On The Optimality of Myopic Sensing in Multi-State Channels [1]

Yi Ouyang[2] and Demosthenis Teneketzis[2]

Abstract

We consider the channel sensing problem arising in opportunistic scheduling over fading channels, cognitive radio networks, and resource constrained jamming. The communication system consists of N channels. Each channel is modeled as a multi-state Markov chain (M.C.). At each time instant a user selects one channel to sense and uses it to transmit information. A reward depending on the state of the selected channel is obtained for each transmission. The objective is to design a channel sensing policy that maximizes the expected total reward collected over a finite or infinite horizon. This problem can be viewed as an instance of a restless bandit problem, for which the form of optimal policies is unknown in general. We discover sets of conditions sufficient to guarantee the optimality of a myopic sensing policy; we show that under one particular set of conditions the myopic policy coincides with the Gittins index rule.

Index Terms

Myopic Sensing, Markov Chain, POMDP, Restless Bandits, Stochastic Order.

I. Introduction and Literature Survey

A. Motivation

Consider a communication system consisting of N independent channels. Each channel is modeled as a K-state Markov chain (M.C.) with known matrix of transition probabilities. At each time period a user selects one channel to sense and uses it to transmit information. A reward depending on the state of the selected channel is obtained for each transmission. The objective is to design a channel sensing policy that maximizes the expected total reward (respectively, the expected total discounted reward) collected over a finite (respectively, infinite) time horizon.

The above channel sensing problem arises in cognitive radio networks, opportunistic scheduling over fading channels, as well as on resource-constrained jamming ([1]). In cognitive radio networks a secondary user may transmit over a channel only when the channel is not occupied by the primary user. Thus, at any time instant, state 1 of the M.C. describing the channel can indicate that the channel is occupied at t by the primary user, and states 2 through K indicate the quality of the channel that is available to the secondary user at t. In opportunistic transmission over fading channels, states 1 through K of the M.C. describe, at any time instant, the quality of the fading channel. In resource-constrained jamming a jammer can only jam one channel at a time, and any given jamming/channel sensing policy results in an expected reward for the jammer due to successful jamming.

The above channel problem is also an instance of a restless bandit problem ([2, 3]). Restless bandit problems arise in many areas, including wired and wireless communication systems, manufacturing systems, economic systems, statistics, etc (see [2, 3]).

B. Related Work

The channel sensing problem has been studied in [4] using a partially observable Markov decision process (POMDP) framework. For the case of two-state channels, the myopic policy was studied in [5], where its optimality was established when the number of channels is two. For more than two channels, the optimality of the myopic policy was proved in [6] under certain conditions on channel parameters. This result for the two-state channel was extended in [7] using a coupling argument to establish the optimality under a relaxed "positively correlated" condition. In [8], under the same "positively correlated" channel condition, the myopic policy was proved to be optimal for two-state channels when the user can select multiple channels at each time instance.

For general restless bandit problems, there is a rich literature; however, very little is known about the structure of optimal policies for this class of problems in general. In [2] it has been shown that the Gittins index rule (see [3],[9] for the definition of the Gittins index rule) is not optimal for a general restless bandit problems. Moreover, this class of problem is PSPACE-hard in general [10]. In [2] Whittle introduced an index policy (referred to as Whittle's index) and an "indexability condition"; the asymptotic optimality of the Whittle index was addressed in [11]. Issues related to Whittle's indexability condition were discussed in [2, 3, 11–13]. For the two-state channel sensing problem, Whittle's index was computed in closed-form in [13],

[1]A preliminary version of this paper appeared in the proceedings of *50th annual Allerton Conference on Communication, Control, and Computing*

[2]Y. Ouyang and D. Teneketzis are with the Department of EECS,University of Michigan, Ann Arbor, MI

where performance simulation of that index was provided. For some special classes of restless bandit problems, the optimality of some index-type policies was established under certain conditions (see [14, 15]). Approximation algorithms for the computation of optimal policies for a class of restless bandit problems similar to the one studied in this paper were investigated in [16].

A preliminary version of this paper appeared in the proceedings of the 50th Allerton conference on Control, Communication, and Computing (see [17]).

C. Contribution of the Paper

In this paper we identify sets of conditions under which the sensing policy that chooses at every time instant the best (in the sense of stochastic dominance [18]) channel maximizes the total expected reward (respectively, the expected total discounted reward) collected over a finite (respectively, infinite) time horizon. We also show that under one particular set of conditions the above-described policy coincides with the Gittins index rule, that is, the rule according to which the user selects at each time instant the channel with the highest Gittins index. Since our model is more general than previously studied models ([7]), our results are a contribution to the state of the art in cognitive radio networks, opportunistic scheduling and resource-constrained jamming. Furthermore, the results of this paper are a contribution to the state of the art of the theory of restless bandits (see for example [2, 3]). The optimization problem formulated in this paper is a restless bandit problem. Restless bandit problems are difficult to solve; very little is known about the nature of the optimal solution of these problems ([3]). Our results reveal instances of restless bandit problems where: (i) the optimal allocation rule is the myopic policy; and (ii) the myopic policy is optimal and coincides with the Gittins index rule.

D. Organization

The rest of this paper is organized as follows. In Section II, we present the model and the formulation of the optimization problem associated with the channel sensing problem. In Section III we discuss the salient features of the optimization problem formulated in Section II and show that it is an instance of a restless bandit problem. In Section IV, we consider the finite horizon problem and identify sets of conditions sufficient to guarantee the optimality of the myopic policy. In Section V, we extend the results of Section IV to the infinite horizon problem. In Section VI, we show that the result for two-state channels in [7] is a special case of the more general results presented in this paper. In Section VII we show that under one particular set of conditions the myopic policy coincides with the Gittins index rule. We conclude in Section VIII. The proofs of several intermediate results needed to establish the optimality of the myopic policy appear in the Appendices A-D.

II. Model and Oprimization Problems

A. The Model

Consider a communication system consisting of N identical channels. Each channel is modeled as a K-state Markov chain (M.C.) with (the same) matrix of transition probabilities P,

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1K} \\ p_{21} & p_{22} & \cdots & p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & \cdots & p_{KK} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_K \end{bmatrix}, \tag{1}$$

where $P_1, P_2, ..., P_K$ are row vectors. The K channel states model the channel's quality. For example, state K may denote the highest quality state, state 1 the lowest quality state, and states $2, 3, ..., K-1$ are medium quality states. We assume that the channel's quality increases as the number of its state increases. We want to use this communication system to transmit information. For that matter, at each time $t = 0, 1, ..., T$, we can select one channel, observe its state, and use it to transmit information.

Let X_t^n denote the state of channel n at time t, and let U_t denote the decision made at time t; $U_t \in \{1, 2, ..., N\}$, where $U_t = n$ means that channel n is chosen for data transmission at time t.

Initially, before any channel selection is made, we assume that we have probabilistic information about the state of each of the N channels. Specifically, we assume that at $t = 0$ the decision-maker (the entity that decides which channel to sense at each time instant) knows the probability mass function (PMF) on the state space of each of the N channels; that is, the decision-maker knows

$$\pi_0 := (\pi_0^1, \pi_0^2, ..., , \pi_0^N), \tag{2}$$

where

$$\pi_0^n := (\pi_0^n(1), \pi_0^n(2), ..., \pi_0^n(K)), n = 1, 2, ..., N, \tag{3}$$

$$\text{and } \pi_0^n(i) := P(X_0^n = i), i = 1, 2, ..., K. \tag{4}$$

Then, in general,

$$U_0 = g_0(\pi_0) \tag{5}$$
$$U_t = g_t(Y^{t-1}, U^{t-1}), t = 1, 2, ... \tag{6}$$

where

$$Y^{t-1} :=(Y_0, Y_1, ..., Y_{t-1}), U^{t-1} := (U_0, U_1, ..., U_{t-1}), \tag{7}$$

and $Y_t = X_t^{U_t}$ denotes the observation at time t; Y_t gives the state of the channel that is chosen at time t (that is, if $U_t = 2$, Y_t gives the state of channel 2 at time t).
Let $R(t)$ denote the reward obtained by the transmission at time t. We assume that $R(t)$ depends on the state of the channel chosen at time t. That is

$$R(t) = R_i, i = 1, 2, ..., K, \tag{8}$$

if the state of the channel chosen at t is i.

B. The Optimization Problems

Under the above assumptions, the objective is to solve:
(i) the finite horizon (T) optimization problem (P1)
Problem (P1)

$$\max_{g\in\mathcal{G}} E^g\{\sum_{t=0}^{T} \beta^t R(t)\}; \tag{9}$$

and (ii) its infinite horizon counterpart, problem (P2)
Problem (P2)

$$\max_{g\in\mathcal{G}} E^g\{\sum_{t=0}^{\infty} \beta^t R(t)\}, \tag{10}$$

where β is the discount factor ($0 < \beta < 1$) and $\mathcal{G}$ is the set of all channel sensing strategies g defined by (5)-(6).
Problems (P1) and (P2) are centralized stochastic optimization problems with imperfect information. Therefore, an information state for the decision-maker at time $t, t = 1, 2, ...$ is the conditional PMF (see [19], Chapter 6)

$$\pi_t := (\pi_t^1, \pi_t^2, ..., \pi_t^N), \tag{11}$$
$$\pi_t^n := (\pi_t^n(1), \pi_t^n(2), ..., \pi_t^n(K)), n = 1, 2, ..., N, \tag{12}$$
$$\pi_t^n(i) := P(X_t^n = i|Y^{t-1}, U^{t-1}), i = 1, 2, ..., K. \tag{13}$$

The information state π_t evolves as follows. If $U_t = n, Y^n = i$, then

$$\pi_{t+1}^n = P_i, \tag{14}$$
$$\pi_{t+1}^j = \pi_t^j P, \tag{15}$$

for all $j \neq n$. From stochastic control theory [19] we know that for problems (P1) and (P2) we can restrict attention (without any loss of optimality) to separated policies, that is, policies of the form

$$g := (g_0, g_1, ...), \tag{16}$$

where $U_t = g_t(\pi_t)$ for all t.
Consequently, problems (P1) and (P2) are equivalent to the following problems (P1') and (P2'), respectively:
Problem (P1')

$$\max_{g\in\mathcal{G}_s} E^g\{\sum_{t=0}^{T} \beta^t R(t)\}, \tag{17}$$

Problem (P2')

$$\max_{g\in\mathcal{G}_s} E^g\{\sum_{t=0}^{\infty} \beta^t R(t)\}, \tag{18}$$

where $\mathcal{G}_s$ is the set of separated policies.

Remark:

One separated policy the performance of which we will analyse in this paper is the "myopic policy" that we define as follows.

Let Π denote the set of PMFs on the state space $S = \{1, 2, ..., K\}$. We define the concept of stochastic dominance/order. Stochastic dominance $\geq_{st}$ between two row vectors $x, y \in \Pi$ is defined as follows:

$x \geq_{st} y$ if

$$\sum_{j=i}^{K} x(j) \geq \sum_{j=i}^{K} y(j) \text{ , for } i = 2, 3, ..., K \tag{19}$$

Note that stochastic order is a partial order, thus, the following facts true (see [18]):

Fact 1 If $x \geq_{st} y$ and $y \geq_{st} z$ then $x \geq_{st} z$.

Fact 2 If $x \geq_{st} y$, $z \in \Pi$ and $a \in \mathbb{R}, a \geq 0$, then $ax + z \geq_{st} ay + z$.

Definition 1. *The myopic policy $g^m := (g_0^m, g_1^m, ..., g_T^m)$ is the policy that selects at each time instant the best(in the sense of stochastic order) channel; that is,*

$$g_t^m(\pi_t) = i \qquad \text{if } \pi_t^i \geq_{st} \pi_t^j \quad \forall j \neq i \tag{20}$$

III. Characteristics of the Optimization Problems

The optimization problems (P1') and (P2') formulated in Section II can be viewed as an instance of a restless bandit problem as follows:

We can view the N channels as N arms with their PMFs as the states of the arms. The decision maker knows perfectly the states of the N arms at every time instant. One arm is operated (selected) at each time t, and an expected reward depending on the state (PMF of the channel) of the selected arm is received. If arm n (channel n) is not selected at t, its PMF π_t^n evolves according to

$$\pi_{t+1}^n = \pi_t^n P; \tag{21}$$

if arm n (channel n) is selected at t, its PMF evolves according to

$$\pi_{t+1}^n = P_{Y_t}, P(Y_t = x) = \pi_t^n(x). \tag{22}$$

The total expected reward for problem (P1') for any sensing policy $g \in \mathcal{G}_s$ can be written as

$$J_{\beta,T}^g := E^g[\sum_{t=0}^{T} \beta^t R(t)] = E^g[\sum_{t=0}^{T} \beta^t \pi_t^{U_t} R]. \tag{23}$$

The total expected reward for problem (P2') for any sensing policy $g \in \mathcal{G}_s$ can be written as

$$J_{\beta}^g := E^g[\sum_{t=0}^{\infty} \beta^t R(t)] = E^g[\sum_{t=0}^{\infty} \beta^t \pi_t^{U_t} R], \tag{24}$$

where $R := [R_1, R_2, ..., R_K]^T$ is the vector of instantaneous rewards.

Since the selected bandit process evolves in a way that differs from the evolution of the non-selected bandit processes, this problem is not a classical multi-armed bandit problem, but a restless bandit problem.

In general, restless bandit problems are difficult to solve because forward induction (the solution methodology for the classical multi-armed bandit problem) does not result in an optimal policy [3]. Consequently, optimal policies may not be of the index type, and the form of optimal policies for general restless bandit problems is still unknown.

IV. Analysis of the Finite Horizon Problem

We will prove the optimality of the myopic policy g^m for Problem (P1) under certain specific assumptions on the structure of the Markov chains describing the channels, on the instantaneous rewards $R = [R_1, R_2, R_3, ..., R_K]^T$ and on the initial PMFs $\pi_0^1, \pi_0^2, ..., \pi_0^N$

A. Key Assumptions/Conditions

We make the following assumptions/conditions

(A1)

$$P_K \geq_{st} P_{K-1} \geq_{st}, ..., \geq_{st} P_1. \tag{25}$$

Note that the quality of a channel state increases as its number increases. Assumption (A1) ensures that the higher the quality of the channel's current state the higher is the likelihood that the next channel state will be of high quality.

(A2) Let ΠP be the set of PMFs on the channel states that can be reached through transitions according to P, i.e.

$$\Pi P := \{\pi P : \pi \in \Pi\}; \tag{26}$$

note that ΠP is the convex hull of $P_1, P_2, ..., P_K$.
At time 0,

$$\pi_0^1, \pi_0^2, ..., \pi_0^N \in \Pi P \tag{27}$$

$$\text{and } \pi_0^1 \leq_{st} \pi_0^2 \leq_{st} ... \leq_{st} \pi_0^N. \tag{28}$$

Assumption (A2) states that initially the channels can be ordered in terms of their quality, expressed by the PMF on S. Moreover, the initial PMFs of the channels are in ΠP. Such a requirement ensures that the initial PMFs on the channel states are in the same space as all subsequent PMFs.

(A3)

$$P_1 P \geq_{st} P_{L-1} \tag{29}$$

$$P_K P \leq_{st} P_L \tag{30}$$

Assumption (A3) along with (A2) ensure that, any PMF π reachable from a non-selected channel has quality between P_{L-1} and P_L, that is $P_L \geq_{st} \pi \geq_{st} P_{L-1}$ (see also Property 2, Section IV-B). Here L is fixed; L can be any number from 2 to K.

