

## INDEX POLICIES FOR A CLASS OF DISCOUNTED RESTLESS BANDITS

K. D. GLAZEBROOK,\* *University of Newcastle upon Tyne*

J. NIÑO-MORA,\*\* *Universitat Pompeu Fabra*

P. S. ANSELL,\*\*\* *University of Newcastle upon Tyne*

### Abstract

The paper concerns a class of discounted restless bandit problems which possess an indexability property. Conservation laws yield an expression for the reward suboptimality of a general policy. These results are utilised to study the closeness to optimality of an index policy for a special class of simple and natural dual speed restless bandits for which indexability is guaranteed. The strong performance of the index policy is confirmed by a computational study.

*Keywords:* Conservation laws; indexability; index policy; restless bandit; suboptimality bound

AMS 2000 Subject Classification: Primary 90B36  
Secondary 93E20

### 1. Introduction

The classical index result of Gittins (1979), (1989) concerned the sequential allocation of effort among a collection of competing projects. In Gittins' solution, each project has a priority index which depends upon its current state and effort is optimally allocated at each decision epoch to whichever project has the highest current index value. In Gittins' classical model, the so-called discounted multiarmed bandit problem, future rewards are discounted and (most crucially) changes of state only occur in projects when effort is allocated to them. Whittle (1988) proposed an important extension to the Gittins' model in which changes of state can also occur in passive projects. He called this the *restless bandit problem* and described a range of applications including the design of clinical trials, aircraft surveillance and worker scheduling. More recent applications reported in the literature include behaviour coordination in robotics (in Faihe and Müller (1998)) and control of a make-to-stock production facility (in Veatch and Wein (1996)).

The restless bandit problem is computationally intractable. Papadimitriou and Tsitsiklis (1999) have proved it to be PSPACE-hard, even in the context of a simple model with deterministic transitions. Hence, the primary research focus concerns the development of heuristic policies which can be shown to be close to optimal. Whittle's (1988) analysis utilised an LP relaxation of his undiscounted problem to produce an index for projects satisfying an *indexability property*. This index is characterised as a Lagrange multiplier associated with a conservation constraint.

---

Received 14 June 2000; revision received 2 January 2002.

\* Current address: School of Management, University of Edinburgh, William Robertson Building, 50 George Square, Edinburgh EH8 9JY, UK. Email address: kevin.glazebrook@ed.ac.uk

\*\* Postal address: Department of Economics and Business, Universitat Pompeu Fabra, E-08005, Barcelona, Spain.

\*\*\* Postal address: School of Mathematics and Statistics, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK.

Weber and Weiss (1990), (1991) took the ideas further by establishing a conjecture of Whittle’s which concerned a form of asymptotic optimality for the policy based on these indices.

Until recently, Whittle’s indexability property has remained something of a mystery in that no simple sufficient conditions for it had been elucidated. Niño-Mora (2001) rectified the situation in a discussion of indexability which drew upon a range of developments of the achievable region account of Gittins indexation given by Bertsimas and Niño-Mora (1996). The current paper takes this work further by developing an approach to policy evaluation for a general class of indexable discounted restless bandit problems. This is then applied to develop an evaluation of an index policy for a special class of dual speed restless bandits which are guaranteed to be indexable.

More specifically, a broad class of discounted restless bandit problems is introduced in Section 2. The conservation laws which these processes respect are given in Section 3 and are shown to form the basis of an approach to policy evaluation. The class of dual speed restless bandit models which is our primary focus is described in Section 4, where their main properties are derived and some simple examples are discussed. For this class, passive projects share the same stochastic dynamics as active projects when they change state, but they change state at a slower (state-dependent) speed. The analysis utilises the results of Section 3 to develop a range of bounds on the degree of reward suboptimality of the index policy for this class. These are then used in Section 5 to derive a performance guarantee for the index policy, expressed in terms of model parameters, most especially the transition speeds of passive projects. This is natural, since when these speeds are all zero, Gittins’ result indicates that the index policy is optimal. The paper concludes in Section 6 with a computational study which evidences a very strong performance of the index policy.

## 2. The discounted restless bandit problem

A single machine or server is available to process  $N$  projects. We denote by  $E$  the finite set which contains all states which any of the projects may occupy during the evolution of the system. At each decision epoch  $t = 0, 1, 2, \dots$ , one project is chosen for processing. We say that the chosen project receives the *active action* ( $a = 1$ ) while the other projects receive the *passive action* ( $a = 0$ ). Projects evolve independently and in a Markovian fashion according to the actions they receive. Should a project be in state  $i$  at time  $t$ , then action  $a$  (either 0 or 1) effects a change of state to  $j$  with probability  $P_{ij}^a$ . This choice of action also earns a reward  $\beta^t r_i^a$ , where  $\beta \in (0, 1)$  is a discount factor. Note that Niño-Mora (2001) explains that we may take  $r_i^0 = 0$ ,  $i \in E$ , without loss of generality. Our goal is to find a scheduling policy  $u$  (a rule for choosing between projects to receive active actions) in the space  $\mathcal{U}$  of stationary policies (which make choices based on current project states only) to maximise the total expected reward earned over an infinite horizon. We write

$$R^{\text{opt}} = \max_{u \in \mathcal{U}} E_u \left\{ \sum_{t=0}^{\infty} \sum_{i \in E} \beta^t r_i^1 I_i^1(t) \right\} \tag{1}$$

where we use  $I_i^a(t)$  for the indicator variable

$$I_i^a(t) = \begin{cases} 1 & \text{if action } a \text{ is taken at a project in state } i \text{ at time } t, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

and where  $a \in \{0, 1\}$  and  $i \in E$ . Plainly, the expectation in (1) is conditional on initial project states. When required, these will be given by the known vector  $\alpha = \{\alpha_i\}_{i \in E}$  with state-indexed

components, where

$$\alpha_i = \begin{cases} 1 & \text{if a project is in state } i \text{ at } t = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

To develop our analysis, we require a matrix of constants, denoted by  $\mathbf{A} := \{A_i^S\}_{i \in E, S \subseteq E}$ . Fix a state  $i \in E$  and a subset  $S \subseteq E$ . Imagine a single project in state  $i$  at time  $t = 0$ , evolving under the  $S$ -active policy  $u_S$  which chooses the active action for the project when in  $S$  and the passive action in  $S^c := E \setminus S$ . In the latter case, it may assist the reader to think of some other project being activated at that point, though this is not crucial in what follows. If we denote by  $\{X(t), t \geq 0\}$  the sequence of states visited by our project under  $u_S$ , then this process is a Markov chain with  $X(0) = i$ ,

$$P\{X(t + 1) = k \mid X(t) = j\} = P_{jk}^1, \quad j \in S,$$

and

$$P\{X(t + 1) = k \mid X(t) = j\} = P_{jk}^0, \quad j \in S^c.$$

Write

$$V_i^S = E_{u_S} \left\{ \sum_{t=0}^{\infty} \beta^t I_S(t) \mid X(0) = i \right\}$$

for the total occupancy time in  $S$  under  $u_S$ , where

$$I_S(t) = \begin{cases} 1 & \text{if } X(t) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$V_i^S = \begin{cases} 1 + \beta \sum_{j \in E} P_{ij}^1 V_j^S, & i \in S, \\ \beta \sum_{j \in E} P_{ij}^0 V_j^S, & i \in S^c. \end{cases} \tag{4}$$

We now define the matrix  $\mathbf{A}$  by means of the quantities  $\{V_i^S\}_{i \in E, S \subseteq E}$  as follows:

$$A_i^S = 1 + \beta \sum_{j \in E} P_{ij}^1 V_j^{S^c} - \beta \sum_{j \in E} P_{ij}^0 V_j^{S^c}, \quad i \in E, S \subseteq E. \tag{5}$$

We impose the following requirement.

**Assumption 1.** For  $i \in E$  and  $S \subseteq E$ ,

$$A_i^S > 0.$$

Observe that Assumption 1 certainly holds for the classical discounted multiarmed bandit problem of Gittins (1979), (1989). If we let  $\mathbf{P}^a$  be the matrix  $\{P_{ij}^a\}_{i, j \in E}$  for  $a \in \{0, 1\}$  and  $\mathbf{I}$  the identity matrix, then this problem is characterised by the condition  $\mathbf{P}^0 = \mathbf{I}$ , namely that passive objects do not change state. Hence, in this case,

$$A_i^S = \begin{cases} 1 + \beta \sum_{j \in S^c} P_{ij}^1 V_j^{S^c} \geq 1, & i \in S, \\ (1 - \beta) V_i^{S^c} \geq 1 - \beta, & i \in S^c, \end{cases} \tag{6}$$

and Assumption 1 holds. Under Assumption 1, we develop a collection of indices  $\{G_i, i \in E\}$ , one for each member of  $E$ . The indices concerned are obtained from the *adaptive greedy algorithm* whose inputs are the active reward rates  $\{r_i^1, i \in E\}$  together with the matrix  $A$ . The algorithm operates as follows:

**Step 1.** Set  $S_{|E|} = E$  and

$$y^{S_{|E|}} = \max \left\{ \frac{r_i^1}{A_i^E}; i \in E \right\}.$$

Select  $\pi_{|E|} \in \arg \max \{r_i^1/A_i^E; i \in E\}$  and set  $G_{\pi_{|E|}} = y^{S_{|E|}}$ .

