Games and Economic Behavior 152 (2025) 396–422

Contents lists available at ScienceDirect

Games and Economic Behavior

journal homepage: www.elsevier.com/locate/geb

A taste for variety ☆

Galit Ashkenazi-Golan [a], Dominik Karos [b,*], Ehud Lehrer [c]

[a] Department of Mathematics, London School of Economics, UK
[b] Center for Mathematical Economics, Bielefeld University, Germany
[c] Tel Aviv University, Israel

ARTICLE INFO

JEL classification:
C61
C73
D01
D91

Keywords:
Repeated decision problem
Intertemporal choice
Time-inconsistent preferences
Habit formation

ABSTRACT

A decision maker repeatedly chooses one of a finite set of actions. In each period, the decision maker's payoff depends on a fixed basic payoff of the chosen action and the frequency with which the action has been chosen in the past. We analyze optimal strategies associated with three types of evaluations of infinite payoffs: discounted present value, the limit inferior, and the limit superior of the partial averages. We show that when the first two are the evaluation schemes (and the discount factor is sufficiently high), a stationary strategy can achieve the best possible outcome. However, for the latter evaluation scheme, a stationary strategy can achieve the best outcome only if all actions that are chosen with strictly positive frequency by an optimal stationary strategy have the same basic payoff.

1. Introduction

When Phil Connors[1] was trapped in a time loop, he initially enjoyed being able to do as he liked without fearing any repercussions. Yet, after a while, he became depressed as the rather limited entertainment options available in Punxsutawney did not measure up to his taste for variety. In this paper we investigate what Phil's optimal long-term payoff would have been, had he not been able to escape his temporal prison. That is, we consider a decision maker who has to repeatedly choose from a finite set of actions and whose stage payoff depends both on the action itself and also on how often she has chosen it in the past.

The model that we propose here looks rather innocuous. There is a finite set of actions, each endowed with a fixed basic payoff, and at each period the decision maker has to choose one of them. Her stage utility from choosing some action a is a's basic payoff multiplied by a factor that depends on the frequency with which a has been played so far and her taste for variety. The greater this frequency, the smaller the utility.

The decision maker is interested in her long-run payoff. We analyze three types of long-term payoff evaluations: the limit inferior and limit superior of the partial averages and the discounted one. It turns out that the limit inferior and discounted evaluations share the important feature that their optimal outcomes can be achieved by stationary strategies. However, the optimal strategy for the limit superior evaluation is stationary only in the degenerate case where all actions chosen with strictly positive frequency by an optimal stationary strategy have the same basic payoff.

☆ This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 461570745.

* Corresponding author.
E-mail address: dominik.karos@uni-bielefeld.de (D. Karos).

[1] Played by Bill Murray in "Groundhog Day", 1993.

https://doi.org/10.1016/j.geb.2025.05.003
Received 23 May 2024

Available online 16 May 2025

From a mathematical perspective, the decision-maker faces a deterministic Markov Decision Process (MDP) with countably many states, which are represented by the empirical distribution of previous actions and the time. In each period, the actual payoff depends on the current state and the decision made, while the transition between states is determined by the effect of the current decision on the empirical distribution.

Compared to MDP models with a finite or compact set of states, a model with countably many states introduces additional technical complexity. Addressing this complexity typically requires methods specifically developed for such models. Indeed, the fixed-point arguments that are typically used to prove the optimality of stationary strategies in problems with discounted evaluation cannot be applied in our case due to the countably infinite state space.

The paper presents two main results: stationary strategies are optimal in cases involving limit inferior and discounted evaluation. To achieve the first result, we introduce a potential function, defined in Equation (19). To the best of our knowledge, this is the first application of a potential function to the analysis of an MDP with a countable state space. It seems that similar techniques, as used for instance for the discounted problem, see Equation (41), could be useful for other such MDPs as well. Our third main result relates to the limit superior and shows that (generically) it cannot be achieved by a stationary strategy. We prove, however, that it can be achieved by some (non-stationary) strategy.

The paper is organized as follows. In Section 3 we introduce the necessary notation and provide some examples that highlight the different ways an infinite history of actions might be evaluated. In particular, we illustrate by means of an example with two actions that the optimal limes superior cannot be achieved by a stationary strategy. In Section 4 we investigate greedy histories, which maximize the stage utility in each period. We observe that such strategies are stationary, but we show that they are far from optimal even within the set of stationary strategies. In Section 5 we show that the optimal limit inferior can be achieved by a stationary strategy. Moreover, we show that the action frequencies of optimal histories are first-order stochastically ordered as the fatigue factor increases: the larger this factor, the more weight the optimal frequency will put on poor actions. Section 6 deals with the optimal limit superior. We show that the sequence of optimal average payoffs after finite time converges against the optimal limit superior and we use this observation to show that the latter cannot be achieved by a stationary strategy unless the optimal stationary strategy chooses the same action in each period. Section 7 deals with two aspects of discounting: discounting future payoffs and discounting the effect of past uses of actions. Discounting future payoffs means that one values future positive payoffs less than present ones. This is because one prefers to have good things now rather than later. Discounting the effect of past uses of actions means the impact of past experience on the present utility diminishes with time. For example, if one eats the same meal every day, one will eventually get tired of it. However, if one had a delicious meal yesterday, he or she would prefer the same meal today less than if he or she had it only a year ago. The main result of this section states that the optimal outcome for a relatively patient decision maker can be obtained with stationary strategies.

2. Related literature

Our model leaves the realm of classical economic theory as the assumptions of static preferences and discounted utility (proposed by Samuelson (1937) and later motivated with an axiomatic foundation by Koopmans (1960)) are dropped. Since then, this approach has been challenged in various contexts. Most closely related to our decision maker's preferences are models of "habit formation" and, in particular, the one of Kaiser and Schwabe (2012). Originally, Becker and Murphy (1988) propose a model of "rational addiction" in which a decision maker maximizes aggregated future utility whereby the stage utility at any time depends on past consumption. In this flavor, axiomatic characterizations of history-dependent consumer preferences over future consumption paths were developed to account for this effect (e.g., He et al., 2013; Rozen, 2010; Rustichini and Siconolfi, 2014). These models play a crucial role in macroeconomic models as they explain some phenomena and fit data better than standard expected utility theory. For instance, Boldrin et al. (2001) introduce habit persistence into a business cycle model, and Constantinides (1990) uses habit persistence to resolve the equity premium puzzle (cf. Mehra and Prescott, 1985).

A less immediate connection can be made to models of reference-dependent utility (Kőszegi and Rabin, 2006; O'Donoghue and Sprenger, 2018) where the reference point is based on the past (Baucells et al., 2011). However, our decision maker does not derive a reference point based on past choices, but rather obtains (or loses) some utility for making the same choice very often.

There is a strand of literature that characterizes limit evaluations (as players become very patient) by using Markov chains with non-standard preferences. This is done, for instance, by using recursive utilities (see Al-Najjar and Shmaya, 2019, Stanca, 2024, Cerreia-Vioglio et al., 2023, and references therein). Similarly, the question whether or not an asymptotic value in a zero-sum stochastic game exists can be phrased as a question about non-standard preferences: The asymptotic value exists when the limit of the (maxmin) value of average of the rewards is equal to the limit of the (maxmin) discounted value as patience increases. The limit of the averages can be thought of as non-standard time preferences. Sorin (2002) Chapter 5, Ziliotto (2016), and Cerreia-Vioglio et al. (2023) provide further discussion.

The arguably most developed branch of literature dealing with history-dependent preferences focuses on situations with incomplete information. Based on the famous example of Allais (1953) dynamic consistency was challenged, and two major branches of the literature emerged: one focused on behavioral aspects and challenged expected utility as a whole (e.g., Machina, 1989; Thaler, 1981); the other focused on optimizing stage decisions based on one's experience from the past (Gilboa and Schmeidler, 1995). This form of "instance-based learning", which has also found its way into cognitive science (Gonzales et al., 2003; Stewart et al., 2006), asserts a causal connection between past and present behavior rather than dynamic consistency. While this paper investigates precisely such a causal connection, it does so in a setting with complete information. Hence, we refrain from providing an extensive overview of the literature here and refer to Etner et al. (2012).

Outside the scope of economic theory, a similar idea has been brought forward in psychology. The “mere exposure effect” (Zajonc, 1968), also called the “familiarity effect”, describes the change in preferences from simply being exposed to some object. Originally, only positive effects were observed in experiments: an object became more popular as the decision maker was exposed to it more often. But there are scenarios where this effect is reversed (Crisp et al., 2008), or the relation is even non-monotonic: increasing, reaching a satiation point, and decreasing again as exposure increases (Williams, 1987; Zajonc et al., 1972). In particular, research on the interdependences between the mere exposure effect and boredom (Bornstein et al., 1990) or the novelty principle (Liao et al., 2011) has provided a range of stage preferences over objects that depend on past exposure.

3. Preliminaries

Let A be a finite set of *actions* that a decision maker has to choose from at each period $t \in \mathbb{N} \setminus \{0\}$ and let $u : A \to (0, \infty)$ be the decision maker's *basic* payoff function. A *finite history of length* T is a map $\vec{a} : \{1, \dots, T\} \to A$, and an *infinite history* is a map $\vec{a} : \mathbb{N} \setminus \{0\} \to A$. For $T \in \mathbb{N}$ we denote the set of histories of length T by A^T, where A^0 only contains the empty history. The set of all finite histories is denoted by $A^{<\infty}$, that is, $A^{<\infty} = \bigcup_{T=0}^{\infty} A^T$, and the set of all infinite histories is denoted by A^∞. For an infinite history $\vec{a} \in A^\infty$ and a non-negative integer $t \in \mathbb{N} \setminus \{0\}$ we write $\vec{a}_t$ for the t-th element of the sequences, $\vec{a}^t$ for the finite history $\left(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_t\right)$, and also $\vec{a}_0 = \vec{a}^0 = \emptyset$. A *strategy* is a map $\sigma : A^{<\infty} \to A$.

