Mathematical Methods of Operations Research (2023) 97:117–133
https://doi.org/10.1007/s00186-022-00806-9

ORIGINAL ARTICLE

An axiomatic approach to Markov decision processes

Adam Jonsson[1]

Received: 21 April 2021 / Revised: 2 September 2022 / Accepted: 21 November 2022 /
Published online: 2 December 2022

Abstract

This paper presents an axiomatic approach to finite Markov decision processes where the discount rate is zero. One of the principal difficulties in the no discounting case is that, even if attention is restricted to stationary policies, a strong overtaking optimal policy need not exists. We provide preference foundations for two criteria that do admit optimal policies: 0-discount optimality and average overtaking optimality. As a corollary of our results, we obtain conditions on a decision maker's preferences which ensure that an optimal policy exists. These results have implications for disciplines where dynamic programming problems arise, including automatic control, dynamic games, and economic development.

Keywords Dynamic programming · Markov decision processes · Axioms · Preferences

Mathematics Subject Classification 60J20 · 62C99

1 Introduction

This paper presents an axiomatic approach to finite Markov decision processes (MPDs) where the discount rate is zero. MDPs comprise a broad class of stochastic dynamic decision problems and they have been studied extensively over the past several decades. To keep the discussion as elementary as possible, we will work within the framework of Blackwell's (1962) classic paper. For extensions of this framework and discussion of its many uses, the reader is referred to Arapostathis et al. (1993), Hernández-Lerma and Vega-Amaya (1998), Rosenberg et al. (2002) and the books by Feinberg and Shwartz (2002), Piunovskiy (2013) and Puterman (1994).

✉ Adam Jonsson
adam.jonsson@ltu.se

[1] Department of Engineering Sciences and Mathematics, Luleå University of Technology, Luleå, Sweden

In its simplest form, a MDP has the following ingredients: A state space $\mathscr{S}$, an action space $\mathscr{A}$, a transition probability function $p_a(s'|s)$ on $\mathscr{S}$ for each $a \in \mathscr{A}$, and a real-valued function $r(s,a)$ on $\mathscr{S} \times \mathscr{A}$. Here $\mathscr{S}$ represents possible states of a system (a manufacturing chain, a biological system, a natural resource, etc.) and $\mathscr{A}$ represents choices available to an agent (the decision maker). Unless stated otherwise, $\mathscr{S}$ and $\mathscr{A}$ are finite sets. At discrete times $t = 1, 2, 3, \ldots$, the agent observes the state and selects an element from $\mathscr{A}$. If the system is in $s \in \mathscr{S}$ and $a \in \mathscr{A}$ is chosen, then a reward of $r(s,a)$ is received and the system moves to s' with probability $p_a(s'|s)$. Rewards are discounted so that a reward of one unit at time t has present value β^t, where $0 < \beta \leq 1$. The problem is to choose a policy (i.e., a rule for selecting actions at all future times) that maximizes the expected net present value of all future rewards.

This problem is particularly difficult when $\beta = 1$. To begin with, it is not clear what it means to maximize net present value in this case. The difficulty is that the total value of a policy is typically infinite if $\beta = 1$. There is a natural sense in which a policy is maximal if it generates a sequence of cumulative expected rewards that eventually dominates that of any other policy. This leads to the intuitive notion of overtaking optimality (formally defined in Sect. 3). It is well known, however, that an overtaking optimal policy need not exist. A less selective criterion is based on the expected long-run average reward of a policy. But this criterion does not differentiate between streams of expected rewards which might have very different appeal to the decision maker.

Blackwell (1962) introduced the *1-optimality* criterion, which evaluates streams of expected rewards on the basis of their Abel means. He also established the existence of 1-optimal policies that are *stationary*, (i.e., for which the action chosen at time t depends only on the state of the system at time t).[1] Subsequently, Veinott (1966) introduced what is often referred to as the *average overtaking* criterion, where Abel means are substituted for Cesàro means. The Blackwell–Veinott criteria are able to select between policies that the average reward criterion does not distinguish. However, the literature has not adressed the following questions:

Q1. Are the Blackwell–Veinott criteria the *only* selective criteria which admit optimal policies in the no discounting case?
Q2. How can these criteria be described axiomatically?
Q3. Under which assumptions on a decision maker's preferences do optimal policies exist?

Our main results are summarized in Theorems 1, 2 and 3. Theorem 1 shows that, subject to certain constraints, Q1 has an affirmative answer. Theorems 2 and 3 provide two sets of axioms that characterize the average overtaking and 1-optimality criterion on the reward streams generated by stationary policies. The second of these two results complements a theorem of Jonsson and Voorneveld (2018) and uses the compensation principle as a key axiom. Finally, we obtain a partial answer to Q3 as a corollary of these results.

[1] More precisely, Blackwell (1962) establishes existence of optimal policies using the criterion now known as *Blackwell optimality*, which is slightly stronger than 1-optimality. He refers to 1-optimality as *near optimality*; other authors use the terms *0-discount optimality* and *bias optimality* (Puterman 1994; Piunovskiy 2013).

2 Preliminaries

Our finite MDP has state space $\mathscr{S}$ and action space $\mathscr{A}$. At times $t = 1, 2, 3, \ldots$, the agent observes the state of and chooses an element a from $\mathscr{A}$. We assume that this choice depends on the history of the system only through its present state. Thus, the action chosen at time t is an element of F, the set of all functions from $\mathscr{S}$ to $\mathscr{A}$. Each $f \in F$ has a corresponding transition matrix, $\mathbf{Q}(f)$, and reward vector, $\mathbf{r}(f)$. With the notation from the introduction, if the system is in $s \in \mathscr{S}$ and f is used, then a reward of $\mathbf{r}(f)_s = r(s, f(s))$ is received and the system moves to s' with probability $\mathbf{Q}(f)_{s,s'} = p_{f(s)}(s'|s)$. Rewards may be interpreted, for example, as payouts of a single good received by an infinitely lived consumer, or as the utilities of future generations.

A *policy* is a sequence $(f_1, f_2, f_3, \ldots)$ in F. Using policy $\pi = (f_1, f_2, f_3, \ldots)$ means that, for each $t = 1, 2, 3, \ldots$, $f_t(s)$ is selected from $\mathscr{A}$ if the system is in state s. A policy is *stationary* if using it implies that the action chosen at time t depends on the state of the system at time t, but not on t itself. Formally, a stationary policy can be written $(f, f, f, \ldots)$ for some $f \in F$.[2] We denote the set of all policies by Π and the set of all stationary policies by Π_F.

Given an initial state $s \in \mathscr{S}$, the sequence of expected rewards that $\pi \in \Pi$ generates is denoted $u(s, \pi)$. If $\pi = (f_1, f_2, f_3, \ldots)$ and $u = (u_1, u_2, u_3, \ldots) = u(s, \pi)$, then

$$\begin{aligned} u_1 &= [\mathbf{r}(f_1)]_s, \\ u_t &= [\mathbf{Q}(f_1) \cdot \ldots \cdot \mathbf{Q}(f_{t-1}) \cdot \mathbf{r}(f_t)]_s, \ t \geq 2. \end{aligned} \tag{1}$$

Let $\mathscr{U}_F$ be the set of sequences generated by stationary policies. That is, $u \in \mathscr{U}_F$ if and only if $u = u(s, \pi)$ for some $s \in \mathscr{S}$ and $\pi \in \Pi_F$.