**Step  $k$ .** For  $k = 2, 3, \dots, |E|$ , set  $S_{|E|-k+1} = S_{|E|-k+2} \setminus \{\pi_{|E|-k+2}\}$  and

$$y^{S_{|E|-k+1}} = \max \left\{ \frac{r_i^1 - \sum_{j=1}^{k-1} A_i^{S_{|E|-j+1}} y^{S_{|E|-j+1}}}{A_i^{S_{|E|-k+1}}}; i \in S_{|E|-k+1} \right\}.$$

Select  $\pi_{|E|-k+1}$  as any maximiser and set  $G_{\pi_{|E|-k+1}} = G_{\pi_{|E|-k+2}} + y^{S_{|E|-k+1}}$ .

Please note that it is straightforward to show that the indices  $\{G_i, i \in E\}$  computed by the algorithm do not depend upon how the  $\pi_k$  are chosen in the event that there is more than one maximiser of the quantity concerned. It follows from the structure of the algorithm that

$$G_{\pi_{|E|}} \geq G_{\pi_{|E|-1}} \geq \dots \geq G_{\pi_1}.$$

Under Assumption 1, these indices coincide with those obtained from the calibration proposed by Whittle (1988) for all choices of the  $r_i^1, i \in E$ , and the system is *indexable* in a sense made precise by Whittle. The indices become the conventional Gittins indices in the multiarmed bandit case. See Niño-Mora (2001) for further discussion of indexability and for proofs of the assertions in this paragraph. Note that Assumption 1 is a sufficient but not necessary condition for indexability.

The policy of prime interest to us is the *index policy* in which, at each decision epoch  $t = 0, 1, 2, \dots$ , the project chosen for processing (i.e. which receives the active action) is that one whose current state has the largest associated index. Such a policy is well known to be optimal for multiarmed bandits (see Gittins (1979), (1989)) and we now proceed to explore its quality for the restless bandit problem.

### 3. Policy evaluation for discounted restless bandit problems

It will simplify our discussion if we introduce the performance variables  $x^a(u) := \{x_i^a(u)\}_{i \in E}$  given by

$$x_i^a(u) = E_u \left\{ \sum_{t=0}^{\infty} \beta^t I_i^a(t) \right\}, \quad a \in \{0, 1\}, i \in E, u \in \mathcal{U}, \tag{7}$$

where the indicator variables  $I_i^a(t)$  are as in (2). For example, (1) may now be written as

$$R^{\text{opt}} = \max_{u \in \mathcal{U}} \left\{ \sum_{i \in E} r_i^1 x_i^1(u) \right\}. \tag{8}$$

Following Bertsimas and Niño-Mora (1996), Glazebrook and Garbe (1999) and Glazebrook and Niño-Mora (2000), we analyse the stochastic optimisation problem in (8) by studying the quantities

$$\sum_{i \in S} A_i^S x_i^1(u), \quad S \subseteq E, \quad u \in \mathcal{U}, \tag{9}$$

where the constants  $A_i^S$  are all components of the matrix  $\mathbf{A}$ , introduced in Section 2. These constants are all positive by Assumption 1. As we shall see, an expression for  $R^u$ , the expected reward earned when control  $u \in \mathcal{U}$  is applied to the restless bandit, may be obtained from the quantities in (9). It is this fact which enables the analysis to proceed. Please note that, in what follows, subscripts  $S$  or  $T$  which are subsets of  $E$  will denote restrictions of the quantities concerned to those subsets. Hence, for example,  $\mathbf{x}_S^a(u) = \{x_i^a(u)\}_{i \in S}$ ,  $\mathbf{V}_S^S = \{V_i^S\}_{i \in S}$ ,  $\mathbf{P}_{ST}^a = \{P_{ij}^a\}_{i \in S, j \in T}$  and so on. The proof of Lemma 1 is a development of that given for Theorem 3 in Niño-Mora (2001). Full details may be found in Glazebrook *et al.* (2000).

**Lemma 1.** (Conservation laws for discounted restless bandits.) *For any control  $u \in \mathcal{U}$  and  $S \subseteq E$ ,*

$$\mathbf{x}_S^1(u) \mathbf{A}_S^S = \left\{ \frac{1}{1 - \beta} - \alpha_E \mathbf{V}_E^{Sc} \right\} + \mathbf{x}_{Sc}^0(u) \mathbf{A}_{Sc}^S. \tag{10}$$

We now use Lemma 1 to develop results which will provide procedures for policy evaluation in restless bandit problems. We firstly observe that, from the structure of the adaptive greedy algorithm in Section 2,

$$r_i^1 = G_{|E|} A_i^{|E|} - \sum_{j=i}^{|E|-1} (G_{j+1} - G_j) A_i^{S_j}, \quad 1 \leq i \leq |E|. \tag{11}$$

In (11), the members of  $E$  have been labelled  $\{1, 2, \dots, |E|\}$  in such a way that

$$G_{|E|} \geq G_{|E|-1} \geq \dots \geq G_2 \geq G_1 \tag{12}$$

and  $S_j = \{j, j - 1, \dots, 1\}$  is the subset of  $E$  of cardinality  $j$  with the lowest indices.

We write

$$R^u = \sum_{i \in E} r_i^1 x_i^1(u)$$

and use  $u^{\text{opt}}$  to denote an optimal control for the problem in (8).

**Corollary 1.** (Policy evaluation.) *For any control  $u \in \mathcal{U}$ ,*

$$R^{\text{opt}} - R^u = \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \{ \mathbf{x}_{S_j}^1(u) \mathbf{A}_{S_j}^{S_j} - \mathbf{x}_{S_j}^1(u^{\text{opt}}) \mathbf{A}_{S_j}^{S_j} \} \tag{13}$$

$$= \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \{ \mathbf{x}_{S_j}^0(u) \mathbf{A}_{S_j}^{S_j} - \mathbf{x}_{S_j}^0(u^{\text{opt}}) \mathbf{A}_{S_j}^{S_j} \}. \tag{14}$$

*Proof.* From (11) we deduce that

$$R^u = G_{|E|} \mathbf{x}_E^1(u) \mathbf{A}_E^E - \sum_{i \in S_{|E|-1}} \sum_{j=i}^{|E|-1} (G_{j+1} - G_j) A_i^{S_j} x_i^1(u). \tag{15}$$

But, by its definition,  $A_i^E = 1$ , for  $i \in E$ , and so the first summand in (15) is equal to  $G_{|E|}/(1 - \beta)$  for all controls  $u \in \mathcal{U}$ . We then deduce from (15) that

$$R^{\text{opt}} - R^u = \sum_{i \in S_{|E|-1}} \left\{ \sum_{j=i}^{|E|-1} (G_{j+1} - G_j) A_i^{S_j} x_i^1(u) - \sum_{j=i}^{|E|-1} (G_{j+1} - G_j) A_i^{S_j} x_i^1(u^{\text{opt}}) \right\}, \tag{16}$$

and we obtain (13) by interchanging the order of summation in (16). Equation (14) is an immediate consequence of Lemma 1.

We now proceed in Section 4 to utilise Corollary 1 in a simple and natural class of restless bandit models for which Assumption 1 is guaranteed to hold.

### 4. Dual speed restless bandit problems

We consider a special case of the general model discussed in Sections 2 and 3 in which

$$P_{ij}^0 = \begin{cases} \varepsilon_i P_{ij}^1, & i \neq j, \\ (1 - \varepsilon_i) + \varepsilon_i P_{ii}^1, & i = j, \end{cases} \tag{17}$$

for some  $\varepsilon_i \in [0, 1]$ ,  $i \in E$ . We write

$$\hat{\varepsilon} = \max_{i \in E} \varepsilon_i. \tag{18}$$

Hence, each project may be thought to have two transition speeds (active and passive) in each state. Under the two actions, the sojourn time in each state  $i$  will be geometric with means  $(1 - P_{ii}^a)^{-1}$ ,  $a \in \{0, 1\}$ . However, conditional upon a change of state, the transition probabilities are the same for both actions. If  $\hat{\varepsilon} = 0$ , then we have a discounted multiarmed bandit in which passive projects are frozen and index policies are optimal. In addition to its intrinsic interest, an advantage of the stochastic structure in (17) is that the degree of restlessness (interpreted as movement under the passive action) is easily measured through the  $\varepsilon_i$  and the model then becomes a natural vehicle for an investigation of the quality of index policies in a restless environment. All active reward rates are nonnegative. In such a set-up, Assumption 1 is always satisfied, as the following calculations indicate.

Firstly note that, upon substitution from (17) into (4) we obtain, for fixed  $S \subseteq E$ ,

$$V_i^{S^c} = \frac{\varepsilon_i \beta}{1 - \beta(1 - \varepsilon_i)} \sum_{j \in E} P_{ij}^1 V_j^{S^c}, \quad i \in S, \tag{19}$$

and

$$V_i^{S^c} = 1 + \beta \sum_{j \in E} P_{ij}^1 V_j^{S^c}, \quad i \in S^c. \tag{20}$$

Hence, we deduce from (5) that

$$A_i^S = 1 + \frac{\beta(1 - \beta)(1 - \varepsilon_i)}{1 - \beta(1 - \varepsilon_i)} \sum_{j \in E} P_{ij}^1 V_j^{S^c}, \quad i \in S, \tag{21}$$

and

$$A_i^S = \varepsilon_i + (1 - \beta)(1 - \varepsilon_i) V_i^{S^c}, \quad i \in S^c. \tag{22}$$

The following result is immediate from (21) and (22).

