Y. Hence, machine costs in state x vary from $2x^2$ to $(2+Y)x^2$. In our study we consider the performance of five heuristic policies for machine maintenance for a range of Y, J, and p.

The first two of our heuristic policies are the first of the index heuristics described in Remark 4, namely the greedy heuristics GI and WGI. Note that, for problems of the size considered here, the knapsack approach to the construction of index heuristics becomes computationally intractable, notwithstanding the ready availability of the indices themselves. The heuristic STATE allocates resource to machines according to their current state value x_j, in descending order, while the heuristic MYOP uses the current cost rate $C_j^b(x_j)$ similarly. If all machines were identical, these four heuristics would coincide. In addition to these, a decentralised round-robin (RR) heuristic is also considered. At each timestep, a batch of consecutively numbered machines (modulo J) is chosen for maintenance by the RR heuristic, the batch being determined sequentially until the next candidate machine cannot be included for lack of resource. At the next timestep, this is the first machine to be included in the following batch.

In Tables 1–3 we present the time average costs incurred when these five heuristics are applied in 36 different problem scenarios. These costs are estimated by Monte Carlo simulation. We provide these results as snapshots which illustrate a generically coherent picture across a much wider range of parameter choices than those actually presented. In each of the tables the final column headed LLB (Lagrangian lower bound) is a lower bound on the optimal cost rate achievable which is obtained from the Lagrangian relaxation in (5). If we write $C^{\text{opt}}(W)$ for the value of the Lagrangian relaxation then

$$\text{LLB} := \max_{W \in \mathbb{R}^+} C^{\text{opt}}(W)$$

is a lower bound on the optimal cost rate C^{opt}. It is not difficult to show that LLB is equal to $\bar{C}^{\text{opt}}$, the value of the relaxation in (4) which enforces average resource usage.

As the number of machines in each problem set grows from 10 to 140, the resource R available also grows, roughly in proportion. Note that in the range $10 \le J \le 30$, in which the number of machines is relatively small, the heuristic GI can make relatively poor use of the

TABLE 1: Cost rate estimates for five heuristics and the LLB for machine maintenance problems with $J = 10$ to $J = 140$, $p = 0.9$, and $Y = 10$. See the text for details.

J	R	GI	WGI	STATE	MYOP	RR	LLB
10	2	144.71	134.25	143.58	137.91	194.80	99.26
20	3	619.79	614.78	639.89	638.03	890.17	544.87
30	5	664.89	661.63	676.31	729.82	885.92	623.32
40	7	773.38	772.90	796.86	848.66	932.94	742.90
50	9	892.23	894.49	931.07	972.29	1030.94	865.72
60	11	1014.90	1019.26	1064.37	1099.34	1186.77	989.87
70	13	1137.97	1144.86	1196.26	1227.49	1328.76	1114.78
80	14	1548.87	1554.82	1603.86	1716.94	1765.82	1523.87
90	16	1668.44	1677.48	1739.66	1840.46	1867.50	1645.39
100	18	1790.11	1801.48	1874.20	1965.57	2009.84	1768.15
120	21	2327.41	2339.91	2413.84	2590.05	2577.05	2304.80
140	24	2869.43	2883.80	2950.35	3222.35	3242.43	2846.81

TABLE 2: Cost rate estimates for five heuristics and the LLB for machine maintenance problems with $J = 10$ to $J = 140$, $p = 0.9$, and $Y = 14$. See the text for details.

J	R	GI	WGI	STATE	MYOP	RR	LLB
10	2	185.84	172.40	186.33	179.02	248.67	127.93
20	3	794.44	789.52	828.85	819.14	1141.16	696.58
30	5	853.16	852.30	874.28	945.07	1136.49	800.05
40	7	990.07	990.94	1032.79	1092.63	1197.49	951.19
50	9	1141.59	1142.94	1208.28	1248.61	1323.55	1107.25
60	11	1297.40	1300.03	1381.74	1410.91	1524.30	1265.67
70	13	1455.36	1459.23	1553.30	1574.43	1707.07	1425.43
80	14	1984.45	1990.32	2081.22	2209.06	2268.54	1952.46
90	16	2136.96	2142.80	2258.13	2364.93	2399.10	2106.82
100	18	2292.08	2298.74	2433.58	2523.91	2582.55	2262.94
120	21	2983.10	2993.12	3132.86	3331.55	3311.18	2953.74
140	24	3682.46	3701.64	3826.38	4151.06	4166.80	3652.82

TABLE 3: Cost rate estimates for five heuristics and the LLB for machine maintenance problems with $J = 10$ to $J = 140$, $p = 0.8$, and $Y = 10$. See the text for details.

J	R	GI	WGI	STATE	MYOP	RR	LLB
10	2	104.11	97.11	107.48	98.04	162.20	74.64
20	3	448.56	440.61	462.87	463.63	683.29	389.37
30	5	480.52	477.03	500.51	510.00	691.72	442.83
40	7	542.79	541.26	565.48	579.25	704.78	504.20
50	9	615.69	615.29	634.60	667.12	830.92	567.52
60	11	695.27	696.33	710.60	764.67	953.02	658.42
70	13	779.50	781.89	792.25	866.77	1068.85	751.64
80	14	1073.55	1079.80	1116.28	1156.87	1384.94	1042.55
90	16	1141.96	1149.22	1178.47	1244.44	1496.57	1104.99
100	18	1216.55	1223.49	1245.79	1339.55	1617.68	1167.63
120	21	1607.16	1621.36	1668.94	1734.13	2047.21	1580.85
140	24	2018.55	2039.64	2113.45	2160.73	2520.09	1996.53

available resource and is generally outperformed by the WGI heuristic. It is also true that, for these relatively small values of J and R, the relaxation which yields LLB is likely to admit a range of unfairly better policies which, in achieving an average resource usage of R, frequently deploy actions which are inadmissible for the original problem. Notwithstanding this relative looseness in the bound for small problems, it is already clear that, for the $J = 30$, $R = 5$ case, the index heuristics are unlikely to be far from optimal. As the value of J increases above 50, say, the propensity of GI to make poor usage of the available resource diminishes in importance and it begins to outperform WGI. It is also true that as J and R grow, the bound LLB becomes tighter such that we observe progressively smaller percentage excesses of GI over LLB with the associated suboptimalities of GI guaranteed to be no larger than around 1% when $J = 140$, $R = 24$. In all cases, whichever is the appropriate choice of index heuristic, it clearly outperforms STATE, MYOP, and RR. The results in Tables 1–3 show very clearly the value of indices in the construction of good policies.

4. Abandoning skip-free dynamics: applications to inventory control

We now abandon the assumption, omnipresent in Section 3, of skip-free dynamics (right or left, as appropriate) under the passive action for project $P = \{\Omega, p^a, p^b, C^a, C^b, R\}$. In this section we shall rely more heavily on project P's cost structure to guarantee the optimality of monotone policies for $P(W)$. Our primary envisaged application area for this material is inventory control and our terminology and some of the notation will reflect that. The multi-project problem may be thought of as one concerning the cost-effective management of a network of stock-holding locations by a central controller. It will, for example, be convenient to write the state space of P as $\Omega = \{0, 1, \dots, M + N\}$, where $M, N \in \mathbb{Z}^+$. In an inventory context, positive integer N is a maximum stock level, while M is a maximum allowed number of backordered customers. The project state may be understood as the inventory level augmented by M, where a negative stock level represents a number of backordered customers.