The agent needs to compare $u(s, \pi)$ and $u(s, \pi')$ for different $s \in \mathscr{S}$ and $\pi, \pi' \in \Pi$. For convenience, we consider (incomplete) preferences on the set of all bounded sequences, which is denoted by $\mathscr{U}$. We reserve the notation $\succsim$ for a preorder on $\mathscr{U}$ (i.e., a reflexive and transitive binary relation), where $u \succsim v$ means that u is at least as good as v. We say that $\succsim$ *compares* u and v if either $u \succsim v$ or $v \succsim u$, and we write $\neg u \succsim v$ to indicate that u is not at least as good as v. As usual, $u \succ v$ denotes strict preference ($u \succsim v$, but $\neg v \succsim u$) and $u \sim v$ denotes indifference ($u \succsim v$ and $v \succsim u$).

In this framework, preferences are thus defined over sequences of expected rewards. That is, it is assumed that preferences over random rewards can be reduced to preferences over expected rewards. The framework is therefore unable to elucidate risk-averse preferences. For risk measures and risk-sensitive control of Markov processes, see Bäuerle and Rieder (2014), Ruszczyński (2010) and the references cited there.

[2] More general definitions of the concepts of a policy and stationary policy allow for randomized actions (see, e.g., Puterman 1994, p. 22). Our results for non-randomized (or pure) stationary policies generalize trivially to randomized stationary policies.

3 A motivating example

For background, we begin by reviewing how different ways of comparing reward streams may fail or succeed to yield optimal policies. The comparisons often involve sums over a finite horizon. For $u \in \mathscr{U}$ and $T \in \mathbb{N}$, we let

$$\sigma_T(u) = \sum_{t=1}^{T} u_t, \quad \sigma(u) = (\sigma_1(u), \sigma_2(u), \sigma_3(u), \ldots). \tag{2}$$

A policy $\pi^* \in \Pi$ is *overtaking optimal* if, for every $\pi \in \Pi$,

$$u(s, \pi^*) \succsim_{\mathrm{O}} u(s, \pi) \text{ for every } s \in \mathscr{S}, \tag{3}$$

where

$$u \succsim_{\mathrm{O}} v \iff \liminf_{T \to \infty} \sigma_T(u - v) \geq 0. \tag{4}$$

This criterion has the advantage of being plausible intuitively. It is also the strongest among the most commonly discussed criteria for undiscounted MDPs. Its drawback is that an optimal policy need not exist (Brown 1965; Gale 1967). The following is a variation of an example from Denardo and Miller (1968). We return to this example in Sect. 6.

Example 1 Figure 1 displays the transition graph of a deterministic MDP with $\mathscr{A} = \{a_1, a_2\}$ and $\mathscr{S} = \{s_1, s_2, s_3\}$. If the system starts in state s_1 and a_1 is chosen, then the system moves to s_2 and a reward of 2 is received; if a_2 is chosen, the system moves to s_3 and a reward of $c \in \mathbb{R}$ is received. Once the system reaches s_2 or s_3, it starts to alternate between these two states, and it does not matter how the agent acts. A reward of 0 is received when the system goes from s_2 to s_3, and a reward of 2 is received when it goes from s_3 to s_2.

Suppose that the system starts in s_1. Let u be the reward stream that is generated if a_1 is chosen, and let v be the stream that obtains if a_2 is chosen. Then

$$u = (2, 0, 2, 0, 2, \ldots) \quad \text{and} \quad v = (c, 2, 0, 2, 0, 2, \ldots).$$

We have $\sigma_T(u - v) = 2 - c$ if T is odd and $\sigma_T(u - v) = -c$ if T is even. Hence, if $0 < c < 2$, then $\neg u \succsim_{\mathrm{O}} v$ and $\neg v \succsim_{\mathrm{O}} u$. This means that there is no overtaking-optimal policy if $0 < c < 2$. □

Note that the MPD in Example 1 still does not admit an overtaking optimal policy if attention is restricted to stationary policies. We remark that it is not only in deterministic MDPs that this limitation of overtaking optimality makes itself known. There are, indeed, ergodic MDPs where no overtaking-optimal policy exists within the class of stationary policies (Nowak and Vega-Amaya 1999).

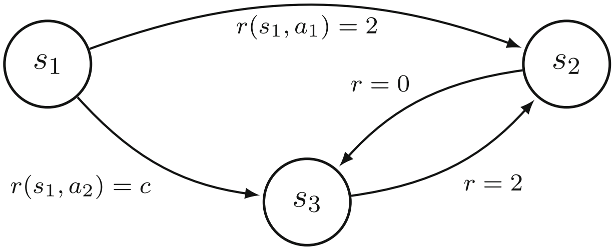

Fig. 1 A deterministic MDP where no overtaking-optimal policy exists

Let us also note that optimal policies often do exist if we adopt an alternative definition of overtaking optimality, according to which $\pi^* \in \Pi$ is optimal if there is no $\pi \in \Pi$ such that

$$u(s, \pi) \succ_O u(s, \pi^*) \text{ for every } s \in \mathscr{S}.$$

(In Example 1, all policies are optimal in this sense if $0 < c < 2$.) This weaker form of overtaking optimality has been used frequently in studies of optimal economic growth (Brock 1970b; Brock and Mirman 1973; Basu and Mitra 2007). It is closely related to the notion of *sporadic overtaking optimality* studied in the operations research literature (Stern 1984; Flesch et al. 2017). Here we have adopted the definition of overtaking optimality that this literature most frequently employs.

Generalizing the definition (4) to an arbitrary preorder $\succsim$, let us say that $\pi^* \in \Pi$ is $\succsim$*-optimal* or *optimal with respect to* $\succsim$ if, for every $\pi \in \Pi$,

$$u(s, \pi^*) \succsim u(s, \pi) \text{ for every } s \in \mathscr{S}. \tag{5}$$

The preorders associated with average reward optimality, average overtaking optimality and 1-optimality are defined as follows:

$$\textbf{(average reward)} \quad u \succsim_{\text{AR}} v \iff \liminf_{T\to\infty} \frac{1}{T}\sigma_T(u - v) \geq 0 \tag{6}$$

$$\textbf{(average overtaking)} \quad u \succsim_{\text{AO}} v \iff \liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sigma_t(u - v) \geq 0 \tag{7}$$

$$\textbf{(1-optimality)} \quad u \succsim_1 v \iff \liminf_{\delta\to 1^-} \sum_{t=1}^{\infty}\delta^t \cdot (u_t - v_t) \geq 0. \tag{8}$$

The average reward criterion is the most studied criterion for undiscounted MDPs. The standard criticism against this criterion concerns the fact that improvements in any finite number of time periods are ignored. In Example 1, for instance, it is average reward-optimal to choose a_1 in state s_1 even if the value of c is very large.

If u and v are the streams in Example 1, the Cesàro sum of $\sum_{t=1}^{\infty}(u_t - v_t)$ is $1 - c$. Hence, it is average overtaking-optimal to choose a_1 if and only if $c \leq 1$. It is well known that average overtaking optimality is equivalent to 1-optimality in finite MDPs (Lippman 1969). In general, any average overtaking-optimal policy is 1-optimal, but a 1-optimal policy need not be average overtaking optimal (see, e.g., Bishop et al. 2014).