We denote the indicator function by $\mathbb{1}$, that is, for a history $\vec{a}$ we have that $\mathbb{1}_{\vec{a}_s = a} = 1$ if $\vec{a}_s = a$ and $\mathbb{1}_{\vec{a}_s = a} = 0$ otherwise. We define the map $\varphi : A \times A^{<\infty} \to \Delta(A)$ as

$$\varphi\left(a \middle| \vec{a}^t\right) = \begin{cases} \frac{1}{t} \sum_{s=1}^{t} \mathbb{1}_{\vec{a}_s = a}, & \text{if } t \geq 1, \\ 0, & \text{if } t = 0. \end{cases}$$

That is, $\varphi\left(a \middle| \vec{a}^{t-1}\right)$ is the *frequency* of a in the history $\vec{a}^{t-1} = \left(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_{t-1}\right)$.

In the repeated decision problem the decision maker experiences some “fatigue” when choosing the same action repeatedly. More precisely, there is $\gamma \in (0, 1]$ such that when taking action $a \in A$ after history $\vec{a}^{t-1}$, the *stage payoff* at stage t is

$$u_{\gamma,t}(a; \vec{a}^{t-1}) = \left(1 - \gamma \varphi\left(a \middle| \vec{a}^{t-1}\right)\right) u(a_t).$$

A large γ represents strong fatigue or a strong “taste for variety”: the stage payoff quickly declines if an action is chosen repeatedly. If $\gamma = 0$, there is no need for variety, and the maximization of stage payoff and basic payoff are equivalent. We exclude this case.

We are interested in the “maximal” payoff a decision maker can obtain in such a repeated decision problem. Specifically, for an infinite history $\vec{a} \in A^\infty$ the decision maker's *average* (*undiscounted*) *utility* at T is

$$U_\gamma^T\left(\vec{a}\right) = \frac{1}{T} \sum_{t=1}^{T} u_{\gamma,t}(\vec{a}_t; \vec{a}^{t-1}) = \frac{1}{T} \sum_{t=1}^{T} \left(1 - \gamma \varphi\left(\vec{a}_t \middle| a^{t-1}\right)\right) u\left(\vec{a}_t\right).$$

Surely, $U_\gamma^T\left(\vec{a}\right) < \infty$ for all $\vec{a} \in A^\infty$ and all $T \in \mathbb{N} \setminus \{0\}$. Yet, in general, the sequence $\left(U_\gamma^T\left(\vec{a}\right)\right)_{T \in \mathbb{N} \setminus \{0\}}$ will not converge.

Example 3.1. Let $A = \{a, b\}$ with $u(a) = 1$ and $u(b) = 10$. Consider the history $\vec{a}$ that is defined by $\vec{a}_1 = a$, $\vec{a}_2 = b$, $\vec{a}_3 = a$ and

$$\vec{a}_t = \begin{cases} a, & \text{if there is an odd } m \in \mathbb{N} \setminus \{0\} \text{ such that } 3 \cdot 2^m + 1 \leq t \leq 3 \cdot 2^{m+1}, \\ b, & \text{if there is an even } m \in \mathbb{N} \setminus \{0\} \text{ such that } 3 \cdot 2^m + 1 \leq t \leq 3 \cdot 2^{m+1}, \end{cases}$$

for $t \geq 4$. That is, $\vec{a} = (a, b, a, b, b, b, a, a, a, a, a, a, b, \dots)$. In this sequence, exponentially increasing blocks of consecutive a's and b's are played alternating. In particular, from $t \geq 4$ onwards each block is as long as the entire history before the block, so that the frequency of either action fluctuates between $1/3$ at the beginning of each block and $2/3$ at the end. The sequence of average utilities of this infinite history does not converge. Intuitively, it will be lowest at the end of any a-block, and highest at the end of any b-block. We shall have a closer look at this behavior later. □

As the sequence $\left(U_\gamma^T\left(\vec{a}\right)\right)_{T \in \mathbb{N} \setminus \{0\}}$ might not converge for all $\vec{a} \in A^\infty$, there is no “obvious” way to compare two infinite histories $\vec{a}, \vec{b} \in A^\infty$. Yet, as every sequence of average utility is bounded, we can use their upper and lower accumulation points for comparisons. To keep notation short, define for any $\vec{a} \in A^\infty$

$$\overline{V}_\gamma\left(\vec{a}\right) = \limsup_{T \to \infty} U_\gamma^T\left(\vec{a}\right) \qquad \text{and} \qquad \underline{V}_\gamma\left(\vec{a}\right) = \liminf_{T \to \infty} U_\gamma^T\left(\vec{a}\right),$$

which are the highest and lowest accumulation points that the sequence of average utilities can reach for the history $\vec{a}$. We define correspondingly

$$\overline{V}_\gamma = \sup\left\{\overline{V}_\gamma\left(\vec{a}\right) \middle| \vec{a} \in A^\infty\right\} \qquad \text{and} \qquad \underline{V}_\gamma = \sup\left\{\underline{V}_\gamma\left(\vec{a}\right) \middle| \vec{a} \in A^\infty\right\}. \tag{1}$$

While $\overline{V}_\gamma$ and $\underline{V}_\gamma$ may initially seem to have similar importance, the latter is, in fact, much more significant from a behavioral perspective. An infinite history that (approximately) achieves $\underline{V}_\gamma$ ensures that, from a certain point onward, the average payoffs are at least close to $\underline{V}_\gamma$. In contrast, an infinite history that (approximately[2]) achieves $\overline{V}_\gamma$ only guarantees that, intermittently and over a sparse set of periods, the average payoffs approach $\overline{V}_\gamma$. To enhance intuition, suppose that $v < \underline{V}_\gamma$. In this case, there exists a history $\vec{a}$ in which the average utility is at least v for all but finitely many periods. However, if $v < \overline{V}_\gamma$, then there exists a history $\vec{a}$ where the average utility is at least v over an infinite, potentially sparse set of periods.

The previous evaluations pertain to infinite histories, while the following evaluation applies to finite histories. For every $T \geq 1$ define

$$v_\gamma^T = \max_{\vec{a} \in A^\infty} U_\gamma^T(\vec{a}). \tag{2}$$

That is, v_γ^T denotes the maximal average payoff that can be obtained from a history of length T. While $\overline{V}_\gamma$ is not important in its own right, we introduce it because, as shown in Proposition 6.2, the limit of v_γ^T as $T \to \infty$ is equal to $\overline{V}_\gamma$.

Example 3.2. Recall the history $\vec{a}$ from Example 3.1. As t gets large, the average frequency of the action that is played in a block is approximated by

$$x = \int_0^1 \left(1 - \frac{2}{3s+3}\right) ds = 1 - \left(\frac{2}{3}\ln(2) - \ln(1)\right) = 1 - \frac{2}{3}\ln(2). \tag{3}$$

Thus, even though the frequencies of a and b do not converge, the average of $\varphi\left(a \middle| a^{t-1}\right)$ taken over all t with $\vec{a}_t = a$ converges towards x, and the same is true for the average of $\varphi\left(b \middle| a^{t-1}\right)$ taken over all t with $\vec{a}_t = b$. Hence, at the end of any block of a's, the average payoff is approximately

$$\begin{aligned} U_\gamma^T(\vec{a}) &\approx \frac{2}{3}\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right)u(a) + \frac{1}{3}\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right)u(b) \\ &= 4\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right). \end{aligned}$$

Observe that for such T it holds that

$$u^T(a;\vec{a}) = \left(1 - \frac{2}{3}\gamma\right) \leq U_\gamma^T(\vec{a}) \qquad \text{and} \qquad u^{T+1}(b;\vec{a}) = 10\left(1 - \frac{1}{3}\gamma\right) \geq U_\gamma^T(\vec{a}),$$

for all $\gamma \in [0,1]$. Thus, $U_\gamma^T(\vec{a})$ is minimized at the end of each a-block, and we find $\underline{V}_\gamma(\vec{a}) = 4\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right)$.