Further project details are as follows. The active action a applied in state x returns the project state instantaneously to the pristine (fully replenished) state $M + N$, from where it moves to state $(M + N - D)^+$, where D is a positive integer-valued random variable with support $\mathbb{N}$ and PMF $\{p_y,\ y \in \mathbb{N}\}$. When the passive action is applied in state x, we observe a state transition $x \to (x - D)^+$. Hence, active dynamics for P are given by

$$p^a(x, y) = p_{M+N-y} I(1 \le y \le M + N) + \bar{F}_{M+N} I(y = 0), \qquad (x, y) \in \Omega^2, \tag{26}$$

while passive dynamics are given by

$$p^b(x, y) = p_{x-y} I(1 \le y \le x) + \bar{F}_x I(y = 0), \qquad (x, y) \in \Omega^2, \tag{27}$$

where $\bar{F}_x = \sum_{y \ge x} p_y$.

In an inventory example, D is the demand for the stock during a single time unit (which we shall take without loss of generality to be one day). We shall write $\mathbb{E}\{D\} = \lambda$, assumed strictly positive, and call λ the *demand rate*. State x represents stock level $x - M$, observed at the start of a day. Under active action a, stock is replenished instantaneously (to its maximum level N) and must then meet that day's demand. The consequential one-step transition probabilities are given in (26). Under the passive action b, there is no replenishment, the stock level is depleted by one day's demand and the one-step transition probabilities are given in (27).

The stochastic dynamics in (26) and (27) model any project P whose deterioration under the passive action is a random process with stationary and independent positive integer-valued increments, while the active action uses resource to restore P to some pristine state. *Inter alia*, this also forms the basis for a natural model of machine maintenance problems with passive dynamics which are not right skip-free.

Project costs are modelled as follows. The active action taken in state x incurs a direct cost $\bar{C}(x)$, say. There are additional costs associated with state deterioration. The latter are determined by the decreasing function $\Phi\colon (-\infty, M + N] \cap \mathbb{Z} \to \mathbb{R}^+$. Expected costs incurred when actions a and b are respectively taken in state x are given by

$$C^a(x) = \bar{C}(x) + \mathbb{E}\{\Phi(M + N - D) - \Phi(M + N)\}, \qquad x \in \Omega,$$

and

$$C^b(x) = \mathbb{E}\{\Phi(x - D) - \Phi(x)\}, \qquad x \in \Omega. \tag{28}$$

In inventory examples, we set $\Phi(M + N) = 0$ and $\Phi(x)$ is interpreted as the total inventory cost incurred in a single cycle of the system (i.e. between successive replenishments) which reaches

state x. If Φ is assumed to be convex then $C^b(x)$ in (28) is decreasing in x, as is appropriate in many applications, including machine maintenance. Finally, $R(x)$ is the resource consumed when the active action is taken in state x. It is natural to suppose that both R and $\bar{C}$ are decreasing.

We shall suppose until further notice that Condition 1 below is satisfied. As we shall see later, for some important examples, the need to meet Condition 1 implies that the range of values of the resource charge W must be constrained. In (29) below we use the notation $x^+ = \max(x, 0)$.

Condition 1. *The function* $\Psi_W : (-\infty, M+N] \cap \mathbb{Z} \to \mathbb{R}^+$ *given by*

$$\Psi_W(x) = \bar{C}(x^+) + WR(x^+) + \Phi(x) \tag{29}$$

is decreasing and convex.

4.1. Monotone policies are optimal for $P(W)$ under Condition 1

We now consider optimisation problem $P(W)$, which is as in (7). We seek a policy to minimise the rate at which an aggregate of replenishment costs, inventory costs, and resource costs are incurred over an infinite horizon from any initial state.

Lemma 2. *Under Condition 1, there exists some monotone policy* x^{A} *which is optimal for* $P(W)$.

Proof. We consider two cases.

Case 1: the passive action is optimal for $P(W)$ *in state 0.* Under the above assumptions concerning stochastic structure, entry into state 0 is guaranteed in finite time under all policies. If $\pi(0) = b$ then, under π, state 0 is absorbing. It follows that any such policy has the same cost rate as $(-1)^{\mathrm{A}}$. Hence, in the case the monotone policy $(-1)^{\mathrm{A}}$ must be optimal.

Case 2: the active action is strictly optimal for $P(W)$ *in state 0.* In such cases we can optimise over S_0, the class of policies which mandate the active replenishment action in 0. Furthermore, it is straightforward to establish that, for any policy in S_0, entry into state 0 in finite time is guaranteed, irrespective of the initial state. Hence, we do not need to be concerned about dependence on the initial state in (7).

Write $V_t(x)$ for the minimum cost incurred by a member of S_0 over a t-day horizon from initial state x. From the above project details, the DP recursion for V_t gives

$$V_t(0) = \bar{C}(0) + WR(0) + \mathbb{E}\{\Phi(M+N-D) - \Phi(M+N) + V_{t-1}((M+N-D)^+)\}$$

and

$$\begin{aligned} V_t(x) = \min[&\bar{C}(x) + WR(x) \\ &+ \mathbb{E}\{\Phi(M+N-D) - \Phi(M+N) + V_{t-1}((M+N-D)^+)\}; \\ &\mathbb{E}\{\Phi(x-D) - \Phi(x) + V_{t-1}((x-D)^+)\}], \qquad 1 \le x \le M+N. \end{aligned} \tag{30}$$

Note that the expectations in (30) are taken with respect to the distribution of the daily demand D. We now write

$$\bar{V}_t(x) = V_t(x) - \bar{C}(x) - WR(x), \qquad 0 \le x \le M+N,\ t \in \mathbb{N}.$$

This derived quantity has an interpretation as the minimum cost (over S_0) over a t-day period from state x when replenishment costs are accounted for incrementally as demand is met rather

than when the active action is taken. From (29) and (30), we have

$$\begin{aligned}\bar{V}_t(0) &= \mathbb{E}\{\Psi_W(M+N-D) - \Psi_W(M+N) + \bar{V}_{t-1}((M+N-D)^+)\},\\ \bar{V}_t(x) &= \min[\bar{V}_t(0); \mathbb{E}\{\Psi_W(x-D) - \Psi_W(x) + \bar{V}_{t-1}((x-D)^+)\}]\end{aligned} \tag{31}$$

for $1 \le x \le M+N$. From Condition 1, it follows that the expectation

$$\mathbb{E}\{\Psi_W(x-D) - \Psi_W(x)\}$$

is positive and decreasing in x. Setting $\bar{V}_0 \equiv 0$ (i.e. $V_0 \equiv \bar{C} + WR$), it then follows from (31) and an induction argument that $\bar{V}_t(x)$ is decreasing in x for all $t \in \mathbb{N}$. Standard arguments from reward-renewal theory guarantee that the bias function given by the limit

$$v(x) := \lim_{t\to\infty}\{\bar{V}_t(x) - \bar{V}_t(M+N)\}, \qquad 0 \le x \le M+N,$$

exists and is finite. The decreasing nature of $\bar{V}_t(x)$ for all t implies that $v(x)$ must also be decreasing in x. From (31) we express the DP recursion for the average cost problem $P(W)$ as

$$\begin{aligned}C(W) + v(0) &= \mathbb{E}\{\Psi_W(M+N-D) - \Psi_W(M+N) + v((M+N-D)^+)\},\\ C(W) + v(x) &= \min[C(W) + v(0); \mathbb{E}\{\Psi_W(x-D) - \Psi_W(x) + v((x-D)^+)\}]\end{aligned} \tag{32}$$

for $1 \le x \le M+N$. In (32), $C(W)$ is the optimal average cost (including resource costs) for $P(W)$. From (32), if the active action is optimal in state x then

$$\mathbb{E}\{\Psi_W(x-D) - \Psi_W(x) + v((x-D)^+)\} \ge C(W) + v(0),$$

from which we infer that

$$\mathbb{E}\{\Psi_W(y-D) - \Psi_W(y) + v((y-D)^+)\} \ge C(W) + v(0), \qquad 1 \le y \le x,$$

and, hence, that the active action is optimal in all states $y \le x$. We infer the optimality of monotone policies of the form x^{A}. This concludes the proof.