(A4)

$$R_i - R_{i-1} \geq \beta(P_i - P_{i-1})M \geq \beta(P_i - P_{i-1})U \geq 0 \text{ for } i \neq L \tag{31}$$

$$R_L - R_{L-1} \geq \beta(h - P_{L-1}R) \geq 0, \tag{32}$$

where M is the vector given by

$$M := U + \beta \sum_{i \geq L} p_{Ki} P U, \tag{33}$$

$$U_i := R_i \text{ for } i = 1, 2, ..., L-1 \tag{34}$$

$$U_i := R_i + \beta(P_i - P_{L-1})U \text{ for } i = L, L+1, ..., K, \tag{35}$$

and h is given by

$$h = \frac{P_K R - \beta \sum_{i<L} p_{Ki} P_i R}{1 - \beta \sum_{i<L} p_{Ki}}. \tag{36}$$

Assumption (A4) states that the instantaneous rewards obtained at different states of the channel are sufficiently separated (see (31)(32)). Such an assumption is essential in establishing the optimality of a myopic policy. For the myopic policy to be optimal, the expected gain incurred by choosing the current best channel (say channel n) versus any other channel (say channel m) must overcompensate future losses in performance resulting in when channel m is chosen instead of channel n. For this to happen, the rewards obtained at different states of the channel must be sufficiently separated.

We note that (A1)-(A4) describe sets of sets of assumptions/conditions; for every value of $L, L = 2, 3, ..., K$, we have a distinct set of conditions.

We now compare the above conditions with those made in [17]. When $L = K$, the above conditions are exactly the same as those in [17]. In [17] we did not address situations where $L \neq K$ that is, situation where the quality of the information state resulting form a non-selected channel is between P_L and P_{L-1} for $L \neq K$. Consequently, the result of this paper subsume the results obtained in [17].

Before we proceed with the analysis of Problem (P1) based on conditions (A1)-(A4), we show that (A1)-(A4) can be simultaneously satisfied. Consider the following situation:

$$K = 5, L = 5, N = 6, \beta = 1 \tag{37}$$

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_5 \end{bmatrix} = \begin{bmatrix} 0.0656 & 0.0458 & 0.1044 & 0.4745 & 0.3096 \\ 0.0655 & 0.0458 & 0.1030 & 0.4454 & 0.3403 \\ 0.0652 & 0.0457 & 0.0966 & 0.4019 & 0.3907 \\ 0.0434 & 0.0336 & 0.1126 & 0.4102 & 0.4001 \\ 0.0206 & 0.0205 & 0.0142 & 0.4475 & 0.4972 \end{bmatrix}, \tag{38}$$

(39)

with

$$R = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \end{bmatrix}^T \tag{40}$$

$$\pi_0^1 = \pi_0^2 = P_1, \pi_0^3 = P_2, \pi_0^4 = P_3, \pi_0^5 = P_4, \pi_0^6 = P_5 \tag{41}$$

By their definition, $P_1, P_2, ..., P_5$ satisfy (A1). By the definition of $\pi_0^1, \pi_0^2, ..., \pi_0^6$ and the definition of ΠP, (A2) is satisfied. By direct computation we can show that

$$P_1 P = \begin{bmatrix} 0.0411 & 0.0322 & 0.0795 & 0.4267 & 0.4205 \end{bmatrix} \tag{42}$$

$$\geq_{st} \begin{bmatrix} 0.0434 & 0.0336 & 0.1126 & 0.4102 & 0.4001 \end{bmatrix} = P_4, \tag{43}$$

Moreover, $P_5 P = p_{51} P_1 + p_{52} P_2 + + ... + p_{55} P_5 \leq_{st} P_5$. Therefore, (A3) is satisfied.
By direct computation, we get

$$U = \begin{bmatrix} 0 & 1 & 2 & 3 & 4.3214 \end{bmatrix}^T \tag{44}$$

$$M = \begin{bmatrix} 1.4997 & 2.5206 & 3.5577 & 4.6003 & 6.0815 \end{bmatrix}^T \tag{45}$$

$$h = 3.7776, \tag{46}$$

So we can compute

$$\beta(P_2 - P_1)M = 0.0470 \leq R_2 - R_1 \tag{47}$$

$$\beta(P_3 - P_2)M = 0.0829 \leq R_3 - R_2 \tag{48}$$

$$\beta(P_4 - P_3)M = 0.0897 \leq R_4 - R_3 \tag{49}$$

$$\beta(h - P_4 R) = 0.7766 \leq R_5 - R_4 \tag{50}$$

Therefore, (A4) is satisfied.
Assumptions (A1)-(A4) are also satisfied when $R, P, \pi_0^1, \pi_0^2, ..., \pi_0^6$, chosen as above, are slightly perturbed. It is also possible to find other ranges of values of $R, P, \pi_0^1, \pi_0^2, ..., \pi_0^6$ which satisfy (A1)-(A4).
Based on the above assumptions, we proceed to establish the optimality of the myopic policy g^m as follows. In sections IV-B-IV-D we develop some preliminary results needed for our purposes. Specifically: In section IV-B we present three properties of the evolution of the PMFs on the channel states. In section IV-C we present a property of the instantaneous expected reward. In section IV-D we define a class of ordering-based channel sensing policies $\mathcal{G}^O$ which includes the myopic policy g^m; using the results of sections IV-B and IV-C we discover four properties of the expected reward resulting from any policy in $\mathcal{G}^O$. In section IV-E we use the results of section IV-D to establish the optimality of a myopic policy for Problem (P1'). We note that all the properties developed in sections IV-B through IV-D are needed to establish the optimality of the myopic policy. We discuss how these properties are used to prove the optimality of the myopic policy in Section IV-F, after we prove the main result of this paper. The proofs of properties 1-9 appear in Appendices A-D.

B. Properties of the Channels' Evolution

Under assumptions/conditions (A1)-(A4) stated in section IV-A, the following properties hold.

Property 1. *Let $x, y \in \Pi$. Under Assumption (A1),*

$$x \geq_{st} y \Longrightarrow xP \geq_{st} yP \tag{51}$$

An implication of Property 1 is the following. If at any time t the information states of two channels (expressed by the PMFs on their state space) are stochastically ordered and none of these channels is sensed at t, then the same stochastic order between the information states at time $t+1$ is maintained.

Property 2. *Let $\pi = xP^2 \in \Pi P^2$, $\Pi P^2 := \{\pi = xP^2, x \in \Pi\}$. Under (A1)-(A3),*

$$P_L \geq_{st} xP^2 \geq_{st} P_{L-1} \tag{52}$$

Property 2 says the following. By condition (A2) a channel's information state (the PMF on its state space) is always in ΠP. If the channel is not sensed at time t, then at time $t+1$ its information state is in ΠP^2, moreover it is stochastically always between P_{L-1} and P_L. If the channel is sensed at time t and its observed state is larger than or equal to L (respectively smaller than L), then at time $t+1$ this channel is in the stochastically largest (respectively stochastically smallest) information state among all channels.

Property 3. *Under (A1)-(A3), we have either $\pi_t^n \leq_{st} \pi_t^m$ or $\pi_t^m \leq_{st} \pi_t^n$ for all $n, m \in \{1, 2, ..., N\}$ for all t.*

Property 3 states that under (A1)-(A3) the information states of all channels can be ordered stochastically at all times.
The proofs of Properties 1-3 appear in Appendix A.

C. A Property of the Instantaneous Expected Reward

A direct consequence of Assumption (A4) is the following Properties of the instantaneous expected reward:

Property 4. *Let $x, y \in \Pi$. Let v be a column vector in increasing order, i.e. $v_i \geq v_{i-1}$ for $i = 2, 3, ..., K$. If $x \geq_{st} y$, we have*

(i) $(x-y)v \geq 0$.

(ii) $(x-y)M \geq (x-y)U \geq (x-y)R \geq 0$, *where M, U, R are defined by eqs (31)-(35).*

(iii) $(x-y)M \geq \beta(x-y)PM$.

(iv) If $x(i) = y(i)$ for all $i \geq L$ or $x(i) = y(i)$ for all $i < L$, we have

$$(x-y)R \geq \beta(x-y)PM \geq \beta(x-y)PU. \tag{53}$$

Part (i) of Property 4 says the following. Consider a reward vector such that the reward increases as the quality of the channel state increases. Then the expected reward increases as the information state of the channel increases stochastically.

Part (ii) is a restatement of part (i) when the reward vector v takes the values $M-U, U-R, R$.

Part (iii) can be interpreted as follows. Consider the reward vector M defined by (33). Consider two channels, channel i and channel j, that have information states x and y respectively, such that $x \geq_{st} y$. Consider the following scenarios: (S1) Sense channel i first, then sense channel j; (S2) Sense channel j first, then sense channel i. Then part (iii) of Property 4 asserts that scenario (S1) is better than scenario (S2), that is, it is better to sense the best (in the sense of stochastic order) channel first.

Part (iv) has an interpretation similar to that of part (iii). Consider any time t and two channels i and j whcih have information states x and y, respectively, such that $x \geq_{st} y$ and x, y satisfy the condition of part (iv). Assume that the reward vector at t is R and the reward vector at $t+1$ is M such that $M_i - R_i$ is increasing in i. Consider scenarios (S1) and (S2) described above. Then part (iv) asserts that the expected reward obtained under scenario (S1) is higher than the expected reward obtained under scenario (S2); that is, it is better to sense the best (in the sense of stochastic order) channel first. Note that Property 4 refers to the situation where we have only two options, described by scenarios (S1) and (S2). Thus, the results of Property 4 do not imply the optimality of the myopic policy, as in Problems (P1) we have more that two options at each time instant.

The proof of Property 4 appears in Appendix B.

D. Properties of the Reward Associated with Ordering-based Channel Sensing Polices

In this section we introduce ordering-based policies and study their Properties. The reason for considering this class of policies is because under conditions (A1)-(A4) we obtain the following: (i) The performance of any sensing policy can be upper-bounded by an appropriately chosen ordering-based policy (see Section IV-E); thus, for the solution of the original optimization problem (Problem (P1)) we can restrict attention to ordering-based policies. (ii) The myopic policy is an optimal ordering-based policy. Combining (i) and (ii) we establish the optimality of the myopic policy for Problem (P1).

We note that Properties 1-4, developed so far, are essential for the discovery of the properties of ordering-based policies that lead eventually to the solution of Problem (P1) (see discussion in Section IV-F).

Let $\mathcal{O}$ be the set of all orderings/permutations of the N channels $\{1, 2, ..., N\}$. Consider the ordering-based selection function $\hat{g} : \mathcal{O} \mapsto \{1, 2, ..., N\}$ and the ordering update mapping $\hat{m} : \mathcal{O} \times \{1, 2, ..., K\} \mapsto \mathcal{O}$ defined as follows.
For every $O := (O(1), O(2), ..., O(N)) \in \mathcal{O}$,

$$\hat{g}(O) = O(N), \tag{54}$$

$$\hat{m}(O, y) = \begin{cases} O & \text{if } y \geq L \\ SO & \text{if } y < L \end{cases}, \tag{55}$$

where S is the cyclic shift operator on $\mathcal{O}$ such that

$$SO =: (O(N), O(1), O(2), ..., O(N-1)) \tag{56}$$

Given a channel ordering $O_t \in \mathcal{O}$ at time t, we define an ordering-based channel sensing policy $g_{t:T}^{O_t} := (g_t^{O_t}, g_{t+1}^{O_t}, ..., g_T^{O_t})$ as follows.

$$U_t = g_t^{O_t}(O_t) = \hat{g}(O_t) = O(N) \tag{57}$$

$$O_s = \hat{m}(O_{s-1}, Y_{s-1}), \text{ for } s = t+1, t+2, ..., T \tag{58}$$

$$U_s = g_s^{O_t}(Y_{t:s-1}, U_{t:s-1}) = g_s^{O_t}(O_s) = \hat{g}(O_s), \text{ for } s = t+1, t+2, ..., T \tag{59}$$

At time $s, t \leq s \leq T$, $g_s^{O_t}$ chooses the last channel in O_s; the ordering O_s is shifted to the right by the update mapping $\hat{m}$ whenever the observed state is less than L, and remains the same otherwise. As a result of the above specification of $g_{t:T}^{O_t}$, if at time t channel n is on the right of channel m in the ordering O_t, channel n will be sensed by policy $g_{t:T}^{O_t}$ before channel m.
Note that, the policy $g_{t:T}^{O_t}$ is not a separated policy in general. However, if the ordering $O_0 = (O_0(1), O_0(2), ..., O_0(N))$ at

time 0 is such that $\pi_0^{O_0(1)} \leq_{st} \pi_0^{O_0(2)} \leq_{st} ... \leq_{st} \pi_0^{O_0(N)}$, then $g_{0:T}^{O_0}$ is the myopic policy g^m, therefore; $g_{0:T}^{O_0} = g^m \in \mathcal{G}_s$, as the following Property shows.

Property 5. *At time $t = 0$ consider the ordering O_0 such that $\pi_0^{O_0(1)} \leq_{st} \pi_0^{O_0(2)} \leq_{st} ... \leq_{st} \pi_0^{O_0(N)}$. Then, the ordering based policy $g_{0:T}^{O_0}$ is just the myopic policy g^m.*

The validity of Property 5 crucially depends on Properties 1 and 2, which say that stochastic order is maintained under the evolution of unobserved channels (Property 1), and the observed channel is either the stochastically best or the stochastically worst among all channels (Property 2). Without Properties 1 and 2 the myopic policy is not an ordering-based policy.
The proof of Property 5 appears in Appendix C.

Define by $V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N)$ to be the expected reward collected from time t up to and including T due to the ordering-based policy $g_{t:T}^{O_t}$. That is,

$$V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) := E^{g_{t:T}^{O_t}}[\sum_{l=t}^{T} \beta^{l-t} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N] \tag{60}$$

Then, $V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N)$ can be written recursively as follows.

$$V_T(O_t, \pi_T^1, \pi_T^2, ..., \pi_T^N) = \pi_T^{O_t(N)} R, \tag{61}$$

$$\begin{aligned} V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) =& \pi_t^{O_t(N)} R + \beta \sum_{i<L} \pi_t^{O_t(N)}(i) V_{t+1}(SO_t, \pi_{t+1}^1, ..., \pi_{t+1}^N) \\ &+ \beta \sum_{i \geq L} \pi_t^{O_t(N)}(i) V_{t+1}(O_t, \pi_{t+1}^1, ..., \pi_{t+1}^N), \end{aligned} \tag{62}$$

$$\text{where } \pi_{t+1}^n = \begin{cases} P_i & \text{for } n = O_t(N) \\ \pi_t^n P & \text{otherwise} \end{cases}. \tag{63}$$

The function $V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N)$ defined above possesses properties 6-9 below. The proof of these Properties appear in Appendix C. We will explain the role of these Properties in Section IV-F after we prove the main result on the optimality of the myopic policy in Section IV-E.

Property 6. *Let $\hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N \in \Pi P$ and $O_t \in \mathcal{O}$.*
Define

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) := V_t(O_t, \hat{\pi}_t^1, \pi_t^2, ..., \pi_t^N) - V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \tag{64}$$

If $\hat{\pi}_t^1 \geq_{st} \pi_t^1$, and $O_t(n) = 1$, then for all $m < n$

$$0 \leq L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) - L_t(S^{-m} O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) \leq (\hat{\pi}_t^1 - \pi_t^1) U, \tag{65}$$

where $S^{-m} O_t$ is the counter-clockwise cyclic shift of O_t by m positions, that is,

$$S^{-m} O_t = (O_t(m+1), O_t(m+2), ..., O_t(N), O_t(1), ..., O_t(m)) \tag{66}$$

Property 7. *For $O_t \in \mathcal{O}$, define the operator W_{nm} as follows.*

$$W_{nm} O_t(i) := \begin{cases} O_t(n) & \textit{for } i = m \\ O_t(m) & \textit{for } i = n \\ O_t(i) & \textit{otherwise} \end{cases}. \tag{67}$$

If $\hat{\pi}_t^1 \geq_{st} \pi_t^1$, and $O_t(n) = 1$, then for $m < n$

$$0 \leq L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) - L_t(W_{nm} O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) \leq (\hat{\pi}_t^1 - \pi_t^1) M \tag{68}$$

The meaning of Properties 6 and 7 is the following. Restrict attention to ordering-based policies. Take any channel, say channel 1. Replace it with a better quality (in the sense of stochastic order) channel. Such a replacement will result in an improvement in performance. This improvement is different for different channel orderings. The earlier channel 1 is used (that is, the closer to the right-most position in the ordering channel 1 is) the higher is the improvement. Properties 6 and 7 also provide bounds on the difference between maximum and minimum improvement. These bounds are useful in proving Properties 6 and 7 by induction.