To sum up, while the average reward criterion is unselective, the overtaking criterion is overselective. One way to formulate the first question (Q1) from the introduction is to ask if the average overtaking criterion is the least selective criterion that admits optimal policies. To state this question in a precise way, we will formulate a set of conditions which we can plausibly require of a selective criterion.

4 Axioms

This section provides five conditions (called axioms) on preorders that are known from the literature. The five conditions are satisfied by the preorders associated with the overtaking criterion, the average overtaking criterion and the 1-optimality criterion (see Jonsson and Voorneveld 2018, p. 28). They may be viewed as conditions that can be plausibly required of a selective criterion.

The first axiom, **A1**, is a standard monotonicity requirement. It asserts that preferences are positively sensitive to improvements in each time period. Preorders that meet this requirement avoid the standard criticism of the average reward criterion.

A1. For all $u, v \in \mathscr{U}$, if $u_t \geq v_t$ for all t and $u_t > v_t$ for some t, then $u \succ v$.

This axiom says, in particular, that the agent prefers a certain reward of 2 units to a certain reward of 1 unit. In the present framework, it also says that the agent disprefers a certain reward of 2 units to a lottery that pays a reward of 1 or 4 units with equal probabilities. As indicated in Sect. 2, such assumptions are inappropriate for risk-averse agents.

The second axiom, **A2**, formalizes the assumption that a reward of one unit at time $t > 1$ is worth the same as a reward of one unit at $t = 1$ (i.e., that $\beta = 1$). In the case when rewards represent utilities (or consumption) of future generations, **A2** is the axiom of *anonymity*, which ensures the equal treatment of generations.

A2. For all $u, v \in \mathscr{U}$, if u can be obtained from v by interchanging two entries of v, then $u \sim v$.

The next axiom is a relaxation of the consistency requirement used in Brock's (1970a) characterization of the overtaking criterion. For $n \geq 1$ and $u \in \mathscr{U}$, let $u_{[n]}$ denote the sequence obtained from u by replacing u_t with 0 for all $t > n$. Our third axiom can then be stated as follows.

A3. For all $u, v \in \mathscr{U}$, if there exists $N > 1$ such that $u_{[n]} \succ v_{[n]}$ for all $n \geq N$, then $u \succsim v$.

That the average reward criterion satisfies **A3** is a trivial consequence of the fact that $u_{[n]} \sim_{\mathrm{AR}} v_{[n]}$ for all $u, v \in \mathscr{U}$ and every $n \geq 1$. The preorders in (4), (7) and (8) have the stronger property that u is at least as good as v if $u_{[n]}$ is merely at least as good as $v_{[n]}$ for all sufficiently large n; this property does *not* hold for the average reward criterion.

The fourth axiom asserts that for reward streams $u, v \in \mathscr{U}$, if both streams are postponed one period and an arbitrary reward of $c \in \mathbb{R}$ is assigned to the first period, then the resulting streams, $(c, u) = (c, u_1, u_2, u_3, \ldots)$ and $(c, v) = (c, v_1, v_2, v_3, \ldots)$, should be ranked in the same way as u and v.

A4. For all $u, v \in \mathscr{U}$ and $c \in \mathbb{R}$, $(c, u) \succsim (c, v)$ if and only if $u \succsim v$.

This axiom was proposed as a fundamental condition by Koopmans (1960) in his pioneering work on intertemporal choice. It is usually referred to as *stationarity* (Asheim et al. 2010; Bleichrodt et al. 2008) or *independent future* (Fleurbaey and Michel 2003; Mitra 2018).

Our last axiom is an adaptation of the standard assumption of interpersonal comparability from social choice theory (see, e.g., d'Aspremont and Gevers 1977). In the intertemporal setting, it asserts that preferences are invariant to changes in the origins of the utility indices used in different periods. This condition has been referred to as *zero independence* (Moulin 1988) and *translation scale invariance* (Asheim et al. 2010).

A5. For all $u, v, \alpha \in \mathscr{U}$, if $u \succsim v$, then $u + \alpha \succsim v + \alpha$.

Note that a preorder $\succsim$ which satisfies **A5** has the property that if $u, v, u', v' \in \mathscr{U}$ are such that $u - v = u' - v'$, then $u \succsim v$ if and only if $u' \succsim v'$. (The converse is also true.) This fact will be used repeatedly below.

5 A rigidity result

If we view the axioms from the previous section as conditions which we expect a selective criterion to satisfy, then the first question from the introduction can be stated as follows: If $\succsim$ satisfies **A1**–**A5**, is every $\succsim$-optimal policy average overtaking-optimal (and hence 1-optimal)?[3] Theorem 1 shows that this question has an affirmative answer if attention is restricted to stationary policies. This restriction does not trivialize any of the questions (Q1–Q3) from the introduction. In fact, replacing Π with Π_F in the preceding discussion would not affect what has been said so far in an essential way.

Theorem 1 *Suppose that $\succsim$ satisfies **A1**–**A5**. If a policy is $\succsim$-optimal within Π_F, then it is average overtaking-optimal within Π_F.*

Proof The proof exploits the fact that under certain conditions on $u \in \mathscr{U}$, if a preorder $\succsim$ satisfies **A1**–**A5**, then

$$u \succsim (0, u) \text{ implies } \bar{u} \geq 0, \tag{9}$$

[3] An alternative way to state Q1 would be to ask if $\succsim_{\mathrm{AO}}$ is the least restrictive extension of $\succsim_{\mathrm{O}}$ that admits optimal policies. This question has a trivial answer, however, because $\succsim_{\mathrm{AO}}$ is not, strictly speaking, even an extension of $\succsim_{\mathrm{O}}$: if $u \succsim_{\mathrm{O}} v$, then $u \succsim_{\mathrm{AO}} v$, but there are $u, v \in \mathscr{U}$ with $u \succ_{\mathrm{O}} v$ and $u \sim_{\mathrm{AO}} v$ (see Jonsson and Voorneveld 2018, p. 28).

where

$$\bar{u} \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} u_t \tag{10}$$

is the average of u. The usefulness of (9) is explained by the fact that if $\succsim$ satisfies **A5** and $u, v \in \mathscr{U}$ are such that $\sigma \equiv \sigma(u - v)$ is bounded, then

$$u \succsim v \text{ if and only } \sigma \succsim (0, \sigma). \tag{11}$$

This is because $u - v = \sigma - (0, \sigma)$. Applying (9) with σ in the role of u, we see that $u \succsim v$ implies $\bar{\sigma} \geq 0$. Since $\bar{\sigma}$ is the Cesàro sum of $\sum_{t=1}^{\infty}(u_t - v_i)$, this means that $u \succsim v$ implies $u \succsim_{\mathrm{AO}} v$.

The conditions on $u \in \mathscr{U}$ which ensure (9) are that (i) the limit (10) exists and (ii) for every $\varepsilon > 0$ there exists an N such that the average of any $n \geq N$ consecutive coordinates of u differs from $\bar{u}$ by at most ε—that is,

$$\left| \frac{1}{n} \sum_{t=t_0}^{t_0+n} u_t - \bar{u} \right| < \varepsilon \text{ for every } t_0 \in \mathbb{N}.$$

We say that $u \in \mathscr{U}$ is *regular* if the two conditions are met.