4.2. Sufficient conditions for the (strict) indexability of P and the related indices

We now progress to consider issues concerning the (strict) indexability of P. These are considerably more challenging than the skip-free models of Section 3. Before proceeding further, we specialise to the case

$$R(x) = R_1 + R_2(M+N-x), \tag{33}$$

where $R_1 = R(M+N)$ and R_2 are both positive constants. Namely, the resource consumed by the act of replenishment in state x is an affine function of the quantity supplied. Further specialising by taking $R_1 = 0$ yields the case where the key resource is the stock to be supplied, while taking $R_2 = 0$ gives the case where the key resource is the availability of replenishment opportunities. From Theorem 2 (more exactly, a modest extension thereof), the crucial step in establishing indexability is the demonstration that $R^{x^{\mathrm{A}}}$ is increasing in x. This is now stated and proved.

Lemma 3. *$R^{x^{\mathrm{A}}}$ is increasing in x over the range $-1 \le x \le M+N$.*

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

Proof. We begin by observing that $R^{(-1)^{\mathrm{A}}} = 0$ and $R^{x^{\mathrm{A}}} \geq 0$, $0 \leq x \leq M + N$. Fix x in the latter range. Let $\{D_n,\ n \in \mathbb{Z}^+\}$ be a sequence of independent and identically distributed random variables (successive daily demands) each with the distribution of D. We introduce the process $\{Y(t),\ t \in \mathbb{N}\}$ by

$$Y(t) := M + N - \sum_{n=1}^{t} D_n,$$

and define the related stopping time by

$$\tau_x^{\mathrm{A}} := \min\{t; t \in \mathbb{Z}^+ \text{ and } Y(t) \leq x\}.$$

If active action a had been taken at time 0 then τ_x^{A} is the first subsequent time at which the active action is taken under policy x^{A}. As before, we write $m^{\mathrm{A}}(x) = \mathbb{E}\{\tau_x^{\mathrm{A}}\}$ and express the distribution of $Y(\tau_x^{\mathrm{A}})$ by

$$q(y \mid x) := \mathbb{P}\{Y(\tau_x^{\mathrm{A}}) = x - y\}, \qquad y \in \mathbb{N}.$$

Using standard reward-renewal theoretic arguments, it is straightforward that the rate of resource used under policy x^{A} is given by

$$R^{x^{\mathrm{A}}} = \frac{R_1 + R_2\mathbb{E}\{M + N - Y^+(\tau_x^{\mathrm{A}})\}}{m^{\mathrm{A}}(x)}, \qquad 0 \leq x \leq M + N. \tag{34}$$

By a simple argument based on realisations of the process $\{Y(t),\ t \in \mathbb{N}\}$ we observe that, for $y \geq 1$ and $x \geq 1$,

$$q(y - 1 \mid x - 1) - q(y \mid x) = q(x)p_y, \tag{35}$$

where

$$q(x) = \mathbb{P}\left\{\bigcup_{t=0}^{\infty}\{Y(t) = x\}\right\}$$

is the probability that the process ever hits state x. Now, the process $\{Y(t) + \lambda t,\ t \in \mathbb{N}\}$ is plainly a martingale and we deduce from the martingale stopping theorem that

$$\mathbb{E}\{Y(\tau_x^{\mathrm{A}})\} + \lambda m^{\mathrm{A}}(x) = M + N, \tag{36}$$

and, hence, that

$$m^{\mathrm{A}}(x) = \lambda^{-1}\left\{M + N - x + \sum_{y=0}^{\infty} yq(y \mid x)\right\}. \tag{37}$$

It follows from (35) and (37) that

$$m^{\mathrm{A}}(x - 1) - m^{\mathrm{A}}(x) = q(x), \tag{38}$$

from which we conclude that, in order to prove that $R^{x^{\mathrm{A}}}$ is increasing in x, it is enough to show that the quantity

$$\alpha(x) := \frac{\mathbb{E}\{Y^+(\tau_x^{\mathrm{A}}) - Y(\tau_x^{\mathrm{A}})\}}{M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\}} = \frac{\sum_{y=1}^{\infty} yq(x + y \mid x)}{M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\}}$$

is decreasing in x over the range $0 \leq x \leq M + N$.

Suppose now that $M + N \geq x \geq 1$. Repeated application of (35) and straightforward algebra yields the identity

$$\begin{aligned}
&[M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\}][M + N - \mathbb{E}\{Y(\tau_{x-1}^{\mathrm{A}})\}]\{\alpha(x-1) - \alpha(x)\} \\
&\quad = q(x)\left\{\left(\sum_{y=1}^{\infty} y p_{x+y}\right)[M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\}] - \lambda \sum_{y=1}^{\infty} y q(x+y \mid x)\right\} \\
&\quad \equiv q(x)\beta(x), \qquad (39)
\end{aligned}$$

say. It is easy to see that $\beta(M + N) = 0$. Further direct calculation yields

$$\beta(x-1) - \beta(x) = \left(\sum_{y=1}^{\infty} p_{x+y-1}\right)[M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\} + \lambda q(x)] > 0,$$

and, hence,

$$\beta(x) > 0, \qquad 0 \leq x \leq M + N - 1,$$

from which we in turn infer that

$$\alpha(x-1) \geq \alpha(x), \qquad 1 \leq x \leq M + N.$$

Hence, $\alpha(x)$ decreases in x over the range $0 \leq x \leq M + N$. This concludes the proof.

At this point in the discussion, and in particular before we explore strict indexability for P, it will be helpful to be more explicit about project costs. We now take

$$\bar{C}(x) = K + C(M + N - x) \qquad (40)$$

as the model for replenishment costs, with inventory costs given by

$$\Phi(x) = b(M - x)I(0 \leq x \leq M) + (bM - lx)I(x < 0), \quad x \in (-\infty, M + N] \cap \mathbb{Z}. \quad (41)$$

In (40) we make the standard choice that $\bar{C}(x)$ is an aggregate of a fixed order cost (K) together with a payment for the items supplied, levied at a per unit cost C. We interpret (41) as follows: b is a cost incurred per backordered customer (of whom there will be $(M - x)^+$ if state x is reached) while l is the cost incurred for each lost sale. All constants K, C, b, and l are positive. We assume that $l \geq b$. Typically, we have $l \gg b$. Note that we exclude inventory holding costs. These are assumed to be negligible in comparison with those associated with shortages. We shall see later that the piecewise-linear model in (41) may be generalised to a convex model for inventory costs without seriously compromising the main results.

From (29), (33), (40), and (41), we have

$$\begin{aligned}
\Psi_W(x) = {} & WR_1 + K + (WR_2 + C)(M + N - x)I(M \leq x \leq M + N) \\
& + \{(WR_2 + C)N + (WR_2 + C + b)(M - x)\}I(0 \leq x \leq M) \\
& + \{(WR_2 + C)N + (WR_2 + C + b)M - lx\}I(x < 0)
\end{aligned}$$

for $x \in (-\infty, M + N] \cap \mathbb{Z}$. For Condition 1 to be met, we need the resource charge W to be constrained to the range

$$W \in [-R_2^{-1}C, R_2^{-1}(l - b - C)].$$

Note that, when $R_2 = 0$ and the key resource is *opportunities for replenishment*, we have a development of the restless bandit model of Archibald *et al.* [2], and the range of W is unconstrained. It is when $R_2 > 0$ that the resource used by the replenishment action is state dependent and the range of W is constrained to a finite interval. Such a constraint on the range of W requires a modification of the notions of indexability/index functions introduced in Definition 1 as follows.