Property 8. *If $\pi_t^{O_t(n)} \geq_{st} \pi_t^{O_t(m)}$, then for $m < n$ then*

$$V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \geq V_t(W_{nm} O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \tag{69}$$

Property 8 states that if the position of two channels in any arbitrary but fixed channel ordering are interchanged so that

the better (in the stochastic order sense) channel comes closer to the right-most position (i.e. it is used earlier) in the new ordering, the performance due to the ordering-based policy improves.

Property 9. *For $O_t \in \mathcal{O}$, define the operator A_{nm} as follows.*

$$A_{nm}O_t(i) := \begin{cases} O_t(n) & \text{for } i = m \\ O_t(i-1) & \text{for } i = m+1, m+2, ..., n \\ O_t(i) & \text{otherwise} \end{cases}. \tag{70}$$

If $\pi_t^1 \leq_{st} \pi_t^1 P$, and $O_t(n) = 1$, then

$$V_t(A_{nm}O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) - V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \leq h - \pi_t^1 P^{N-n} R \tag{71}$$

Property 9 states the following. Suppose that a channel, say channel 1, is such that as long as it is not sensed its quality is continuously improving (i.e. its PMF is continuously increasing stochastically). Then, no matter how late this channel is sensed (that is, no matter how much we move the channel to the left from its initial position in the original channel ordering) the change in performance due to an ordering-based policy can not exceed a certain bound.

E. Optimality of a Myopic Policy

The main result of this paper is summarized by the following theorem

Theorem 1. *Under assumptions (A1)-(A4), the myopic policy g^m, that is, the policy that picks at every time instant the best (in the sense of stochastic order) channel is optimal for Problem (P1).*

Proof: We proceed by induction.
At T, the expected reward is the instantaneous expected reward. Since by part (ii) of Property 4 a better channel (in the sense of stochastic order) gives larger instantaneous expected reward, the myopic policy g^m is optimal at T. This establishes the basis of induction.

The induction hypothesis is that the myopic policy g^m is optimal at $t+1, t+1, ..., T$. To complete the induction we need to prove that g^m is optimal at t (induction step).

Without loss of generality, we assume $\pi_t^1 \leq_{st} \pi_t^2 \leq_{st} ... \leq_{st} \pi_t^N$.
Consider any policy g. If g picks channel n at time t, then the expected reward collected from t on due to the policy g is given by

$$E^g[\sum_{l=t}^{T} \beta^{l-t} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N] = \pi^n R + \sum_{i=1}^{K} \pi_t^n(i) E^g[\sum_{l=t+1}^{T} \beta^{l-t} R(l) | \pi_{t+1}^n = P_i, \pi_{t+1}^m = \pi_t^m P \text{ for } m \neq n]. \tag{72}$$

By the induction hypothesis we have

$$\begin{aligned} &E^g[\sum_{l=t+1}^{T} R(l) | \pi_{t+1}^n = P_i, \pi_{t+1}^m = \pi_t^m P \text{ for } m \neq n] \\ \leq &E^{g^m}[\sum_{l=t+1}^{T} R(l) | \pi_{t+1}^n = P_i, \pi_{t+1}^m = \pi_t^m P \text{ for } m \neq n]. \end{aligned} \tag{73}$$

Using (73) in (72) we get

$$\begin{aligned} &E^g[\sum_{l=t}^{T} \beta^{l-t} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N] \\ \leq &\pi_t^n R + \sum_{i=1}^{K} \pi_t^n(i) E^{g^m}[\sum_{l=t+1}^{T} \beta^{l-t} R(l) | \pi_{t+1}^n = P_i, \pi_{t+1}^m = \pi_t^m P \text{ for } m \neq n] \\ = &\pi_t^n R + \beta \sum_{i<L} \pi_t^n(i) V_{t+1}(SO_t, \pi_{t+1}^1, ..., \pi_{t+1}^N) + \beta \sum_{i \geq L} \pi_t^n(i) V_{t+1}(O_t, \pi_{t+1}^1, ..., \pi_{t+1}^N) \\ = &V_t(O_t, \pi_t^1, ..., \pi_t^N), \end{aligned} \tag{74}$$

where

$$O_t = (1, 2, ..., n-1, n+1, ..., N, n), \tag{75}$$

$$SO_t = (n, 1, 2, ..., n-1, n+1, ..., N). \tag{76}$$

The inequality in (74) follows by (73); the first equality in (74) is true because of Property 5, for $s = t+1, t+2, ..., T$, $g_s^m = g_s^{SO_t}$ when $\pi_{t+1}^n = P_i, i < L$ and $g_s^m = g_s^{O_t}$ when $\pi_{t+1}^n = P_i, i \geq L$; the last equality follows from equation (62) for V_t.
Since $\pi_t^n \leq_{st} \pi_t^m$ for all $m = n+1, n+2, ..., N$, repeatedly applying Property 8 we get

$$\begin{aligned} V_t(O_t, \pi_t^1, ..., \pi_t^N) \leq & V_t((1, 2, ..., n-1, n+1, ..., N-1, n, N), \pi_t^1, ..., \pi_t^N) \\ & \vdots \\ \leq & V_t((1, 2, ..., n-1, n+1, n, n+2, ..., N), \pi_t^1, ..., \pi_t^N) \\ \leq & V_t((1, 2, ..., n-1, n, n+1, ..., N), \pi_t^1, ..., \pi_t^N) \\ = & E^{g^m}[\sum_{l=t}^{T} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N] \end{aligned} \tag{77}$$

Combing (74) (77) we obtain

$$E^{g}[\sum_{l=t}^{T} \beta^{l-t} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N] \leq E^{g^m}[\sum_{l=t}^{T} \beta^{l-t} R(l) | \pi_t^1, \pi_t^2, ..., \pi_t^N], \tag{78}$$

which completes the proof.

F. Discussion

The key steps in establishing the optimality of the myopic policy, under the assumptions made in the problem formulation, are the following:

(K1) The assertion that the performance of any separated policy can be upper-bounded by the performance of an ordering-based policy. Consequently, for the solution of the original optimization problem, one can restrict attention to ordering-based policies.

(K2) The assertion that the performance of an ordering-based policy improves when a better (in the sense of stochastic order) channel is used earlier. This assertion implies the optimality of the myopic policy.

The assertion of (K1) is established in Theorem 1 (its induction step). The assertion of (K2) is established by Property 8, provided that the myopic policy is an ordering-based policy, and that stochastic order is maintained among all channels at every time. The fact that the myopic policy is an ordering-based policy is ensured by Property 5. The existence of a stochastic ordering among all channels at any time t is ensured by Property 3. To establish these properties we need Properties 1-9.

We now elaborate on the interdependence of Properties 1-9. Property 3, which asserts that channels can be ordered stochastically, is a consequence of Properties 1 and 2 for the unobserved channels and the observed channel, respectively. Properties 1 and 2 also ensure that the myopic policy g^m belongs to the class of ordering-based policies (Property 5). Property 8 is a special case of Property 7 when $\hat{\pi}_t^1 = \pi_t^{O_t(m)} \geq_{st} \pi_t^1 = \pi_t^{O_t(n)}$. Property 7 is coupled with Properties 6 and 9, that is, Properties 6, 7 and 9 need to be proven simultaneously. The proof of Properties 6, 7 and 9 requires Property 4.

The upper bounds that appear in Properties 6, 7 and 9 are essential in establishing the optimality of the myopic policy. These bounds along with condition (A4) ensure that the instantaneous advantage in expected reward obtained by the use of the myopic policy g^m over any other policy g, overcompensates any future possible expected reward losses of g^m as compared to g.

V. The Infinite Horizon Problem

For the infinite horizon Problem (P2) we have the following theorem.

Theorem 2. *Under assumptions (A1)-(A4), the myopic policy g^m is optimal for Problem (P2).*

Proof: From the theory of stochastic control [19] we know that for Problem (P2) there exists a separated stationary policy g^* that maximizes the total expected discounted reward.
Let $\pi := (\pi^1, \pi^2, ..., \pi^N)$; for any stationary separated policy g let

$$J_\beta^g(\pi) := E^g\{\sum_{t=0}^{\infty} \beta^t R(t) | \pi_0 = \pi\}. \tag{79}$$

Then the dynamic program for Problem (P2) is

$$J_\beta^{g^*}(\pi) = \max_{n=1,2,...,N} \left\{ \pi^n R + \beta E\{J_\beta^{g^*}(\pi_1) | \pi_0 = \pi, U_0 = n\} \right\}, \tag{80}$$

where π_0, π_1 are defined by (11)-(13). The myopic policy g^m that is optimal for the finite horizon T problem (by Theorem 1) satisfies the dynamic program

$$J^{g^m}_{\beta,T}(\pi) = \max_{n\in\{1,2,\ldots,N\}} \left\{\pi^n R + \beta E\{J^{g^m}_{\beta,T-1}(\pi_1)|\pi_0 = \pi, U_0 = n\}\right\}, \tag{81}$$

where

$$J^{g^m}_{\beta,T}(\pi) := E^{g^m}\{\sum_{t=0}^{T} \beta^t R(t)|\pi_0 = \pi\}. \tag{82}$$

Since the reward $R(t) \leq R_K$ is bounded, by the bounded convergence theorem we get

$$\begin{aligned} J^{g^m}_{\beta}(\pi) =& E^{g}\{\sum_{t=0}^{\infty} \beta^t R(t)|\pi_0 = \pi\} \\ =& \lim_{T\to\infty} E^{g}\{\sum_{t=0}^{T} \beta^t R(t)|\pi_0 = \pi\} \\ =& \lim_{T\to\infty} J^{g^m}_{\beta,T}(\pi), \end{aligned} \tag{83}$$

Letting $T \to \infty$ in (81) and using the bounded convergence theorem we obtain

$$J^{g^m}_{\beta}(\pi) = \max_{n\in\{1,2,\ldots,N\}} \left\{\pi^n R + \beta E\{J^{g^m}_{\beta}(\hat{\pi}(\pi, n))\}\right\}, \tag{84}$$

Notice that (84) is exactly the dynamic programming equation (80); therefore,

$$J^{g^m}_{\beta}(\pi) = J^{g^*}_{\beta}(\pi); \tag{85}$$

consequently, the myopic policy g^m is optimal for the infinite horizon problem (P2).

VI. Comparison with the Result of the Two-State Channel Model

The situation where each channel has two states, i.e. $K = 2$, has been previously investigated in the literature (e.g. [7]). In this section we show that when $K = 2$ our conditions (A1)-(A4) reduce to the assumptions made in [7].

When $K = 2$, then L has to be two, and the matrix of transition probabilities is given by

$$P_1 = (p_{1,1}, p_{1,2}) = (1 - p_{1,2}, p_{1,2}), \tag{86}$$

$$P_2 = (p_{2,1}, p_{2,2}) = (1 - p_{2,2}, p_{2,2}). \tag{87}$$

In this case, for any two PMF $x, y \in \Pi$, let $x = (1 - a, a), y = (1 - b, b)$; then we have

$$x \geq_{st} y \iff a \geq b. \tag{88}$$

Without loss of generality, let $R_1 = 0, R_2 = 1$, then our conditions reduce to the following conditions.

For (A1), we get

$$P_2 \geq_{st} P_1 \iff p_{2,2} \geq p_{1,2} \tag{89}$$

For (A2) note that

$$\Pi = \{(1 - p, p) : 0 \leq p \leq 1\}; \tag{90}$$

$$\Pi P = \{(1 - p, p) : p_{1,2} \leq p \leq p_{2,2}\}. \tag{91}$$

Consequently, (A2) reduces to

$$\pi^n_0 = (1 - p^n, p^n), p_{1,2} \leq p^n \leq p_{2,2} \text{ for } n = 1, 2, ..., N (cf.(27)) \tag{92}$$

$$\text{and } p^1 \leq p^2 \leq ... \leq p^N (cf.(28)). \tag{93}$$

Using (89) we get

$$P_1 P = p_{1,1}P_1 + p_{1,2}P_2 \geq_{st} P_1, \tag{94}$$

$$P_2 P = p_{2,1}P_1 + p_{2,2}P_2 \leq_{st} P_2, \tag{95}$$

thus (A3) is automatically satisfied.
For (A4), we have

$$h = \frac{p_{2,2} - \beta p_{2,1} p_{1,2}}{1 - \beta p_{2,1}}. \tag{96}$$

Therefore,

$$\beta(h - P_1 R) = \beta \frac{p_{2,2} - p_{1,2}}{1 - \beta p_{2,1}} \leq \frac{p_{2,2} - p_{1,2}}{p_{2,2}} \leq 1 = R_2 - R_1. \tag{97}$$

Consequently, (A4) is automatically satisfied.

As a result of the above analysis, our conditions (A1)-(A4) for the special case $K = 2$ reduce to

$$p_{2,2} \geq p_{1,2} \tag{98}$$
$$\pi_0^n = (1 - p^n, p^n), p_{1,2} \leq p^n \leq p_{2,2} \text{ for } n = 1, 2, ..., N \tag{99}$$
$$p^1 \leq p^2 \leq ... \leq p^N. \tag{100}$$

Condition (98) is precisely the "positively correlated" condition in [7]. Condition (99) is satisfied, if the channels evolve before we begin sensing them (before time $t = 0$). Condition (100) is always satisfied by renumbering of the channels.

VII. Myopic policy vs. Gittins index rule

In this section we investigate conditions under which the myopic policy coincides with the Gittins index rule.

Select a channel, say channel $n, n = 1, 2, ..., N$. For PMF $\pi \in \Pi$, the Gittins index ([3, 9]) of channel n is defined is defined by

$$\nu^n(\pi) := \max_{\tau} \frac{E^{g^\tau}[\sum_{t=0}^{\tau-1} \beta^t \pi_t^n R | \pi_0^n = \pi]}{E^{g^\tau}[\sum_{t=0}^{\tau-1} \beta^t | \pi_0^n = \pi]}, \tag{101}$$

where τ is any stopping time with respect to $\{\pi_t^n, t = 0, 1, ...\}$ and g^τ chooses channel n from $t = 0$ up to $t = \tau - 1$. The Gittins index rule ([3, 9]) chooses the channel with the highest Gittins index at every time instant t.

In condition (A3) (Section IV-A) L is fixed; it can be any number form 2 to K. In this section we show that when $L = K$, under conditions (A1)-(A4), after time 0 the myopic policy coincides with the Gittins index rule. We establish this result via Theorem 3 and 4.

Theorem 3. *(i) For $\pi \in \Pi P$, $P_{K-1} \leq_{st} \pi \leq_{st} P_K$, the Gittins index $\nu(\pi)$ is given by*

$$\nu(\pi) = \frac{\pi R + \beta \pi(K) \frac{P_K R}{1 - \beta p_{KK}}}{1 + \beta \pi(K) \frac{1}{1 - \beta p_{KK}}}. \tag{102}$$

(ii) If $\pi_x, \pi_y \in \Pi P$, $P_{K-1} \leq_{st} \pi_y \leq_{st} \pi_x \leq_{st} P_K$, then $\nu(\pi_x) \geq \nu(\pi_y)$
(iii) If $\pi \in \Pi P$, $P_{K-1} \leq_{st} \pi \leq_{st} P_K$, then $\nu(\pi) \geq \nu(P_i)$ for $i < K$.