In order to find $\overline{V}_\gamma(\vec{a})$ we show that $U_\gamma^T(\vec{a})$ achieves its maxima always at the end of b-blocks. So, consider a (large) b-block. We want to show that the average utility of $\vec{a}$ is increasing throughout the entire block. So, keeping in mind that the block is large, let $x \in [0,1]$ and consider the period after a fraction x of the block has passed. The frequencies of a and b at this point in time are given by $f_a \approx \frac{2}{3x+3}$ and $f_b = 1 - f_a \approx \frac{3x+1}{3x+3}$. Hence, the stage utility is given by

$$v(x) = 10\left(1 - \gamma f_b\right) \approx \frac{30(1-\gamma)x + 30 - 10\gamma}{3x+3}.$$

The average frequency of a at x (taken over the periods where a has been chosen) is still given in (3). The average frequency of b at $x > 0$ is given by

$$\frac{1}{x}\int_0^x 1 - \frac{2}{3s+3} ds = 1 - \frac{1}{x}\left(\frac{2}{3}\ln(3x+3) - \frac{2}{3}\ln(3)\right) = 1 - \frac{2}{3x}\ln(x+1).$$

Thus, the average utility at x is approximated by

$$\begin{aligned} U(x) &= f_a\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right)u(a) + f_b\left(1 - \gamma\left(1 - \frac{2}{3x}\ln(x+1)\right)\right)u(b) \\ &= \frac{2}{3x+3}\left(1 - \gamma\left(1 - \frac{2}{3}\ln(2)\right)\right) + \frac{3x+1}{3x+3}\left(1 - \gamma\left(1 - \frac{2}{3x}\ln(x+1)\right)\right)10. \end{aligned}$$

In particular, $U(1) < v(1)$ for all $\gamma \in [0,1]$. As U is increasing at x if and only if $v(x) > U(x)$, and v is falling in x, this implies that U reaches its maximum at $x = 1$. Thus, we obtain

[2] We show later that, even though $\underline{V}_\gamma$ and $\overline{V}_\gamma$ are non-continuous functions defined on a compact set, the supremum in the definitions of $\underline{V}_\gamma$ and $\overline{V}_\gamma$ can be replaced by maximum.

$$\overline{V}_\gamma\left(\vec{a}\right) = U(1) = 7\left(1-\gamma\left(1-\frac{2}{3}\ln(2)\right)\right)$$

for the highest limit point that $U_\gamma^T\left(\vec{a}\right)$ can reach. □

4. Greedy behavior and stationary strategies

A simple strategy σ that a decision maker might follow is to maximize her stage utility at each t, that is, choose her action at t according to

$$\vec{a}_t = \sigma\left(\vec{a}^{t-1}\right) \in \arg\max_{a\in A}\left(1-\gamma\varphi\left(a\middle|\vec{a}^{t-1}\right)\right)u(a).$$

We call such a strategy a *greedy strategy*. In this case the frequency $\varphi\left(a\middle|\vec{a}^t\right)$ necessarily converges for all $a\in A$.

Proposition 4.1. *Let $\vec{a}\in A^\infty$ be the history evolving from a greedy strategy. Then $\varphi\left(a\middle|\vec{a}^t\right)$ converges for all $a\in A$ and*

$$\lim_{t\to\infty}\varphi\left(a\middle|\vec{a}^t\right) = \frac{\gamma - \left|A^*\right| + \sum_{b\in A^*}\frac{u(a)}{u(b)}}{\gamma\sum_{b\in A^*}\frac{u(a)}{u(b)}} \tag{4}$$

for all $a\in A^*$, where A^* is the set of actions that are chosen infinitely often.

Proof. For each $\varepsilon>0$ there is $T\in\mathbb{N}\setminus\{0\}$ such that

$$\left|\left(1-\gamma\varphi\left(a\middle|\vec{a}^{t-1}\right)\right)u(a) - \left(1-\gamma\varphi\left(b\middle|\vec{a}^{t-1}\right)\right)u(b)\right| < \varepsilon$$

for all $a,b\in A^*$ and all $t\geq T$. As $\sum_{a\in A^*}\varphi\left(a\middle|\vec{a}^t\right)=1$ for all $t\geq 1$, the frequencies converge. Let $f_a=\lim_{t\to\infty}\varphi\left(a\middle|\vec{a}^t\right)$. Then $\left(1-\gamma f_a\right)u(a) = \left(1-\gamma f_b\right)u(b)$ for all $a,b\in A^*$. Solving for b and summing over all b we find that

$$1=\sum_{b\in A^*}f_b = \frac{1}{\gamma}\sum_{b\in A^*}\left(1-\frac{u(a)}{u(b)}\left(1-\gamma f_a\right)\right) = \frac{1}{\gamma}\left(\left|A^*\right| - \left(1-\gamma f_a\right)\sum_{b\in A^*}\frac{u(a)}{u(b)}\right)$$

Solving for f_a delivers (4). □

The expression in (4) provides a bound on the number of actions that can be played with positive probability. In particular, for $\gamma<1$ it is possible that the greedy strategy will only choose a single action that is played at every t.

As seen in Example 3.1, frequencies do not converge for all $\vec{a}\in A^\infty$. Yet, if they do, as for the greedy strategy above, the average utility converges as well. We say that a history $\vec{a}\in A^\infty$ is *stationary* if $\lim_{t\to\infty}\varphi\left(a\middle|\vec{a}^{t-1}\right)$ exists for all $a\in A$. In this case we write $\varphi\left(a\middle|\vec{a}\right)=\lim_{t\to\infty}\varphi\left(a\middle|\vec{a}^{t-1}\right)$. If there is no risk of confusion, we will even write $\varphi(a)=\varphi\left(a\middle|\vec{a}\right)$. The limit of the average utilities is then given by

$$\overline{V}_\gamma\left(\vec{a}\right) = \underline{V}_\gamma\left(\vec{a}\right) = \lim_{T\to\infty}U_\gamma^T\left(\vec{a}\right) = \sum_{a\in A}\varphi(a)\left(1-\gamma\varphi(a)\right)u(a). \tag{5}$$

We denote the optimal limit that can be achieved by any stationary history by

$$V_\gamma^* = \sup\left\{\underline{V}_\gamma\left(\vec{a}\right) \mid \vec{a}\in A^\infty \text{ is stationary}\right\}.$$

Finally, we say that a strategy is *stationary* if it generates a stationary history.

Example 4.2. Let $A=\{a,b\}$ with $u(a)=1$ and $u(b)=10$. If $\gamma\leq 0.9$, the greedy strategy will choose b for all t. If $\gamma>0.9$, then the frequencies achieved by the greedy strategy are $\varphi(a)=\frac{10\gamma-9}{11\gamma}$ and $\varphi(b)=\frac{\gamma+9}{11\gamma}$. Thus,

$$\underline{V}_\gamma\left(\vec{a}\right) = \frac{10\gamma-9}{11\gamma}\left(1-\gamma\frac{10\gamma-9}{11\gamma}\right)u(a) + \frac{\gamma+9}{11\gamma}\left(1-\gamma\frac{\gamma+9}{11\gamma}\right)u(b) = \frac{20-10\gamma}{11}.$$

In particular, for $\gamma=0.9$, only b will be chosen and its stage payoff converges towards 1. □

The previous example illustrates that the greedy strategy does not deliver particularly high payoffs. Indeed, the "good" actions are overused so that their stage payoffs become very low, resulting in a low average payoff. Finding V_γ^* is indeed not very difficult; by (5), it is given by

$$V_\gamma^* = \max_{x\in\Delta(A)} \sum_{a\in A} x_a \left(1-\gamma x_a\right) u(a), \tag{6}$$

where $\Delta(A)$ denotes the set of probability measures over A. As the objective function is strictly quasi-concave for all $\gamma > 0$, the maximization problem in (6) has a unique solution $x^* \in \Delta(A)$. In particular, every stationary history $\vec{a}$ with $\varphi\left(\cdot|\vec{a}\right) = x^*$ is optimal. The next proposition specifies these optimal frequencies.

Proposition 4.3. *Let $\vec{a} \in A^\infty$ be the history evolving from an optimal stationary strategy. Then*

$$\varphi(a) = \frac{2\gamma - |A^*| + \sum_{b\in A^*} \frac{u(a)}{u(b)}}{2\gamma \sum_{b\in A^*} \frac{u(a)}{u(b)}} \tag{7}$$

for all $a \in A^*$, where $A^* \subseteq A$ is the largest subset of A with $u(a) \le \frac{2\gamma - |A^*|}{\sum_{b\in A^*} \frac{1}{u(b)}}$ and $u(a^*) \ge u(a)$ for all $a \in A \setminus A^*$.

Proof. The first-order conditions of the maximization problem in (6) are

$$(1-2\gamma\varphi(a))u(a) = (1-2\gamma\varphi(b))u(b).$$

for all $a, b \in A^*$. With the same steps as in the proof of Proposition 4.1 one obtains (7) for all a with $\varphi(a) > 0$. DK(2): Optimality requires that if A^* contains some action a, it also contains all actions with higher basic payoff. This proves the second inequality. The best action $a \in A \setminus A^*$ must satisfy $(1-2\gamma\varphi(a^*))u(a^*) \ge u(a)$ for all $a^* \in A^*$, as otherwise optimality would require $a \in A^*$ as well, at least with some small frequency. Substitution $\varphi(a)$ by the right hand side of (7) completes the proof. □

Example 4.4. Let $A = \{a, b\}$ with $u(a) = 1$ and $u(b) = 10$. Let $\vec{a}$ be the history evolving from an optimal stationary strategy. For $\gamma \le \frac{9}{20}$ action a will not be played with positive probability. For $\gamma > \frac{9}{20}$, the optimal frequencies are $\varphi(a) = \frac{20\gamma-9}{22\gamma}$ and $\varphi(b) = \frac{2\gamma+9}{22\gamma}$. Thus,

$$\underline{V}_\gamma\left(\vec{a}\right) = \frac{20\gamma-9}{22\gamma}\left(1-\gamma\frac{20\gamma-9}{22\gamma}\right)u(a) + \frac{2\gamma+9}{22\gamma}\left(1-\gamma\frac{2\gamma+9}{22\gamma}\right)u(b) = \frac{-40\gamma^2+80\gamma+81}{44\gamma}.$$

In particular, this expression is strictly larger than the average utility of the greedy strategy in Example 4.2. □

A special case of optimal stationary histories emerges if A contains exactly two elements and $\gamma = 1$. In this case Proposition 4.3 immediately implies the following corollary.

Corollary 4.5. *Let $\gamma = 1$, let $A = \{a, b\}$, and let $\vec{a} \in A^\infty$ be an optimal stationary history. Then $\varphi(a) = \varphi(b) = \frac{1}{2}$. In particular, $V_1^* = \frac{1}{4}(u(a) + u(b))$.*

At this point it has become clear that defining what an optimal strategy is crucially depends on how the evolving histories are evaluated. Finding optimal stationary strategies is rather simple, as shown in Proposition 4.3, yet Examples 3.2 and 4.4 illustrate that stationary strategies might not be able to achieve $\overline{V}_\gamma$ as an average utility. Indeed, for $\gamma = 1$, the strategy in Example 3.2 achieves an average utility of $\frac{14}{3}\ln(2) \approx 3.23$, while the best stationary strategy in Example 4.4 achieves only $\frac{11}{4} = 2.75$.