Definition 3. Project P is indexable over the range $[W_1, W_2]$ if there exists a collection of policies $\{\pi(W),\ W \in [W_1, W_2]\}$ such that

(i) $\pi(W)$ is optimal for $P(W)$ for all $W \in [W_1, W_2]$; and

(ii) $b(\pi(W))$ is increasing in W over the range $W_1 \le W \le W_2$.

The related index function $W\colon \Omega \to \mathbb{R} \cup \{-\infty, \infty\}$ is given by

$$W(x) = \begin{cases} W_1 & \text{if } x \in b(\pi(W)) \text{ for all } W \in [W_1, W_2], \\ W_2 & \text{if } x \notin b(\pi(W)) \text{ for all } W \in [W_1, W_2], \\ \inf\{W; x \in b(\pi(W))\} & \text{otherwise.} \end{cases}$$

If project P is indexable over $[W_1, W_2]$ then it is strictly indexable over $[W_1, W_2]$ if, for any $x \neq y$ for which $W(x) \notin \{W_1, W_2\}$ and $W(y) \notin \{W_1, W_2\}$, we have $W(x) \neq W(y)$.

In what follows we shall establish (strict) indexability over $[-R_2^{-1}C, R_2^{-1}(l-b-C)]$ for the project under given conditions. We shall see that this in turn implies strict indexability for the project over $\mathbb{R}^+$, and so will be sufficient for our purposes. See Remark 3 following Definition 1. We begin by the elucidation of indexability of the project over $[-R_2^{-1}C, R_2^{-1}(l-b-C)]$.

Theorem 3. *Project P with resource function (33) and costs given by (40) and (41) is indexable over $[-R_2^{-1}C, R_2^{-1}(l-b-C)]$.*

Proof. It is easy to show that, under the conditions of the theorem, Condition 1 is satisfied over the range $W \in [-R_2^{-1}C, R_2^{-1}(l-b-C)]$. It then follows from Lemma 2 that over this range there exists some policy x^{A} which is optimal for $P(W)$. We thus have

$$C(W) = \min_{-1 \le x \le M+N} (C^{x^{\mathrm{A}}} + W R^{x^{\mathrm{A}}}), \qquad W \in [-R_2^{-1}C, R_2^{-1}(l-b-C)]. \tag{42}$$

From Lemma 3, $R^{x^{\mathrm{A}}}$ is increasing in x over the range $-1 \le x \le M+N$. An appropriate variant of the argument in the proof of Theorem 1 now establishes the existence of a minimising x_W in (42) which decreases as W increases through the interval $[-R_2^{-1}C, R_2^{-1}(l-b-C)]$. This establishes indexability for P over $[-R_2^{-1}C, R_2^{-1}(l-b-C)]$ and concludes the proof.

We now focus on strict indexability and on the specification of the index function. It will simplify calculations if we introduce the quantities

$$\rho^{x^{\mathrm{A}}} := \frac{\mathbb{E}\{M+N-Y^+(\tau_x^{\mathrm{A}})\}}{m^{\mathrm{A}}(x)}$$

and

$$\omega^{\mathrm{A}}(x) := \frac{C^{(x-1)^{\mathrm{A}}} - C^{x^{\mathrm{A}}}}{\rho^{x^{\mathrm{A}}} - \rho^{(x-1)^{\mathrm{A}}}}, \qquad 0 \le x \le M+N. \tag{43}$$

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

Note that $\rho^{x^{\mathrm{A}}}$ is the long-run average rate at which stock is consumed under policy x^{A}. From Theorem 2, key quantities for strict indexability are

$$W^{\mathrm{A}}(x) = \frac{C^{(x-1)^{\mathrm{A}}} - C^{x^{\mathrm{A}}}}{R^{x^{\mathrm{A}}} - R^{(x-1)^{\mathrm{A}}}} = \omega^{\mathrm{A}}(x)\left\{\frac{\rho^{x^{\mathrm{A}}} - \rho^{(x-1)^{\mathrm{A}}}}{R^{x^{\mathrm{A}}} - R^{(x-1)^{\mathrm{A}}}}\right\}, \qquad 0 \le x \le M + N. \tag{44}$$

Lemma 4. *Under the hypotheses of Theorem 3, $W^{\mathrm{A}}(x)$ is strictly decreasing over the range $0 \le x \le M + N$ when $K > bN$ and $l > C + b$.*

Proof. It is straightforward that, for the cost model in (40) and (41), we have

$$\begin{aligned} C^{x^{\mathrm{A}}} = {} & (K + C\mathbb{E}\{M + N - Y^{+}(\tau_x^{\mathrm{A}})\} + b\mathbb{E}\{\{M - Y(\tau_x^{\mathrm{A}})\}^{+} - Y^{-}(\tau_x^{\mathrm{A}})\} \\ & + l\mathbb{E}\{Y^{-}(\tau_x^{\mathrm{A}})\})\{m^{\mathrm{A}}(x)\}^{-1}, \qquad 0 \le x \le M + N. \end{aligned} \tag{45}$$

Note that $C^{(-1)^{\mathrm{A}}} = l\lambda$. The expression in (45) may be simplified by observing that

$$Y(\tau_x^{\mathrm{A}}) = Y^{+}(\tau_x^{\mathrm{A}}) - Y^{-}(\tau_x^{\mathrm{A}}),$$

from which it follows via (36) that

$$\begin{aligned} \lambda m^{\mathrm{A}}(x) - \mathbb{E}\{M + N - Y^{+}(\tau_x^{\mathrm{A}})\} &= \mathbb{E}\{M + N - Y(\tau_x^{\mathrm{A}})\} - \mathbb{E}\{M + N - Y^{+}(\tau_x^{\mathrm{A}})\} \\ &= \mathbb{E}\{Y^{-}(\tau_x^{\mathrm{A}})\}. \end{aligned}$$

Using this identity in (45), we infer that

$$C^{x^{\mathrm{A}}} = \lambda(l - b) + (C + b - l)\rho^{x^{\mathrm{A}}} + [K + b\mathbb{E}\{M - Y(\tau_x^{\mathrm{A}})\}^{+}]\{m^{\mathrm{A}}(x)\}^{-1}. \tag{46}$$

It follows from (43) and (46) that

$$\begin{aligned} \omega^{\mathrm{A}}(x) = {} & (l - C - b) \\ & + \frac{m^{\mathrm{A}}(x)[K + b\mathbb{E}\{M - Y(\tau_{x-1}^{\mathrm{A}})\}^{+}] - m^{\mathrm{A}}(x-1)[K + bE\{M - Y(\tau_x^{\mathrm{A}})\}^{+}]}{m^{\mathrm{A}}(x-1)\mathbb{E}\{M + N - Y^{+}(\tau_x^{\mathrm{A}})\} - m^{\mathrm{A}}(x)\mathbb{E}\{M + N - Y^{+}(\tau_{x-1}^{\mathrm{A}})\}} \end{aligned} \tag{47}$$

for $0 \le x \le M + N$. In order to simplify the expression in (47), we recall from (38) that

$$m^{\mathrm{A}}(x) - m^{\mathrm{A}}(x-1) = -q(x) \tag{48}$$

and from the material around (39) that

$$\begin{aligned} & m^{\mathrm{A}}(x-1)\mathbb{E}\{M + N - Y^{+}(\tau_x^{\mathrm{A}})\} - m^{\mathrm{A}}(x)\mathbb{E}\{M + N - Y^{+}(\tau_{x-1}^{\mathrm{A}})\} \\ & \quad = \lambda^{-1} m^{\mathrm{A}}(x-1) m^{\mathrm{A}}(x)\{\alpha(x-1) - \alpha(x)\} \\ & \quad = \lambda^{-1} q(x)\beta(x). \end{aligned} \tag{49}$$