Proof: (i). From Properties 2 and part (ii) of 4 we know that

$$\pi R \leq P_K R \text{ for all } \pi \in \Pi P. \tag{103}$$

Using (103) in the definition of Gittins index (101) we get

$$\nu(\pi) \leq P_K R \text{ for all } \pi \in \Pi P. \tag{104}$$

Letting $\tau = 1$ in (101), we get an lower bound on the Gittins index of P_K

$$\nu(P_K) \geq E[R(\pi_0) | \pi_0 = P_K] = P_K R. \tag{105}$$

Combing (104) and (105) we obtain

$$\nu(P_K) \geq P_K R \geq \nu(\pi) \text{ for all } \pi \in \Pi P. \tag{106}$$

Consequently, the PMF P_K has the largest Gittins index among all PMFs.
From Theorem 4.1 in [20] we know that the second largest Gittens index among PMFs $\{\pi, P_1, P_2, .., P_{K-1}, P_K\}$ is given by

$$\max_{x = \{\pi, P_1, P_2, .., P_{K-1}\}} \nu_K(x), \tag{107}$$

where

$$\nu_K(x) := \frac{A_K(x)}{B_K(x)}, \tag{108}$$

$$A_K(x) := xR + \beta x(K) A_K(P_K), A_K(P_K) = \frac{P_K R}{1-\beta P_{KK}}, \tag{109}$$

$$B_K(x) := 1 + \beta x(K) B_K(P_K), B_K(P_K) = \frac{1}{1-\beta P_{KK}}. \tag{110}$$

We now show that for $P_{K-1} \leq_{st} \pi \leq_{st} P_K$

$$\nu_K(\pi) = \max_{x=\{\pi, P_1, P_2, .., P_{K-1}\}} \nu_K(x). \tag{111}$$

For that matter we need to show that $\nu(\pi_x) \geq \nu(\pi_y)$ whenever $\pi_x \geq_{st} \pi_y, \pi_x, \pi_y \in \Pi P$. From (108),

$$\begin{aligned}
\nu_K(\pi_x) =& \frac{\pi_x R + \beta \pi_x(K) A_K(P_K)}{1 + \beta \pi_x(K) B_K(P_K)} \\
=& \frac{A_K(P_K)}{B_K(P_K)} + \frac{\pi_x R - \frac{A_K(P_K)}{B_K(P_K)}}{1 + \beta \pi_x(K) B_K(P_K)} \\
=& P_K R + \frac{\pi_x R - P_K R}{1 + \beta \pi_x(K) B_K(P_K)} \\
\geq& P_K R + \frac{\pi_y R - P_K R}{1 + \beta \pi_x(K) B_K(P_K)} \\
\geq& P_K R + \frac{\pi_y R - P_K R}{1 + \beta \pi_y(K) B_K(P_K)} \\
=& \nu_K(\pi_y).
\end{aligned} \tag{112}$$

The first inequality in (112) follows from part (ii) of Property 4 and $\pi_x \geq_{st} \pi_y$. The last inequality in (112) holds because $\pi_y R - P_K R \leq 0$ as $\pi_y \leq_{st} P_K$.
Since $\pi \geq_{st} P_i$ for $i = 1, 2, ..., K-1$, (112) ensures that $\nu_K(\pi) \geq \nu_K(P_i)$ for $i = 1, 2, ..., K-1$. Thus, π is the PMF with the second largest Gittins index among $\{\pi, P_1, P_2, .., P_{K-1}, P_K\}$.
The Gittins index for $\pi \in \Pi P, P_{K-1} \leq_{st} \pi \leq_{st} P_K$ is given by

$$\nu(\pi) = \nu_K(\pi) = \frac{\pi R + \beta \pi(K) \frac{P_K R}{1-\beta p_{KK}}}{1 + \beta \pi(K) \frac{1}{1-\beta p_{KK}}}. \tag{113}$$

This completes the proof of (i).
(ii). If $\pi_x, \pi_y \in \Pi P$, $P_{K-1} \leq_{st} \pi_y \leq_{st} \pi_x \leq_{st} P_K$, by (112) and (113), we get

$$\nu(\pi_y) = \nu_K(\pi_y) \leq \nu_K(\pi_x) = \nu(\pi_x). \tag{114}$$

(iii). From part (i) we know that for $\pi \in \Pi P, P_{K-1} \leq_{st} \pi \leq_{st} P_K$, π gives the second largest Gittins index among $\{\pi, P_1, P_2, .., P_{K-1}, P_K\}$. Consequently, $\nu(\pi) \geq \nu(P_i)$ for $i < K$.

Theorem 4. *Under conditions (A1)-(A4) and $L = K$, after time $t = 0$ the Gittins index rule is an optimal channel sensing policy for Problems (P1) and (P2).*

Proof: Consider any time $t > 0$. If the channel observed at time $t-1$ is in state K then the PMF of that channel at t is P_K. The myopic policy senses this channel at t. The Gittins index rule senses the same channel at t as P_K is the PMF with the largest Gittins index by Theorem 3, part (ii).
If the channel observed at time $t-1$ is in state $i, i < K$, then the PMF of that channel at t is P_i and the PMFs of all other channels are stochastically ordered and are stochastically larger than P_{K-1} and stochastically smaller than P_K by Property 2. The myopic policy will choose the channel with the stochastically largest PMF (among all channels that are not observed at $t-1$). By Theorem 3 (ii), the Gittins index of the same channel is the largest among the Gittins indices of all channels that are not observed at $t-1$. By Theorem 3 (iii), the Gittins index of the channel observed at time $t-1$ is $\nu(P_i) \leq \nu(\pi)$ for all $P_{K-1} \leq_{st} \pi \leq_{st} P_K$. Therefore, the Gittins index chooses the same channel as the myopic policy. From the optimality of the myopic policy, under conditions (A1)-(A4) (Theorem 1 and 2) and the condition $L = K$, after time $t = 0$ the Gittins index rule is an optimal channel sensing strategy for problem (P1) and (P2). Note that, if two channels, say channel 1 and 2 are such that $\pi_0^1, \pi_0^2 \in \{P_1, P_2, ..., P_{K-1}\}$ then $\pi_0^1, \pi_0^2 \in \Pi P$ and thus, (A2) is satisfied. Nevertheless π_0^1, π_0^2 do not necessarily satisfy the condition $P_{k-1} \leq_{st} \pi_0^i \leq_{st} P_K$ of Theorem 3. Thus, at $t = 0$, the assertion of Theorem 3 may not be true for channels 1 and 2, thus the Gittins index rule may not be optimal at time 0.

VIII. Conclusion

We investigated a channel sensing problem where each channel has more than two states. We formulated an optimization problem which is an instance of the restless bandit problem. For this problem, we identified conditions sufficient to guarantee the optimality of the myopic policy, the policy that selects at each time instant the channel with the stochastically largest PMF on its states. We also identified conditions under which the Gittins index rule coincides with the myopic policy (and is optimal).

Our results on the optimality of the myopic policy extend previously existing results on the same problem when each channel has two states. In our opinion such an extension is non-trivial for the following reason. When each channel has two states, the information states of the channels can always be totally ordered (as each information state is described by a single number); on the other hand, when each channel has more than two states, the information states of the channels (expressed by their PMF on the states) are not even guaranteed to be partially ordered. Such a lack of order creates serious technical problems, and requires significant insight into the nature of the problem (so as to identify the appropriate assumptions), and much more careful and complicated analysis (so as to establish the optimality of the myopic policy).

Our results on the optimality of the Gittins index rule rely on : (i) the fact that the information state of any channel after $t > 0$ lies stochastically between P_{K-1} and P_K, i.e. $P_{K-1} \leq_{st} \pi \leq_{st} P_K$; and (ii) the fact that $\nu(\hat{\pi}) \geq \nu(\pi)$ whenever $\hat{\pi} \geq_{st} \pi$ and both $\hat{\pi}$ and π are stochastically ordered between P_{K-1} and P_K. We have not been able to prove whether or not the Gittins index rule coincides with the myopic policy when conditions (A1)-(A4) are valid and $L \neq K$ in (A3).

Appendix A

Proof of Property 1 :

$$\begin{aligned} xP - yP &= \sum_{i=1}^{K} (x(i) - y(i)) P_i \\ &= \sum_{i=2}^{K} \left[\left(\sum_{j=i}^{K} (x(j) - y(j)) \right) (P_i - P_{i-1}) \right]. \end{aligned} \tag{115}$$

The last equality follows from a standard identity on the summation by parts of two sequence $\{(x(i) - y(i)), i = 1, 2, ..., K\}$ and $\}P_i, i = 1, 2, ..., K\}$. Note that $\sum_{j=i}^{K} (x(j) - y(j)) \geq 0$ since $x \geq_{st} y$, and by assumption (A1) $P_i \geq_{st} P_{i-1}$. Consequently, $\left(\sum_{j=i}^{K} (x(j) - y(j)) \right) (P_i - P_{i-1}) \geq_{st} \mathbf{0}$, where $\mathbf{0} := (0, 0, ..., 0)$ is the zero vector. Thus by (115)

$$xP - yP \geq_{st} \sum_{i=1}^{K} \mathbf{0} = \mathbf{0}, \tag{116}$$

Hence, $xP \geq_{st} yP$.

Proof of Property 2 :

$$xP^2 = \sum_{i=1}^{K} x(i) P_i P \tag{117}$$

Then, from Property 1 , (A1) and (A3) we obtain

$$P_i P \leq_{st} P_K P \leq_{st} P_L \tag{118}$$

$$P_i P \geq_{st} P_1 P \geq_{st} P_{L-1} \tag{119}$$

The first inequality in (118) and the first inequality in (119) are true because of Property 1 and the fact that $P_1 \leq_{st} P_i \leq_{st} P_K$ (condition (A1)). The second inequality in (118) and the second inequality in (119) are true because of condition (A3). Therefore, (117) along with (118) and (119) give

$$P_{L-1} \leq_{st} xP^2 \leq_{st} P_L \tag{120}$$

Proof of Property 3 : We prove this Property by induction. The Property is true at $t = 0$ by (A2).
Now assume the Property is true at t.
If n, m are not selected at t, $\pi_{t+1}^n = \pi_t^n P$, $\pi_{t+1}^m = \pi_t^m P$.
By the induction hypothesis we have $\pi_t^n \leq_{st} \pi_t^m$ or $\pi_t^m \leq_{st} \pi_t^n$. Then by Property 1, we obtain $\pi_t^n P \leq_{st} \pi_t^m P$ or $\pi_t^m P \leq_{st} \pi_t^n P$, consequently, $\pi_{t+1}^n \leq_{st} \pi_{t+1}^m$ or $\pi_{t+1}^m \leq_{st} \pi_{t+1}^n$.
Suppose, without loss of generality, that channel n is selected at t.

Since channel m is not selected at t, $\pi_{t+1}^m = \pi_t^m P \in \Pi P^2$.
If the observed state is $i \geq L$, then by Property 2, $\pi_{t+1}^n = P_i \geq_{st} P_L \geq_{st} \pi_{t+1}^m$.
If the observed state is $i < L$, then, again by Property 2, $\pi_{t+1}^n = P_i \leq_{st} P_{L-1} \leq_{st} \pi_{t+1}^m$. Consequently, $\pi_{t+1}^n \leq_{st} \pi_{t+1}^m$ or $\pi_{t+1}^m \leq_{st} \pi_{t+1}^n$.

APPENDIX B

Proof of Property 4:

(i) By summation by parts we have

$$\begin{aligned}(x-y)v =& \sum_{i=1}^{K}(x(i)-y(i))v_i \\ =& \sum_{i=2}^{K}\left[\left(\sum_{j=i}^{K}(x(j)-y(j))\right)(v_i - v_{i-1})\right]. \end{aligned} \tag{121}$$

Since $x \geq_{st} y$,

$$\sum_{j=i}^{K}(x(j)-y(j)) \geq 0. \tag{122}$$

The condition $v_i \geq v_{i-1}$, $i = 2, 3, ..., K-1$ in the statement of Property 4, and (122) give

$$\left(\sum_{j=i}^{K}(x(j)-y(j))\right)(v_i - v_{i-1}) \geq 0 \text{ for all } i = 2, 3, ..., K. \tag{123}$$

Then (123) and (121) result in

$$(x-y)v \geq 0. \tag{124}$$

(ii) From the definition of U we have:

$$\text{For } i < L, U_i - U_{i-1} = R_i - R_{i-1}. \tag{125}$$

$$\text{For } i \geq L, U_i - U_{i-1} = R_i - R_{i-1} + \beta(P_i - P_{i-1})U \geq R_i - R_{i-1}. \tag{126}$$

Then, for all i, by the definition of M we obtain

$$\begin{aligned} M_i - M_{i-1} =& U_i - U_{i-1} + \sum_{i \geq L} p_{Ki}(P_i - P_{i-1})U \\ \geq& U_i - U_{i-1} \\ \geq& R_i - R_{i-1} \geq 0. \end{aligned} \tag{127}$$

The first inequality in (127) holds because of condition (A4) (eq. (31)). The second inequality in (127) follows from (125) and (126). From (127), it follows that $M - U$ and $U - R$ are in increasing order (i.e. $M_i - U_i$ and $U_i - R_i$ increase as i increases).
Since $x \geq_{st} y$, from (127) and the result of part (i) we have

$$(x-y)M \geq (x-y)U \geq (x-y)R \geq 0. \tag{128}$$

(iii) Because of Assumption (A4) and the result of part (ii) we have:

$$\text{For } i < L, U_i - U_{i-1} = R_i - R_{i-1} \geq \beta(P_i - P_{i-1})M \geq \beta(P_i - P_{i-1})U. \tag{129}$$

$$\text{For } i \geq L, U_i - U_{i-1} = R_i - R_{i-1} + \beta(P_i - P_{i-1})U \geq \beta(P_i - P_{i-1})U. \tag{130}$$

Then, (129) and (130) imply that $U - \beta PU$ is in increasing order, consequently by the result of part (i) we obtain

$$(x-y)U \geq \beta(x-y)PU. \tag{131}$$

Since $M = U + \beta \sum_{i \geq L} p_{Ki} PU$,

$$\begin{aligned}(x-y)M =&(x-y)(U + \beta \sum_{i \geq L} p_{Ki} PU)\\ =&(x-y)U + \beta \sum_{i \geq L} p_{Ki}(xP - yP)U\\ \geq&\beta(x-y)PU + \beta \sum_{i \geq L} p_{Ki}\beta(xP - yP)PU\\ =&\beta(x-y)PM, \end{aligned} \tag{132}$$

where the inequality in (132) is a consequence of (131).

(iv) If $x(i) = y(i)$ for all $i \geq L$, then $x(i) - y(i) = 0$ for $i \geq L$.
Define $v := (v_1, v_2, ..., v_K)$ such that

$$v_i = R_i - \beta P_i M \text{ for } i = 1, 2, ..., L-1, \tag{133}$$

$$v_i = v_{L-1} \text{ for } i \geq L. \tag{134}$$

From assumption (31) in (A4) we know that $v_i - v_{i-1} = R_i - R_{i-1} - \beta(P_i - P_{i-1})M \geq 0$ for $i \leq L-1$ and $v_i - v_{i-1} = 0$ for $i \geq L$. Then by the result of part (i) we obtain

$$\begin{aligned}(x-y)(R - \beta PM) =& \sum_{i=1}^{L-1}(x(i) - y(i))(R_i - \beta P_i M)\\ =& \sum_{i=1}^{L-1}(x(i) - y(i))v_i + \sum_{i \geq L}(x(i) - y(i))v_i\\ =&(x-y)v \geq 0. \end{aligned} \tag{135}$$

The second equality in (135) follows from the definition of v_i (eq. (133)) and the fact that $x(i) - y(i) = 0$ for $i \geq L$. The inequality in (135) is true by the result of part (i).
Since $M = U + \beta \sum_{i \geq L} p_{Ki} PU$ and $x \geq_{st} y$, it follows that

$$\beta(x-y)PU \leq \beta(x-y)P(U + \beta \sum_{i \geq L} p_{Ki} PU) = \beta(x-y)PM \leq (x-y)R, \tag{136}$$

where the first inequality in (136) follows from the fact that $xP^2 \geq_{st} yP^2$, the fact that U_i is increasing with i, and the result of part (i); and the last inequality in (136) follows from (135).
The case where $x(i) = y(i)$ for all $i < L$ can be proved in the same way.