**Lemma 2.** Under the dual speed restless bandit model, Assumption 1 holds.

**Lemma 3.** For the dual speed restless bandit model,  $\mathbf{V}_E^{S^c \cup \{i\}} \geq \mathbf{V}_E^{S^c}$  for  $i \in S$  and  $S \subseteq E$ .

*Proof.* Fix  $i \in S$ . From (19) and (20),  $\mathbf{V}_E^{S^c}$  is obtained by solving

$$\begin{aligned} V_i^{S^c} &= \frac{\varepsilon_i \beta}{1 - \beta(1 - \varepsilon_i)} \sum_{k \in E} P_{ik}^1 V_k^{S^c}, \\ V_j^{S^c} &= \frac{\varepsilon_j \beta}{1 - \beta(1 - \varepsilon_j)} \sum_{k \in E} P_{jk}^1 V_k^{S^c}, \quad j \in S \setminus \{i\}, \\ V_j^{S^c} &= 1 + \beta P_{ji}^1 V_i^{S^c} + \beta \sum_{k \in S \setminus \{i\}} P_{jk}^1 V_k^{S^c} + \beta \sum_{k \in S^c} P_{jk}^1 V_k^{S^c}, \quad j \in S^c. \end{aligned} \tag{23}$$

We further develop the transform  $T_1 : \mathbb{R}^{|E|} \rightarrow \mathbb{R}^{|E|}$  by defining

$$\begin{aligned} \{T_1(\mathbf{v})\}_i &= 1 + \beta P_{ii}^1 v_i + \beta \sum_{k \in S \setminus \{i\}} P_{ik}^1 v_k + \beta \sum_{k \in S^c} P_{ik}^1 v_k, \\ \{T_1(\mathbf{v})\}_j &= \frac{\varepsilon_j \beta}{1 - \beta(1 - \varepsilon_j)} \sum_{k \in E} P_{jk}^1 v_k, \quad j \in S \setminus \{i\}, \\ \{T_1(\mathbf{v})\}_j &= 1 + \beta P_{ji}^1 v_i + \beta \sum_{k \in S \setminus \{i\}} P_{jk}^1 v_k + \beta \sum_{k \in S^c} P_{jk}^1 v_k, \quad j \in S^c, \end{aligned} \tag{24}$$

where, in (24),  $\mathbf{v} = \{v_k\}_{k \in E}$  is an arbitrarily chosen  $|E|$ -vector. Please note that  $T_1$  is nondecreasing in that

$$\mathbf{v} \geq \mathbf{w} \implies T_1(\mathbf{v}) \geq T_1(\mathbf{w}) \tag{25}$$

and also that it follows from (19) and (20) that  $\mathbf{V}_E^{S^c \cup \{i\}}$  is a fixed point of  $T_1$ . It is a consequence of the inequality

$$\frac{\varepsilon_i \beta}{1 - \beta(1 - \varepsilon_i)} \leq \beta$$

together with (23) and (24) that

$$V_i^{S^c} < \{T_1(\mathbf{V}_E^{S^c})\}_i. \tag{26}$$

In addition, we have

$$V_j^{S^c} = \{T_1(\mathbf{V}_E^{S^c})\}_j, \quad j \neq i. \tag{27}$$

We conclude from (26) and (27) that

$$\mathbf{V}_E^{S^c} \leq T_1(\mathbf{V}_E^{S^c}). \tag{28}$$

Now introduce  $T_n$  as the  $n$ th iterate of  $T_1$ , namely

$$T_n(\mathbf{v}) = T_1\{T_{n-1}(\mathbf{v})\}, \quad n > 1.$$

Finally, an appeal to (25), (28) and the fact that all components of  $\mathbf{V}_E^{S^c \cup \{i\}}$  are bounded above by  $(1 - \beta)^{-1}$  yields

$$\mathbf{V}_E^{S^c} \leq \lim_{n \rightarrow \infty} T_n(\mathbf{V}_E^{S^c}) = \mathbf{V}_E^{S^c \cup \{i\}}.$$

The result follows.

**Corollary 2.** For the dual speed restless bandit model,  $A_{S \setminus \{i\}}^{S \setminus \{i\}} \geq A_{S \setminus \{i\}}^S$  for  $i \in S$  and  $S \subseteq E$ .

*Proof.* The result is an immediate consequence of (21) and Lemma 3.

**Theorem 1.** For the dual speed restless bandit model, all indices are nonnegative.

*Proof.* We use an induction argument based on the operation of the adaptive greedy algorithm. We firstly observe that  $A_i^E = 1$ ,  $i \in E$ , and hence

$$G_{|E|} = r_{|E|}^1 = \max_{i \in E} r_i^1 \geq 0,$$

where we continue to use the numbering of the states in (12) and the notation  $S_j = \{j, j - 1, \dots, 1\}$ .

We now suppose that  $G_i \geq 0$  for  $k + 1 \leq i \leq |E|$  and consider  $G_k$ . From the operation of the algorithm, we deduce that

$$G_k - G_{k+1} = \frac{r_k^1 - \sum_{j=k+1}^{|E|-1} A_k^{S_j} (G_j - G_{j+1}) - G_{|E|}}{A_k^{S_k}}. \quad (29)$$

However,  $G_k - G_{k+1} \leq 0$  and, further, it is a consequence of Corollary 2 that  $A_k^{S_{k+1}} \leq A_k^{S_k}$ . Using these facts in (29), we conclude that

$$\begin{aligned} G_k &\geq G_{k+1} + \frac{r_k^1 - \sum_{j=k+1}^{|E|-1} A_k^{S_j} (G_j - G_{j+1}) - G_{|E|}}{A_k^{S_{k+1}}} \\ &= \frac{r_k^1}{A_k^{S_{k+1}}} + \frac{\sum_{j=k+2}^{|E|} G_j (A_k^{S_{j-1}} - A_k^{S_j})}{A_k^{S_{k+1}}} \\ &\geq 0. \end{aligned} \quad (30)$$

Note that the inequality (30) utilises the inductive hypothesis ( $G_j \geq 0$ ,  $k + 2 \leq j \leq |E|$ ), Corollary 2 ( $A_k^{S_{j-1}} - A_k^{S_j} \geq 0$ ,  $k + 2 \leq j \leq |E|$ ) and Lemma 2 ( $A_k^{S_{k+1}} \geq 0$ ). The induction goes through and the result is proved.

**Corollary 3.** (Policy evaluation for dual speed restless bandit problems.) For the dual speed restless bandit model,

$$\begin{aligned} R^{\text{opt}} - R^u &\leq \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \left\{ \mathbf{x}_{S_j}^1(u) A_{S_j}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j}^1(v) A_{S_j}^{S_j} \right] \right\} \\ &= \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \left\{ \mathbf{x}_{S_j^c}^0(u) A_{S_j^c}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j^c}^0(v) A_{S_j^c}^{S_j} \right] \right\} \end{aligned} \quad (31)$$

$$\begin{aligned} &\leq \left( \max_{i \in E} r_i^1 \right) \max_{1 \leq j \leq |E|-1} \left\{ \mathbf{x}_{S_j}^1(u) A_{S_j}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j}^1(v) A_{S_j}^{S_j} \right] \right\} \\ &= \left( \max_{i \in E} r_i^1 \right) \max_{1 \leq j \leq |E|-1} \left\{ \mathbf{x}_{S_j^c}^0(u) A_{S_j^c}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j^c}^0(v) A_{S_j^c}^{S_j} \right] \right\}, \end{aligned} \quad (32)$$

the above infima being taken over  $v \in \mathcal{U}$ .

*Proof.* The inequality and equation up to (31) follow from Lemma 1 and Corollary 1. From Theorem 1 and the structure of the adaptive greedy algorithm,

$$\sum_{j=1}^{|E|-1} (G_{j+1} - G_j) = G_{|E|} - G_1 \leq G_{|E|} = \max_{i \in E} r_i^1, \tag{33}$$

with all terms in the summation on the left-hand side of (33) nonnegative. The inequality and equation up to (32) now follow from (31) and (33).

**Remark 1.** The reader should note that, in going from the exact expression for  $R^{\text{opt}} - R^u$  in Corollary 1 to the upper bound in Corollary 3, the quantities

$$\mathbf{x}_{S_j}^1(u^{\text{opt}}) \mathbf{A}_{S_j}^{S_j} \quad \text{and} \quad \mathbf{x}_{S_j^c}^0(u^{\text{opt}}) \mathbf{A}_{S_j^c}^{S_j}$$

have been replaced by the (possibly) smaller

$$\inf_v \left[ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right] \quad \text{and} \quad \inf_v \left[ \mathbf{x}_{S_j^c}^0(v) \mathbf{A}_{S_j^c}^{S_j} \right]$$

respectively. This is motivated as follows. Firstly, we do not know what the optimal policy is and, hence, the form of the expression in Corollary 1 is not of immediate use. Second, in the case of discounted multiarmed bandits ( $\hat{\varepsilon} = 0$ ), the first two bounds given in Corollary 3 are equal to the exact expressions of Corollary 1 for all  $u \in \mathcal{U}$ . This arises from the fact, established by Bertsimas and Niño-Mora (1996), that the infima in Corollary 3 are achieved for multiarmed bandits by all controls  $v : S_j^c \rightarrow S_j$  which give priority to  $S_j^c$ . However, the index policy is such a control and is indeed optimal in the multiarmed bandit case. Hence,

$$\mathbf{x}_{S_j}^1(u^{\text{opt}}) \mathbf{A}_{S_j}^{S_j} = \inf_v \left\{ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right\} \tag{34}$$

when  $\hat{\varepsilon} = 0$ . Equation (34) would certainly lead us to expect that the expressions (31) should provide reasonably tight upper bounds on  $R^{\text{opt}} - R^u$  for our class of dual speed restless models, most especially when  $\hat{\varepsilon}$  is small. Finally, performance guarantees based on conservation laws of the general type given in Corollary 3 have proved effective as analytical tools in other problem contexts. See, for example, Glazebrook and Wilkinson (2000).