Lemma 1 (Jonsson and Voorneveld 2018, Proposition 1) *Suppose that $\succsim$ satisfies **A1**–**A5**. If $u \in \mathscr{U}$ is regular and $c \in \mathbb{R}$, then*

$$(c, u) \succsim u \textit{ implies } c \geq \bar{u}$$

and

$$u \succsim (c, u) \textit{ implies } c \leq \bar{u}.$$

Now, for every $\pi \in \Pi_F$, $u(s, \pi)$ is regular for each $s \in \mathscr{S}$. This follows from the well known fact that the reward stream generated by a stationary policy can be written as the sum of a periodic sequence and a summable sequence. (The stream generated by $(f, f, f, \ldots)$ is defined by powers of $\mathbf{Q}(f)$ acting on $\mathbf{r}(f)$—see (1). By the Perron-Frobenius theorem for non-negative matrices, the sequence $\mathbf{Q}(f) \cdot \mathbf{r}(f), \mathbf{Q}(f)^2 \cdot \mathbf{r}(f), \mathbf{Q}(f)^3 \cdot \mathbf{r}(f), \ldots$ approaches a periodic orbit at exponential rate.) To apply the arguments preceding Lemma 1, we need to know that $\sigma(u - v)$ is bounded and regular if u and v are generated by stationary policies. We have the following result.

Lemma 2 *Suppose that u and v are generated by stationary policies, and let $\sigma \equiv \sigma(u - v)$ be defined as in (2). If $\bar{u} = \bar{v}$, then $\sigma \in \mathscr{U}$ is regular.*

Proof Write

$$u = x^{(u)} + y^{(u)}, \quad v = x^{(v)} + y^{(v)}, \tag{12}$$

where $x^{(u)}$ and $x^{(v)}$ are periodic and where $y^{(u)}$ and $y^{(v)}$ are summable. Let p be the product of the periods of $x^{(u)}$ and $x^{(v)}$. Then $\bar{u} = \bar{x}^{(u)} = \sigma_p(x^{(u)})/p$ and $\bar{v} = \bar{x}^{(v)} = \sigma_p(x^{(v)})/p$. So, if $\bar{u} = \bar{v}$, then $\sigma_p(x^{(u)} - x^{(v)}) = 0$. This means that $\sigma(x^{(u)} - x^{(v)})$ is periodic. The sequence $\sigma(y^{(u)} - y^{(v)})$ is convergent by our choice of $y^{(u)}$ and $y^{(v)}$. Hence, $\sigma = \sigma(u - v)$ is the sum of a periodic sequence and a convergent sequence. This means that $\sigma \in \mathscr{U}$ is regular. □

To complete the proof of Theorem 1, let $\succsim$ be a preorder that satisfies **A1**–**A5**, and suppose that π^* is $\succsim$-optimal within Π_F. Let $u = u(s, \pi^*)$ and $v = u(s, \pi)$, where $\pi \in \Pi_F$ and $s \in \mathscr{S}$ are arbitrary, and let $\sigma \equiv \sigma(u - v)$ be defined as in (2). Since π^* is $\succsim$-optimal within Π_F, $u \succsim v$. We need to show that $u \succsim_{\mathrm{AO}} v$. If $\bar{u} = \bar{v}$, then this follows from Lemmas 1 and 2 and the remarks preceding Lemma 1. It remains to show that $u \succsim_{\mathrm{AO}} v$ if $\bar{u} \neq \bar{v}$. It is enough to show that $\bar{u} > \bar{v}$, since this clearly implies $u \succ_{AO} v$. Given any preorder $\succsim'$ that satisfies **A1**–**A5**, if $x \in \mathscr{U}$ and $y \in \mathscr{U}$ are such that $\bar{x} > \bar{y}$, then $x \succ' y$ (see Basu and Mitra 2007 or Jonsson and Voorneveld 2015). Thus, if $\bar{u} \neq \bar{v}$, then we must have $\bar{u} > \bar{v}$. (If it were the case that $\bar{v} > \bar{u}$, then we would have $v \succ u$, which contradicts the assumption that $u \succsim v$.) We can therefore conclude that $u \succ_{\mathrm{AO}} v$, and the proof of Theorem 1 is thereby complete. □

6 Characterizations

One goal of this paper is to provide a preference foundation for finite MDPs. In the case of a positive discount rate, the well known preference foundation of Koopmans (1960, 1972) is easily adapted to the present setting. The literature provides characterizations of two criteria for the no discounting case: the overtaking criterion (Asheim and Tungodden 2004; Basu and Mitra 2007; Brock 1970a) and the average reward criterion (Kothiyal et al. 2014; Marinacci 1998; Khan and Stinchcombe 2018; Pivato 2022). The overtaking criterion is characterized by axioms that are similar to those in Sect. 4. The characterizations of the average reward criterion, which does not satisfy **A1**, involve further conditions of permutability and numeric representability. These conditions are well known to be incompatible with **A1** in the no discounting case (Basu and Mitra 2003; Fleurbaey and Michel 2003).

In this section, we axiomatize the preorders associated with the average overtaking criterion and the 1-optimality criterion. As in the previous section, we restrict attention to stationary policies.

6.1 First characterization

The axioms from Sect. 4 do not characterize $\succsim_{\mathrm{AO}}$. Indeed, the preorder associated with the overtaking criterion satisfies **A1**–**A5**, and $\succsim_{\mathrm{O}}$ does not agree with $\succsim_{\mathrm{AO}}$ on $\mathscr{U}_F$. As illustrated in Example 1, for $\succsim_{\mathrm{AO}}$-optimality to imply $\succsim$-optimality, it is necessary that $\succsim$ compares at least some pairs of streams that $\succsim_{\mathrm{O}}$ does not compare.

Insisting that all pairs $u, v \in \mathscr{U}$ be comparable has unwanted consequences. In fact, it is not possible to give an explicit definition of a preorder, satisfying **A1** and **A2**, that compares all pairs of sequences of 0s and 1s (Lauwers 2010). On the other hand,

$\succsim_{\text{AO}}$ compares each pair $u, v \in \mathscr{U}_F$ and coincides with $\succsim_1$ on this domain. Thus, the following condition is compatible with **A1**–**A5**:

A6. For all $u, v \in \mathscr{U}_F$, $\succsim$ compares u and v.

If $\succsim$ satisfies **A1**–**A6** and $u, v \in \mathscr{U}_F$, then $u \succ v$ if and only if $u \succ_{\text{AO}} v$. To conclude that the symmetric parts of $\succsim$ and $\succsim_{\text{AO}}$ agree, further assumptions are needed. A sufficient condition asserts that, for all $u, v \in \mathscr{U}$, if $(\varepsilon + u_1, u_2, u_3, \ldots) \succsim v$ for every $\varepsilon > 0$, then $u \succsim v$. This condition can be formalized by defining a metric on $\mathscr{U}$ and demanding that $\{v \in \mathscr{U} : u \succsim v\}$ be a closed subset of $\mathscr{U}$ for every $u \in \mathscr{U}$. Almost any metric from the literature will do (e.g., Banerjee and Mitra 2008, p. 5). For example, let $d(u, v) = \min\{1, \sum_{i=1}^{\infty} |u_i - v_i|\}$. The continuity requirement can then be stated as follows.