APPENDIX C

Proof of Property 5: We want to show that under $g^{O_0}_{0:T}$, at any time t the ordering O_t has the property that $\pi_t^{O_t(1)} \leq_{st} \pi_t^{O_t(2)} \leq_{st} ... \leq_{st} \pi_t^{O_t(N)}$.
At $t = 0$, by the statement of Property 5, the initial ordering O_0 is such that $\pi_0^{O_0(1)} \leq_{st} \pi_0^{O_0(2)} \leq_{st} ... \leq_{st} \pi_0^{O_0(N)}$.
Suppose at time t, the ordering O_t is such that $\pi_t^{O_t(1)} \leq_{st} \pi_t^{O_t(2)} \leq_{st} ... \leq_{st} \pi_t^{O_t(N)}$.
If the observation is $Y_t \geq L$, the new ordering is $O_{t+1} = \hat{m}(O_t, Y_t) = O_t$ and the PMFs of the channels evolves to

$$\pi_{t+1}^n = \pi_t^n P \text{ for } n \neq O_t(N), \tag{137}$$

$$\pi_{t+1}^{O_t(N)} = P_{Y_t} \geq_{st} P_L. \tag{138}$$

From Properties 1 and 2 we know that

$$\pi_t^{O_t(1)} P \leq_{st} \pi_t^{O_t(2)} P \leq_{st} ... \leq_{st} \pi_t^{O_t(N-1)} P \leq_{st} P_L \leq_{st} P_{Y_t}, \tag{139}$$

therefore,

$$\pi_{t+1}^{O_{t+1}(1)} \leq_{st} \pi_{t+1}^{O_{t+1}(2)} \leq_{st} ... \leq_{st} \pi_{t+1}^{O_{t+1}(N)}. \tag{140}$$

On the other hand, if the observation is $Y_t < L$, the new ordering is $O_{t+1} = \hat{m}(O_t, Y_t) = SO_t$ and the PMFs of the channels become

$$\pi_{t+1}^n = \pi_t^n P \text{ for } n \neq O_t(N), \tag{141}$$

$$\pi_{t+1}^{O_t(N)} = P_{Y_t} \leq_{st} P_{L-1}. \tag{142}$$

Again, from Properties 1 and 2 we get

$$P_{Y_t} \leq_{st} P_{L-1} \leq_{st} \pi_t^{O_t(1)} P \leq_{st} \pi_t^{O_t(2)} P \leq_{st} ... \leq_{st} \pi_t^{O_t(N-1)} P, \tag{143}$$

hence,

$$\pi_{t+1}^{O_{t+1}(1)} \leq_{st} \pi_{t+1}^{O_{t+1}(2)} \leq_{st} ... \leq_{st} \pi_{t+1}^{O_{t+1}(N)}. \tag{144}$$

Thus, the ordering-based policy $g_{0:T}^{O_0}$ selects at any time t the channel $O_t(N)$ from the ordering O_t with $\pi_t^{O_t(1)} \leq_{st} \pi_t^{O_t(2)} \leq_{st} ... \leq_{st} \pi_t^{O_t(N)}$. This ordering-based policy is exactly the same as the myopic policy g^m.

APPENDIX D

We first establish a lemma that is needed for the proof of Properties 6-9.

Lemma 1. *The functions $V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N)$, $t = 1, 2, ..., T$ (defined by eq. (60)), are linear in every component $\pi_t^n, n = 1, 2, ..., N$.*
That is, for all $n = 1, 2, ..., N$

$$V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) = \sum_{i=1}^{K} \pi^n(i) V_t(O_t, \pi_t^1, ..., \pi_t^{n-1}, e_i, \pi_t^{n+1}, ..., \pi_t^N), \tag{145}$$

where e_i is the vector with 1 in the ith position and 0 otherwise, i.e. $e_i = [0, ..., 0, \underset{\uparrow\, i\text{th position}}{1}, 0, ..., 0]$.

Furthermore, $L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N)$ satisfies for $n = 2, 3, ..., N$

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) = \sum_{i=1}^{K} \pi^n(i) L_t(O_t, \hat{\pi}_t^1, \pi_t^1, ..., \pi_t^{n-1}, e_i, \pi_t^{n+1}, ..., \pi_t^N), \tag{146}$$

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^2, ..., \pi_t^N) = \sum_{i=1}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)) V_t(O_t, e_i, \pi_t^2, ..., \pi_t^N). \tag{147}$$

Proof: By definition of V_t (eq (60)) we have

$$\begin{aligned} V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) :=& E^{g_{t:T}^{O_t}}[\sum_{s=t}^{T} \beta^{s-t} R(s) | \pi_t^1, \pi_t^2, ..., \pi_t^N] \\ =& \sum_{i=1}^{K} \pi_t^n(i) E^{g_{t:T}^{O_t}}[\sum_{s=t}^{T} \beta^{s-t} R(s) | \pi_t^1, \pi_t^2, ..., \pi_t^N, X_t^n = i]. \end{aligned} \tag{148}$$

Because of the specification of the ordering-based policy $g_{t:T}^{O_t}$ and the fact that conditional on $\{X_t^n = i, \pi_t^n\}$ the evolution of channel n is the same as that conditional on $\{\pi_t^n = e_i\}$, we have

$$\begin{aligned} & E^{g_{t:T}^{O_t}}[\sum_{s=t}^{T} \beta^{s-t} R(s) | \pi_t^1, \pi_t^2, ..., \pi_t^N, X_t^n = i] \\ =& E^{g_{t:T}^{O_t}}[\sum_{s=t}^{T} \beta^{s-t} R(s) | \pi_t^1, ..., \pi_t^{n-1}, \pi_t^{n+1}, ..., \pi_t^N, \pi_t^n = e_i]. \end{aligned} \tag{149}$$

Then from (148) and (149) we obtain

$$\begin{aligned} & V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \\ =& \sum_{i=1}^{K} \pi_t^n(i) E^{g_{t:T}^{O_t}}[\sum_{s=t}^{T} \beta^{s-t} R(s) | \pi_t^1, ..., \pi_t^{n-1}, \pi_t^{n+1}, ..., \pi_t^N, \pi_t^n = e_i] \\ =& \sum_{i=1}^{K} \pi_t^n(i) V_t(O_t, \pi_t^1, ..., \pi_t^{n-1}, e_i, \pi_t^{n+1}, ..., \pi_t^N). \end{aligned} \tag{150}$$

Furthermore, L_t is the difference of two V_t's, so the linearity of V_t leads directly to equations (146) and (147). We Proceed now with the proof of Properties 6-9. In the following proofs, we use the notation

$$\pi_t^{k_1:k_2} := (\pi_t^{k_1}, \pi_t^{k_1+1}, ..., \pi_t^{k_2}) \tag{151}$$

$$\pi_t^{k_1:k_2} P := (\pi_t^{k_1} P, \pi_t^{k_1+1} P, ..., \pi_t^{k_2} P) \tag{152}$$

Proof of Properties 6-9: First note that Property 8 is a special case of Property 7. This can be seen as follows. Without loss of generality, let $O_t(n) = 1, O_t(m) = 2$, and $\pi_t^1 \geq_{st} \pi_t^2$. Note that

$$V_t(O_t, \pi_t^2, \pi_t^2, ..., \pi_t^N) = V_t(W_{nm}O_t, \pi_t^2, \pi_t^2, ..., \pi_t^N). \tag{153}$$

Applying Property 7 at time t, we have

$$\begin{aligned} &V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) - V_t(W_{nm}O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \\ =&V_t(O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) - V_t(O_t, \pi_t^2, \pi_t^2, ..., \pi_t^N) + V_t(W_{nm}O_t, \pi_t^2, \pi_t^2, ..., \pi_t^N) - V_t(W_{nm}O_t, \pi_t^1, \pi_t^2, ..., \pi_t^N) \\ =&L_t(O_t, \pi_t^1, \pi_t^2, \pi_t^2, ..., \pi_t^N) - L_t(W_{nm}O_t, \pi_t^1, \pi_t^2, \pi_t^2, ..., \pi_t^N) \geq 0. \end{aligned} \tag{154}$$

The first equality in (154) holds because of (153). The second equality is a consequence of the definition of L_t (eq (64)). The inequality follows from Property 7 at t.
Therefore, Property 8 is true at time t once Property 7 is true at time t.

We will prove all three Properties 6, 7 and 9 simultaneously by induction.
We remind the reader that for Properties 6, 7 and 9 $O_t \in \mathcal{O}$ with $O_t(n) = 1$, $1 \leq m < n \leq N$ and

$$S^{-m}O_t = (O_t(m+1), O_t(m+2), ..., O_t(N), O_t(1), ..., O_t(m)), \tag{155}$$

$$W_{nm}O_t(i) = \begin{cases} O_t(n) & \text{for } i = m \\ O_t(m) & \text{for } i = n \\ O_t(i) & \text{otherwise} \end{cases}, \tag{156}$$

$$A_{nm}O_t(i) = \begin{cases} O_t(n) & \text{for } i = m \\ O_t(i-1) & \text{for } i = m+1, m+2, ..., n \\ O_t(i) & \text{otherwise} \end{cases}. \tag{157}$$

For both the basis of induction and the induction we consider two cases.

(i) When channel 1 is not the right-most channel in O_t (i.e. $n \neq N$ and $O_t(N) \neq 1$).
(ii) When channel 1 is the right-most channel in O_t (i.e. $n = N$ and $O_t(N) = 1$).

Basis of induction

For Property 6:
(i) If $O_T(N) \neq 1$ (i.e. $n \neq N$),

$$L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(S^{-m}O_T, \hat{\pi}_T^1, \pi_T^{1:N}) = (\pi_T^{O_T(N)}R - \pi_T^{O_T(N)}R) - (\pi_T^{O_T(m)}R - \pi_T^{O_T(m)}R) = 0. \tag{158}$$

(ii) If $O_T(N) = 1$ (i.e. $n = N$), then

$$L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(S^{-m}O_T, \hat{\pi}_T^1, \pi_T^{1:N}) = (\hat{\pi}_T^1 R - \pi_T^1 R) - (\pi_T^{O_T(m)}R - \pi_T^{O_T(m)}R) = (\hat{\pi}_T^1 - \pi_T^1)R. \tag{159}$$

By part (ii) of Property 4 and $\hat{\pi}_t^1 \geq_{st} \pi_t^1$ we get

$$(\hat{\pi}_T^1 - \pi_T^1)U \geq (\hat{\pi}_T^1 - \pi_T^1)R \geq 0. \tag{160}$$

Combing (159) with (160) we obtain

$$(\hat{\pi}_T^1 - \pi_T^1)U \geq (\hat{\pi}_T^1 - \pi_T^1)R = L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(S^{-m}O_T, \hat{\pi}_T^1, \pi_T^{1:N}) \geq 0. \tag{161}$$

For Property 7:
(i) If $O_T(N) \neq 1$ (i.e. $n \neq N$),

$$L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(W_{nm}O_T, \hat{\pi}_T^1, \pi_T^{1:N}) = (\pi_T^{O_T(N)}R - \pi_T^{O_T(N)}R) - (\pi_T^{O_T(m)}R - \pi_T^{O_T(m)}R) = 0. \tag{162}$$

(ii) If $O_T(N) = 1$ (i.e. $n = N$), then

$$L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(W_{nm}O_t, \hat{\pi}_t^1, \pi_T^{1:N}) = (\hat{\pi}_T^1 R - \pi_T^1 R) - (\pi_T^{O_T(m)}R - \pi_T^{O_T(m)}R) = (\hat{\pi}_T^1 - \pi_T^1)R. \tag{163}$$

By part (ii) of Property 4 and $\hat{\pi}_t^1 \geq_{st} \pi_t^1$ we get

$$(\hat{\pi}_T^1 - \pi_T^1)M \geq (\hat{\pi}_T^1 - \pi_T^1)R \geq 0. \tag{164}$$

Combing (163) with (164) we obtain

$$(\hat{\pi}_T^1 - \pi_T^1)M \geq (\hat{\pi}_T^1 - \pi_T^1)R = L_T(O_T, \hat{\pi}_T^1, \pi_T^{1:N}) - L_T(S^{-m}O_t, \hat{\pi}_t^1, \pi_T^{1:N}) \geq 0. \tag{165}$$

For Property 9:

Since $P_K \geq P_i$, by part (ii) of Property 4, we get

$$h := \frac{P_K R - \beta \sum_{i<L} p_{Ki} P_i R}{1 - \beta \sum_{i<L} p_{Ki}} \geq \frac{P_K R - \beta \sum_{i<L} p_{Ki} P_K R}{1 - \beta \sum_{i<L} p_{Ki}} = P_K R \tag{166}$$

Consequently, part (ii) of Property 4 ensures that

$$\pi R \leq P_K R \leq h \text{ for all } \pi \in \Pi P. \tag{167}$$

Then:
(i) If $O_T(N) \neq 1$ (i.e. $n \neq N$), we have

$$V_T(A_{nm} O_T, \pi_T^{1:N}) - V_t(O_T, \pi_T^{1:N}) = \pi_T^{O_T(N)} R - \pi_T^{O_T(N)} R = 0 \leq h - \pi_T^1 P^{N-n} R. \tag{168}$$

The inequality in (168) follows from (167) and the fact that $\pi_T^1 P^{N-n} \in \pi P$.
(ii) If $O_T(N) = 1$(i.e. $n = N$), we have

$$V_T(A_{nm} O_T, \pi_T^{1:N}) - V_t(O_T, \pi_T^{1:N}) = \pi_T^{O_T(N-1)} R - \pi_T^1 R \leq h - \pi_T^1 R. \tag{169}$$

The inequality in (169) follows from (167).
This completes the basis of induction.

Induction hypothesis
Assume that the assertions of Properties 6, 7 and 9 are true for time $t+1, t+2, ..., T$.
Induction step
We prove here Properties 6, 7 and 9 for t.
We first develop five expressions (175),(178),(179), (180) and (184) for L_t and L_{t+1} defined by eq. (64), that will be useful in the sequel.
For any PMF $\pi \in \Pi$ we define

$$\underline{\pi} := (\pi(1), \pi(2), ..., \pi(L-2), \sum_{i=L-1}^{K} \pi(i), 0, ..., 0), \tag{170}$$

$$\bar{\pi} := (0, ..., 0, \sum_{i=1}^{L} \pi(i), \pi(L+1), ..., \pi(K)) \tag{171}$$

Then, $\underline{\pi}, \bar{\pi} \in \Pi$, and

$$\pi = \underline{\pi} + \bar{\pi} - e_L + \sum_{i=L}^{K} \pi(i)(e_L - e_{L-1}) \tag{172}$$

Furthermore, if $\hat{\pi} \geq_{st} \pi$, it follows that

$$\underline{\hat{\pi}} \geq_{st} \underline{\pi}, \tag{173}$$

$$\bar{\hat{\pi}} \geq_{st} \bar{\pi}. \tag{174}$$

Consider any arbitrary ordering $O \in \mathcal{O}$. When $O(N) \neq 1$, assume $O(N) = 2$ without any loss of generality. Then,

$$\begin{aligned}
&L_t(O, \hat{\pi}_t^1, \pi_t^1, \pi_t^{2:N}) \\
:=&V_t(O, \hat{\pi}_t^1, \pi_t^{2:N}) - V_t(O, \pi_t^1, \pi_t^{2:N}) \\
=&(\pi_t^2 R - \pi_t^2 R) + \beta \sum_{i<L} \pi_t^2(i)(V_{t+1}(SO, \hat{\pi}_t^1 P, P_i, \pi_t^{3:N} P) - V_{t+1}(SO, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
&+ \beta \sum_{i \geq L} \pi_t^2(i)(V_{t+1}(O, \hat{\pi}_t^1 P, P_i, \pi_t^{3:N} P) - V_{t+1}(O, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
=&\beta \sum_{i<L} \pi_t^2(i) L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) + \beta \sum_{i \geq L} \pi_t^2(i) L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P).
\end{aligned} \tag{175}$$

The second equality in (175) follows from the recursive equation for V_t (eq. (62)). The last equality in (175) follows from the definition of L_t (eq. 64).
Furthermore, by the induction hypothesis for Property 6, we get, for all $i = 1, 2, ..., K$,

$$L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \geq L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P). \tag{176}$$

Therefore,

$$\begin{aligned}
&\beta L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^{1:N} P) \\
=&\beta \sum_{i=1}^{L} \pi_t^2(i) L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \\
\geq&\beta \sum_{i<L} \pi_t^2(i) L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) + \beta \sum_{i\geq L} \pi_t^2(i) L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \\
\geq&\beta \sum_{i=1}^{L} \pi_t^2(i) L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \\
=&\beta L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^{1:N} P). \qquad (177)
\end{aligned}$$