**Example 1.** Consider a problem in which  $E = \{1, 2, 3, 4\}$  with  $P_{12}^1 = P_{22}^1 = P_{34}^1 = P_{44}^1 = 1$ ,  $r_1^1 = 0.95$ ,  $r_2^1 = 0$ ,  $r_3^1 = 1$ ,  $r_4^1 = 0$ ,  $\varepsilon_1 = 0.1$ ,  $\varepsilon_3 = 0.05$ ,  $\beta = 0.95$  and  $N = 2$ . Project 1 is initially in state 1 and because of the stochastic dynamics can only occupy states 1 and 2, while project 2 is initially in state 3 and can only occupy states 3 and 4. The indices corresponding to states 1 and 3 are 0.95 and 1 respectively and an index policy will process project 2 at time 0 and project 1 at time 1. The corresponding expected reward is

$$r_3^1 + \beta(1 - \varepsilon_1)r_1^1 = 1.81225.$$

However, the expected reward for a policy which processes project 1 at time 0 and project 2 at time 1 is

$$r_1^1 + \beta(1 - \varepsilon_3)r_3^1 = 1.8525$$

and, hence, outperforms the index policy. The issue here is that project 1 is more attractive at time 0, primarily because of its greater transition speed under the passive action ( $\varepsilon_1 > \varepsilon_3$ ). An imperative in this example is to make a choice at time 0 which minimises the probability of a passive transition. The indices of states 1 and 3 do not reflect this imperative.

**Example 2.** The reader might speculate that there are simply stated conditions on model parameters which render the indices for our restless model equal to their multiarmed bandit equivalents (i.e. when  $\hat{\varepsilon} = 0$ ) and that the index policy might perform well under such conditions. The following example is strongly suggestive of the fact that this line of enquiry is unlikely to bear fruit. Let  $E = \{1, 2, 3, 4\}$  with  $P_{ij}^1 = 0$  when  $i \in \{1, 2\}$ ,  $j \in \{3, 4\}$  and when  $i \in \{3, 4\}$ ,  $j \in \{1, 2\}$ . We have two project types, with one type occupying states 1 and 2 and the other occupying states 3 and 4. Consider projects of the first type and suppose that  $r_1^1 > r_2^1$ . Deployment of (21) and the adaptive greedy algorithm serve to show that the index for state 1 is  $r_1^1$  while that for state 2 is given by the expression

$$r_1^1 + \frac{1 - \beta P_{11}^1 + \beta \varepsilon_2 P_{21}^1}{1 - \beta P_{11}^1 + \beta P_{21}^1} (r_2^1 - r_1^1),$$

which can only equal the multiarmed bandit equivalent when either  $\varepsilon_2 = 0$  or  $P_{22}^1 = 1$ . Further, when the latter is the case, Example 1 furnishes us with an example for which the index policy is not optimal.

### 5. On the closeness to optimality of the index policy

In this section, we shall utilise Corollary 3 to analyse the closeness to optimality of the index policy  $u_G$  (described in Section 2). The inequality (31) with  $u = u_G$  yields

$$R^{\text{opt}} - R^{u_G} \leq \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \left\{ \mathbf{x}_{S_j}^1(u_G) \mathbf{A}_{S_j}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right] \right\} =: B^{u_G}. \quad (35)$$

Fix  $j$  and consider the quantity

$$\mathbf{x}_{S_j}^1(u_G) \mathbf{A}_{S_j}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right]. \quad (36)$$

The policy  $u_G$  enforces the priority  $S_j^c \rightarrow S_j$ . Suppose, in the light of Remark 1, that there exists a policy attaining the infimum in (36) which also enforces this priority. This will certainly be the case if  $\hat{\varepsilon}$  is small enough; see the proof of Theorem 2. Under these assumptions,

$$\mathbf{x}_{S_j}^1(u_G) \mathbf{A}_{S_j}^{S_j} - \inf_v \left[ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right] \leq \sup_{u_1, u_2} \left\{ \mathbf{x}_{S_j}^1(u_1) \mathbf{A}_{S_j}^{S_j} - \mathbf{x}_{S_j}^1(u_2) \mathbf{A}_{S_j}^{S_j} \right\}, \quad (37)$$

where the supremum is over all policies  $u_1, u_2$  which enforce the priority  $S_j^c \rightarrow S_j$ . To take these ideas further, we now require a short digression on the discounted multiarmed bandit problem ( $\hat{\varepsilon} = 0$ ).

#### 5.1. Digression: multiarmed bandit problem

Suppose that  $\varepsilon_i = 0$  for  $i \in E$ . Fix a subset  $S \subseteq E$  and consider the vector  $\mathbf{W}_E^{S^c}$  defined by

$$\mathbf{W}_E^{S^c} = \mathbf{1}_E + \beta \mathbf{P}_{E S^c}^1 \mathbf{W}_{S^c}^{S^c}, \quad (38)$$

where  $\mathbf{1}_E$  is a vector of 1s.

From (6),  $\mathbf{W}_S^{S^c}$  coincides with the appropriate form of  $\mathbf{A}_S^S$  for this multiarmed bandit case. Further, and as indicated in Remark 1, when  $\hat{\varepsilon} = 0$  the infimum

$$\inf_v \left[ \mathbf{x}_S^1(v) \mathbf{A}_S^S \right] = \inf_v \left[ \mathbf{x}_S^1(v) \mathbf{W}_S^{S^c} \right] \quad (39)$$

is attained by all controls enforcing the priority  $S^c \rightarrow S$  and takes the value

$$\prod_{i \in S^c} \frac{1 - \alpha_i(1 - \beta)W_i^{S^c}}{1 - \beta}. \tag{40}$$

See Bertsimas and Niño-Mora (1996). Now introduce the conventional state descriptor  $\mathbf{i} = (i_1, i_2, \dots, i_N)$  in which  $i_m$  is the state of project  $m$ ,  $1 \leq m \leq N$ . Hence, the system has initial state  $\mathbf{i}$  if and only if  $\alpha_{i_m} = 1$  for  $1 \leq m \leq N$ . We shall refer to the quantity in (40) as  $W^S(\mathbf{i})$ .

In light of (39) and (40), we develop (37) further by writing

$$\sup_{u_1, u_2} \left\{ \mathbf{x}_{S_j}^1(u_1)A_{S_j}^{S_j} - \mathbf{x}_{S_j}^1(u_2)A_{S_j}^{S_j} \right\} \leq 2 \sup_u \left\{ \left| \mathbf{x}_{S_j}^1(u)A_{S_j}^{S_j} - W^{S_j}(\mathbf{i}) \right| \right\}, \tag{41}$$

where the supremum on the right-hand side is over all controls enforcing the priority  $S_j^c \rightarrow S_j$ . Lemma 5 provides an upper bound on the right-hand side of (41) which enables us to bound  $B^{u^G}$  above via (35).

**5.2. Digression over**

Before proceeding to the statement and proof of Lemma 5 we require the information in Lemma 4 concerning how well  $W_S^{S^c}$  approximates  $A_S^S$  for our dual speed restless bandit problem. The proof may be found in Appendix A.

**Lemma 4.** *For the dual speed restless bandit model,*

$$W_S^{S^c} - \frac{\hat{\epsilon}\beta}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \mathbf{1}_S \leq A_S^S \leq W_S^{S^c} \left\{ 1 + \frac{\hat{\epsilon}\beta^2}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \right\}, \quad S \subseteq E.$$

In the proof of Lemma 5, we shall write  $\mathbf{i} = (i_1, \mathbf{i}^1)$  when we need to focus on the dynamics of a particular project, 1 say. Since we now need to emphasise dependence on the initial state in much of the notation, we write, for each  $S \subseteq E$  and  $u \in \mathcal{U}$ ,

$$\mathbf{x}_S^1(u)A_S^S = A_u^S(\mathbf{i}),$$

with  $\mathbf{i}$  a specified initial state.

**Lemma 5.** *For the dual speed restless bandit model and a subset  $S \subseteq E$ ,*

$$\sup_u \left\{ \left| A_u^S(\mathbf{i}) - W^S(\mathbf{i}) \right| \right\} \leq \frac{(N - 1)\hat{\epsilon}\beta}{(1 - \beta)^2} + \frac{\hat{\epsilon}\beta}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} W^S(\mathbf{i}) + O\{\hat{\epsilon}^2\}, \tag{42}$$

for any initial state  $\mathbf{i}$ , where the supremum is over all controls  $u : S^c \rightarrow S$ .