In the remainder of the paper we shall investigate how the three possible values, that is, the optimal highest accumulation point, the optimal lowest accumulation point, and the optimal limit (if it exists) compare. They must satisfy

$$V_\gamma^* \le \underline{V}_\gamma \le \overline{V}_\gamma.$$

Our main results will be that here the first inequality is actually an equality, while the second inequality is strict if there are at least two actions $a, b \in A$ with $u(a) \neq u(b)$ that are chosen with positive frequency in an optimal stationary history.

5. Stationary strategies achieve $\underline{V}_\gamma$

Stationary strategies in dynamic problems offer several advantages that make them appealing. The first advantage is their simplicity in implementation and analysis compared to non-stationary strategies, which may require tracking the entire history.

Suppose that $x^* \in \Delta(A)$ is such that

$$V_\gamma^* = \sum_{a\in A} x_a^*\left(1-\gamma x_a^*\right)u(a).$$

It is relatively straightforward to construct an infinite history that asymptotically follows the distribution x^*. This can be achieved, for instance, by selecting an action at each period randomly according to the distribution x^*, independently of the history. The strong law of large numbers guarantees that, with probability 1, the frequency of actions will converge to x^*, and thereby support V^*_γ. If x^* consists of rational numbers, an infinite history that follows the distribution x^* can be achieved through periodic behavior, such as having pizza every Sunday evening. This illustrates why people often exhibit periodic behavior in their routines.

For many types of Markov Decision Processes, including those with discounted rewards and a compact set of actions, stationary strategies are proven to be optimal. This implies that more complex strategies are unnecessary to achieve the best possible outcome. Furthermore, the simplicity of stationary strategies allows for an efficient computation, as algorithms designed to find optimal strategies can often do so more quickly when only stationary strategies are considered. The existence of optimal stationary strategies also simplifies the theoretical analysis of MDPs, allowing us to focus on a smaller class of strategies when proving specific properties, as demonstrated in Subsection 5.2.

In our case, finding a strategy such that the evolving history $\vec{a}$ achieves $\underline{V}_\gamma(\vec{a}) = \underline{V}_\gamma$ essentially constitutes a dynamic programming problem on a countable state space. Unfortunately, these problems typically lack a tractable structure, so there are no general results applicable in the current context. Therefore, we will develop specific tools in Subsection 5.1 to obtain our result. In Subsection 5.2, we will then investigate how the optimal frequencies change as the parameter γ varies.

5.1. The optimality of V^*_γ

Let $\vec{a}$ be a history. For any $t_1, t_2 \in \mathbb{N} \setminus \{0\}$ with $t_2 > t_1$ let the *block* from t_1 to t_2 in $\vec{a}$ be the sequence of actions $\left(\vec{a}_{t_1+1}, \vec{a}_{t_1+2}, \dots, \vec{a}_{t_2}\right)$. The average utility within this block is given by

$$W_\gamma = W_\gamma\left(\vec{a}, t_1, t_2\right) = \frac{1}{t_2 - t_1} \sum_{s=t_1+1}^{t_2} \left(1 - \gamma\varphi\left(a_s \mid \vec{a}^{s-1}\right)\right) u(a_s). \tag{8}$$

Let $p(a)$ be the frequency with which a is played in the block, that is,

$$p(a) = p\left(a; \vec{a}, t_1, t_2\right) = \frac{1}{t_2 - t_1} \sum_{s=t_1+1}^{t_2} \mathbb{1}_{\vec{a}_s = a}. \tag{9}$$

If such a block is "not too long", the frequencies will not change much between t_1 and t_2. We want to use this observation to derive an approximation of W_γ by means of p and the frequency at the beginning, that is, $\varphi\left(\cdot \mid \vec{a}^{t_1}\right)$. In particular, we define $\widetilde{U}_\gamma$ as

$$\widetilde{U}_\gamma = \widetilde{U}_\gamma\left(\vec{a}, t_1, t_2\right) = \sum_{a \in A} p(a)\left(1 - \gamma\varphi\left(a|\vec{a}^{t_1}\right)\right) u(a). \tag{10}$$

We show that $\widetilde{U}_\gamma$ is close to W_γ if t_2 is relatively close to t_1, that is, if $\frac{t_2 - t_1}{t_1}$ is small, the proof of the following lemma can be found in the appendix.

Lemma 5.1. *Let $\vec{a} \in A^\infty$ be a history and let $t_1, t_2 \in \mathbb{N} \setminus \{0\}$ with $t_2 > t_1$. Then*

$$\left|W_\gamma - \widetilde{U}_\gamma\right| \le 2\frac{t_2 - t_1}{t_1}\gamma \sum_{a \in A} u(a). \tag{11}$$

With Lemma 5.1 we can now prove our first main result, namely that there is a stationary strategy such that the evolving history $\vec{a}$ satisfies $\underline{V}_\gamma = \underline{V}_\gamma\left(\vec{a}\right) = V^*_\gamma$. The idea of the proof is to suppose by contradiction that there is a non-stationary history $\vec{a}$ with $\underline{V}\left(\vec{a}\right) = V^*_\gamma + 4c$ for some strictly positive constant c. This infinite history is divided into blocks, whose lengths depend on a parameter α that will be adjusted to c and γ. The lengths of the blocks are designed to increase over time while remaining sufficiently short so that, relative to the total history up to that point, each block's length remains small. This structure enables the application of Lemma 5.1.

To be more specific, let $\varphi^k(a) = \varphi\left(a \mid \vec{a}^{t_k}\right)$ denote the frequency of a at the beginning of the k-th block. For each block k, consider the quantity

$$x_k = \sum_{a \in A} \left(1 - \varphi^k(a)\right)^2 u(a).$$

We will show that these numbers exhibit impossible behavior if $\underline{V}_\gamma$ were bounded away from V^*_γ. To this end, we define a potential function $H(K, \alpha)$ as the weighted average of $x_1, \dots, x_K$, where the weight of x_k corresponds to the relative length of the k-th block within the first K blocks. This step lies at the heart of the proof.

The main idea behind the definition of $H(K, \alpha)$ is to derive a recursive formula, given by (20), that relates $H(K, \alpha)$ and $H(K-1, \alpha)$. Since $H(K, \alpha)$ converges as $K \to \infty$, the gap $H(K, \alpha) - H(K-1, \alpha)$ diminishes. However, if $\underline{V}_\gamma$ were bounded away from V^*_γ, this gap would remain strictly negative, leading to a contradiction, as shown in (23).

Theorem 5.2. *It holds that $\underline{V}_\gamma = V^*_\gamma$.*

Proof. Assume, by contradiction, that $\underline{V}_\gamma > V_\gamma^*$. Then there is $\vec{a} \in A^\infty$ such that $\underline{V}_\gamma\left(\vec{a}\right) = V_\gamma^* + 4c$ for some constant $c > 0$. Thus,

$$U_\gamma^T\left(\vec{a}\right) \geq V_\gamma^* + 3c \tag{12}$$

for all sufficiently large T. Let T_1 be such that (12) holds for all $T \geq T_1$.

Let $\alpha \in (0,1)$. We divide the set of periods into blocks. To that end let $t_0 = 0$, and for each integer $k \geq 1$ the let t_k be defined by $t_k = \left\lceil (1+\alpha)^{k-1} T_1 \right\rceil$, which is the smallest integer larger than or equal to $(1+\alpha)^{k-1}T_1$. The *k-th block* starts at $t_{k-1}+1$ and ends at t_k. For each block k, denote the average payoff, the frequency, and the approximation by $W_\gamma^k = W_\gamma\left(\vec{a}, t_{k-1}, t_k\right)$, $p^k(a) = p\left(a; \vec{a}, t_{k-1}, t_k\right)$, and $\widetilde{U}_\gamma^k = \widetilde{U}_\gamma\left(\vec{a}, t_{k-1}, t_k\right)$, respectively, as in Equations (8), (9), and (10). In particular, $W_\gamma^1 = U_\gamma^{T_1}$.