We further observe that

$$\{M - Y(\tau_x^{\mathrm{A}})\}^{+} = M - Y(\tau_x^{\mathrm{A}}), \qquad 0 \le x \le M,$$

and, hence, from (36) that

$$\mathbb{E}\{M - Y(\tau_x^{\mathrm{A}})\}^{+} = \lambda m^{\mathrm{A}}(x) - N, \qquad 0 \le x \le M. \tag{50}$$

It follows from (50) and the material around (39) that

$$\frac{m^{\mathrm{A}}(x)\mathbb{E}\{M - Y(\tau_{x-1}^{\mathrm{A}})\}^+ - m^{\mathrm{A}}(x-1)\mathbb{E}\{M - Y(\tau_x^{\mathrm{A}})\}^+}{m^{\mathrm{A}}(x-1)\mathbb{E}\{M + N - Y^+(\tau_x^{\mathrm{A}})\} - m^{\mathrm{A}}(x)\mathbb{E}\{M + N - Y^+(\tau_{x-1}^{\mathrm{A}})\}} = \frac{\lambda N}{\beta(x)} \tag{51}$$

for $0 \le x \le M$. We now consider the range $M \le x \le M + N$ and introduce the quantity

$$\beta_M(x) := \left(\sum_{y=1}^{\infty} y p_{x-M+y}\right)[M + N - \mathbb{E}\{Y(\tau_x^{\mathrm{A}})\}] - \lambda \sum_{y=1}^{\infty} y q(x - M + y \mid x).$$

Compare with the expression for $\beta(x)$ in (39). By calculations similar to those around (39) we can show that $\beta_M(x)$ is strictly decreasing in x over the range $M \le x \le M + N$. Furthermore, straightforward algebra utilising (35) yields

$$\frac{m^{\mathrm{A}}(x)\mathbb{E}\{M - Y(\tau_{x-1}^{\mathrm{A}})\}^+ - m^{\mathrm{A}}(x-1)\mathbb{E}\{M - Y(\tau_x^{\mathrm{A}})\}^+}{m^{\mathrm{A}}(x-1)\mathbb{E}\{M + N - Y^+(\tau_x^{\mathrm{A}})\} - m^{\mathrm{A}}(x)\mathbb{E}\{M + N - Y^+(\tau_{x-1}^{\mathrm{A}})\}} = \frac{\beta_M(x)}{\beta(x)} \tag{52}$$

for $M \le x \le M + N$. Compare with (51) and note that it is easy to check that $\beta_M(M) = \lambda N$. Combining (47)–(52) yields

$$\omega^{\mathrm{A}}(x) = \begin{cases} (l - C - b) - \lambda(K - bN)\{\beta(x)\}^{-1}, & 0 \le x \le M, \\ (l - C - b) - \{\lambda K - b\beta_M(x)\}\{\beta(x)\}^{-1}, & M \le x \le M + N. \end{cases} \tag{53}$$

It is straightforward from earlier calculations in and around (34) and (39) that

$$\frac{\rho^{x^{\mathrm{A}}} - \rho^{(x-1)^{\mathrm{A}}}}{R^{x^{\mathrm{A}}} - R^{(x-1)^{\mathrm{A}}}} = \frac{\beta(x)}{R_1\lambda + R_2\beta(x)}, \qquad 0 \le x \le M + N,$$

and, hence, from (44) and (53) we have

$$W^{\mathrm{A}}(x) = \begin{cases} \{(l - C - b)\beta(x) - \lambda(K - bN)\}\{R_1\lambda + R_2\beta(x)\}^{-1}, & 0 \le x \le M, \\ ((l - C - b)\beta(x) - \{K\lambda - b\beta_M(x)\})\{R_1\lambda + R_2\beta(x)\}^{-1}, & M \le x \le M + N. \end{cases} \tag{54}$$

The two expressions in (54) are indeed in agreement at $x = M$. The strictly decreasing nature of $\beta_M(x)$ over the range $M \le x \le M + N$ together with the condition $K > bN$ guarantees that

$$K\lambda > bN\lambda > b\beta_M(x), \qquad M + 1 \le x \le M + N.$$

That $W^{\mathrm{A}}(x)$ is strictly decreasing in x over the range $0 \le x \le M + N$ under the hypotheses of the lemma then follows easily from (54). This concludes the proof.

The following result is now an immediate consequence of the above lemma.

Theorem 4. *Under the hypotheses of Lemma 4, project P is strictly indexable over $[-R_2^{-1}C, R_2^{-1}(l - b - C)]$ with related index function given by*

$$W(x) = \max\{W^{\mathrm{A}}(x), -R_2^{-1}C\}, \qquad 0 \le x \le M + N,$$

where $W^{\mathrm{A}}(x)$ is as in (54).

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

Proof. We define x^* by

$$x^* := \min_{0 \le x \le M+N} \{x;\, W^{\mathrm{A}}(x) \le -R_2^{-1}C\},$$

and note from Lemma 4 that the index introduced in the statement of the theorem satisfies $W(y) > -R_2^{-1}C$, $0 \le y \le x^* - 1$, and is strictly decreasing over this range. Note also from (54), together with the argument following, that

$$W(x) < R_2^{-1}(l - b - C), \qquad 0 \le x \le M + N.$$

Furthermore, following the statement and proof of Theorem 3, we can easily infer that, under the conditions of the theorem, there exists a collection of policies $\{\pi(W), W \in [-R_2^{-1}C, R_2^{-1}(l - b - C)]\}$ such that $\pi(W)$ is optimal for $P(W)$ and is given by

$$\pi(W) = \begin{cases} (x^* - 1)^{\mathrm{A}}, & -R_2^{-1}C \le W \le W(x^* - 1), \\ x^{\mathrm{A}}, & W(x+1) < W \le W(x),\ 0 \le x \le x^* - 2, \\ (-1)^{\mathrm{A}}, & W(0) < W \le R_2^{-1}(l - b - C). \end{cases}$$

The strict indexability of P over the range $[-R_2^{-1}C, R_2^{-1}(l - b - C)]$ with W the related index function now follows easily. This concludes the proof.

The following is the central result of this section and follows straightforwardly from the foregoing analysis.

Theorem 5. *Under the hypotheses of Lemma 4, project P is strictly indexable over $\mathbb{R}^+$. The value of the related index function at x is given by $W(x) = \{W^{\mathrm{A}}(x)\}^+$, $0 \le x \le M + N$, where $W^{\mathrm{A}}(x)$ is as in (54).*

Proof. From the proof of Theorem 4 we note that the policy $(-1)^{\mathrm{A}}$ which never replenishes stock is optimal for $P(W)$ over the range $\max\{W^{\mathrm{A}}(0), -R_2^{-1}C\} < W \le R_2^{-1}(l - b - C)$. It follows that the associated optimal cost rate is $C(W) = l\lambda$, over this W-range. However, $C(W)$ is plainly nondecreasing in W and, hence, it must follow that $C(W) = l\lambda$, $\max\{W^{\mathrm{A}}(0), -R_2^{-1}C\} < W < \infty$, with the idling policy $(-1)^{\mathrm{A}}$ optimal whenever W exceeds $\max\{W^{\mathrm{A}}(0), -R_2^{-1}C\}$. The result now follows by a simple modification of the argument in the proof of Theorem 4.