The equalities in (177) are true because of the linearity of L_t (Lemma 1). The inequalities in (177) are true because of (176). Combing (175) and (177) we get

$$\beta L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^{1:N} P) \geq L_t(O, \hat{\pi}_t^1, \pi_t^{1:N}). \qquad (178)$$

$$L_t(O, \hat{\pi}_t^1, \pi_t^{1:N}) \geq \beta L_{t+1}(O, \hat{\pi}_t^1 P, \pi_t^{1:N} P). \qquad (179)$$

When $O(N) = 1$,

$$\begin{aligned}
&L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
:=&V_t(O_t, \hat{\pi}_t^1, \pi_t^{2:N}) - V_t(O_t, \pi_t^1, \pi_t^{2:N}) \\
=&(\hat{\pi}_t^1 R - \pi_t^1 R) + \beta \sum_{i<L} (\hat{\pi}_t^1(i) - \pi_t^1(i)) V_{t+1}(SO_t, P_i, \pi_t^{2:N} P) + \beta \sum_{i\geq L} (\hat{\pi}_t^1(i) - \pi_t^1(i)) V_{t+1}(O_t, P_i, \pi_t^{2:N} P) \\
=&(\hat{\pi}_t^1 - \pi_t^1) R + \beta \sum_{i=1}^{K} (\underline{\hat{\pi}}_t^1(i) - \underline{\pi}_t^1(i)) V_{t+1}(SO_t, P_i, \pi_t^{2:N} P) + \beta \sum_{i=1}^{K} (\bar{\hat{\pi}}_t^1(i) - \bar{\pi}_t^1(i)) V_{t+1}(O_t, P_i, \pi_t^{2:N} P) \\
&+ \beta (V_{t+1}(O_t, P_L, \pi_t^{2:N} P) - V_{t+1}(SO_t, P_{L-1}, \pi_t^{2:N} P)) \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)) \\
=&(\hat{\pi}_t^1 - \pi_t^1) R + \beta L_{t+1}(SO_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) + \beta L_{t+1}(O_t, \bar{\hat{\pi}}_t^1 P, \bar{\pi}_t^1 P, \pi_t^{2:N} P) \\
&+ \beta (V_{t+1}(O_t, P_L, \pi_t^{2:N} P) - V_{t+1}(SO_t, P_{L-1}, \pi_t^{2:N} P)) \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)). \qquad (180)
\end{aligned}$$

The second equality in (180) follows from the recursive equation for V_t (eq. (62)). The third equality in (180) is true because of the definition of $\underline{\pi}, \bar{\pi}$ given by (170) and (171). The last equality in (180) follows from the linearity of L_t (Lemma 1). Furthermore, using (180) we get

$$\begin{aligned}
&L_t(O, \hat{\pi}_t^1, \pi_t^{1:N}) - \beta L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^{1:N} P) \\
=&(\hat{\pi}_t^1 - \pi_t^1) R + \beta L_{t+1}(SO, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) + \beta L_{t+1}(O, \bar{\hat{\pi}}_t^1 P, \bar{\pi}_t^1 P, \pi_t^{2:N} P) \\
&+ \beta (V_{t+1}(O, P_L, \pi_t^{2:N} P) - V_{t+1}(SO, P_{L-1}, \pi_t^{2:N} P)) \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)) \\
&- \beta L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, \pi_t^{2:N} P) \\
=&(\hat{\pi}_t^1 - \pi_t^1) R + \beta L_{t+1}(O, \bar{\hat{\pi}}_t^1 P, \bar{\pi}_t^1 P, \pi_t^{2:N} P) - \beta L_{t+1}(SO, \bar{\hat{\pi}}_t^1 P, \bar{\pi}_t^1 P, \pi_t^{2:N} P) \\
&+ \beta (V_{t+1}(O, P_L, \pi_t^{2:N} P) - V_{t+1}(SO, P_L, \pi_t^{2:N} P) \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)) \\
\leq&(\hat{\pi}_t^1 - \pi_t^1) R + \beta (\bar{\hat{\pi}}_t^1 - \bar{\pi}_t^1) PU \\
&+ \beta (V_{t+1}(O, P_L, \pi_t^{2:N} P) - V_{t+1}(SO, P_L, \pi_t^{2:N} P) \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i)). \qquad (181)
\end{aligned}$$

The first equality in (181) follows from (180). The second equality in (181) follows from (172) and the linearity of L_t (Lemma 1). The inequality in (181) follows from the induction hypothesis for the upper bound of Property 6 at $t+1$ and the fact that $\bar{\hat{\pi}}_t^1 P \geq_{st} \bar{\pi}_t^1 P$.

For the last term in (181), because $\bar{\hat{\pi}}_t^1 \geq_{st} \bar{\pi}_t^1$, we have

$$\sum_{i=L}^{K}(\hat{\pi}_t^1(i) - \pi_t^1(i)) \geq 0. \tag{182}$$

Moreover,

$$\begin{aligned}
&V_{t+1}(O, P_L, \pi_t^{2:N}P) - V_{t+1}(SO, P_L, \pi_t^{2:N}P)\\
=&L_{t+1}(O, P_L, P_{L-1}, \pi_t^{2:N}P) - L_{t+1}(SO, P_L, P_{L-1}, \pi_t^{2:N}P)\\
&+ V_{t+1}(O, P_{L-1}, \pi_t^{2:N}P) - V_{t+1}(SO, P_{L-1}, \pi_t^{2:N}P)\\
=&L_{t+1}(O, P_L, P_{L-1}, \pi_t^{2:N}P) - L_{t+1}(SO, P_L, P_{L-1}, \pi_t^{2:N}P)\\
&+ V_{t+1}(O, P_{L-1}, \pi_t^{2:N}P) - V_{t+1}(W_{12}...W_{(N-1)(N-2)}W_{N(N-1)}O, P_{L-1}, \pi_t^{2:N}P)\\
\leq&L_{t+1}(O, P_L, P_{L-1}, \pi_t^{2:N}P) - L_{t+1}(SO, P_L, P_{L-1}, \pi_t^{2:N}P)\\
\leq&(P_L - P_{L-1})U.
\end{aligned} \tag{183}$$

The first equality in (183) follows from the definition of L_{t+1}. The second equality in (183) is true because $SO = W_{12}...W_{(N-1)(N-2)}W_{N(N-1)}O$. The first inequality in (183) follows by repeatedly using Property 8 at $t+1$ and the fact that $\pi_t^m P \geq_{st} P_{L-1}$ for all $m = 2, 3, ..., N$. The second inequality in (183) follows from the induction hypothesis for the upper bound of Property 6 at $t+1$ and the fact that $P_L \geq_{st} P_{L-1}$.
Therefore, using (182) and (183) in (181) give

$$\begin{aligned}
&L_t(O, \hat{\pi}_t^1, \pi_t^1, \pi_t^{2:N}) - \beta L_{t+1}(SO, \hat{\pi}_t^1 P, \pi_t^1 P, \pi_t^{2:N}P)\\
\leq&(\hat{\pi}_t^1 - \pi_t^1)R + \beta(\bar{\hat{\pi}}_t^1 - \bar{\pi}_t^1)PU + \beta(P_L - P_{L-1})U\sum_{i=L}^{K}(\hat{\pi}_t^1(i) - \pi_t^1(i))\\
=&(\hat{\pi}_t^1 - \pi_t^1)R + \beta\sum_{i\geq L}(\hat{\pi}_t^1(i) - \pi_t^1(i))P_iU + \beta\sum_{i<L}(\hat{\pi}_t^1(i) - \pi_t^1(i))P_{L-1}U\\
=&(\hat{\pi}_t^1 - \pi_t^1)U.
\end{aligned} \tag{184}$$

The inequality in (184) follows from (181), (182) and (183). The first equality in (184) follows from the definition of $\bar{\hat{\pi}}_t^1$ and $\bar{\pi}_t^1$ given by (171). The last equality in (184) follows from the definition of U.

Induction step for Property 6:
We first consider the lower bound of Property 6. We want to show that

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \geq L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N}). \tag{185}$$

(i) When $O_t(N) \neq 1$ (i.e. $n \neq N$), we also have $S^{-m}O_t(N) = O_t(m) \neq 1$. Then,

$$\begin{aligned}
L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \geq&\beta L_{t+1}(O_t, \hat{\pi}_t^1 P, \pi_t^{1:N}P)\\
=&\beta L_{t+1}(S^m S^{-m}O_t, \hat{\pi}_t^1 P, \pi_t^{1:N}P)\\
\geq&\beta L_{t+1}(S^{1-m}O_t, \hat{\pi}_t^1 P, \pi_t^{1:N}P)\\
\geq&L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N}).
\end{aligned} \tag{186}$$

The first inequality in (186) follows from (178) and the fact that $O_t(N) \neq 1$. The second inequality in (186) follows from the induction hypothesis for Property 6 at $t+1$. The last inequality in (186) follows from (179) and the fact that $S^{-m}O_t(N) \neq 1$. This completes the proof of the lower bound of Property 6 for case (i).

(ii) When $O_t(N) = 1$ (i.e. $n = N$).
Since $S^{-m}O_t(N) = O_t(m) \neq 1$, we get

$$\begin{aligned}
&L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N})\\
\leq&\beta L_{t+1}(S^{1-m}O_t, \hat{\pi}_t^1 P, \pi_t^{1:N}P)\\
=&\beta L_{t+1}(S^{1-m}O_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N}P) + \beta L_{t+1}(S^{1-m}O_t, \bar{\hat{\pi}}_t^1 P, \bar{\pi}_t^1 P, \pi_t^{2:N}P)\\
&+ \beta\sum_{i=L}^{K}(\hat{\pi}_t^1(i) - \pi_t^1(i))L_{t+1}(S^{1-m}O_t, P_L, P_{L-1}, \pi_t^{2:N}P)
\end{aligned} \tag{187}$$

The inequality in (187) follows from (179) and the fact that $S^{-m}O_t(N) \neq 1$. The equality in (187) follows from (172) and

the linearity of L_t (Lemma 1).
Since $O_t(N)=1$, applying (180) we obtain

$$\begin{aligned}
&L_t(O_t,\hat{\pi}_t^1,\pi_t^{1:N})-L_t(S^{-m}O_t,\hat{\pi}_t^1,\pi_t^{1:N})\\
=&(\hat{\pi}_t^1-\pi_t^1)R+\beta L_{t+1}(SO_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)+\beta L_{t+1}(O_t,\bar{\hat{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P)\\
&+\beta(V_{t+1}(O_t,P_L,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P))\sum_{i=L}^{K}(\hat{\pi}_t^1(i)-\pi_t^1(i))-L_t(S^{-m}O_t,\hat{\pi}_t^1,\pi_t^1,\pi_t^{2:N})\\
\geq&(\hat{\pi}_t^1-\pi_t^1)R+\beta L_{t+1}(SO_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)-\beta L_{t+1}(S^{1-m}O_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)\\
&+\beta L_{t+1}(O_t,\bar{\hat{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P)-\beta L_{t+1}(S^{1-m}O_t,\bar{\hat{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P)\\
&+\beta(V_{t+1}(O_t,P_L,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P)-L_{t+1}(S^{1-m}O_t,P_L,P_{L-1},\pi_t^{2:N}P))\sum_{i=L}^{K}(\hat{\pi}_t^1(i)-\pi_t^1(i))\\
\geq&(\hat{\pi}_t^1-\pi_t^1)R+\beta L_{t+1}(SO_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)-\beta L_{t+1}(S^{1-m}O_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)\\
&+\beta(V_{t+1}(O_t,P_L,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P)-L_{t+1}(S^{1-m}O_t,P_L,P_{L-1},\pi_t^{2:N}P))\sum_{i=L}^{K}(\hat{\pi}_t^1(i)-\pi_t^1(i)).
\end{aligned}\tag{188}$$

The equality in (188) follows from (180) and the fact that $O_t(N)=1$. The first inequality in (188) follows from (187). The second inequality in (188) follows from the induction hypothesis for the lower bound of Property 6 at $t+1$ and the fact that $\bar{\hat{\pi}}_t^1P\geq_{st}\bar{\pi}_t^1P$.
Letting $\underline{O}_{t+1}:=S^{1-m}O_t$ and $\underline{n}:=N+1-m,\underline{m}:=N-m$, we have $\underline{m}<\underline{n}$ and

$$\underline{O}_{t+1}(\underline{n})=S^{1-m}O_t(\underline{n})=1,\tag{189}$$

$$SO_t=S^{-(\underline{m})}\underline{O}_{t+1}.\tag{190}$$

Consequently, the induction hypothesis for the upper bound of Property 6 at $t+1$ gives

$$\begin{aligned}
&L_{t+1}(S^{1-m}O_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)-L_{t+1}(SO_t,\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)\\
=&L_{t+1}(\underline{O}_{t+1},\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)-L_{t+1}(S^{-(\underline{m})}\underline{O}_{t+1},\underline{\hat{\pi}}_t^1P,\underline{\pi}_t^1P,\pi_t^{2:N}P)\leq(\underline{\hat{\pi}}_t^1P-\underline{\pi}_t^1P)U.
\end{aligned}\tag{191}$$

Letting $\underline{m}':=1$, we have $\underline{m}'<n=N$ and

$$A_{\underline{m}'n}O_t=SO_t.\tag{192}$$

Therefore,

$$\begin{aligned}
&V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P)-V_{t+1}(O_t,P_L,\pi_t^{2:N}P)+L_{t+1}(S^{1-m}O_t,P_L,P_{L-1},\pi_t^{2:N}P)\\
\leq&V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P)-V_{t+1}(O_t,P_L,\pi_t^{2:N}P)+L_{t+1}(O_t,P_L,P_{L-1},\pi_t^{2:N}P)\\
=&V_{t+1}(A_{\underline{m}'n}O_t,P_{L-1},\pi_t^{2:N}P)-V_{t+1}(O_t,P_{L-1},\pi_t^{2:N}P)\\
\leq&h-P_{L-1}R.
\end{aligned}\tag{193}$$

The first inequality in (193) follows from the induction hypothesis for the lower bound of Property 6 at $t+1$ and the fact that $P_L\geq_{st}P_{L-1}$. The equality in (193) follows from the definition of L_{t+1} and (192). The last inequality in (193) follows from the induction hypothesis for Property 9 at $t+1$ and the fact that $P_{L-1}\in\pi P$, therefore $P_{L-1}\leq_{st}P_{L-1}P$ by Property 2.

Using (191) and (193) in (188) we obtain

$$\begin{aligned}
&L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
\geq&(\hat{\pi}_t^1 - \pi_t^1)R - \beta(\underline{\hat{\pi}}_t^1 P - \underline{\pi}_t^1 P)U - \beta \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))(h - P_{L-1}R) \\
=&(\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)R + (\bar{\hat{\pi}}_t^1 - \bar{\pi}_t^1)R + \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))(R_L - R_{L-1}) \\
&- \beta(\underline{\hat{\pi}}_t^1 P - \underline{\pi}_t^1 P)U - \beta \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))(h - P_{L-1}R) \\
=&(\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)(R - \beta U) + (\bar{\hat{\pi}}_t^1 - \bar{\pi}_t^1)R + \sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))(R_L - R_{L-1} - \beta(h - P_{L-1}R)) \\
\geq&0. \qquad (194)
\end{aligned}$$

The first inequality in (194) follows from eqs (191) and (193) and the fact that $\sum_{i=L}^{K}(\hat{\pi}_t^1(i) - \pi_t^1(i)) \geq 0$ (since $\hat{\pi}_t^1(i) \geq_{st} \pi_t^1$). The first equality in (194) follows from (172). The last inequality in (194) is true because: the terms $(\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)(R - \beta U)$ and $(\bar{\hat{\pi}}_t^1 - \bar{\pi}_t^1)R$ are positive by parts (iv) and (ii) of Property 4 and the fact that $\underline{\hat{\pi}}_t^1 \geq_{st} \underline{\pi}_t^1$ and $\bar{\hat{\pi}}_t^1 \geq_{st} \bar{\pi}_t^1$; the term $(R_L - R_{L-1} - \beta(h - P_{L-1}R))$ is positive by condition (A4).
The proof of the lower bound of Property 6 is now complete.