A7. For every $u \in \mathscr{U}$, $\{v \in \mathscr{U} : u \succsim v\}$ is a closed subset of $\mathscr{U}$.

Theorem 2 *If $\succsim$ satisfies **A1**–**A7**, then $\succsim$ and $\succsim_{AO}$ coincide on $\mathscr{U}_F$.*

Proof Let $\succsim$ satisfy **A1**–**A7**, and let $u, v \in \mathscr{U}_F$. We know that $u \succsim_{\text{AO}} v$ if $u \succsim v$ (Theorem 1). So it is enough to show that $u \succsim_{\text{AO}} v$ implies $u \succsim v$.

If $u \succ_{\text{AO}} v$, then either (i) $\bar{u} > \bar{v}$ or (ii) $\bar{u} = \bar{v}$ and $\bar{\sigma} > 0$, where $\sigma = \sigma(u - v)$. In case (i), we get $u \succ v$ as a consequence of the fact that $\succsim$ satisfies **A1**–**A5**. In case (ii), $\neg(0, \sigma) \succsim \sigma$ by Lemma 1, so $\neg v \succsim u$ by **A5**. By **A6**, $u \succ v$. Conclude that $u \succ_{\text{AO}} v$ implies $u \succ v$.

Now suppose that $u \sim_{\text{AO}} v$. Let $u^{(\varepsilon)} = (\varepsilon + u_1, u_2, u_3, \ldots)$. Then $u^{(\varepsilon)} \succ_{\text{AO}} v$ for every $\varepsilon > 0$, so (by the above conclusion) $u^{(\varepsilon)} \succ v$ for every $\varepsilon > 0$. By **A7**, $u \succsim v$. The same argument shows that $v \succsim u$. □

6.2 Second characterization

Axioms **A6** and **A7** were motivated by necessity rather than some normative or economic reason. In our second characterization, these axioms are replaced by the *compensation principle*.

As an illustration of this principle, imagine that the decision maker is faced with two options. The first option yields some sequence of expected rewards $u \in \mathscr{U}$. The second option is to obtain a one-period postponement of u and a compensation of $c \in \mathbb{R}$ in the first period. Which value of c should make the agent indifferent?

In some cases, this value will be zero. This is the case if u has at most finitely nonzero entries—then $(0, u)$ and u are equally good by **A2**. However, the agent will not always be indifferent if $c = 0$. For instance, if $u = (r, r, r, \ldots)$ is constant and c is less than $r > 0$, then (c, u) is worse than u by **A1**. The compensation principle says that u and (c, u) are equally good if $c = \bar{u}$ (compare Lemma 1). Its precise statement is as follows:

A8. For every $u \in \mathscr{U}$, if $\bar{u}$ is well defined, then $(\bar{u}, u) \sim u$.

For a case of the two options described above, consider again the system in Fig. 1, and suppose that the system starts in s_1. The agent then has two options. If a_1 is chosen, then $u = (2, 0, 2, 0, 2, \ldots)$ obtains. If a_2 is chosen, then u is delayed one period, and a reward of c is obtained in the first period. Thus, the two feasible alternatives are u and $v = (c, u)$. Since $\bar{u} = 1$, **A8** says that u and v are equally good if $c = 1$.

Example 1 illustrates the fact that $\succsim_{\text{O}}$ violates **A8**. It is easy to check that $\succsim_{\text{AO}}$ satisfies **A8**, and the same is true of $\succsim_1$ (Jonsson and Voorneveld 2018). To see that the average reward criterion also satisfies **A8**, note that if $d = (c, u) - u$, then we have $\sigma_T(d) = c - u_T$ and therefore $\liminf_{T\to\infty} \frac{1}{T}\sigma_T(d) = \liminf_{T\to\infty} \frac{1}{T}\sigma_T(-d) = 0$. It follows that $(c, u) \sim_{\text{AR}} u$ for *every* $c \in \mathbb{R}$ and $u \in \mathscr{U}$.

Like (9), the usefulness of **A8** stems from the fact that if $\succsim$ satisfies **A5** and $u, v \in \mathscr{U}$ are such that $\sigma \equiv \sigma(u - v)$ is bounded, then $u \succsim v$ if and only $\sigma \succsim (0, \sigma)$. Thus, if $\succsim$ satisfies **A1**, **A5** and **A8**, then $u \succsim v$ if and only $\bar{\sigma} \geq 0$. In Jonsson and Voorneveld (2018), this observation is used to characterize $\succsim_1$ on the set of streams that are summable or eventually periodic. Theorem 3 extends this result to streams that can be decomposed according to (12).

Theorem 3 *If $\succsim$ satisfies A1, A5 and A8, then $\succsim$ and $\succsim_{\text{AO}}$ coincide on $\mathscr{U}_F$.*

Proof Let $\succsim$ be a preorder that satisfies **A1**, **A5** and **A8**. For $u, v \in \mathscr{U}_F$, let $\sigma = \sigma(u - v)$. Suppose that $\bar{u} = \bar{v}$. Then $\sigma \in \mathscr{U}$ is regular (Lemma 2), which means that $\bar{\sigma}$ is well defined. By **A1** and **A8**, $\sigma \succsim (0, \sigma)$ if and only if $\bar{\sigma} \geq 0$. By **A5**, $u \succsim v$ if and only if $\sigma \succsim (0, \sigma)$. Hence, $u \succsim v$ if and only if $\bar{\sigma} \geq 0$. Since $\bar{\sigma}$ is the Cesàro sum of $\sum_{t=1}^{\infty}(u_t - v_i)$, we see that $u \succsim v$ if and only if $u \succsim_{\text{AO}} v$.

Now suppose (without loss of generality) that $\bar{u} > \bar{v}$. Then $u \succ_{\text{AO}} v$. We show that $u \succ v$. For $T > 1$, define $z \in \mathscr{U}$ by setting $z_t = u_t$ for $t \leq T$ and $z_t = u_t - c$ for $t > T$. Then z is the sum of periodic sequence and a summable sequence, and $u \succ z$ by **A1**. Since $\bar{u} > \bar{v}$, we can choose T so that $\sigma_t(u - z) \geq 0$ for all $t \geq T$. Since $\bar{z} = \bar{v}$, the preceding argument gives that $z \succsim v$, so $u \succ v$ by transitivity. □

We can obtain a characterization of average overtaking optimality in general discrete time MDPs by generalizing **A8**. This result, which concerns optimality within the class of all policies, is provided in the appendix. There we also verify that the axioms in Theorem 3 are logically independent.

Theorems 2 and 3 provide two axioms sets that characterize $\succsim_{\text{AO}}$ on $\mathscr{U}_F$. As a corollary of these results, we obtain a partial answer to the third question (Q3) from the introduction: If $\succsim$ satisfies the axioms in any one of these axiom sets, then a policy is $\succsim$-optimal within Π_F if and only if it is $\succsim_{\text{AO}}$-optimal within Π_F. In particular, a $\succsim$-optimal policy exists within Π_F.