By construction, $\frac{t_{k+1}-t_k}{t_k} \leq \alpha + \frac{1}{(1+\alpha)^{k-1}T_1}$ for all $k \geq 1$. Thus, by Lemma 5.1,

$$\left| W_\gamma^k - \widetilde{U}_\gamma^k \right| \leq 2\left(\alpha + \frac{1}{(1+\alpha)^{k-2}T_1}\right)\gamma \sum_{a\in A} u(a),$$

for every $k \geq 2$. Denote $\beta = \frac{\alpha}{1+\alpha}$ and also $d^k(a) = p^k(a) - \varphi^k(a)$ for each $k \in \mathbb{N}\setminus\{0\}$ and $a \in A$. Recall from (12) and the definition of T_1 that

$$V_\gamma^* + 3c \leq U_\gamma^{t_K}\left(\vec{a}\right) \tag{13}$$

for all $K \geq 1$. In particular,

$$\begin{aligned}
V_\gamma^* + 3c &\leq U_\gamma^{t_K}\left(\vec{a}\right) \\
&= \frac{t_1}{t_K} W_\gamma^1 + \sum_{k=2}^{K} \frac{t_k - t_{k-1}}{t_K} W_\gamma^k \\
&= \frac{T_1}{\left\lceil (1+\alpha)^{K-1}T_1\right\rceil} W_\gamma^1 + \sum_{k=2}^{K} \frac{\left\lceil (1+\alpha)^{k-1}T_1\right\rceil - \left\lceil (1+\alpha)^{k-2}T_1\right\rceil}{\left\lceil (1+\alpha)^{K-1}T_1\right\rceil} W_\gamma^k \\
&\leq \frac{T_1}{(1+\alpha)^{K-1}T_1} W_\gamma^1 + \sum_{k=2}^{K} \frac{\left\lceil (1+\alpha)^{k-2}T_1 + \alpha(1+\alpha)^{k-2}T_1\right\rceil - \left\lceil (1+\alpha)^{k-2}T_1\right\rceil}{\left\lceil (1+\alpha)^{K-1}T_1\right\rceil} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\left\lceil (1+\alpha)^{k-2}T_1\right\rceil + \left\lceil \alpha(1+\alpha)^{k-2}T_1\right\rceil - \left\lceil (1+\alpha)^{k-2}T_1\right\rceil}{\left\lceil (1+\alpha)^{K-1}T_1\right\rceil} W_\gamma^k \\
&= \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\left\lceil \alpha(1+\alpha)^{k-2}T_1\right\rceil}{\left\lceil (1+\alpha)^{K-1}T_1\right\rceil} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\left\lceil \alpha(1+\alpha)^{k-2}\right\rceil \cdot T_1}{(1+\alpha)^{K-1}T_1} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\alpha(1+\alpha)^{k-2}+1}{(1+\alpha)^{K-1}} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\alpha(1+\alpha)^{k-2}+1}{(1+\alpha)^{K-1}} \left(\tilde{U}_\gamma^k + 2\left(\alpha + \frac{1}{(1+\alpha)^{k-2}T_1}\right)\gamma\sum_{a\in A}u(a)\right) \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}} \widetilde{U}_\gamma^k \\
&\qquad + 2\gamma\sum_{a\in A}u(a)\sum_{k=2}^{K}\frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}}\left(\alpha + \frac{1}{(1+\alpha)^{k-2}}\right),
\end{aligned} \tag{14}$$

for every $K \geq 2$.

Let $\bar{\alpha} = c\left(4\gamma\sum_{a\in A}u(a)\right)^{-1}$. Then, for every $0 < \alpha < \bar{\alpha}$ there is $K(\alpha)$ such that for all $K \geq K(\alpha)$ it holds that

$$2\gamma\sum_{a\in A}u(a)\sum_{k=2}^{K}\frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}}\left(\alpha + \frac{1}{(1+\alpha)^{k-2}}\right) < c.$$

This is so because

$$\sum_{k=2}^{K}\frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}}\left(\alpha + \frac{1}{(1+\alpha)^{k-2}}\right)\xrightarrow[K\to\infty]{}\alpha.$$

By (14) we, hence, have

$$V_\gamma^* + 2c \le W_\gamma^1 + \sum_{k=2}^{K} \frac{(1+\alpha)^{k-2}\alpha + 1}{(1+\alpha)^{K-1}} \widetilde{U}_\gamma^k. \tag{15}$$

For every $x, y \in \mathbb{R}^A$, denote

$$\langle x, y\rangle := \sum_{a\in A} x_a y_a u(a) \qquad \text{and} \qquad \|x\|^2 := \langle x, x\rangle. \tag{16}$$

For each k and $a \in A$ define $d^k(a) = p^k(a) - \varphi^k(a)$, and define $\beta = \frac{\alpha}{1+\alpha}$. Since $\varphi^k(\cdot)$ is a convex combination of $\varphi^{k-1}(\cdot)$ and $p^{k-1}(\cdot)$, with weights $\frac{t_{k-1}}{t_k}$ and $\frac{t_k - t_{k-1}}{t_k}$, we have

$$\begin{aligned}
\varphi^k(a) - \varphi^{k-1}(a) &= \frac{t_{k-1}}{t_k}\varphi^{k-1}(a) + \frac{t_k - t_{k-1}}{t_k} p^{k-1}(a) - \varphi^{k-1}(a) \\
&= \frac{t_k - t_{k-1}}{t_k}\left(p^{k-1}(a) - \varphi^{k-1}(a)\right) \\
&\ge \frac{(1+\alpha)^k - (1+\alpha)^{k-1} - 1}{(1+\alpha)^k} d^k(a) \\
&\ge \beta d^k(a) - \frac{1}{(1+\alpha)^k}.
\end{aligned} \tag{17}$$

Let $p^k = \left(p^k(a)\right)_a$, $\varphi^k = \left(\varphi^k(a)\right)_a$ and $d^k = \left(d^k(a)\right)_a$, so that $p^k, \varphi^k, d^k \in \mathbb{R}^A$. Moreover, denote by $\mathbf{1} \in \mathbb{R}^A$ the vector with 1 in each entry. By (6)

$$V_\gamma^* = \sup_{x\in\Delta(A)} \langle x, \mathbf{1} - \gamma x\rangle \ge \left\langle \varphi^{t_k}, \mathbf{1} - \gamma\varphi^{t_k}\right\rangle$$

for all k. Since $\widetilde{U}_\gamma^k = \left\langle p^k, \mathbf{1} - \gamma\varphi^k\right\rangle$, we obtain from the definition of d^k and Inequality (15) that for all $0 < \alpha < \bar{\alpha}$ and $K \ge K(\alpha)$

$$\begin{aligned}
V_\gamma^* + 2c &\le \sum_{k=1}^{K} \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle p^k, \mathbf{1} - \gamma\varphi^k\rangle \\
&= \sum_{k=1}^{K} \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle \varphi^{t_k}, \mathbf{1} - \gamma\varphi^k\rangle + \sum_{k=1}^{K} \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle \\
&\le V_\gamma^* + \sum_{k=1}^{K} \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle.
\end{aligned}$$

Hence,

$$\begin{aligned}
2c &\le \sum_{k=1}^{K} \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle = \sum_{k=1}^{K} \frac{\alpha}{(1+\alpha)^{K-k+1}}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle \\
&= \frac{1}{1+\alpha}\sum_{k=1}^{K} \frac{\alpha}{(1+\alpha)^{K-k}}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle.
\end{aligned} \tag{18}$$

For any K and α define

$$H(K, \alpha) = \sum_{k=1}^{K} \frac{\alpha}{(1+\alpha)^{K-k+1}} \left\|\mathbf{1} - \gamma\varphi^k\right\|^2. \tag{19}$$

Note that $\sum_{k=1}^{K} \frac{\alpha}{(1+\alpha)^{K-k+1}} \le 1$, so that $H(K, \alpha)$ is bounded by some weighted average of $\left\|1 - \gamma\varphi^k\right\|^2$ where $k = 1, 2, ..., K$. Furthermore,

$$\begin{aligned}
H(K, \alpha) &= \sum_{k=1}^{K-1} \frac{\alpha}{(1+\alpha)^{K-k+1}} \left\|\mathbf{1} - \gamma\varphi^k\right\|^2 + \frac{\alpha}{(1+\alpha)}\left\|\mathbf{1} - \gamma\varphi^K\right\|^2 \\
&= (1-\beta)\, H(K-1, \alpha) + \beta\left\|\mathbf{1} - \gamma\varphi^K\right\|^2.
\end{aligned}$$

Therefore,

$$H(K, \alpha) - H(K-1, \alpha) = -\beta H(K-1, \alpha) + \beta\left\|\mathbf{1} - \gamma\varphi^K\right\|^2. \tag{20}$$

Equation (20) is crucial, as it establishes a connection between $H(K,\alpha)$ and $H(K-1,\alpha)$. The remainder of the proof focuses on the gap $H(K,\alpha)-H(K-1,\alpha)$. On one hand, this gap should diminish, but on the other hand, due to (13), it must remain strictly negative, leading to a contradiction.

Define

$$\varepsilon_{K,\alpha} = \frac{\alpha}{(1+\alpha)^K}\left\|\mathbf{1}-\gamma\varphi^1\right\|^2.$$

Then

$$\begin{aligned}
H(K,\alpha) &= \varepsilon_{K,\alpha} + \sum_{k=2}^{K}\frac{\alpha}{(1+\alpha)^{K-k+1}}\left\|\mathbf{1}-\gamma\varphi^k\right\|^2 \\
&\le \varepsilon_{K,\alpha} + \sum_{k=2}^{K}\frac{\alpha}{(1+\alpha)^{K-k+1}}\left\|\mathbf{1}-\gamma\varphi^{k-1}-\gamma\beta d^{k-1}\right\|^2 + \sum_{k=2}^{K}\frac{\alpha}{(1+\alpha)^{K-k+1}}\frac{1}{(1+\alpha)^k} \\
&= \varepsilon_{K,\alpha} + \sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|\mathbf{1}-\gamma\varphi^k-\gamma\beta d^k\right\|^2 + \frac{(K-2)\alpha}{(1+\alpha)^{K+1}} \\
&= \varepsilon_{K,\alpha} + \sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left(\left\|\mathbf{1}-\gamma\varphi^k\right\|^2 - 2\gamma\beta\left\langle d^k,\mathbf{1}-\gamma\varphi^k\right\rangle + \gamma^2\beta^2\left\|d^k\right\|^2\right) + \frac{(K-2)\alpha}{(1+\alpha)^{K+1}} \\
&= \varepsilon_{K,\alpha} + H(K-1,\alpha) - 2\gamma\beta\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\langle d^k,\mathbf{1}-\gamma\varphi^k\right\rangle + \gamma^2\beta^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 \\
&\qquad + \frac{(K-2)\alpha}{(1+\alpha)^{K+1}}.
\end{aligned}$$