Now consider the upper bound of Property 6. We want to show that

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \leq (\hat{\pi}_t^1 - \pi_t^1)U. \qquad (195)$$

Let $O'_t := S^{N-n}O_t$;, then $O'_t(N) = 1$ and $SO'_t(1) = 1$. Consequently,

$$\begin{aligned}
L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(S^{-m}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \leq& L_t(O'_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(SO'_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
\leq& L_t(O'_t, \hat{\pi}_t^1, \pi_t^{1:N}) - \beta L_{t+1}(SO'_t, \hat{\pi}_t^1 P, \pi_t^{1:N} P) \\
\leq& (\hat{\pi}_t^1 - \pi_t^1)U. \qquad (196)
\end{aligned}$$

The first inequality in (196) is true because of the lower bound of Property 6 at t. The second inequality in (196) follows from (179) and the fact that $SO'_t(N) \neq 1$. The third inequality in (196) follows from (184) and the fact that $O'_t(N) = 1$.
This completes the proof of Property 6 at time t.

Induction step for Property 7:
(i) When $O_t(N) \neq 1$ (i.e. $n \neq N$), assume $O_t(N) = 2$ without loss of generality. Then because of (175),

$$\begin{aligned}
&L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
=&\beta \sum_{i<L} \pi^2(i)(L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(S(W_{nm}O_t), \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
&+ \beta \sum_{i \geq L} \pi^2(i)(L_{t+1}(O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{nm}O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
=&\beta \sum_{i<L} \pi^2(i)(L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{(n+1)(m+1)}(SO_t), \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
&+ \beta \sum_{i \geq L} \pi^2(i)(L_{t+1}(O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{nm}O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)). \qquad (197)
\end{aligned}$$

The first equality in (197) follows from (175). The second equality is true because $S(W_{nm}O_t) = W_{(n+1)(m+1)}(SO_t)$.

By the induction hypothesis for Property 7, each term in (197) is positive and smaller than $(\hat{\pi}_t^1 P - \pi_t^1 P)M$. Thus,

$$\begin{aligned}
0 \leq & L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
= & \beta \sum_{i<L} \pi^2(i)(L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{(n+1)(m+1)}(SO_t), \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
& + \beta \sum_{i \geq L} \pi^2(i)(L_{t+1}(O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{nm}O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
\leq & \beta(\hat{\pi}_t^1 P - \pi_t^1 P)M \\
\leq & (\hat{\pi}_t^1 - \pi_t^1)M. \qquad (198)
\end{aligned}$$

The first and second inequalities in (198) follow from the induction hypothesis for Property 7. The equality in (198) follow from (197). The last inequality in (198) holds by part (iii) of Property 4 and the fact that $\hat{\pi}_t^1 \geq_{st} \pi_t^1$.
The proof of Property 7 is now complete when $O_t(N) \neq 1$.

(ii) $O_t(N) = 1$ (i.e. $n = N$).
We first consider the lower-bound. We want to show that

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \geq 0. \qquad (199)$$

Using (172) and the linearity of L_t (Lemma 1) we get

$$\begin{aligned}
& L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
= & L_t(O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) \\
& + L_t(O_t, \tilde{\hat{\pi}}_t^1, \hat{\pi}_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \tilde{\hat{\pi}}_t^1, \hat{\pi}_t^1, \pi_t^{2:N}) \\
& + \left[\sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))\right] [L_t(O_t, e_L, e_{L-1}, \pi_t^{2:N}) - L_t(W_{Nm}O_t, e_L, e_{L-1}, \pi_t^{2:N})]. \qquad (200)
\end{aligned}$$

We consider each of the terms
(a) $L_t(O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N})$.
(b) $L_t(O_t, \tilde{\hat{\pi}}_t^1, \hat{\pi}_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \tilde{\hat{\pi}}_t^1, \hat{\pi}_t^1, \pi_t^{2:N})$.
(c) $\left[\sum_{i=L}^{K} (\hat{\pi}_t^1(i) - \pi_t^1(i))\right] [L_t(O_t, e_L, e_{L-1}, \pi_t^{2:N}) - L_t(W_{Nm}O_t, e_L, e_{L-1}, \pi_t^{2:N})]$.
that appear in the right hand side of (200) separately. We do this because the channel orderings are different in each of the tree terms, different methods are needed to establish the bounds.

(a) Consider the first term.
Let $O'_t = S(W_{Nm}O_t) = W_{1m+1}(SO_t)$, then $O'_t(m+1) = 1$ and $W_{m+1,1}O'_t = SO_t$. Therefore,

$$\begin{aligned}
& L_t(O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) \\
= & (\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)R + \beta L_{t+1}(SO_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) - L_t(W_{Nm}O_t, \underline{\hat{\pi}}_t^1, \underline{\pi}_t^1, \pi_t^{2:N}) \\
\geq & (\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)R + \beta L_{t+1}(SO_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) - \beta L_{t+1}(S(W_{Nm}O_t), \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) \\
= & (\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)R - \beta(L_{t+1}(O'_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P) - L_{t+1}(W_{m+1,1}O'_t, \underline{\hat{\pi}}_t^1 P, \underline{\pi}_t^1 P, \pi_t^{2:N} P)) \\
\geq & (\underline{\hat{\pi}}_t^1 - \underline{\pi}_t^1)R - \beta(\underline{\hat{\pi}}_t^1 P - \underline{\pi}_t^1 P)M \\
\geq & 0. \qquad (201)
\end{aligned}$$

The first equality in (201) follows from (180), the fact that $O_t(N) = 1$ and that fact that $\underline{\hat{\pi}}_t^1(i) = \underline{\pi}_t^1(i) = 0$ for $i \geq L$. The first inequality in (201) follows from (178) and that fact that $W_{Nm}O_t(N) \neq 1$. The second inequality in (201) follows from the induction hypothesis for the upper bound of Property 7 at $t+1$ and the fact that $\underline{\hat{\pi}}_t^1 P \geq_{st} \underline{\pi}_t^1 P$ (since $\underline{\hat{\pi}}_t^1 \geq_{st} \underline{\pi}_t^1$ and Property 1). The last inequality in (201) holds by part (iv) of Property 4, the fact that $\underline{\hat{\pi}}_t^1 \geq_{st} \underline{\pi}_t^1$ and that fact that $\underline{\hat{\pi}}_t^1(i) = \underline{\pi}_t^1(i) = 0$ for $i \geq L$.

(b) Consider the second term.

Similar to case (a), we have

$$\begin{aligned}
&L_t(O_t,\bar{\tilde{\pi}}_t^1,\bar{\pi}_t^1,\pi_t^{2:N}) - L_t(W_{Nm}O_t,\bar{\tilde{\pi}}_t^1,\bar{\pi}_t^1,\pi_t^{2:N})\\
=&(\bar{\tilde{\pi}}_t^1-\bar{\pi}_t^1)R+\beta L_{t+1}(O_t,\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P) - L_t(W_{Nm}O_t,\bar{\tilde{\pi}}_t^1,\bar{\pi}_t^1,\pi_t^{2:N})\\
\geq&(\bar{\tilde{\pi}}_t^1-\bar{\pi}_t^1)R+\beta L_{t+1}(SO_t,\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P) - L_t(W_{Nm}O_t,\bar{\tilde{\pi}}_t^1,\bar{\pi}_t^1,\pi_t^{2:N})\\
\geq&(\bar{\tilde{\pi}}_t^1-\bar{\pi}_t^1)R+\beta L_{t+1}(SO_t,\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P) - \beta L_{t+1}(S(W_{Nm}O_t),\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P)\\
=&(\underline{\hat{\pi}}_t^1-\underline{\pi}_t^1)R-\beta(L_{t+1}(O_t',\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P) - L_{t+1}(W_{m+1,1}O_t',\bar{\tilde{\pi}}_t^1P,\bar{\pi}_t^1P,\pi_t^{2:N}P))\\
\geq&(\bar{\tilde{\pi}}_t^1-\bar{\pi}_t^1)R-\beta(\bar{\tilde{\pi}}_t^1P-\bar{\pi}_t^1P)M\\
\geq&0.
\end{aligned}\tag{202}$$

The first equality in (202) follows from (180), the fact that $O_t(N)=1$ and that fact that $\bar{\tilde{\pi}}_t^1(i)=\bar{\pi}_t^1(i)=0$ for $i<L$. The first inequality in (202) follows from the induction hypothesis for the lower bound of Property 6 at $t+1$, the fact that $\bar{\tilde{\pi}}_t^1P\geq_{st}\bar{\pi}_t^1P$ (since $\bar{\tilde{\pi}}_t^1\geq_{st}\bar{\pi}_t^1$ and Property 1) and the fact that $SO_t=S^{-(N-1)}O_t$ and $O_t(N)=1$. The second inequality in (202) follows from (178) and that fact that $W_{Nm}O_t(N)\neq 1$. The third inequality in (202) follows from the induction hypothesis for the upper bound of Property 7 at $t+1$ and the fact that $\bar{\tilde{\pi}}_t^1P\geq_{st}\bar{\pi}_t^1P$. The last inequality in (202) holds by part (iv) of Property 4, the fact that $\bar{\tilde{\pi}}_t^1P\geq_{st}\bar{\pi}_t^1P$ and that fact that $\bar{\tilde{\pi}}_t^1(i)=\bar{\pi}_t^1(i)=0$ for $i<L$.

(c) Consider the third part.
Assume $O_t(m)=2$ without any loss of generality. Then $W_{Nm}O_t(N)=2$. Therefore,

$$\begin{aligned}
&L_t(O_t,e_L,e_{L-1},\pi_t^{2:N}) - L_t(W_{Nm}O_t,e_L,e_{L-1},\pi_t^{2:N})\\
=&R_L-R_{L-1}+\beta[V_{t+1}(O_t,P_L,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_{L-1},\pi_t^{2:N}P)]\\
&-\beta\sum_{i<L}\pi_t^2(i)L_{t+1}(SW_{Nm}O_t,P_L,P_{L-1},P_i,\pi_t^{3:N}P)-\beta\sum_{i\geq L}\pi_t^2(i)L_{t+1}(W_{Nm}O_t,P_L,P_{L-1},P_i,\pi_t^{3:N}P)\\
=&R_L-R_{L-1}\\
&+\beta\sum_{i<L}\pi^2(i)[V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)-V_{t+1}(SO_t,P_{L-1},P_i,\pi_t^{3:N}P)\\
&\quad-L_{t+1}(SW_{Nm}O_t,P_L,P_{L-1},P_i,\pi_t^{3:N}P)]\\
&+\beta\sum_{i\geq L}\pi^2(i)[V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)-V_{t+1}(SO_t,P_{L-1},P_i,\pi_t^{3:N}P)\\
&\quad-L_{t+1}(W_{Nm}O_t,P_L,P_{L-1},P_i,\pi_t^{3:N}P)].
\end{aligned}\tag{203}$$

The first equality in (203) follows from (175) and (180). The last equality in (203) holds because of Lemma 1.
Let $O_t':=S(W_{Nm}O_t)=W_{1m+1}(SO_t)$; then $O_t'(m+1)=1$ and $W_{m+1,1}O_t'=SO_t$.
For each term in the first sum in (203), we have $P_{L-1}\geq_{st}P_i$ ($i<L$ in the first sum in (203)). Therefore,

$$\begin{aligned}
&V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)-V_{t+1}(SO_t,P_{L-1},P_i,\pi_t^{3:N}P)-L_{t+1}(SW_{Nm}O_t,P_L,P_{L-1},P_i,\pi_t^{3:N}P)\\
=&V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)-V_{t+1}(W_{m+1,1}O_t',P_{L-1},P_i,\pi_t^{3:N}P)\\
&-V_{t+1}(O_t',P_L,P_i,\pi_t^{3:N}P)+V_{t+1}(O_t',P_{L-1},P_i,\pi_t^{3:N}P)\\
\geq&V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)-V_{t+1}(O_t',P_L,P_i,\pi_t^{3:N}P).
\end{aligned}\tag{204}$$

The equality in (204) follows from the definition of L_{t+1}. The inequality in (204) follows from the induction hypothesis for the lower bound of Property 8 at $t+1$ and the fact that $P_{L-1}\geq_{st}P_i$.
Furthermore, since $P_L\geq_{st}\pi_t^{O_t(l)}P$ for all $l=1,2,...,N$ by Property 2, repeatedly applying Property 8 at $t+1$ we obtain

$$\begin{aligned}
&V_{t+1}(O_t,P_L,P_i,\pi_t^{3:N}P)\\
\geq&V_{t+1}(W_{(m+2)(m+1)}...W_{N(N-1)}O_t,P_L,P_i,\pi_t^{3:N}P)\\
=&V_{t+1}(A_{N(m+1)}O_t,P_L,P_i,\pi_t^{3:N}P),
\end{aligned}\tag{205}$$

where A_{Nm+1} is the operator defined by (70). The equality in (205) is true because $W_{(m+2)(m+1)}...W_{N(N-1)}O_t=A_{N(m+1)}O_t$.
Note that

$$A_{m1}(A_{N(m+1)}O_t)=S(W_{Nm}O_t)=O_t',A_{N(m+1)}O_t(m)=O_t(m)=2.\tag{206}$$

Consequently,

$$\begin{aligned}
&V_{t+1}(O_t, P_L, P_i, \pi_t^{3:N}P) - V_{t+1}(SO_t, P_{L-1}, P_i, \pi_t^{3:N}P) - L_{t+1}(SW_{Nm}O_t, P_L, P_{L-1}, P_i, \pi_t^{3:N}P)\\
\geq& V_{t+1}(A_{N(m+1)}O_t, P_L, P_i, \pi_t^{3:N}P) - V_{t+1}(O'_t, P_L, P_i, \pi_t^{3:N}P)\\
=& V_{t+1}(A_{N(m+)1}O_t, P_L, P_i, \pi_t^{3:N}P) - V_{t+1}(A_{m1}(A_{N(m+1)}O_t), P_L, P_i, \pi_t^{3:N}P)\\
\geq& -(h - P_iP^{N-m}R). \qquad (207)
\end{aligned}$$

The first inequality in (207) follows from (204) and (205). The equality in (207) follows from (206). The second inequality in (207) follows from the induction hypothesis for Property 9 at $t+1$ and the fact that $P_i \in \pi P$, therefore $P_i \leq_{st} P_{L-1} \leq_{st} P_iP$ for $i < L$ by Property 2.
For each term in the second sum in (203), we have

$$\begin{aligned}
&V_{t+1}(O_t, P_L, P_i, \pi_t^{3:N}P) - V_{t+1}(SO_t, P_{L-1}, P_i, \pi_t^{3:N}P) - L_{t+1}(W_{Nm}O_t, P_L, P_{L-1}, P_i, \pi_t^{3:N}P)\\
\geq& V_{t+1}(O_t, P_L, P_i, \pi_t^{3:N}P) - V_{t+1}(SO_t, P_{L-1}, P_i, \pi_t^{3:N}P) - L_{t+1}(O_t, P_L, P_{L-1}, P_i, \pi_t^{3:N}P)\\
=& V_{t+1}(O_t, P_{L-1}, P_i, \pi_t^{3:N}P) - V_{t+1}(SO_t, P_{L-1}, P_i, \pi_t^{3:N}P)\\
=& V_{t+1}(O_t, P_{L-1}, P_i, \pi_t^{3:N}P) - V_{t+1}(A_{N1}O_t, P_{L-1}, P_i, \pi_t^{3:N}P)\\
\geq& -(h - P_{L-1}R). \qquad (208)
\end{aligned}$$

The first inequality in (208) follows from the induction hypothesis for the lower bound of Property 7 at $t+1$ and the fact that $P_L \geq_{st} P_{L-1}$. The fist equality in (208) follows from the definition of L_{t+1} (eq. (64)). The second equality in (208) follows from the fact that $SO_t = A_{N1}O_t$. The last inequality in (208) follows from the induction hypothesis for Property 9 at $t+1$ and the fact that $P_{L-1} \leq_{st} P_{L-1}P$.
Using the lower bounds provided by (207) and (208) for terms in (203), we obtain

$$\begin{aligned}
&L_t(O_t, e_L, e_{L-1}, \pi_t^{2:N}) - L_t(W_{Nm}O_t, e_L, e_{L-1}, \pi_t^{2:N})\\
\geq& R_L - R_{L-1} - \beta\sum_{i<L}\pi_t^2(i)(h - P_iP^{N-m}R) - \beta\sum_{i\geq L}\pi_t^2(i)(h - P_{L-1}R)\\
\geq& R_L - R_{L-1} - \beta(h - P_{L-1}R) \geq 0. \qquad (209)
\end{aligned}$$

The first inequality in (209) follows from (207) and (208). The second inequality in (209) follows from part (ii) of Property 4 and the fact that $P_i \in \pi P$, therefore $P_iP^{N-m} \in \pi P^2$, thus $P_iP^{N-m} \geq_{st} P_{L-1}$ by Property 2. The last inequality in (209) holds by condition (A4).