Acknowledgements Thanks to Henrik Hult for a discussion on the topic and to Eugene Feinberg for very helpful feedback on an earlier version of the paper. This version has greatly benefitted from comments by two anonymous referees.

Funding Open access funding provided by Lulea University of Technology.

Appendices

Appendix A contains a characterization result on average overtaking optimality within the set of all policies. Appendix B establishes that the axioms used in Theorems 1 and 3 are logically independent.

A Average overtaking optimality within the set of all policies

Theorem 4 below provides a characterization of average overtaking optimality in general discrete time MDPs. In particular, we make no assumptions on the state and action spaces.

To allow for unbounded reward functions, let us substitute $\mathscr{U}$, the set of bounded sequences, for $\mathscr{V} = \mathbb{R}^{\mathbb{N}}$, the set of all real sequences. The reward stream generated by a non-stationary policy need not be regular, so its average may be undefined. For $u \in \mathscr{V}$, we let

$$\bar{u}_* = \liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} u_t, \quad \bar{u}^* = \limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} u_t. \tag{13}$$

Our characterization result for discrete time MDPs uses the following three properties of the average overtaking criterion:

A1′. For all $u, v \in \mathscr{V}$, if $u_t \geq v_t$ for all t and $u_t > v_t$ for some t, then $u \succ v$.
A5′. For all $u, v, \alpha \in \mathscr{V}$, if $u \succsim v$, then $u + \alpha \succsim v + \alpha$.
A8′. For every $u \in \mathscr{V}$, if $\bar{u}^*$ is finite, then $(\bar{u}^*, u) \succsim u$. If $\bar{u}^* = +\infty$, then $u \succsim (0, u)$.

That $\succsim_{\mathrm{AO}}$ satisfies **A8′** is easy to see once we observe that for $u \in \mathscr{V}, c \in \mathbb{R}$, if $v = (c, u)$, then we have $\sigma_t(u - v) = u_t - c$ and $\sigma_t(v - u) = c - u_t$. Hence,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \sigma_t(v - u) = c - \bar{u}^*.$$

It follows that $(\bar{u}^*, u) \succsim_{\mathrm{AO}} u$ if $\bar{u}^*$ is finite, and that $u \succsim (0, u)$ if $\bar{u}^* = +\infty$.

Let us also note that every preorder that satisfies **A8′** and **A5′** also has the property that for all $u \in \mathscr{V}$,

$$\text{if } \bar{u}_* \text{ is finite, then } u \succsim (\bar{u}_*, u). \tag{14}$$

To see this, let $x = -u$. Then $\bar{x}^* = -\bar{u}_*$, so **A8′** implies $(-\bar{u}_*, -u) \succsim -u$. By **A5′** (adding $\alpha = u + (\bar{u}_*, u)$), this means that $u \succsim (\bar{u}_*, u)$. In particular,

$$u \succsim_{\mathrm{AO}} (\bar{u}_*, u) \quad \text{and} \quad (\bar{u}^*, u) \succsim_{\mathrm{AO}} u. \tag{15}$$

Theorem 4 *Let $\succsim$ be a preorder on $\mathscr{V}$ that satisfies* **A1′**, **A5′** *and* **A8′**. *If a policy is $\succsim_{AO}$-optimal, then it is also $\succsim$-optimal.*

Note that Theorem 4 concerns the implication from $\succsim_{\text{AO}}$-optimality to $\succsim$-optimality whereas Theorem 1 concerns the reverse implication. As indicated above, Theorem 1 does not hold in non-finite MDPs. For example, the 1-optimality criterion satisfies **A1′**, **A5′** and **A8′**, and a 1-optimal policy need not be average overtaking optimal. Whether or not Theorem 1 holds in finite MDPs, without restricting to stationary policies, is a question that we have not been able to answer.

Proof Let $\succsim$ be a preorder on $\mathscr{V}$ that satisfies **A1′**, **A5′** and **A8′**. Let $u \in \mathscr{V}$ be the stream of expected rewards generated by a $\succsim_{\text{AO}}$-optimal policy, given some initial state, and let $v \in \mathscr{V}$ be generated by some other policy for the same initial state. We need to show that $u \succsim v$.

Let $\sigma_n = \sigma_n(u - v)$, $n \geq 1$, and let $\sigma = (\sigma_1, \sigma_2, \sigma_3 \ldots) \in \mathscr{V}$. That u is generated by a $\succsim_{\text{AO}}$-optimal policy means that $u \succsim_{\text{AO}} v$. By the definition of $\succsim_{\text{AO}}$, this implies that $\bar{\sigma}_* \geq 0$. Suppose first that $\bar{\sigma}_* < +\infty$. Since $\succsim$ satisfies **A5′** and **A8′**, we then have $\sigma \succsim (\bar{\sigma}_*, \sigma)$ (see (14)). By **A1′** and transitivity, we thus have $\sigma \succsim (0, \sigma)$. By **A5′** and the fact that $u - v = \sigma - (0, \sigma)$, this entails $u \succsim v$. If $\bar{\sigma}_*$ equals $+\infty$, then so does $\bar{\sigma}^*$. We then have $\sigma \succsim (0, \sigma)$ by **A8′** and hence $u \succsim v$ by **A5′**. Conclude that $u \succsim v$. Since v was generated by an arbitrary policy, this shows that any $\succsim_{\text{AO}}$-optimal policy is $\succsim$-optimal. □

B Logical independence

Independence of the axioms in Theorem 1

The following binary relations, defined for all $u, v \in \mathscr{U}$, fail to satisfy precisely one of the axioms used in Theorem 1:

$$u \succsim_{\neg\mathbf{A1}} v \iff u, v \in \mathscr{U} \text{ (all streams are equivalent)}$$

$$u \succsim_{\neg\mathbf{A2}} v \iff \sum_{t=1}^{\infty} 2^{-t}(u_t - v_t)$$

$$u \succsim_{\neg\mathbf{A3}} v \iff \exists T_0 \in \mathbb{N} \text{ s.t. } u_t \geq v_t \text{ for all } t > T_0 \text{ and } \sigma_{T_0}(u - v) \geq 0$$

$$u \succsim_{\neg\mathbf{A4}} v \iff \liminf_{T \to \infty} \sigma_{2 \cdot T}(u - v) \geq 0 \text{ (cf. Fleurbaey and Michel 2003, p. 786)}$$

$$u \succsim_{\neg\mathbf{A5}} v \iff \liminf_{T \to \infty} \sigma_T(u^3 - v^3) \geq 0.$$

We omit the proofs for the first three of these five preorders. The fourth clearly satisfies **A1**, **A2** and **A5**. To verify **A3**, note that for any $u, v \in \mathscr{U}$ and $n \in \mathbb{N}$, we have $u_{[n]} \succ_{\neg\mathbf{A4}} v_{[n]}$ if and only if $\sigma_n(u - v) > 0$. (Recall that $u_{[n]}$ is the stream obtained from u by replacing u_t with 0 for $t > n$.) Thus, if $u_{[n]} \succ_{\neg\mathbf{A4}} v_{[n]}$ for all sufficiently large n, then $\sigma_n(u - v) > 0$ for all sufficiently large n. In particular, $\sigma_{2 \cdot n}(u - v) > 0$ for all sufficiently large n, which means that $u \succsim_{\neg\mathbf{A4}} v$. Conclude that **A3** holds.