Remark 5. As explained in Remark 3 following Definition 1, Theorem 5 enables us to construct index heuristics for our inventory control problems after the fashion described in Remark 4. Our numerical experience, some of which is reported below, is that the resulting policies perform strongly, especially when penalties from lost sales are a major concern, as is usually the case. The form of W^{A} given in (54), together with the fact that $\beta(M + N) = 0$, implies that $W^{\mathrm{A}}(M + N) < 0$ and, hence, that the related index for state $M + N$ given in Theorem 5 is zero. It then follows that our index heuristics never replenish a stock-holding location when its stock level is maximal. For all other states x, it is clear from (54) that the value of $W^{\mathrm{A}}(x)$ is strictly increasing in the lost sales rate l and so (in particular) the location index in Theorem 5 can be forced positive for large enough l and will be dominated by a measure of the additional costs incurred by lost sales in the absence of any stock replenishment.

We now consider a generalisation of the piecewise-linear inventory costs in (41) to

$$\Phi(x) = \phi(x) I(0 \le x \le M) + (\phi(0) - lx) I(x < 0), \qquad x \in (-\infty, M + N] \cap \mathbb{Z}, \quad (55)$$

with ϕ decreasing and convex such that $\phi(M) = 0$. Hence, we have a general convex model for backorder costs (which are notoriously difficult to quantify) while retaining a linear model for the costs from lost sales. To achieve indexability, we restrict to the class of stationary policies S_M for which $\pi(x) = a$, $0 \le x \le M$. Namely, a replenishment is mandated as soon as a stockout occurs. We shall refer to S_M-indexability, S_M-strict indexability, etc. in what follows. This restriction of policies to S_M can be thought of as reducing the state space of P to $\{M+1, M+2, \dots, M+N\}$.

The analysis of this case rests on the fact that Φ may be expressed as

$$\begin{aligned}\Phi(x) = \sum_{m=1}^{M} &\{\phi(M-m) - 2\phi(M-m+1) + \phi(M-m+2)\} \\ &\times (M-x) I(0 \le x \le M-m+1) \\ &+ (\phi(0) - lx) I(x < 0), \qquad x \in (-\infty, M+N] \cap \mathbb{Z}, \end{aligned} \tag{56}$$

where we take $\phi(M+1) = 0$. The superposition of linear terms in (56) permits an analysis akin to that in Theorem 3/Lemma 4/Theorem 4/Theorem 5. The proof of Theorem 6 below is available from the authors.

Theorem 6. *Project P with resource function (33) and costs given by (40) and (56) is S_M-indexable over $\mathbb{R}^+$. It is also S_M-strictly indexable over $\mathbb{R}^+$ when $K > \{\phi(0) - \phi(1)\}N$ and $l > C + \phi(0) - \phi(1)$ with related index function given by*

$$W(x) = \{W^{\mathrm{A}}(x)\}^+, \qquad M+1 \le x \le M+N,$$

where

$$\begin{aligned} W^{\mathrm{A}}(x) = \Bigg(&\{l - C - \phi(0) + \phi(1)\}\beta(x) \\ &- \left[K\lambda - \sum_{m=1}^{M} \{\phi(M-m) - 2\phi(M-m+1) + \phi(M-m+2)\}\beta_{M-m+1}(x) \right] \Bigg) \\ &\times \{R_1\lambda + R_2\beta(x)\}^{-1}, \qquad M+1 \le x \le M+N. \end{aligned}$$

4.3. Numerical study

We explore the performance of the greedy index heuristic GI in a two-phase numerical study. In the design of GI, we use the indices given in Theorem 5 throughout. In the first phase we explore its performance in 640 problems, each involving three inventory-holding locations. These problems are sufficiently small that direct comparisons with the optimal of the cost performance of GI and other natural heuristics for the problems is possible. In the second phase of the study, the cost performance of GI is compared with a range of natural heuristics for networks consisting of between 10 and 100 inventory-holding locations. These problems are sufficiently large that direct application of DP to obtain optimal costs is out of the question.

We take $R(x) = M+N-x$ throughout the study and so the resource used in replenishment is always identified with the items supplied. The heuristics which we place in competition with GI utilise, for some choice of state x^*, an index of the form $\mathrm{Index}(x) = (\lambda/x) I(x^* \ge x \ge 0)$, with the resulting heuristic denoted by $H(x^*)$. In words, $H(x^*)$ is a policy which (i) only replenishes locations whose state is no more than x^*, and (ii) among qualifying locations, greedily allocates resource to locations according to (small values of) the index x/λ, namely the mean number

of days remaining at the location before lost sales will occur in the absence of an intervening replenishment. In the tables below, we consider the cost rate performances of an optimal policy (for $J = 3$ only) and GI along with policies $H(x^*)$ for $x^* = 0.25M, 0.5M, 0.75M, M, M+\lambda, M+2\lambda, M+N$. These latter choices reflect a wide range from $x^* = 0.25M$ when 75% of the available backorders must be used before replenishment is sought, through to $x^* = M + N$ in which the index x/λ is used without qualification. The cost model in (40) and (41) is applied throughout. In what follows, optimal cost rates are computed via DP value iteration and average costs for the heuristics are estimated by means of Monte Carlo simulation.

The 640 three-location problems studied in the first phase are chosen as five blocks containing 128 problems each. The choices $C = 10$, $b = 5$ are made throughout. In block 1, $\bar{N} = (N_1, N_2, N_3)$, the maximum stock levels at the three locations, $\bar{M} = (M_1, M_2, M_3)$, the maximum number of backorders, and R, the amount available resource, are set at $\bar{N} = (20, 20, 20)$, $\bar{M} = (10, 10, 10)$, and $R = 60$, respectively. For block 2, we set $\bar{N} = (10, 15, 20)$, $\bar{M} = (10, 15, 20)$, $R = 70$; for block 3, we set $\bar{N} = (15, 15, 15)$, $\bar{M} = (10, 15, 20)$, $R = 65$; for block 4, we set $\bar{N} = (20, 20, 20)$, $\bar{M} = (10, 15, 20)$, $R = 75$; and, for block 5, we set $\bar{N} = (10, 15, 20)$, $\bar{M} = (15, 15, 15)$, $R = 65$. Within each block, further parameter choices are made as follows: demand rates $\bar{\lambda} \in (\bar{M}/4, \bar{M}/3, \bar{M}/2, 2\bar{M}/3)$, order costs $\bar{K} \in ((500, 500, 500), (100, 500, 900), (2000, 2000, 2000), (1500, 2000, 2500))$, and lost sales penalties $\bar{l} \in ((100, 100, 100), (300, 300, 300), (500, 500, 500), (700, 700, 700), (900, 900, 900), (100, 300, 500), (300, 500, 700), (500, 700, 900))$. This makes $4 \times 4 \times 8 = 128$ parameter combinations for the problems in each block. For each of the 128 problems within each block, the optimal cost rate was computed along with that of GI and the seven $H(x^*)$ heuristics. Furthermore, for each problem, the strongest performing of the $H(x^*)$ policies was noted along with its cost rate. Call this cost minimising heuristic H. In Table 4 find, for each block of problems and each GI, the seven heuristics $H(x^*)$ and H, and the median (MED) and maximum (MAX) percentages suboptimality across the 128 problems. The median suboptimality of GI was well below 1% for all five blocks and among all 640 problems it was never more than 8% suboptimal.