*Proof.* Consider first an initial state  $\mathbf{i}$  for which  $i_1 \in S^c$  and a control  $u : S^c \rightarrow S$  which activates project 1 at time  $t = 0$ . According to the objective  $A_u^S$ , rewards are earned only when an  $S$ -state project is activated, and so we have

$$A_u^S(\mathbf{i}) = \beta \sum_{j \in E} P_{i_1 j}^1 E_{X^1} \{ A_u^S(j, X^1) \}. \tag{43}$$

Here,  $X^1$  denotes the random state in which projects  $2, \dots, N$  find themselves at time  $t = 1$ , following a passive transition in each from initial state  $\mathbf{i}^1$ . In (43), we will now utilise the fact that

$$W^S(\mathbf{i}) = \beta \sum_{j \in E} P_{i_1 j}^1 W^S(j, \mathbf{i}^1). \tag{44}$$

This equality follows from the fact that, when  $\hat{\varepsilon} = 0$ , the infimum in (39) is achieved by any control giving priority to  $S^c$ . We now use (44) to rewrite (43) as

$$A_u^S(\mathbf{i}) - W^S(\mathbf{i}) = \beta \sum_{j \in E} P_{i_1 j}^1 \mathbb{E}_{X^1} \{W^S(j, \mathbf{X}^1) - W^S(j, \mathbf{i}^1)\} + \beta \sum_{j \in E} P_{i_1 j}^1 \mathbb{E}_{X^1} \{A_u^S(j, \mathbf{X}^1) - W^S(j, \mathbf{X}^1)\}. \quad (45)$$

Consider now an initial state  $\mathbf{i}$  for which  $i_m \in S$ ,  $1 \leq m \leq N$ , and a control  $u : S^c \rightarrow S$  which activates project 1 at time  $t = 0$ . In this case,

$$A_u^S(\mathbf{i}) = A_{i_1}^S + \beta \sum_{j \in E} P_{i_1 j}^1 \mathbb{E}_{X^1} \{A_u^S(j, \mathbf{X}^1)\}, \quad (46)$$

and, correspondingly, for the multiarmed bandit problem ( $\hat{\varepsilon} = 0$ ),

$$W^S(\mathbf{i}) = W_{i_1}^{S^c} + \beta \sum_{j \in E} P_{i_1 j}^1 W^S(j, \mathbf{i}^1). \quad (47)$$

Combining (46) and (47) we deduce that

$$A_u^S(\mathbf{i}) - W^S(\mathbf{i}) = A_{i_1}^S - W_{i_1}^{S^c} + \beta \sum_{j \in E} P_{i_1 j}^1 \mathbb{E}_{X^1} \{W^S(j, \mathbf{X}^1) - W^S(j, \mathbf{i}^1)\} + \beta \sum_{j \in E} P_{i_1 j}^1 \mathbb{E}_{X^1} \{A_u^S(j, \mathbf{X}^1) - W^S(j, \mathbf{X}^1)\}. \quad (48)$$

From (45) and (48), it is possible to regard  $A_u^S(\mathbf{i}) - W^S(\mathbf{i})$  as a discounted reward earned by our dual speed restless bandit over an infinite horizon from initial state  $\mathbf{i}$  under the control  $u : S^c \rightarrow S$ . To achieve this, we need to suppose that the reward earned when control  $u$  activates project  $m$  with the process in state  $\mathbf{j}$  is given (in an obvious notation) by

$$\delta_m(\mathbf{j}) = (A_{j_m}^S - W_{j_m}^{S^c})I(j_m \in S) + \beta \sum_{k \in E} P_{j_m k}^1 \mathbb{E}_{X^m} \{W^S(k, \mathbf{X}^m) - W^S(k, \mathbf{j}^m)\}. \quad (49)$$

We now bound  $A_u^S(\mathbf{i}) - W^S(\mathbf{i})$  by means of bounds on the right-hand side of (49). Note firstly that Lemma 4 furnishes us with bounds on the first term on the right-hand side. We now proceed to consider the second term via the expression in (40).

To utilise (40), we introduce the constants

$$w_i^S = \begin{cases} 1 - (1 - \beta)W_i^{S^c}, & i \in S^c, \\ 1, & i \in S. \end{cases}$$

By straightforward algebra, we have that, for given  $\mathbf{j}$  and  $k \in E_m$ ,

$$\mathbb{E}_{X^m} \{W^S(k, \mathbf{X}^m) - W^S(k, \mathbf{j}^m)\} = w_k^S \sum_{n \neq m} \hat{\delta}_n \prod_{n' \neq m, n} w_{j_{n'}}^S + O\{(\hat{\varepsilon})^2\}, \quad (50)$$

where

$$\hat{\delta}_n = \begin{cases} \varepsilon_{j_n} \left( W_{j_n}^{S^c} - \sum_{l \in S^c} P_{j_n l}^1 W_l^{S^c} \right), & j_n \in S^c, \\ -\varepsilon_{j_n} \sum_{l \in S^c} P_{j_n l}^1 W_l^{S^c}, & j_n \in S. \end{cases} \quad (51)$$

Substituting from (50) and (51) into the second term on the right-hand side of (49) yields

$$\beta \left| \sum_{k \in E} P_{jmk}^1 \mathbb{E}_{X^m} \{W^S(k, X^m) - W^S(k, j^m)\} \right| \leq \frac{(N-1)\hat{\epsilon}\beta}{1-\beta} + O\{\{\hat{\epsilon}\}^2\}. \tag{52}$$

Further, we can infer from Lemma 4 that

$$\begin{aligned} |A_{j_m}^S - W_{j_m}^{S^c}| I(j_m \in S) &\leq \frac{\hat{\epsilon}\beta}{\{1-\beta(1-\hat{\epsilon})\}(1-\beta)} W_{j_m}^{S^c} I(j_m \in S) \\ &\leq \frac{\hat{\epsilon}\beta}{\{1-\beta(1-\hat{\epsilon})\}(1-\beta)} A_{j_m}^S I(j_m \in S) + O\{\{\hat{\epsilon}\}^2\}. \end{aligned} \tag{53}$$

Combining (49) and the material preceding with (52) and (53) we infer that

$$|A_u^S(\mathbf{i}) - W^S(\mathbf{i})| \leq \frac{(N-1)\hat{\epsilon}\beta}{(1-\beta)^2} + \frac{\hat{\epsilon}\beta}{\{1-\beta(1-\hat{\epsilon})\}(1-\beta)} A_u^S(\mathbf{i}) + O\{\{\hat{\epsilon}\}^2\}, \tag{54}$$

from which we conclude that

$$A_u^S(\mathbf{i}) \leq |A_u^S(\mathbf{i}) - W^S(\mathbf{i})| + W^S(\mathbf{i}) \leq W^S(\mathbf{i}) + O(\hat{\epsilon}). \tag{55}$$

The result now follows by combining (54) and (55).

**Theorem 2.** (Evaluation of the index policy.) *For the dual speed restless bandit model,*

$$\begin{aligned} R^{\text{opt}}(\mathbf{i}) - R^{u_G}(\mathbf{i}) &\leq 2 \left( \max_{i \in E} r_i^1 \right) \frac{(N-1)\hat{\epsilon}\beta}{(1-\beta)^2} \\ &\quad + 2 \sum_{j=1}^{|E|-1} (G_{j+1} - G_j) \frac{\hat{\epsilon}\beta}{\{1-\beta(1-\hat{\epsilon})\}(1-\beta)} W^{S_j}(\mathbf{i}) + O\{\{\hat{\epsilon}\}^2\}. \end{aligned} \tag{56}$$

*Proof.* We introduce the notation

$$W_u^S(\mathbf{i}) = \mathbf{x}_S^1(u) W_S^{S^c},$$

where  $u \in \mathcal{U}$ ,  $S \subseteq E$ ,  $\mathbf{i}$  is a general initial state and multiarmed bandit dynamics ( $\hat{\epsilon} = 0$ ) are assumed to apply. The term  $A_u^S(\mathbf{i})$  is as before. A much simplified version of the proof of Lemma 5 serves to show that

$$|A_u^S(\mathbf{i}) - W_u^S(\mathbf{i})| \leq O(\hat{\epsilon}). \tag{57}$$

We now note that a calculation for the multiarmed bandit case shows that, when  $S^c$ -states are present in  $\mathbf{i}$ , any control  $u$  which fails to implement the priority  $S^c \rightarrow S$  at  $t = 0$  satisfies

$$W_u^S(\mathbf{i}) \geq W^S(\mathbf{i}) + (1-\beta). \tag{58}$$

It now follows from (57), (58) and standard dynamic programming arguments that there must exist an  $\epsilon > 0$  such that, when  $\hat{\epsilon} \in (0, \epsilon)$ , no control for the dual speed restless bandit model which fails to implement the priority  $S^c \rightarrow S$  can minimise  $A_u^S(\mathbf{i})$  for any  $\mathbf{i}$ .

Henceforth, suppose that we are considering a model for which  $\hat{\epsilon} \in (0, \epsilon)$  and consider the first bound on  $R^{\text{opt}} - R^u$  described in Corollary 3. From the above argument, we may assume that the infimum

$$\inf_v \{A_v^{S_j}(\mathbf{i})\}$$

over all admissible controls  $\mathcal{U}$  is attained by a control  $\hat{u}_j$ , say, which gives priority to  $S_j^c$ . Further, we note that, by construction, the index policy also gives priority to  $S_j^c$ .