To see that $\succsim_{\neg\mathbf{A4}}$ violates **A4**, let u and v be the streams in Example 1 with $c = 1$, so that $u = (2, 1, 2, \ldots)$ and $v = (1, 2, 1, 2, \ldots)$. By the definition of $\succsim_{\neg\mathbf{A4}}$, we have $u \sim_{\neg\mathbf{A4}} v$ and $(2, u) \succ_{\neg\mathbf{A4}} (2, v)$. This shows that $\succsim_{\neg\mathbf{A4}}$ fails to satisfy **A4**.

It is straightforward to check that $\succsim_{\neg\mathbf{A5}}$ satisfies **A1**–**A4**. To show that **A5** fails, let $u = (3, 0, 0, 0, \ldots)$, $v = (2, 2, 0, 0, 0, \ldots)$, $\alpha = (0, 1, 0, 0, 0, \ldots)$. Define $x = u+\alpha$, $y = v + \alpha$. Then $x = (3, 1, 0, 0, 0, \ldots)$ and $y = (2, 3, 0, 0, \ldots)$. So, for $T \geq 2$, we have $\sigma_T(u^3 - v^3) = 11$ and $\sigma_T(x^3 - y^3) = -7$. Hence, $u \succsim_{\neg\mathbf{A5}} v$, but $u + \alpha \neg \succsim_{\neg\mathbf{A5}} v + \alpha$. This shows that **A5** fails.

The five preorders ($\succsim_{\neg\mathbf{A1}}$ to $\succsim_{\neg\mathbf{A5}}$) establish logical independence of **A1**– **A5**. The first, second, fourth and fifth preorder show that Theorem 1 fails if we drop any one of **A1**, **A2**, **A4** and **A5**. We have been unable to find a preorder which shows that the theorem fails if **A3** is dropped.

Independence of the axioms in Theorem 3

We show that the axioms in Theorem 3 are logically independent by providing three preorders, each violating precisely one of the three axioms. The overtaking criterion satisfies **A1** and **A5**, but not **A8** (see Example 1). The preorder $\succsim_{\neg\mathbf{A1}}$ satisfies **A5** and **A8**, but not **A1**. (The average reward criterion provides another example.) It remains to find a preorder that satisfies **A1** and **A8**, but violates **A5**.

For $u, v \in \mathscr{U}$, let us we say that u *dominates* if $u_t \geq v_t$ for all $t \in \mathbb{N}$, and that u is a *finite permutation* of v if u can be obtained from v by permuting finitely many entries of v. Consider the following binary relation:

$$u \succsim_{\mathrm{SS}} v \iff \text{some finite permutation of } u \text{ dominates } v.$$

This is the Suppes-Sen grading principle. It is the weakest preorder satisfying **A1** and **A2** (see, e.g., Basu and Mitra 2007, p. 356). Note that $\succsim_{\mathrm{SS}}$ satisfies **A4**.

For $c \in \mathbb{R}$, $n \in \mathbb{Z}_+ = \{0, 1, 2, 3, \ldots\}$ and $u \in \mathscr{U}$, let

$$([c]_n, u) = \begin{cases} (\overbrace{c, c, \ldots, c}^{n \text{ times}}, u) & \text{if } n \geq 1, \\ u & \text{if } n = 0. \end{cases} \tag{16}$$

Our last preorder is defined as follows for all $u, v \in \mathscr{U}$:

$$u \succsim_* v \iff ([\bar{u}_*]_n, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_m, v) \text{ for some } n, m \in \mathbb{Z}_+.$$

Note that $\bar{u}_*$ and $\bar{u}^*$ (see (13)) are finite for each $u \in \mathscr{U}$.

We first check that $\succsim_*$ is indeed a preorder. Reflexivity is obvious. To show that $\succsim_*$ is transitive, let $u, v, w \in \mathscr{U}$ be such that $u \succsim_* v$ and $v \succsim_* w$. That is, $([\bar{u}_*]_n, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_m, v)$ and $([\bar{v}_*]_k, v) \succsim_{\mathrm{SS}} ([\bar{w}^*]_l, w)$ for some $n, m, k, l \in \mathbb{Z}_+$. By the definition of $\succsim_{\mathrm{SS}}$, we must then have that

$$\bar{u}^* \geq \bar{v}^* \geq \bar{w}^* \text{ and } \bar{u}_* \geq \bar{v}_* \geq \bar{w}_*.$$

Since $\succsim_{\mathrm{SS}}$ satisfies **A4**, $([\bar{u}_*]_n, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_m, v)$ implies

$$([\bar{v}_*]_k, [\bar{u}_*]_n, u) \succsim_{\mathrm{SS}} \overbrace{([\bar{v}_*]_k, [\bar{v}^*]_m, v)}^{x}$$

and $([\bar{v}_*]_k, v) \succsim_{\mathrm{SS}} ([\bar{w}^*]_l, w)$ implies

$$\overbrace{([\bar{v}^*]_m, [\bar{v}_*]_k, v)}^{y} \succsim_{\mathrm{SS}} ([\bar{v}^*]_m, [\bar{w}^*]_l, w).$$

By **A2** and transitivity, $x \sim_{\mathrm{SS}} y$. By transitivity,

$$([\bar{v}_*]_k, [\bar{u}_*]_n, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_m, [\bar{w}^*]_l, w). \tag{17}$$

By **A1** and transitivity, (17) implies $([\bar{u}_*]_{n+k}, u) \succsim_{\mathrm{SS}} ([\bar{w}_*]_{l+m}, w)$, which means that $u \succsim_* w$. Conclude that $\succsim_*$ is transitive and hence a preorder.

To see that $\succsim_*$ satisfies **A8**, let $u \in \mathscr{U}$ be such that $\bar{u}$ is well defined, and let $v = (\bar{u}, u)$. Then $\bar{v}$ is well defined and $\bar{u} = \bar{v}$. Since $([\bar{u}_*]_1, u) = ([\bar{v}^*]_0, v)$, we have $([\bar{u}_*]_1, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_0, v)$ and therefore $u \succsim_* v$. Since $([\bar{v}_*]_0, v) = ([\bar{u}^*]_1, u)$, we have $([\bar{v}_*]_0, v) \succsim_{\mathrm{SS}} ([\bar{u}^*]_1, u)$ and therefore $v \succsim_* u$. Thus, $v = (\bar{u}, u) \sim_* u$, which shows that $\succsim_*$ satisfies **A8**.

To verify **A1**, suppose u is strictly better than v by **A1**. Then $([\bar{u}_*]_0, u) \succsim_{\mathrm{SS}} ([\bar{v}^*]_0, v)$, so $u \succsim_* v$. To show that $v\neg \succsim_* u$, we need to rule out the possibility that $([\bar{v}_*]_n, v) \succsim_{\mathrm{SS}} ([\bar{u}^*]_m, u)$ for some $n, m \in \mathbb{Z}_+$. Suppose for contradiction that $([\bar{v}_*]_n, v) \succsim_{\mathrm{SS}} ([\bar{u}^*]_m, u)$ for $n, m \in \mathbb{Z}_+$. Since $\succsim_{\mathrm{AO}}$ satisfies **A1** and **A2**, this implies that $([\bar{v}_*]_n, v) \succsim_{\mathrm{AO}} ([\bar{u}^*]_m, u)$. By (15) and transitivity, we have $v \succsim_{\mathrm{AO}} ([\bar{v}_*]_n, v)$ and $([\bar{u}^*]_m, u) \succsim_{\mathrm{AO}} u$. By transitivity, this means that $v \succsim_{\mathrm{AO}} u$, contradicting that $\succsim_{\mathrm{AO}}$ satisfies **A1**. "We can therefore conclude that $v\neg \succsim_* u$, so that $u \succ_* v$ This shows that $\succsim_*$ satisfies **A1**".