Using the lower bounds given by (201), (202) and (209) for the three terms (a), (b) and (c), respectively, in 200, we obtain

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^{2:N}) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^1, \pi_t^{2:N}) \geq 0. \qquad (210)$$

This completes the proof for the lower bound of Property 7 when $O_t(N) = 1$ (case (ii)).

We now proceed to establish the upper bound of Property 7 when $O_t(N) = 1$ (case (ii)). We want to show that

$$L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \leq (\hat{\pi}_t^1 - \pi_t^1)M. \qquad (211)$$

Assume $O_t(m) = 2$ without any loss of generality; then $W_{Nm}O_t(N) = 2$. Therefore,

$$\begin{aligned}
&L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
=&L_t(O_t, \hat{\pi}_t^1, \pi_t^{1:N}) - \beta L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^{1:N} P) \\
&+ \beta L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^{1:N} P) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
\leq&(\hat{\pi}_t^1 - \pi_t^1)U + \beta L_{t+1}(SO_t, \hat{\pi}_t^1 P, \pi_t^{1:N} P) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
\leq&(\hat{\pi}_t^1 - \pi_t^1)U + \beta L_{t+1}(S(W_{Nm}O_t), \hat{\pi}_t^1 P, \pi_t^{1:N} P) - L_t(W_{Nm}O_t, \hat{\pi}_t^1, \pi_t^{1:N}) \\
=&(\hat{\pi}_t^1 - \pi_t^1)U + \beta L_{t+1}(S(W_{Nm}O_t), \hat{\pi}_t^1 P, \pi_t^{1:N} P) - \beta \sum_{i<L} \pi_t^2(i) L_t(S(W_{Nm}O_t), \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \\
&- \beta \sum_{i \geq L} \pi_t^2(i) L_{t+1}(W_{Nm}O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) \\
=&(\hat{\pi}_t^1 - \pi_t^1)U + \beta \sum_{i \geq L} \pi_t^2(i)(L_{t+1}(S(W_{Nm}O_t), \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P) - L_{t+1}(W_{Nm}O_t, \hat{\pi}_t^1 P, \pi_t^1 P, P_i, \pi_t^{3:N} P)) \\
\leq&(\hat{\pi}_t^1 - \pi_t^1)U + \beta \sum_{i \geq L} \pi_t^2(i)(\hat{\pi}_t^1 P - \pi_t^1 P)U \\
\leq&(\hat{\pi}_t^1 - \pi_t^1)U + \beta \sum_{i \geq L} p_{Ki}(\hat{\pi}_t^1 P - \pi_t^1 P)U \\
=&(\hat{\pi}_t^1 - \pi_t^1)M. \qquad (212)
\end{aligned}$$

The first inequality in (212) follows from (184). The second inequality in (212) follows from the induction hypothesis for the lower bound of Property 7 at $t+1$, the fact that $SO_t = W_{(m+1),1}(S(W_{Nm}O_t))$ and the fact that $\hat{\pi}_t^1 \geq_{st} \pi_t^1$. The second equality in (212) follows from (175). The third equality in (212) follows from the linearity of the function L_t (Lemma 1). The third inequality in (212) follows from the induction hypothesis for the upper bound of Property 6 and the fact that $\hat{\pi}_t^1 P \geq_{st} \pi_t^1 P$ (since $\hat{\pi}_t^1 \geq_{st} \pi_t^1$ and Property 1). The last inequality in (212) is true because $\pi_t^2 \leq_{st} P_K$. The last equality in (212) follows from the definition of M.

The proof of the upper bound of Property 7 at t is now complete. The proof of the induction step for Property 7 at t is also complete.

Induction step for Property 9:

(i) When $O_t(N) \neq 1$ (i.e. $n \neq N$), assume $O_t(N) = N$ without loss of generality. Then,

$$\begin{aligned}
&V_t(A_{nm}O_t, \pi_t^{1:N}) - V_t(O_t, \pi_t^{1:N}) \\
=&\sum_{i<L} \pi_t^N(i)[V_{t+1}(S(A_{nm}O_t), \pi_t^{1:N-1} P, P_i) - V_{t+1}(SO_t, \pi_t^{1:N-1} P, P_i)] \\
&+ \sum_{i \geq L} \pi_t^N(i)[V_{t+1}(A_{nm}O_t, \pi_t^{1:N-1} P, P_i) - V_{t+1}(O_t, \pi_t^{1:N-1} P, P_i)] \\
=&\sum_{i<L} \pi_t^N(i)[V_{t+1}(A_{(n+1),(m+1)}(SO_t), \pi_t^{1:N-1} P, P_i) - V_{t+1}(SO_t, \pi_t^{1:N-1} P, P_i)] \\
&+ \sum_{i \geq L} \pi_t^N(i)[V_{t+1}(A_{nm}O_t, \pi_t^{1:N-1} P, P_i) - V_{t+1}(O_t, \pi_t^{1:N-1} P, P_i)] \\
\leq&\sum_{i<L} \pi_t^N(i)(h - \pi_t^1 P(P^{N-n-1}R)) + \sum_{i \geq L} \pi_t^N(i)(h - \pi_t^1 P(P^{N-n}R)) \\
\leq&h - \pi_t^1 P^{N-n} R. \qquad (213)
\end{aligned}$$

The first equality in (213) follows from the recursive equation for V_t (eq. (62)). The second equality in (213) is true because $S(A_{nm}O_t) = A_{(n+1),(m+1)}(SO_t)$. The first inequality in (213) follows from the induction hypothesis for Property 9 and the fact that $\pi_t^1 P \leq_{st} \pi_t^1 P^2$ (Property 1). The last inequality in (213) follows from part (ii) of Property 4 and the fact that $\pi_t^1 P^{N-n} \leq_{st} \pi_t^1 P^{N-n+1}$ (Property 1).

(i) When $O_t(N) = 1$ (i.e. $n = N$), assume $O_t(N-1) = N$ without loss of generality. Then $A_{Nm}O_t(N) = O_t(N-1) = N$.

Therefore,

$$
\begin{aligned}
&V_t(A_{Nm}O_t,\pi_t^{1:N})-V_t(O_t,\pi_t^{1:N})\\
=&(\pi_t^N-\pi_t^1)R+\beta\sum_{i<L}\pi_t^N(i)V_{t+1}(S(A_{Nm}O_t),\pi_t^{1:N-1}P,P_i)+\beta\sum_{i\geq L}\pi_t^N(i)V_{t+1}(A_{Nm}O_t,\pi_t^{1:N-1}P,P_i)\\
&-\beta\sum_{i<L}\pi_t^1(i)V_{t+1}(SO_t,P_i,\pi_t^{2:N}P)-\beta\sum_{i\geq L}\pi_t^1(i)V_{t+1}(O_t,P_i,\pi_t^{2:N}P)\\
=&(\pi_t^N-\pi_t^1)R+\beta\sum_{i<L}\pi_t^N(i)[V_{t+1}(S(A_{Nm}O_t),\pi_t^{1:N-1}P,P_i)-V_{t+1}(A_{Nm}O_t,\pi_t^{1:N-1}P,P_i)]+\beta V_{t+1}(A_{Nm}O_t,\pi_t^{1:N}P)\\
&-\beta\sum_{i<L}\pi_t^1(i)V_{t+1}(SO_t,P_i,\pi_t^{2:N}P)-\beta\sum_{i\geq L}\pi_t^1(i)V_{t+1}(O_t,P_i,\pi_t^{2:N}P)\\
=&(\pi_t^N-\pi_t^1)R+\beta\sum_{i<L}\pi_t^N(i)[V_{t+1}(S(A_{Nm}O_t),\pi_t^{1:N-1}P,P_i)-V_{t+1}(A_{Nm}O_t,\pi_t^{1:N-1}P,P_i)]\\
&+\beta\sum_{i<L}\pi_t^1(i)[V_{t+1}(A_{Nm}O_t,P_i,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_i,\pi_t^{2:N}P)]\\
&+\beta\sum_{i\geq L}\pi_t^1(i)[V_{t+1}(A_{Nm}O_t,P_i,\pi_t^{2:N}P)-V_{t+1}(O_t,P_i,\pi_t^{2:N}P)]\\
\leq&(\pi_t^N-\pi_t^1)R+\beta\sum_{i<L}\pi_t^N(i)[V_{t+1}(S(A_{Nm}O_t),\pi_t^{1:N-1}P,P_i)-V_{t+1}(A_{Nm}O_t,\pi_t^{1:N-1}P,P_i)]\\
\leq&(\pi_t^N-\pi_t^1)R+\beta\sum_{i<L}\pi_t^N(i)(h-P_iR).
\end{aligned}
\tag{214}
$$

The three equalities in (214) follow from the recursive equation and the linearity of the function V_{t+1} ((62) and Lemma 1). The last inequality in (214) follows from the induction hypothesis for Property 9, the fact that $S(A_{Nm}O_t)=A_{N1}(A_{Nm}O_t)$, and $A_{Nm}O_t(N)=O_t(N-1)=N$ and the fact that $P_i\leq_{st}P_iP$ for $i<L$ by Property 2.
The first inequality in (214) is true because of the following:
For $i<L$, $P_i\leq_{st}P_{L-1}\leq_{st}\pi_t^lP$ for all l by Property 2. Then,

$$
\begin{aligned}
&V_{t+1}(A_{Nm}O_t,P_i,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_i,\pi_t^{2:N}P)\\
=&V_{t+1}(W_{m(m-1)}...W_{32}W_{21}SO_t,P_i,\pi_t^{2:N}P)-V_{t+1}(SO_t,P_i,\pi_t^{2:N}P)\leq 0.
\end{aligned}
\tag{215}
$$

The equality in (215) is true because $A_{Nm}O_t=W_{m(m-1)}...W_{32}W_{21}SO_t$. The inequality in (215) follows by repeatedly using Property 8 at $t+1$ and the fact that for $i<L$, $P_i\leq_{st}\pi_t^lP$ for all l.
For $i\geq L$, $P_i\geq_{st}P_L\geq_{st}\pi_t^lP$ for all l by Property 2. Then,

$$
\begin{aligned}
&V_{t+1}(A_{Nm}O_t,P_i,\pi_t^{2:N}P)-V_{t+1}(O_t,P_i,\pi_t^{2:N}P)\\
=&V_{t+1}(W_{m(m+1)}...W_{(N-1)(N-2)}W_{N(N-1)}O_t,P_i,\pi_t^{2:N}P)-V_{t+1}(O_t,P_i,\pi_t^{2:N}P)\leq 0.
\end{aligned}
\tag{216}
$$

The equality in (216) is true because $A_{Nm}O_t=W_{m(m+1)}...W_{(N-1)(N-2)}W_{N(N-1)}O_t$. The inequality in (216) follows by repeatedly using Property 8 at $t+1$ and the fact that for $i\geq L$, $P_i\geq_{st}\pi_t^lP$ for all l.

Let v be the vector such that

$$
v_i=\begin{cases}R_i+\beta(h-P_iR), & \text{for } i<L\\ R_i, & \text{for } i\geq L\end{cases}.
\tag{217}
$$

For $i\geq L$ we have

$$
v_{i+1}-v_i=R_{i+1}-R_i\geq 0.
\tag{218}
$$

For $i=L-1$,

$$
v_L-v_{L-1}=R_L-R_{L-1}-\beta(h-P_{L-1}R)\geq 0;
\tag{219}
$$

the inequality if (219) holds because of condition (A4).
For $i<L-1$, we have

$$
\begin{aligned}
v_{i+1}-v_i=&R_{i+1}-R_i-\beta(P_{i+1}-P_i)R\\
\geq&R_{i+1}-R_i-\beta(P_{i+1}-P_i)M\\
\geq&0.
\end{aligned}
\tag{220}
$$

The first inequality in (220) follows from part (ii) of Property 4; the last inequality in (220) follows from condition (A4). Consequently, v_i increases with i. Then, from part (i) of Property 4 and the fact that $\pi_t^N \leq_{st} P_K$ we obtain

$$\begin{aligned}
&V_t(A_{Nm}O_t, \pi_t^{1:N}) - V_t(O_t, \pi_t^{1:N}) \\
\leq&(\pi_t^N - \pi_t^1)R + \beta \sum_{i<L} \pi_t^N(i)(h - P_i R) \\
=&\pi_t^N v - \pi_t^1 R \\
\leq&P_K v - \pi_t^1 R \\
=&h - \pi_t^1 R
\end{aligned} \tag{221}$$

The first inequality in (221) follows from (214). The second inequality in (221) follows from part (i) of Property 4, the fact that v_i increases with i, and the fact that $\pi_t^N \leq_{st} P_K$. The last equality in (221) follows from the observation that

$$P_K v = P_K R + \beta \sum_{i<L} p_{Ki}(h - P_i R) = h. \tag{222}$$

This completes the proof of the induction step for Property 9 at t, and the proof of the entire induction step.

ACKNOWLEDGMENT

This work was supported in part by National Science Foundation (NSF) Grant CCF-1111061 and NASA grant NNX12A0546.

REFERENCES

[1] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[2] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of Applied Probability*, pp. 287–298, 1988.

[3] J. Gittins, R. Weber, and K. Glazebrook, *Multi-Armed Bandit Allocation Indices*. WileyBlackwell, 2011.

[4] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589–600, 2007.

[5] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5431–5440, 2008.

[6] T. Javidi, B. Krishnamachari, Q. Zhao, and M. Liu, "Optimality of myopic sensing in multi-channel opportunistic access," in *2008. ICC'08. IEEE International Conference on Communications*. IEEE, 2008, pp. 2107–2112.

[7] S. Ahmad, M. Liu, T. Javidi, Q. Zhao, and B. Krishnamachari, "Optimality of myopic sensing in multichannel opportunistic access," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4040–4050, 2009.

[8] S. Ahmad and M. Liu, "Multi-channel opportunistic access: A case of restless bandits with multiple plays," in *2009. Allerton 2009. 47th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2009, pp. 1361–1368.

[9] J. Gittins, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 148–177, 1979.

[10] C. Papadimitriou and J. Tsitsiklis, "The complexity of optimal queueing network control," in *Proceedings of the Ninth Annual Structure in Complexity Theory Conference, 1994.* IEEE, 1994, pp. 318–322.

[11] R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, pp. 637–648, 1990.

[12] J. Niño-Mora, "Dynamic priority allocation via restless bandit marginal productivity indices," *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, vol. 15, no. 2, pp. 161–198, 2007.

[13] K. Liu and Q. Zhao, "Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5547–5567, 2010.

[14] C. Lott and D. Teneketzis, "On the optimality of an index rule in multi-channel allocation for single-hop mobile networks with multiple service classes," *Probab. Eng. Inf. Sci*, vol. 14, p. 259, 2000.

[15] N. Ehsan and M. Liu, "Server allocation with delayed state observation: Sufficient conditions for the optimality of an index policy," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1693–1705, 2009.

[16] S. Guha, K. Munagala, and P. Shi, "Approximation algorithms for restless bandit problems," *Journal of the ACM (JACM)*, vol. 58, no. 1, p. 3, 2010.

[17] Y. Ouyang and D. Teneketzis, "On the optimality of a myopic policy in multi-state channel probing," *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012.

[18] A. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Verlag, 2010.

[19] P. Kumar and P. Varaiya, *Stochastic Systems :Estimation Identification and Adaptive Control*. Prentice-Hall, Inc., 1986.

[20] P. Varaiya, J. Walrand, and C. Buyukkoc, "Extensions of the multiarmed bandit problem: the discounted case," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 426–439, 1985.