Thus, we obtain an upper bound on the gap $H(K,\alpha)-H(K-1,\alpha)$, as follows.

$$\begin{aligned}
H(K,\alpha)-H(K-1,\alpha) \le \varepsilon_{K,\alpha} &- 2\beta\gamma\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\langle d^k,\mathbf{1}-\gamma\varphi^k\rangle \\
&+ \beta^2\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\|d^k\|^2 + \frac{(K-2)\alpha}{(1+\alpha)^{K+1}}
\end{aligned}$$

and together with (18) and (20), we get

$$\begin{aligned}
\beta H(K-1,\alpha) - \beta\left\|\mathbf{1}-\gamma\varphi^{t_K}\right\|^2 &\ge -\varepsilon_{K,\alpha} + 2\beta\gamma\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\langle d^k,\mathbf{1}-\gamma\varphi^k\right\rangle \\
&\qquad - \beta^2\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 - \frac{(K-2)\alpha}{(1+\alpha)^{K+1}} \\
&> -\varepsilon_{K,\alpha} + 4\beta\gamma(1+\alpha)c \\
&\qquad - \beta^2\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 - \frac{(K-2)\alpha}{(1+\alpha)^{K+1}},
\end{aligned}$$

or equivalently,

$$\begin{aligned}
H(K-1,\alpha) - \left\|\mathbf{1}-\gamma\varphi^K\right\|^2 &> -\frac{\varepsilon_{K,\alpha}}{\beta} + 4\gamma(1+\alpha)c - \beta\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 - \frac{(K-2)\alpha}{\beta(1+\alpha)^{K+1}} \\
&= 4\gamma(1+\alpha)c - \frac{\alpha}{1+\alpha}\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 - \left(\frac{\varepsilon_{K,\alpha}}{\beta} + \frac{(K-2)}{(1+\alpha)^K}\right). \qquad (21)
\end{aligned}$$

We now proceed to show that the right-hand side of (21) is bounded away from zero.

Since $\left\|d^k\right\|^2$ are all uniformly bounded, the sum on the right-hand side is bounded. Thus,

$$\begin{aligned}
\frac{\alpha}{1+\alpha}\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\left\|d^k\right\|^2 &\le \frac{\alpha}{1+\alpha}\gamma^2\sup_{k\in\mathbb{N}}\left\|d^k\right\|^2\sum_{k=0}^{\infty}\frac{\alpha}{(1+\alpha)^k} \\
&= \frac{\alpha}{1+\alpha}\gamma^2\sup_{k\in\mathbb{N}}\left\|d^k\right\|^2
\end{aligned}$$

$$< c\gamma$$

for all K and all sufficiently small $\alpha > 0$. Moreover, there are α^* and $K^* \geq K(\alpha^*)$ such that for all $K \geq K^*$

$$\frac{\varepsilon_{K,\alpha^*}}{\beta^*} + \frac{(K-2)}{(1+\alpha^*)^K} = \frac{1}{(1+\alpha^*)^{K-1}} \left\| \mathbf{1} - \gamma\varphi^1 \right\|^2 + \frac{(K-2)}{(1+\alpha^*)^K} < c\gamma,$$

where $\beta^* = \frac{\alpha^*}{1+\alpha^*}$. This implies that

$$H(K-1,\alpha) - \left\| \mathbf{1} - \gamma\varphi^K \right\|^2 > 4\gamma(1+\alpha)c - \gamma c - \gamma c > 2\gamma c \tag{22}$$

for all $\alpha \leq \alpha^*$ and all $K \geq K^*$.

We are now ready to obtain a contradiction. Due to (20), for $\alpha \leq \alpha^*$ we get from (22),

$$\begin{aligned} 0 &= \limsup_{K\to\infty}(H(K,\alpha) - H(K-1,\alpha)) \\ &= \limsup_{K\to\infty} \beta(\left\| \mathbf{1} - \gamma\varphi^K \right\|^2 - H(K-1,\alpha)) \leq -2\beta\gamma c < 0 \end{aligned} \tag{23}$$

and have reached a contradiction. □

Observe that from Equation (19) it is not obvious that H should be called a "potential function". However, as can be seen from (20) and (22), H is decreasing in K for sufficiently large K and sufficiently small α, which makes it a potential function in the original sense.

5.2. Increasing fatigue

As we have shown that $\underline{V}_\gamma$ can be achieved using a stationary strategy, we shall now have a closer look into how the frequencies of optimal histories change as γ varies. Intuitively, a larger γ forces good actions to be used less often so that their stage payoff does not wear down too much. The following lemma makes this formal. As γ increases, the aggregated weight on the top actions is decreasing.

Lemma 5.3. *Let $A = \{a_1, \ldots, a_m\}$ with $u(a_1) \geq u(a_2) \geq \cdots \geq u(a_m)$. For each $\gamma \in (0,1]$ let $x_\gamma \in \Delta(A)$ be the (unique) solution to the maximization problem in (6). Then*

$$\sum_{i=1}^{k} \frac{d}{d\gamma} x_\gamma(a_i) \leq 0 \tag{24}$$

for all $k = 1, \ldots, m$.

The proof of Lemma 5.3 is a straightforward application of standard optimization tools and therefore deferred to the appendix

As $\sum_{a\in A} x_\gamma(a) = 1$ for all $\gamma \in (0,1]$, an immediate consequence of Lemma 5.3 is that the aggregated weight on the poor actions is increasing as γ increases. Hence, we immediately obtain the following proposition, which says that a decision maker who exhibits a higher fatigue parameter γ will choose actions with poorer basic utility more frequently.

Proposition 5.4. *Let $A = \{a_1, \ldots, a_m\}$ with $u(a_1) \geq u(a_2) \geq \cdots \geq u(a_m)$, let $0 < \gamma < \gamma' \leq 1$, and let $\vec{a}, \vec{b} \in A^\infty$ be two optimal stationary histories with respect to γ and γ', respectively. Then $\varphi(\cdot|\vec{a})$ first order stochastically dominates $\varphi(\cdot|\vec{b})$.*

As the proposition follows immediately from Lemma 5.3, its formal proof is omitted.

6. The values of the finite decision problems and $\overline{V}_\gamma$

In Examples 3.2 and 4.4 we have seen a set of actions for which $\overline{V}_\gamma > V_\gamma^*$. This relation is quite robust, as we will show in this section: whenever A contains at least two actions with different basic payoffs, it holds true. The rough idea of the proof is to show first that the sequence $\left(v_\gamma^T\right)_{T\in\mathbb{N}}$ (which is the sequence of maximal average payoffs that can be achieved by histories of length T, see Equation (2)) converges with $\lim_{T\to\infty} v_\gamma^T = \overline{V}_\gamma$, and second that $\lim_{T\to\infty} v_\gamma^T$ is bounded away from V_γ^*.

6.1. The sequence $\left(v_\gamma^T\right)_{T\in\mathbb{N}}$ converges to $\overline{V}_\gamma$

In this subsection we show that the sequence $\left(v_\gamma^T\right)_{T\in\mathbb{N}}$ converges. The idea of the proof is to show that for a history $\vec{a}$ of sufficient length T and any $S > T$ we can find a history $\vec{b}$ of length S that has an average payoff at S that is close to the one of $\vec{a}$ at T. This implies that $\liminf_S v_\gamma^S$ is at least close to $U_\gamma^T\left(\vec{a}\right)$. As $\vec{a}$ can be chosen such that $U_\gamma^T\left(\vec{a}\right)$ is close to $\limsup_T v_\gamma^T$, one finds that the limit inferior and limit superior of $\left(v_\gamma^T\right)_T$ must be identical.

The construction of $\vec{b}$ relies on the division of $\vec{a}$ into blocks such that the length of any block is a fraction α of the previous history, similar to the construction in the proof of Theorem 5.2. These blocks are then "stretched" by some factor $\delta > 1$ such that $S = \delta T$. Using the same approximations as in Equations (8), (9), and (10) in the proof of Theorem 5.2, we show that in each block k the average utility of $\vec{b}$, denoted Y_γ^t, is close to the average utility in the k-th block of $\vec{a}$, denoted W_γ^t. The average payoff of $\vec{a}$ at time T is a convex combination of these average block utilities, where each W^k is weighted by $\frac{t_{k+1}-t_k}{T}$. Similarly, the average payoff of $\vec{b}$ at time S is a convex combination, where each Y^k is weighted by $\frac{s_{k+1}-s_k}{S}$. By the construction of $\vec{b}$, these two fractions are close as well (recall that $\vec{b}$ is merely a stretched version of $\vec{a}$). Hence, $U_\gamma^S(\vec{b})$ and $U_\gamma^T\left(\vec{a}\right)$ are close as well.

Proposition 6.1. *The sequence* $\left(v_\gamma^T\right)_{T\in\mathbb{N}}$ *converges.*

Proof. Clearly, the sequence is bounded, so that $\limsup_{T\to\infty} v_\gamma^T$ and $\liminf_{T\to\infty} v_\gamma^T$ exist. We show that for each $\varepsilon > 0$ there is $T^* \in \mathbb{N}$ such that if $v_{T^*} \geq \limsup_{T\to\infty} v_\gamma^T - \varepsilon$, then $v_S \geq \limsup_{T\to\infty} v_\gamma^T - 2\varepsilon$ for all $S \geq T^*$. This implies that for every $\varepsilon > 0$ it holds that $\limsup_{T\to\infty} v_T - \liminf_{T\to\infty} v_T < 2\varepsilon$, so that $\limsup_{T\to\infty} v_\gamma^T = \liminf_{T\to\infty} v_\gamma^T = \lim_{T\to\infty} v_\gamma^T$.