We now proceed to describe the second phase of the numerical study involving larger problems. Table 5 contains estimates of the average costs incurred by collections of J locations, each of which has $N = 20, M = 10, \lambda = 5, C = 10, b = 5$, and $l = 500$, but which have widely differing fixed delivery costs K. We suppose that each K lies in the set $\{200, 450, 700, 950, 1200\}$ with $J/5$ locations taking each candidate K-value. The amount of available resource R is scaled with J. For the results given in Table 6, we repeated the numerical experiments but with a reduced lost sales cost of $l = 100$. The strongest competition to GI in cost performance comes from $H(0.75M)$ in Table 5 and from $H(0.5M)$ in Table 6. Hence, unsurprisingly, good policies become more conservative (replenish earlier) when lost sales are more costly. To give the reader a greater feel for the greedy index heuristic, a set of location indices $W(x)$, $0 \le x \le 9$, for a Table 5 location with $K = 200$ is as follows.

x	0	1	2	3	4	5	6	7	8	9
$W(x)$	481.67	480.70	479.14	476.47	471.62	462.20	442.77	399.91	298.88	44.24

For this location, we have $W(x) = 0$, $10 \le x \le 30$. We see that index values for this location are close to the maximum value ($W(0)$) when there is a good chance of lost sales in the coming day. Table 7 contains the average costs incurred by collections of J locations which may differ markedly in the scale of their operations. We now suppose that all locations have

TABLE 4: Median and maximum percentage suboptimalities for nine replenishment heuristics for five blocks of 128 small (three-location) problems. See the text for details.

Block	GI	$H(0.25M)$	$H(0.50M)$	$H(0.75M)$	$H(M)$	$H(M+\lambda)$	$H(M+2\lambda)$	$H(M+N)$	H
					MED				
1	0.26	9.84	2.28	4.00	11.38	31.85	61.65	399.96	0.93
2	0.48	12.01	4.39	9.37	25.42	78.43	133.54	231.03	2.05
3	0.48	12.15	4.22	10.38	26.93	83.89	149.08	219.90	2.05
4	0.42	12.54	3.59	7.65	19.12	61.41	115.48	262.18	1.61
5	0.58	13.07	4.41	9.30	24.74	78.77	130.71	227.94	1.96
					MAX				
1	1.73	146.59	41.74	18.29	32.29	57.66	160.18	842.82	3.02
2	3.91	185.64	61.47	38.22	66.95	155.90	253.86	576.62	8.69
3	6.57	191.05	58.83	36.31	66.08	148.67	266.71	563.29	8.29
4	5.91	186.54	58.98	27.77	48.76	115.75	241.00	653.46	5.40
5	7.98	213.99	74.61	36.86	68.37	138.07	251.81	567.06	5.96

TABLE 5: Cost rate estimates for eight replenishment heuristics for a collection of problems with $l = 500$ comprising identical locations that differ only in the fixed delivery cost K. See the text for details.

J	R	GI	$H(0.25M)$	$H(0.50M)$	$H(0.75M)$	$H(M)$	$H(M+\lambda)$	$H(M+2\lambda)$	$H(M+N)$
10	125	1956.74	2624.12	2022.29	1968.13	2084.12	2502.00	3297.94	7452.90
20	250	3907.18	5205.66	4022.78	3934.04	4173.03	5006.44	6598.29	14 906.34
30	375	5861.29	7799.77	6031.83	5902.35	6260.71	7510.47	9898.53	22 358.98
40	500	7810.54	10 400.45	8042.54	7867.72	8345.36	10 012.32	13 194.30	29 813.12
50	625	9763.47	13 005.74	10 051.91	9833.48	10 431.27	12 513.26	16 493.46	37 262.62
60	750	11 717.10	15 601.78	12 064.84	11 801.56	12 517.49	15 016.75	19 793.28	44 717.31
70	875	13 668.02	18 192.83	14 070.93	13 766.21	14 603.26	17 518.08	23 092.64	52 169.89
80	1000	15 621.09	20 794.41	16 079.43	15 732.08	16 688.73	20 021.70	26 391.47	59 624.83
90	1125	17 575.08	23 402.95	18 092.18	17 698.85	18 776.62	22 524.27	29 689.29	67 074.70
100	1250	19 532.63	26 004.86	20 109.98	19 671.32	20 864.75	25 030.50	32 993.10	74 528.59

TABLE 6: Cost rate estimates for eight replenishment heuristics for a collection of problems with $l = 100$ comprising identical locations that differ only in the fixed delivery cost K. See the text for details.

J	R	GI	$H(0.25M)$	$H(0.50M)$	$H(0.75M)$	$H(M)$	$H(M+\lambda)$	$H(M+2\lambda)$	$H(M+N)$
10	125	1844.71	1888.04	1870.32	1930.78	2081.82	2501.98	3297.94	7452.90
20	250	3687.43	3771.79	3741.72	3867.25	4168.40	5006.42	6598.29	14 906.34
30	375	5532.25	5657.54	5613.47	5802.39	6253.81	7510.43	9898.53	22 358.98
40	500	7374.06	7542.02	7482.84	7734.75	8336.08	10 012.28	13 194.30	29 813.12
50	625	9217.08	9427.80	9353.11	9667.66	10 419.55	12 513.22	16 493.46	37 262.62
60	750	11 061.60	11 312.97	11 223.60	11 601.42	12 503.63	15 016.70	19 793.28	44 717.31
70	875	12 904.07	13 197.05	13 094.01	13 534.64	14 587.00	17 518.03	23 092.64	52 169.89
80	1000	14 747.42	15 082.91	14 964.55	15 467.90	16 670.81	20 021.67	26 391.47	59 624.83
90	1125	16 591.00	16 969.61	16 835.37	17 401.39	18 755.83	22 524.20	29 689.29	67 074.70
100	1250	18 438.86	18 857.95	18 709.91	19 338.54	20 841.85	25 030.45	32 993.10	74 528.59

TABLE 7: Cost rate estimates for eight replenishment heuristics for a collection of problems comprising locations that differ markedly in the scale of their operations. See the text for details.

J	R	GI	$H(0.25M)$	$H(0.50M)$	$H(0.75M)$	$H(M)$	$H(M+\lambda)$	$H(M+2\lambda)$	$H(M+N)$
10	600	6164.87	13 301.02	9677.42	7701.31	6541.14	6317.01	7275.42	17 999.49
20	1200	11 957.00	24 745.45	18 109.56	14 339.94	12 285.79	12 673.07	14 588.32	35 999.25
30	1800	17 884.58	36 620.81	26 841.16	21 235.89	18 274.40	19 015.99	21 888.06	53 998.58
40	2400	23 836.13	48 640.82	35 683.36	28 210.77	24 328.42	25 361.82	29 188.84	71 998.97
50	3000	29 791.72	60 740.57	44 560.81	35 220.11	30 397.49	31 703.40	36 485.46	89 998.91
60	3600	35 753.77	72 872.99	53 456.30	42 262.90	36 469.48	38 047.16	43 784.89	10 7999.53
70	4200	41 704.98	84 968.92	62 325.54	49 290.49	42 535.04	44 383.08	51 078.53	12 5997.63
80	4800	47 667.18	97 151.92	71 221.99	56 326.58	48 617.43	50 727.26	58 379.19	14 3999.24
90	5400	53 622.57	109 286.00	80 144.49	63 367.59	54 699.34	57 068.57	65 677.01	16 1998.75
100	6000	59 588.28	121 397.55	89 043.23	70 428.64	60 773.08	63 412.63	72 974.91	17 9999.08

$K = 1500$, $C = 10$, $b = 5$, and $l = 500$, but the triple (N, M, λ) lies in the set {(100, 10, 10), (150, 20, 20), (200, 30, 30), (250, 40, 40), (300, 50, 50)}, with $J/5$ locations assigned to each. Hence, each location can accommodate around one week's worth of demand, but have widely varying demand rates.