We can now utilise Lemma 5 and write

$$\begin{aligned} & A_{u_G}^{S_j}(\mathbf{i}) - \inf_v \{A_v^{S_j}(\mathbf{i})\} \\ &= A_{u_G}^{S_j}(\mathbf{i}) - A_{\hat{u}_j}^{S_j}(\mathbf{i}) \\ &\leq |A_{u_G}^{S_j}(\mathbf{i}) - W^{S_j}(\mathbf{i})| + |W^{S_j}(\mathbf{i}) - A_{\hat{u}_j}^{S_j}(\mathbf{i})| \\ &\leq 2 \sup_u \{|A_u^{S_j}(\mathbf{i}) - W^{S_j}(\mathbf{i})|\} \\ &\leq 2 \frac{(N-1)\hat{\varepsilon}\beta}{(1-\beta)^2} + 2 \frac{\hat{\varepsilon}\beta}{\{1-\beta(1-\hat{\varepsilon})\}(1-\beta)} W^{S_j}(\mathbf{i}) + O\{\{\hat{\varepsilon}\}^2\}, \end{aligned} \tag{59}$$

where the supremum is over all controls  $u : S_j^c \rightarrow S_j$ . The result now follows by substituting from (59) into (31) and utilising (33).

**Remark 2.** The proof of Theorem 2 makes rigorous the approach to the bounding of  $B^{uG}$  sketched out at the beginning of this section.

### 6. Computational study

To complement the theoretical analysis of Sections 4 and 5, we have conducted extensive numerical investigations into the performance of index policies for the dual speed restless bandit model, including cases in which  $\hat{\varepsilon}$  is not close to zero. Some of our results are summarised in Tables 1–4. Table 1 concerns problems which are constructed as follows:

- (i) All problems concern models with two constituent four-state projects. All eight members of  $E$  share a common value of  $\varepsilon_i$ , that is,  $\varepsilon_i = \hat{\varepsilon}$  for  $i \in E$ . We consider 36 different values of the pair  $(\hat{\varepsilon}, \beta)$  for which  $\hat{\varepsilon}$  varies between 0.01 and 0.75 and  $\beta$  between 0.80 and 0.99. Hence, we consider cases which, on one hand are close to the multiarmed bandit case ( $\hat{\varepsilon} = 0$ ) for which the index policy is known to be optimal and, on the other, are moderately close to the  $\hat{\varepsilon} = 1$  case for which all (nonidling) policies are optimal. The eight numbers in Table 1 for each  $(\hat{\varepsilon}, \beta)$  pair summarise results from 500 problems chosen at random by the mechanisms in (ii) and (iii) below. Please note that a new set of problems is generated for each  $(\hat{\varepsilon}, \beta)$ -pair. Hence, this part of the study concerns 18 000 independently generated problems in all.
- (ii) An active one-step transition matrix is constructed for each project by sampling independently from a uniform[0.1, 0.9] distribution with a subsequent normalisation across rows.
- (iii) For each problem, the eight active rewards are drawn independently from a uniform[1, 5] distribution.

For each of the 18 000 problems, the matrix  $A$  was computed from (4) and (5) and indices were obtained from the adaptive greedy algorithm described in Section 2 and the index policy thereby constructed. The expected rewards  $R^{\text{opt}}$  and  $R^{uG}$  were then computed, using our own code to implement a standard DP value iteration. In addition, the upper bound  $B^{uG}$  in (35),

TABLE 1: The performance of the index policy for a class of dual speed restless bandit problems with  $\hat{\varepsilon}_1 = \hat{\varepsilon}_2 = \hat{\varepsilon}$ .

$\hat{\varepsilon}$	$\beta$	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
0.010	0.80	100.00	100.00	0.0000	0.0000	0.0000	47.00	0.0375	0.0074
0.025		99.40	99.60	0.0120	0.0099	0.0060	18.20	0.0977	0.0116
0.050		95.20	100.00	0.0000	0.0000	0.0480	7.60	0.1820	0.0208
0.100		86.80	99.60	0.0424	0.0015	0.1320	0.60	0.3070	0.0390
0.150		85.40	93.20	0.0900	0.0047	0.1580	3.40	0.4159	0.0602
0.200		79.40	90.80	0.2646	0.0075	0.2220	3.80	0.5160	0.0767
0.250		75.00	88.80	0.4192	0.0257	0.2640	2.60	0.5963	0.0936
0.500		39.80	99.80	0.0028	0.0016	0.7480	0.00	0.7987	0.1377
0.750	28.40	90.20	0.3360	0.0180	1.0580	0.00	0.6742	0.1083	
0.010	0.90	100.00	100.00	0.0000	0.0000	0.0000	21.60	0.0307	0.0046
0.025		98.40	98.60	0.0204	0.0101	0.0160	8.40	0.0786	0.0113
0.050		92.80	94.40	0.0689	0.0014	0.0720	3.80	0.1624	0.0237
0.100		88.00	85.40	0.0806	0.0063	0.1200	0.00	0.2776	0.0484
0.150		84.00	79.80	0.1970	0.0125	0.1600	0.00	0.4146	0.0754
0.200		75.20	75.20	0.2255	0.0125	0.2480	0.00	0.4976	0.0979
0.250		69.40	78.60	0.2536	0.0220	0.3060	0.00	0.5873	0.1213
0.500		34.80	88.60	0.0822	0.0216	0.8160	0.00	0.7278	0.1871
0.750	25.20	88.00	0.3055	0.0329	1.1680	0.00	0.5212	0.1475	
0.010	0.95	93.00	98.40	0.0066	0.0045	0.0700	12.60	0.0222	0.0036
0.025		92.00	95.80	0.0287	0.0007	0.0800	5.20	0.0566	0.0099
0.050		87.80	77.00	0.0420	0.0027	0.1220	5.00	0.1174	0.0231
0.100		82.80	63.20	0.1105	0.0043	0.2080	3.40	0.2387	0.0479
0.150		72.20	65.00	0.1800	0.0116	0.3120	0.00	0.3729	0.0788
0.200		68.60	67.60	0.1862	0.0130	0.3480	0.00	0.4731	0.1114
0.250		61.80	68.00	0.2079	0.0142	0.4160	0.00	0.5352	0.1355
0.500		36.20	77.80	0.1654	0.0245	0.9120	0.00	0.7507	0.2146
0.750	23.00	85.80	0.1435	0.0258	1.2960	0.00	0.5117	0.1673	
0.010	0.99	84.40	91.20	0.0034	0.0008	0.1560	5.00	0.0072	0.0017
0.025		72.40	69.80	0.0085	0.0010	0.2900	5.00	0.0240	0.0073
0.050		70.80	67.80	0.0250	0.0024	0.3060	5.00	0.0654	0.0219
0.100		59.80	65.00	0.0490	0.0060	0.4980	4.80	0.1790	0.0526
0.150		60.80	48.60	0.1244	0.0113	0.4760	0.00	0.2952	0.0908
0.200		53.80	57.20	0.0909	0.0152	0.5420	0.00	0.3957	0.1221
0.250		49.60	60.40	0.1116	0.0238	0.5880	0.00	0.4867	0.1569
0.500		30.80	70.40	0.1088	0.0461	0.9480	0.00	0.7074	0.2523
0.750	21.60	89.20	0.1713	0.0431	1.3980	0.00	0.4764	0.1722	

which played a central rôle in the discussion of Section 5, was also computed. To achieve this, the indices were used to identify the subsets  $S_1, \dots, S_7$ . For each  $S_j$ ,

$$\inf_v \left[ \mathbf{x}_{S_j}^1(v) \mathbf{A}_{S_j}^{S_j} \right] \quad \text{and} \quad \mathbf{x}_{S_j}^1(u_G) \mathbf{A}_{S_j}^{S_j}$$

were computed using value iteration. Finally, the index policy for the same problem but with  $\hat{\varepsilon} = 0$  was also constructed and compared with  $u_G$ . Denote this new index policy by  $u_G^0$ .

TABLE 2: The performance of the index policy for a class of dual speed restless bandit problems with  $\hat{\epsilon}_1 = 0.0$ .