So, let $\varepsilon > 0$ be sufficiently small. Let

$$t_1 \geq \max\left\{\frac{1}{\varepsilon^3}\left(|A| + 8\gamma\sum_{a\in A} u(a)\right), \frac{1+2\varepsilon^2}{\varepsilon^3}16\gamma\sum_{a\in A} u(a)\right\},$$
$$\alpha = \frac{\varepsilon}{16\gamma\sum_a u(a)} - \frac{2}{t_1},$$

and let $T^* \geq \frac{4t_1\sum_a u(a)}{\varepsilon}$ be such that $v_{T^*} \geq \limsup_{T\to\infty} v_T - \varepsilon$. Observe that for sufficiently small ε we have

$$1 \geq \alpha = \frac{\varepsilon}{16\gamma\sum_a u(a)} - \frac{2}{t_1} \geq \frac{1+2\varepsilon^2}{\varepsilon^2 t_1} - \frac{2}{t_1} = \frac{1}{\varepsilon^2 t_1} > \frac{1}{t_1}. \tag{25}$$

For $k \geq 2$, let r^k be the smallest integer such that $r^k \geq (1+\alpha)^{k-1} t_1$, and let K be the smallest integer with $r^K \geq T^*$. Let $t^0 = 0$, for $k = 2, \dots, K-1$ let $t_k = r^k$, and let $t_K = T^* > t_{K-1}$. Let $\vec{a} \in A^\infty$ be such that $U_\gamma^{T^*}\left(\vec{a}\right) = v_\gamma^{T^*}$. As in the proof of Theorem 5.2, let the k-th block of $\vec{a}$ be the finite sequence $\left(\vec{a}_{t_k+1}, \dots, \vec{a}_{t_{k+1}}\right)$. For $k = 0, \dots, K-1$, denote the average payoff, the frequency, and the approximation by $W_\gamma^k = W_\gamma\left(\vec{a}, t_k, t_{k+1}\right)$, $p^k(a) = p\left(a; \vec{a}, t_k, t_{k+1}\right)$, and $\widetilde{U}_\gamma^k = \widetilde{U}_\gamma\left(\vec{a}, t_k, t_{k+1}\right)$, respectively, as in Equations (8), (9), and (10). In particular, $W_\gamma^0 = U^{t_1}$ and $p^0 = \varphi\left(.\middle|\vec{a}^{t_1}\right)$. By construction, $\frac{t_{k+1}-t_k}{t_k} \leq \alpha + \frac{1}{t_1}$ for all $k = 1, \dots, K$. Thus, by Lemma 5.1 and the definition of α,

$$\left|W_\gamma^k - \widetilde{U}_\gamma^k\right| \leq 2\left(\alpha + \frac{1}{t_1}\right)\gamma\sum_{a\in A} u(a) \leq 2\frac{\varepsilon}{16\gamma\sum_a u(a)}\gamma\sum_{a\in A} u(a) = \frac{1}{8}\varepsilon \tag{26}$$

for $k = 1, \dots, K$.

Let $S \geq T^*$ and define $\delta = \frac{S}{T^*}$. For each $k \geq 1$, let s_k be the largest integer smaller than or equal to δt_k. Let $\vec{b} \in A^\infty$ be such that $\vec{b}^{t_1} = \vec{a}^{t_1}$ and

$$\vec{b}_s = \begin{cases} \arg\min_{a\in A} \varphi\left(a \mid b^{s-1}\right) - \varphi\left(a \mid a^{t_1}\right), & \text{if } s = t_1+1, \dots, s^1, \\ \arg\min_{a\in A} \frac{1}{s-s_k}\sum_{s'=s_k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a} - p^k(a), & \text{if } s_k+1 \leq s \leq s_{k+1}, \text{ where } k \geq 1. \end{cases} \tag{27}$$

That is, the individual blocks of history $\vec{b}$ are longer than those of $\vec{a}$, stretched by the factor δ, and in the k-th block of $\vec{b}$ actions are chosen to minimize the difference between the frequencies in the k-th block of $\vec{a}$ and $\vec{b}$. For $k = 0, \dots, K-1$ denote the average payoff, the frequency, and the approximation in $\vec{b}$ by $Y_\gamma^k = W_\gamma\left(\vec{b}, s_k, s_{k+1}\right)$, $q^k(a) = p\left(a; \vec{b}, s_k, s_{k+1}\right)$, and $\widetilde{V}_\gamma^k = \widetilde{U}_\gamma\left(\vec{b}, s_k, s_{k+1}\right)$, respectively, as in Equations (8), (9), and (10). As $\frac{s_{k+1}-s_k}{s_k} \leq \alpha + \frac{2}{t_1}$ for all $k \geq 1$, Lemma 5.1 together with the definition of α gives

$$\left|Y^k - \widetilde{V}_\gamma^k\right| \leq 2\left(\alpha + \frac{2}{t_1}\right)\gamma\sum_{a\in A} u(a) = 2\frac{\varepsilon}{16\gamma\sum_a u(a)}\gamma\sum_{a\in A} u(a) = \frac{1}{8}\varepsilon. \tag{28}$$

By construction, $\varphi\left(.\middle|\vec{a}^{t_1}\right) = \varphi\left(.\middle|\vec{b}^{t_1}\right)$. By (27), for all $t_1+1 \leq s \leq s^1$, action a is only chosen if $\varphi\left(a\middle|\vec{b}^{s-1}\right) \leq \varphi\left(a\middle|\vec{a}^{t_1}\right)$. Thus, $\varphi\left(a\middle|\vec{b}^{s}\right) \leq \varphi\left(a\middle|\vec{a}^{t_1}\right) + \frac{1}{s}$ for all $s \leq s_1$. Since $\sum_{a\in A}\varphi\left(a\middle|\vec{b}^{s}\right) = 1$, this implies that $\varphi\left(a\middle|\vec{b}^{s}\right) \geq \varphi\left(a\middle|\vec{a}^{t_1}\right) - \frac{|A|-1}{s}$ for all $s \leq s_1$. Thus, for sufficiently small $\varepsilon > 0$

$$\left|\varphi\left(a\middle|\vec{a}^{t_1}\right) - \varphi\left(a\middle|\vec{b}^{s}\right)\right| \leq \frac{|A|-1}{s} \leq \frac{|A|-1}{t_1} \leq \varepsilon^3\frac{|A|-1}{|A| + 8\gamma\sum_a u(a)} \leq \varepsilon^2 \tag{29}$$

for all $s \le s_1$. Let $k \ge 1$ and $s_k + 1 \le s \le s_{k+1}$. By (27), action a is only being played at s if $\frac{1}{s-s_k}\sum_{s'=s_k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a} \le p^k(a)$. Thus, $p^k(a) \le \frac{1}{s-s^k} + \frac{1}{s-s^k}\sum_{s'=s^k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a}$. In particular, for $s = s^{k+1}$ it holds that

$$p^k(a) \le \frac{1}{s_{k+1}-s_k} + \frac{1}{s_{k+1}-s_k}\sum_{s'=s_k+1}^{s_{k+1}} \mathbb{1}_{\vec{b}_{s'}=a} = \frac{1}{s_{k+1}-s_k} + q^k(a).$$

Since $\sum_{a\in A} q^k(a) = \sum_{a\in A} p^k(a) = 1$, this implies $p^k(a) \ge q^k(a) + \frac{|A|-1}{s_{k+1}-s_k}$, so that for sufficiently small $\varepsilon > 0$

$$\begin{aligned}\left|p^k(a) - q^k(a)\right| &\le \frac{|A|-1}{s_{k+1}-s_k} \le \frac{|A|-1}{t_{k+1}-t_k} \\ &\le \frac{|A|-1}{\alpha t_k - 1} \le \frac{|A|-1}{\alpha(1+\alpha)^k t_1 - 1} \\ &\le \frac{|A|-1}{\frac{1}{\varepsilon^2}(1+\alpha)^k - k} \le \varepsilon^2 \frac{|A|-1}{(1+\alpha)^k - \varepsilon^2} \\ &\le \frac{\varepsilon}{8|A|u(a)}\end{aligned} \tag{30}$$

for all $a \in A$ and all $k \ge 1$. In particular,

$$\sum_{a\in A}\left|p^k(a) - q^k(a)\right|u(a) \le \sum_{a\in A}\frac{\varepsilon}{8|A|u(a)}u(a) = \frac{1}{8}\varepsilon.$$