It remains to show that $\succsim_*$ violates **A5**. Define $u = (3, 5, 1, 1, 1, \ldots)$, $v = (4, 2, 1, 1, 1, \ldots)$. Since $u \succsim_{\mathrm{SS}} v$, we have $u \succsim_* v$. Let $\alpha = (-1, 1, 0, 0, 0, \ldots)$, let $x = u + \alpha$, and let $y = v + \alpha$. Then $x = (2, 6, 1, 1, 1, \ldots)$ and $y = (3, 3, 1, 1, 1, \ldots)$. Since $([\bar{x}_*]_n, x) = ([1]_n, 2, 6, 1, 1, 1, \ldots)$ and $([\bar{y}^*]_m, y) = ([1]_m, 3, 3, 1, 1, 1, \ldots)$, there are no $n, m \in \mathbb{Z}_+$ with $([\bar{x}_*]_n, x) \succsim_{\mathrm{SS}} ([\bar{y}]^*_m, y)$. Conclude that $x\neg \succsim_* y$. This shows that $\succsim_*$ violates **A5**.

The three preorders ($\succsim_{\neg\mathbf{A1}}$, $\succsim_*$ and $\succsim_{\mathrm{O}}$) establish logical independence of **A1**, **A5** and **A8**. These preorders also show that Theorem 3 fails if any one of these three axioms is dropped.

References

Arapostathis A, Borkar VS, Fernández-Gaucherand E, Ghosh MK, Marcus SI (1993) Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM J Control Optim 31(2):282–344

Asheim G, Tungodden B (2004) Resolving distributional conflicts between generations. Econ Theor 24(1):221–230

Asheim GB, d'Aspremont C, Banerjee K (2010) Generalized time-invariant overtaking. J Math Econ 46(4):519–533

Banerjee K, Mitra T (2008) On the continuity of ethical social welfare orders on infinite utility streams. Soc Choice Welfare 30(1):1–12

Basu K, Mitra T (2003) Aggregating infinite utility streams with intergenerational equity: the impossibility of being Paretian. Econometrica 71(5):1557–1563

Basu K, Mitra T (2007) Utilitarianism for infinite utility streams: a new welfare criterion and its axiomatic characterization. J Econ Theory 133(1):350–373

Bäuerle N, Rieder U (2014) More risk-sensitive Markov decision processes. Math Oper Res 39(1):105–120

Bishop CJ, Feinberg EA, Zhang J (2014) Examples concerning Abel and Cesàro limits. J Math Anal Appl 420(2):1654–1661

Blackwell D (1962) Discrete dynamic programming. Ann Math Stat 33(2):719–726

Bleichrodt H, Rohde KIM, Wakker PP (2008) Koopmans' constant discounting for intertemporal choice: a simplification and a generalization. J Math Psychol 52:341–347

Brock WA (1970) An axiomatic basis for the Ramsey-Weizsäcker overtaking criterion. Econometrica 38(6):927–929

Brock WA (1970) On existence of weakly maximal programmes in a multi-sector economy. Rev Econ Stud 37(2):275–280

Brock WA, Mirman LJ (1973) Optimal economic growth and uncertainty: the no discounting case. Int Econ Rev 14(3):560–573

Brown BW (1965) On the iterative method of dynamic programming on a finite space discrete time Markov processes. Ann Math Stat 36:1279–1285

d'Aspremont C, Gevers L (1977) Equity and the informational basis of collective choice. Rev Econ Stud 44(2):199–209

Denardo EV, Miller BL (1968) An optimality condition for discrete dynamic programming with no discounting. Ann Math Stat 39(4):1220–1227

Feinberg EA, Shwartz A (2002) Handbook of Markov decision processes: methods and applications. International series in operations research and management science. Kluwer Academic Publishers, London

Flesch J, Predtetchinski A, Solan E (2017) Sporadic overtaking optimality in Markov decision problems. Dyn Games Appl 7:212–228

Fleurbaey M, Michel P (2003) Intertemporal equity and the extension of the Ramsey criterion. J Math Econ 39(7):777–802

Gale D (1967) On optimal development in a multi-sector economy. Rev Econ Stud 34(1):1–18

Hernández-Lerma O, Vega-Amaya O (1998) Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality. Appl Math 25(2):153–178

Jonsson A, Voorneveld M (2015) Utilitarianism on infinite utility streams: summable differences and finite averages. Econ Theory Bull 3:19–31

Jonsson A, Voorneveld M (2018) The limit of discounted utilitarianism. Theor Econ 13(1):19–37

Khan U, Stinchcombe M (2018) Planning for the long run: programming with patient, Pareto responsive preferences. J Econ Theory 176:444–478

Koopmans TC (1960) Stationary ordinal utility and impatience. Econometrica 28(2):287–309

Koopmans TC (1972) Representations of preference orderings with independent components of consumption, and representations of preference orderings over time. In: McGuire CB, Radner R (eds) Decision and organization. University of Minnesota Press, Minneapolis

Kothiyal A, Spinu V, Wakker PP (2014) Average utility maximization: a preference foundation. Oper Res 62(1):207–218

Lauwers L (2010) Ordering infinite utility streams comes at the cost of a non-Ramsey set. J Math Econ 46(1):32–37

Lippman SA (1969) Letter to the Editor: criterion equivalence in discrete dynamic programming. Oper Res 17(5):920–923

Marinacci M (1998) An axiomatic approach to complete patience and time invariance. J Econ Theory 83:105–144

Mitra T (2018) Sensitivity of stationary equitable preferences. In: Mishra A, Ray T (eds) Markets, governance, and institutions in the process of economic development. Oxford University Press

Moulin H (1988) Axioms of cooperative decision making, vol 15. Econometric society monographs. Cambridge University Press, Cambridge

Nowak AS, Vega-Amaya O (1999) A counterexample on overtaking optimality. Math Methods Oper Res 49:435–439

Piunovskiy AB (2013) Examples in Markov decision processes. World Scientific, Singapore

Pivato M (2022) A characterization of Cesàro average utility. J Econ Theory 201:1054408

Puterman ML (1994) Markov Decision processes: discrete stochastic dynamic programming. Wiley, London

Rosenberg D, Solan E, Vieille N (2002) Blackwell optimality in Markov decision processes with partial observation. Ann Stat 30(4):1178–1193

Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. Math Program 125:235–261

Stern LE (1984) Criteria of optimality in the infinite-time optimal control problem. J Optim Theory Appl 44(3):497–508

Veinott AF (1966) On finding optimal policies in discrete dynamic programming with no discounting. Ann Math Stat 37(5):1284–1294