The numerical results presented in Tables 4–7 provide powerful evidence of the value to the decision-maker of the indices $W(x)$, $0 \le x \le M + N$, in the design of strongly performing replenishment policies in a wide range of problems. The index heuristic GI has strong performance in comparison to optimal in small problems, and in comparison to a class of natural replenishment heuristics for both small and large problems whose size leaves them well beyond the reach of conventional DP approaches.

4.4. Projects living on $\mathbb{N}$

A natural generalisation of the above material is to descendant projects with state space $\mathbb{N}$. We now give a brief account and omit all proofs. These are available from the authors. In such cases, the stochastic dynamics in (26) and (27) are replaced by

$$p^a(x, y) = p^b(x, x + y) = p_y, \qquad (x, y) \in \mathbb{N}^2.$$

We continue to use D for an integer-valued random variable with PMF $\{p_y,\ y \in \mathbb{N}\}$. Hence, under the active action, the project is returned to pristine state 0 before further deterioration takes place. A natural application area for such models is to *machine maintenance* problems where we wish to abandon an assumption of skip-free evolution under the passive action in the interests of modelling sudden rapid deterioration in machine condition. Machine costs are captured by the increasing and convex function $\Phi\colon \mathbb{N} \to \mathbb{R}^+$, where $\Phi(0) = 0$. Expected costs incurred when actions a and b are taken in state x are given by

$$C^a(x) = \mathbb{E}\{\Phi(D)\}, \qquad x \in \mathbb{N},$$

and

$$C^b(x) = \mathbb{E}\{\Phi(x + D) - \Phi(x)\}, \qquad x \in \mathbb{N},$$

which are positive and increasing in x. The function $R\colon \mathbb{N} \to \mathbb{R}^+$ yields resource consumption under the active action. We need the following condition.

Condition 2. *The function* $\Psi_W : \mathbb{N} \to \mathbb{R}^+$ *given by*

$$\Psi_W = WR(x) + \Phi(x), \qquad x \in \mathbb{N},$$

is increasing and convex.

Remark 6. If R is affine and increasing then Condition 2 will be satisfied at all positive W. More generally, if R is increasing and concave then Condition 2 may be satisfied over some finite range of positive W.

The proof of Lemma 5 below is along the lines of Lemma 3.

Lemma 5. *Under Condition 2 there exists some monotone policy* x^{D} *which is optimal for* $P(W)$.

To achieve indexability requires $R^{x^{\mathrm{D}}}$ to be decreasing in x. See Theorem 1. Condition 3 below is sufficient.

https://doi.org/10.1239/aap/1363354103 Published online by Cambridge University Press

Condition 3. *The resource function* $R\colon \mathbb{N} \to \mathbb{R}^+$ *is increasing and concave such that, for all* $x \in \mathbb{N}$,

$$\frac{\mathbb{E}\{R(D+r) \mid D \geq x-r\}}{\mathbb{E}\{D+r \mid D \geq x-r\}}$$

is decreasing in r *over the range* $0 \leq r \leq x$.

Remark 7. Cases which satisfy Condition 3 include

- D has a geometric distribution with support $\mathbb{N}$ and R is any increasing and concave function;
- R is affine and increasing and the hazard rate of D, namely,

$$\mathbb{P}(D = x \mid D \geq x),$$

  is increasing in x. This includes, *inter alia*, the classical case in which D is Poisson.

Lemma 6. *Under Condition 3,* $R^{x^{\mathrm{D}}}$ *is decreasing in* x.

In Corollary 1 we write $[0, W_1]$ for the range of positive W (assumed nonempty) for which Condition 2 holds. Its proof is via Lemmas 5 and 6 and (a minor extension of) Theorem 1.

Corollary 1. *Under Conditions 2 and 3, project* P, *with resource function* R *and costs determined by* Φ, *is indexable over* $[0, W_1]$.

Index values for P may be obtained by suitable application of Algorithm 1.

Acknowledgements

The first two authors gratefully acknowledge the support of EPSRC grant EP/E049265/01, and the third author the support of an RCUK Fellowship. All authors would like to thank an anonymous referee whose thoughtful comments led to a considerable strengthening of the paper.

References

[1] Ansell, P., Glazebrook, K. D., Niño-Mora, J. and O'Keeffe, M. (2003). Whittle's index policy for a multi-class queueing system with convex holding costs. *Math. Meth. Operat. Res.* **57,** 21–39.

[2] Archibald, T. W., Black, D. P. and Glazebrook, K. D. (2009). Indexability and index heuristics for a simple class of inventory routing problems. *Operat. Res.* **57,** 314–326.

[3] Dacre, M., Glazebrook, K. and Niño-Mora, J. (1999). The achievable region approach to the optimal control of stochastic systems (with discussion). *J. R. Statist. Soc. B* **61,** 747–791.

[4] Gittins, J. C. (1979). Bandit processes and dynamic allocation indices (with discussion). *J. R. Statist. Soc. B* **41,** 148–177.

[5] Gittins, J. C. (1989). *Multi-Armed Bandit Allocation Indices*. John Wiley, Chichester.

[6] Glazebrook, K. D. and Minty, R. (2009). A generalized Gittins index for a class of multiarmed bandits with general resource requirements. *Math. Operat. Res.* **34,** 26–44.

[7] Glazebrook, K. D., Kirkbride, C. and Ruiz-Hernandez, D. (2006). Spinning plates and squad systems: policies for bidirectional restless bandits. *Adv. Appl. Prob.* **38,** 95–115.

[8] Glazebrook, K. D., Niño-Mora, J. and Ansell, P. S. (2002). Index policies for a class of discounted restless bandits. *Adv. Appl. Prob.* **34,** 754–774.

[9] Jacko, P. (2009). Marginal productivity index policies for dynamic priority allocation in restless bandit models. Doctoral thesis, Universidad Carlos III de Madrid.

[10] Le Ny, J., Dahleh, M. and Feron, E. (2008). Multi-UAV dynamic routing with partial observations using restless bandit allocation indices. In *2008 American Control Conference,* pp. 4220–4225.

[11] Liu, K. and Zhao, Q. (2010). Indexability of restless bandit problems and optimality of Whittle index for dynamic multichannel access. *IEEE Trans. Inf. Theory* **56,** 5547–5567.

[12] Niño-Mora, J. (2001). Restless bandits, partial conservation laws and indexability. *Adv. Appl. Prob.* **33,** 76–98.
[13] Niño-Mora, J. (2007). Dynamic priority allocation via restless bandit marginal productivity indices. *TOP* **15,** 161–198.
[14] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.
[15] Weber, R. R. and Weiss, G. (1990). On an index policy for restless bandits. *J. Appl. Prob.* **27,** 637–648.
[16] Weber, R. R. and Weiss, G. (1991). Addendum to 'On an index policy for restless bandits'. *Adv. Appl. Prob.* **23,** 429–430.
[17] Whittle, P. (1988). Restless bandits: activity allocation in a changing world. In *A Celebration of Applied Probability* (*J. Appl. Prob. Spec. Vol.* **25A**), ed. J. Gani, Applied Probability Trust, Sheffield, pp. 287–298.
[18] Whittle, P. (1996). *Optimal Control*. John Wiley, Chichester.