$\hat{\epsilon}_2$	$\beta$	(b)	(c)	(d)	(f)	(g)	(h)
0.010	0.80	100.00	0.0000	0.0000	25.00	0.1069	0.0240
0.025		98.80	0.0574	0.0052	22.00	0.2611	0.0604
0.050		95.20	0.2695	0.0017	19.80	0.5172	0.1180
0.100		51.80	0.4107	0.0056	19.80	0.9349	0.2541
0.150		33.80	1.4638	0.0148	17.60	1.7017	0.4085
0.200		21.20	1.3513	0.0319	15.40	2.2350	0.5350
0.250		12.80	3.2983	0.0499	11.80	3.7339	0.7205
0.010	0.90	98.20	0.0277	0.0045	26.00	0.0836	0.0193
0.025		91.40	0.1216	0.0077	22.60	0.2000	0.0501
0.050		59.60	0.1525	0.0031	21.20	0.4025	0.1035
0.100		28.60	0.4755	0.0101	21.20	0.7781	0.2044
0.150		22.40	1.276	0.0270	19.40	1.671	0.3209
0.200		15.60	1.7356	0.0690	15.60	2.2089	0.4995
0.250		11.40	1.8113	0.1646	11.40	2.5341	0.6485
0.010	0.95	92.00	0.0193	0.0049	28.20	0.0535	0.0120
0.025		71.20	0.0955	0.0025	24.00	0.1342	0.0328
0.050		34.20	0.2080	0.0043	22.00	0.2687	0.0740
0.100		21.40	0.4321	0.0315	21.00	0.5491	0.1693
0.150		18.20	1.2200	0.0627	18.20	1.4178	0.2358
0.200		15.60	1.3443	0.1537	15.60	2.0065	0.4107
0.250		12.20	1.7472	0.2908	12.20	2.2864	0.5795
0.010	0.99	58.60	0.0074	0.0011	34.80	0.0151	0.0041
0.025		26.60	0.0249	0.0046	24.80	0.0419	0.0131
0.050		24.60	0.0707	0.0148	24.60	0.1032	0.0311
0.100		21.60	0.1905	0.0488	21.60	0.2459	0.0740
0.150		17.00	0.6006	0.0931	17.00	0.6608	0.1157
0.200		16.20	1.0375	0.1804	16.20	1.2547	0.2267
0.250		12.20	1.3555	0.3375	12.20	1.5695	0.3951

For each  $(\hat{\epsilon}, \beta)$ -pair, eight numbers were computed from the corresponding 500 problems and are presented in Table 1 as follows:

- (a) is the percentage of problems for which the index policy  $u_G$  is identical to the index policy  $u_G^0$  (i.e. takes the same action in all states);
- (b) is the percentage of problems for which  $R^{u_G} = R^{\text{opt}}$ ;
- (c) is the maximum value of  $100(R^{\text{opt}} - R^{u_G})/R^{\text{opt}}$  (percentage reward suboptimality of  $u_G$ );
- (d) is the median value of  $100(R^{\text{opt}} - R^{u_G})/R^{\text{opt}}$  for those problems for which this quantity is positive;
- (e) is the mean number of states of the system (out of the 16 possible) for which the index policies  $u_G$  and  $u_G^0$  choose different actions;
- (f) is the percentage of problems for which  $B^{u_G} = 0$ ;

TABLE 3: The performance of the index policy for a class of dual speed restless bandit problems with  $\hat{\epsilon}_1 = 0.1$ .

$\hat{\epsilon}_2$	$\beta$	(b)	(c)	(d)	(f)	(g)	(h)
0.010	0.80	94.00	0.2189	0.0047	9.80	0.8766	0.1665
0.025		92.40	0.3400	0.0216	1.20	0.6799	0.1254
0.050		100.00	0.0026	0.0019	0.00	0.5430	0.0721
0.100		99.20	0.0441	0.0013	1.20	0.3104	0.0382
0.150		83.00	0.0579	0.0071	1.60	0.7416	0.1034
0.200		49.40	0.7345	0.0126	4.80	1.1349	0.2518
0.250		26.00	0.9824	0.0374	4.80	1.7597	0.4275
0.010	0.90	72.80	0.2938	0.0229	5.00	0.7149	0.1445
0.025		80.20	0.1903	0.1830	5.00	0.5814	0.1161
0.050		82.80	0.1910	0.0175	3.40	0.5227	0.0791
0.100		86.00	0.0813	0.0064	0.00	0.2878	0.0486
0.150		48.00	0.3605	0.0156	0.00	0.6643	0.1256
0.200		30.60	0.6442	0.0228	3.00	1.0219	0.2699
0.250		18.20	1.3272	0.0434	5.00	1.8682	0.4529
0.010	0.95	23.60	0.3386	0.0090	4.60	0.5965	0.1165
0.025		49.60	0.2392	0.0133	4.60	0.5385	0.1033
0.050		73.00	0.1978	0.0223	4.80	0.4226	0.0750
0.100		62.20	0.1136	0.0046	3.40	0.2381	0.0493
0.150		44.00	0.2974	0.0217	0.00	0.5570	0.1146
0.200		22.00	0.4494	0.0296	0.00	0.8491	0.2411
0.250		6.80	1.1509	0.0501	1.80	1.6911	0.4073
0.010	0.99	5.00	0.1849	0.0177	5.00	0.3212	0.0589
0.025		9.80	0.1388	0.0107	5.00	0.3276	0.0654
0.050		53.80	0.1000	0.0166	5.00	0.2710	0.0533
0.100		65.60	0.0495	0.0064	5.00	0.1717	0.0504
0.150		36.80	0.1144	0.0205	0.00	0.4033	0.1112
0.200		11.60	0.2159	0.0339	0.00	0.6061	0.2302
0.250		7.40	0.6468	0.0721	0.00	1.2654	0.3663

(g) is the maximum value of  $100B^{u_G}/R^{\text{opt}}$ ;

(h) is the median value of  $100B^{u_G}/R^{\text{opt}}$  for those problems for which this quantity is positive.

The entries (a) and (e) should be read together; they give information on the closeness of the policies  $u_G$  and  $u_G^0$ . As we might expect, the policies become more dissimilar as  $\hat{\epsilon}$  grows from 0 for fixed  $\beta$ . When the indices for the multiarmed bandit case  $\hat{\epsilon} = 0$  are computed, the  $W_S^{\text{sc}}$  replace the  $A_S^{\hat{\epsilon}}$  as inputs to the adaptive greedy algorithm. Lemma 4 suggests quite strongly that these will differ more as  $\beta$  grows towards 1 for fixed  $\hat{\epsilon}$ . Hence, it is no surprise that the evidence of Table 1 is that  $u_G$  and  $u_G^0$  become more dissimilar as  $\beta$  increases for fixed  $\hat{\epsilon}$ .

The values of (b), (c) and (d) shed direct light on the quality of performance of index policy  $u_G$  while comparisons between the triples (b), (c), (d) and (f), (g), (h) give information on the tightness of the bound  $B^{u_G}$  given in (35). So far as the former is concerned, the overwhelming evidence of Table 1 is that the performance of the index policy  $u_G$  is remarkably strong throughout. In none of the 18 000 cases studied did the degree of reward suboptimality

TABLE 4: The performance of the index policy for a class of dual speed restless bandit problems with  $\hat{\epsilon}_1 = 0.2$ .

$\hat{\epsilon}_2$	$\beta$	(b)	(c)	(d)	(f)	(g)	(h)
0.010	0.80	70.00	0.6342	0.0270	0.00	1.7601	0.3174
0.025		75.20	0.3584	0.0411	0.00	1.5895	0.2980
0.050		77.60	0.6216	0.0434	0.00	1.2303	0.2544
0.100		84.00	0.2247	0.0202	0.00	0.9923	0.1611
0.150		95.40	0.3330	0.0261	0.00	0.7831	0.0924
0.200		90.80	0.2635	0.0073	4.00	0.5111	0.0761
0.250		67.80	0.4886	0.0199	0.00	0.8559	0.1345
0.010	0.90	31.40	0.7940	0.0162	5.00	1.4493	0.3076
0.025		36.20	0.3658	0.0115	3.40	1.3416	0.3206
0.050		64.60	0.7195	0.0250	3.20	1.1763	0.2788
0.100		80.60	0.4790	0.0719	0.00	1.021	0.1749
0.150		95.00	0.3807	0.0652	0.00	0.7027	0.1131
0.200		76.20	0.3119	0.0160	0.00	0.4929	0.0995
0.250		50.60	0.4540	0.0169	0.00	0.8590	0.1529
0.010	0.95	13.80	0.8919	0.0408	5.00	1.1227	0.2412
0.025		13.20	0.5442	0.0303	4.80	1.0982	0.2689
0.050		19.20	0.4897	0.0166	5.00	1.0460	0.2522
0.100		66.60	0.3965	0.0356	3.40	0.8545	0.1785
0.150		73.60	0.3118	0.0337	0.00	0.5840	0.1204
0.200		67.00	0.1844	0.0128	0.00	0.4686	0.1115
0.250		47.20	0.3433	0.0338	0.00	0.7473	0.1755
0.010	0.99	5.00	0.5002	0.0579	4.60	0.7300	0.1162
0.025		5.60	0.4231	0.0452	5.00	0.7713	0.1475
0.050		10.60	0.3269	0.0274	5.00	0.7711	0.1634
0.100		57.80	0.1374	0.0302	5.00	0.6700	0.1563
0.150		57.20	0.1160	0.0130	0.00	0.4040	0.1271
0.200		57.00	0.0968	0.0152	0.00	0.4006	0.1223
0.250		42.40	0.1639	0.0570	0.00	0.6289	0.1773

exceed 0.42%. Within that very strong overall performance, the index policy performs best for  $\hat{\epsilon}$  close to 0 and 1 for fixed  $\beta$ , as would be expected. If we fix  $\hat{\epsilon}$  and allow  $\beta$  to decrease, then the index policy has more chance of being optimal. To understand why this is so, observe that, as  $\beta$  decreases, both the optimal policy and the index policy tend to become myopic. So far as the tightness of the bound  $B^{UG}$  is concerned, its mode of construction is such that it will tend to increase with  $\hat{\epsilon}$ . Note that this is clearly true (for small  $\hat{\epsilon}$ ) of the simple bound developed in Theorem 2. As  $\hat{\epsilon}$  goes to 0,  $B^{UG}$  approaches 0 along with the suboptimality gap  $R^{\text{opt}} - R^{UG}$ . See also Remark 1. The bound is then at its tightest for small  $\hat{\epsilon}$ , as we would expect. Any impression of how tight  $B^{UG}$  is when the performance of the index policy is (relatively) weak can be gained by comparing (c) and (g). The evidence that the table provides on the tightness of the bounds is encouraging and, we would argue, good enough to suggest that Corollary 3 is an effective tool of analysis.

We now consider Tables 2–4. This part of the computational study is motivated by the fact, demonstrated by Example 1, that (relative) weakness in the performance of the index policy

can be brought about by differing transition speeds under the passive action for the states of distinct projects. Hence, we develop the models considered in Table 1 by allowing the four states of project 1 to have a common transition speed ( $\hat{\varepsilon}_1$ ) which may be distinct from the common transition speed of the four states of project 2 ( $\hat{\varepsilon}_2$ ). Tables 2–4 report a study based on 84 values of the triple  $(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \beta)$ . As before, 500 problems were generated for each of these combinations, yielding 42 000 independently generated problems in all. The details of the problem generation are as in (ii) and (iii) above.

For each  $(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \beta)$ -triple six numbers, (b), (c), (d), (f), (g), (h) described above, were computed from the corresponding 500 problems and presented in Tables 2–4.

While the performance of the index policy remains strong overall (with reward never worse than 3.30% below the optimum), a striking feature of Tables 2–4 is the tendency of this performance to improve for problems in which  $\hat{\varepsilon}_1$  and  $\hat{\varepsilon}_2$  are close. Hence, the intuition developed on the basis of Example 1 seems to be borne out more generally.

**Appendix A. Proof of Lemma 4**

Observe from (19), together with the inequalities  $V_k^{Sc} \leq (1 - \beta)^{-1}$  for  $k \in E$ , that

$$\begin{aligned}
 V_j^{Sc} &= \frac{\varepsilon_j \beta}{1 - \beta(1 - \varepsilon_j)} \sum_{k \in E} P_{jk}^1 V_k^{Sc} \\
 &\leq \frac{\hat{\varepsilon} \beta}{\{1 - \beta(1 - \hat{\varepsilon})\}(1 - \beta)}, \quad j \in S.
 \end{aligned}
 \tag{60}$$

From (20) and (60) we have that

$$\begin{aligned}
 V_{Sc}^{Sc} &= \mathbf{1}_{Sc} + \beta P_{ScS}^1 V_S^{Sc} + \beta P_{ScSc}^1 V_{Sc}^{Sc} \\
 &\leq \left\{ 1 + \frac{\hat{\varepsilon} \beta^2}{\{1 - \beta(1 - \hat{\varepsilon})\}(1 - \beta)} \right\} \mathbf{1}_{Sc} + \beta P_{ScSc}^1 V_{Sc}^{Sc}
 \end{aligned}
 \tag{61}$$

and hence that

$$\mathbf{1}_{Sc} + \beta P_{ScSc}^1 V_{Sc}^{Sc} \leq V_{Sc}^{Sc} \leq \left\{ 1 + \frac{\hat{\varepsilon} \beta^2}{\{1 - \beta(1 - \hat{\varepsilon})\}(1 - \beta)} \right\} \mathbf{1}_{Sc} + \beta P_{ScSc}^1 V_{Sc}^{Sc}.
 \tag{62}$$

However,

$$\mathbf{W}_E^{Sc} = \mathbf{1}_E + \beta P_{ESc}^1 \mathbf{W}_{Sc}^{Sc}.
 \tag{63}$$

From (62) and (63) it follows that

$$\mathbf{W}_{Sc}^{Sc} \leq V_{Sc}^{Sc} \leq \mathbf{W}_{Sc}^{Sc} \left\{ 1 + \frac{\hat{\varepsilon} \beta^2}{\{1 - \beta(1 - \hat{\varepsilon})\}(1 - \beta)} \right\}.
 \tag{64}$$

Now, from (4) and the definition of the matrix  $A$  in (5), we have that

$$\begin{aligned}
 A_S^S &= \mathbf{1}_S + \beta P_{SSc}^1 V_{Sc}^{Sc} + \beta P_{SSS}^1 V_S^{Sc} - \beta P_{SE}^0 V_E^{Sc} \\
 &= \mathbf{1}_S + \beta P_{SSc}^1 V_{Sc}^{Sc} + \beta P_{SSS}^1 V_S^{Sc} - V_S^{Sc}.
 \end{aligned}
 \tag{65}$$

Fix  $i \in S$  and focus on the last two terms on the last line of (65). Firstly, observe that

$$\beta \sum_{j \in S} P_{ij}^1 V_j^{Sc} - V_i^{Sc} \geq -V_i^{Sc} \geq \frac{-\hat{\varepsilon} \beta}{\{1 - \beta(1 - \hat{\varepsilon})\}(1 - \beta)}
 \tag{66}$$

by (60). Similarly, we conclude that

$$\beta \sum_{j \in S} P_{ij}^1 V_j^{Sc} - V_i^{Sc} \leq \beta \sum_{j \in S} P_{ij}^1 V_j^{Sc} \leq \frac{\hat{\epsilon} \beta^2}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)}. \quad (67)$$

Combining (63)–(66), we deduce that

$$\begin{aligned} A_S^S &\geq \left\{ 1 - \frac{\hat{\epsilon} \beta}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \right\} \mathbf{1}_S + \beta \mathbf{P}_{SS^c}^1 V_{S^c}^{Sc} \\ &\geq \mathbf{1}_S + \beta \mathbf{P}_{SS^c}^1 \mathbf{W}_{S^c}^{Sc} - \frac{\hat{\epsilon} \beta}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \mathbf{1}_S \\ &= \mathbf{W}_S^{Sc} - \frac{\hat{\epsilon} \beta}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \mathbf{1}_S. \end{aligned} \quad (68)$$

Combining (63)–(65) and (67), we deduce that

$$\begin{aligned} A_S^S &\leq \left\{ 1 + \frac{\hat{\epsilon} \beta^2}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \right\} \mathbf{1}_S + \beta \mathbf{P}_{SS^c} V_{S^c}^{Sc} \\ &\leq \left\{ 1 + \frac{\hat{\epsilon} \beta^2}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \right\} (\mathbf{1}_S + \beta \mathbf{P}_{SS^c} \mathbf{W}_{S^c}^{Sc}) \\ &= \left\{ 1 + \frac{\hat{\epsilon} \beta^2}{\{1 - \beta(1 - \hat{\epsilon})\}(1 - \beta)} \right\} \mathbf{W}_S^{Sc}. \end{aligned} \quad (69)$$

The inequalities (68) and (69) together yield the result.

### Acknowledgements

Dr Ansell and Professor Glazebrook would like to express appreciation to the Engineering and Physical Sciences Research Council for supporting their work through the award of grant GR/M09308 while Professors Glazebrook and Niño-Mora were also supported by NATO under Collaborative Linkage Grant PST.CLG976568. The authors express gratitude to Ryan Dunn and Richard Lumley for implementing the computations in Section 6 and to a referee, whose comments and suggestions resulted in a range of improvements to the paper.

### References

- BERTSIMAS, D. AND NIÑO-MORA, J. (1996). Conservation laws, extended polymatroids and multi-armed bandit problems: a polyhedral approach to indexable systems. *Math. Operat. Res.* **21**, 257–306.
- FAIHE, Y. AND MÜLLER, J.-P. (1998). Behaviors coordination using restless bandit allocation indices. In *From Animals to Animals 5* (Proc. 5th Internat. Conf. Simulation of Adaptive Behavior, Zürich), eds R. Pfeifer *et al.*, MIT Press, Cambridge, MA.
- GITTINS, J. C. (1979). Bandit processes and dynamic allocation indices (with discussion). *J. R. Statist. Soc. B* **41**, 148–177.
- GITTINS, J. C. (1989). *Multi-Armed Bandit Allocation Indices*. John Wiley, New York.
- GLAZEBROOK, K. D. AND GARBE, R. (1999). Almost optimal policies for stochastic systems which almost satisfy conservation laws. *Ann. Operat. Res.* **92**, 19–43.
- GLAZEBROOK, K. D. AND NIÑO-MORA, J. (2001). Parallel scheduling of multiclass  $M/M/m$  queues: approximate and heavy-traffic optimization of achievable performance. *Operat. Res.* **49**, 609–623.
- GLAZEBROOK, K. D. AND WILKINSON, D. J. (2000). Index-based policies for discounted multi-armed bandits on parallel machines. *Ann. Appl. Prob.* **10**, 877–896.
- GLAZEBROOK, K. D., NIÑO-MORA, J. AND ANSELL, P. S. (2000). Index policies for a class of discounted restless bandits. Tech. Rep., University of Newcastle upon Tyne.

- NIÑO-MORA, J. (1999). Restless bandits, partial conservation laws and indexability. Working paper 435, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona.
- PAPADIMITRIOU, C. H. AND TSITSIKLIS, J. N. (1999). The complexity of optimal queueing network control. *Math. Operat. Res.* **24**, 293–305.
- VEATCH, M. AND WEIN, L. M. (1996). Scheduling a make-to-stock queue: index policies and hedging points. *Operat. Res.* **44**, 634–647.
- WEBER, R. R. AND WEISS, G. (1990). On an index policy for restless bandits. *J. Appl. Prob.* **27**, 637–648.
- WEBER, R. R. AND WEISS, G. (1991). Addendum to ‘On an index policy for restless bandits’. *Adv. Appl. Prob.* **23**, 429–430.
- 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.