Further, by using (30) we find for $k \ge 2$ that

$$\begin{aligned}\left|\varphi\left(a\middle|\vec{a}^{t_k}\right) - \varphi\left(a\middle|\vec{b}^{s_k}\right)\right| &= \left|\frac{t_{k-1}}{t_k}\varphi\left(a\middle|\vec{a}^{t_{k-1}}\right) + \frac{t_k - t_{k-1}}{t_k}p^k(a) - \frac{s_{k-1}}{s_k}\varphi\left(a\middle|\vec{b}^{s_{k-1}}\right) - \frac{s_k - s_{k-1}}{s_k}q^k(a)\right| \\ &\le \left|\frac{t_{k-1}}{t_k}\varphi\left(a\middle|\vec{a}^{t_{k-1}}\right) - \frac{\delta t_{k-1} - x_1}{\delta t_k - x_2}\varphi\left(a\middle|\vec{b}^{s_{k-1}}\right)\right| \\ &\qquad + \left|\frac{t_k - t_{k-1}}{t_k}p^k(a) - \frac{\delta t_k - x_2 - \delta t_{k-1} + x_1}{\delta t_k - x_2}q^k(a)\right| \\ &\le \frac{t_{k-1}}{t_k}\left|\varphi\left(a\middle|\vec{a}^{t_{k-1}}\right) - \varphi\left(a\middle|\vec{b}^{s_{k-1}}\right)\right| + \frac{x_1}{s_k}\varphi\left(a\middle|\vec{b}^{s_{k-1}}\right) \\ &\qquad + \frac{t_k - t_{k-1}}{t_k}\left|p^k(a) - q^k(a)\right| + \frac{|x_1 - x_2|}{s^k}q^k(a) \\ &\le \frac{t_{k-1}}{t_k}\left|\varphi\left(a\middle|\vec{a}^{t_{k-1}}\right) - \varphi\left(a\middle|\vec{b}^{s_{k-1}}\right)\right| + \frac{2}{t_k} + \frac{t_k - t_{k-1}}{t_k}\frac{|A|-1}{t_{k+1}-t_k} \\ &\le \frac{t_{k-1}}{t_k}\left|\varphi\left(a\middle|\vec{a}^{t_{k-1}}\right) - \varphi\left(a\middle|\vec{b}^{s_{k-1}}\right)\right| + \frac{|A|+1}{t_k},\end{aligned}$$

where $x_1, x_2 \le 1$ are such that $s_{k-1} = \delta t_{k-1} - x_1$ and $s_k = \delta t_k - x_2$. We thus find inductively that, for sufficiently small ε,

$$\begin{aligned}\left|\varphi\left(a\middle|\vec{a}^{t_k}\right) - \varphi\left(a\middle|\vec{b}^{s_k}\right)\right| &\le \frac{t_1}{t_k}\left|\varphi\left(a\middle|\vec{a}^{t_1}\right) - \varphi\left(a\middle|\vec{b}^{s_1}\right)\right| + (|A|+1)\sum_{l=2}^{k}\frac{1}{t^l} \\ &\le \frac{t_1}{t_k}\varepsilon^2 + \frac{|A|+1}{t_1}\sum_{l=2}^{k}\left(\frac{1}{1+\alpha}\right)^l \\ &\le \frac{1}{(1+\alpha)^k}\varepsilon^2 + \frac{|A|+1}{\alpha t_1} \\ &\le \frac{1}{(1+\alpha)^k}\varepsilon^2 + (|A|+1)\varepsilon^2 \\ &\le (|A|+2)\varepsilon^2 \\ &\le \frac{\varepsilon}{8|A|u(a)}\end{aligned}$$

for all $k \ge 1$, so that

$$\sum_{a\in A}\left|\varphi\left(a\middle|\vec{a}^{t_k}\right) - \varphi\left(a\middle|\vec{b}^{s_k}\right)\right| \le \frac{1}{8}\varepsilon$$

for all $k \ge 0$. (The case $k = 0$ follows from (29).) Hence,

$$\begin{aligned}\left|\widetilde{U}_{\gamma}^{k}-\widetilde{V}_{\gamma}^{k}\right| &=\left|\sum_{a\in A}p^{k}(a)\left(1-\varphi\left(a\middle|\vec{a}^{t_k}\right)\right)u(a)-\sum_{a\in A}q^{k}(a)\left(1-\varphi^{k}\left(a\middle|\vec{b}^{s_k}\right)\right)u(a)\right|\\ &=\left|\sum_{a\in A}\left(p^{k}(a)-q^{k}(a)\right)\left(1-\varphi\left(a\middle|\vec{a}^{t_k}\right)\right)u(a)-\sum_{a\in A}q^{k}(a)\left(\varphi\left(a\middle|\vec{a}^{t_k}\right)-\varphi\left(a\middle|\vec{b}^{s_k}\right)\right)u(a)\right|\\ &\leq\sum_{a\in A}\left|p^{k}(a)-q^{k}(a)\right|u(a)-\sum_{a\in A}\left|\varphi\left(a\middle|\vec{a}^{t_k}\right)-\varphi\left(a\middle|\vec{b}^{s_k}\right)\right|u(a)\\ &\leq\frac{1}{4}\varepsilon\end{aligned}\tag{31}$$

for all $k\geq 1$. From (26), (28), an (31) we find

$$\left|W_{\gamma}^{k}-Y_{\gamma}^{k}\right|\leq\left|W_{\gamma}^{k}-\tilde{U}_{\gamma}^{k}\right|+\left|\tilde{U}_{\gamma}^{k}-\tilde{V}_{\gamma}^{k}\right|+\left|\tilde{V}_{\gamma}^{k}-Y_{\gamma}^{k}\right|\leq\frac{1}{2}\varepsilon$$

for all $k\geq 1$. Thus, recalling that $t_K=T^*$ and $s_K=S$, we have

$$\begin{aligned}\left|U_{\gamma}^{T}\left(\vec{a}\right)-U_{\gamma}^{S}\left(\vec{b}\right)\right| &\leq\sum_{k=0}^{K-1}\left|\frac{t_{k+1}-t_k}{T}W_{\gamma}^{k}-\frac{s_{k+1}-s_k}{S}Y_{\gamma}^{k}\right|\\ &\leq\sum_{k=0}^{K-1}\left|\frac{\delta t_{k+1}-\delta t_k}{\delta T}W_{\gamma}^{k}-\frac{s_{k+1}-s_k}{\delta T}Y_{\gamma}^{k}\right|\\ &\leq\sum_{k=0}^{K-1}\left(\left|\frac{\delta t_{k+1}-\delta t_k}{\delta T}\left(W_{\gamma}^{k}-Y_{\gamma}^{k}\right)\right|+\left|\frac{1}{\delta T}Y_{\gamma}^{k}\right|\right)\\ &\leq\sum_{k=0}^{K-1}\frac{(1+\alpha)^{k+1}t_1-(1+\alpha)^{k}t_1+1}{(1+\alpha)^{K-1}t_1}\left|W_{\gamma}^{k}-Y_{\gamma}^{k}\right|+\frac{K}{\delta(1+\alpha)^{K-1}t_1}\sum_{a\in A}u(a)\\ &\leq\frac{\varepsilon}{2}\sum_{k=0}^{K-1}\left(\alpha(1+\alpha)^{k-K+1}+\frac{1}{(1+\alpha)^{K-1}t_1}\right)+\frac{K}{(1+\alpha)^{K-1}t_1}\sum_{a\in A}u(a)\\ &\leq\frac{\varepsilon}{2}\left(1+\frac{K}{(1+\alpha)^{K}t_1}\right)+\frac{K}{(1+\alpha)^{K-1}t_1}\sum_{a\in A}u(a).\end{aligned}$$

Using that

$$\frac{K}{(1+\alpha)^{K}t_1}\leq\frac{K}{(1+\alpha)^{K-1}t_1}\leq\frac{K}{(1+\alpha(K-1))t_1}\leq\frac{K}{(\alpha+\alpha K-\alpha)t_1}=\frac{1}{\alpha t_1}\leq\varepsilon^2$$

we conclude that, for sufficiently small ε,

$$\left|U_{\gamma}^{T}\left(\vec{b}\right)-U_{\gamma}^{S}\left(\vec{a}\right)\right|\leq\frac{\varepsilon}{2}\left(1+\varepsilon^2\right)+\varepsilon^2\sum_{a\in A}u(a)\leq\varepsilon.$$

Thus,

$$v_{\gamma}^{S}\geq U_{\gamma}^{S}\left(\vec{b}\right)\geq U_{\gamma}^{T}\left(\vec{a}\right)-\varepsilon\geq\limsup_{T'}v_{T'}-2\varepsilon$$

as required. □

After establishing that $\lim_{T\to\infty}v_{\gamma}^{T}$ is well-defined, we will now show that the sequence converges to $\overline{V}_{\gamma}$. As explained earlier, the value $\overline{V}_{\gamma}$ itself may not be behaviorally sound. However, the convergence of $\lim_{T\to\infty}v_{\gamma}^{T}$ to $\overline{V}_{\gamma}$ grants it a certain significance.

Proving that v_{γ}^{T} converges towards $\overline{V}_{\gamma}$ is now fairly simple: as there is a sequence $\vec{a}$ whose average payoff gets infinitely often arbitrarily close to $\overline{V}_{\gamma}$, the sequence v_{γ}^{T} gets infinitely often at least close to $\overline{V}_{\gamma}$ as well. Hence, $\lim_{T\to\infty}v_{\gamma}^{T}\geq\overline{V}_{\gamma}$. For the other inequality note that for any history $\vec{a}$ the number $\overline{V}_{\gamma}\left(\vec{a}\right)$ gets arbitrarily close to $U_{\gamma}^{T}\left(\vec{a}\right)$ for sufficiently large T. This is true, in particular, for the optimal $\vec{a}$, for which $U_{\gamma}^{T}\left(\vec{a}\right)$ gets arbitrarily close to $\lim_{T\to\infty}v_{\gamma}^{T}$.

Proposition 6.2. *It holds that* $\lim_{T\to\infty}v_{\gamma}^{T}=\overline{V}_{\gamma}$.

Proof. Let $\varepsilon>0$ and let $\vec{a}$ be such that $\overline{V}_{\gamma}\left(\vec{a}\right)\geq\overline{V}_{\gamma}-\varepsilon$. Then there is a sequence $\left(T_k\right)_{k\in\mathbb{N}}$ such that

$$v_{\gamma}^{T_k}\geq U_{\gamma}^{T_k}\left(\vec{a}\right)\geq\overline{V}_{\gamma}\left(\vec{a}\right)-\varepsilon\geq\overline{V}_{\gamma}-2\varepsilon$$

for all $k \in \mathbb{N}$. Thus, $\lim_{T\to\infty} v_\gamma^T = \lim_{k\to\infty} v_\gamma^{T_k} \geq \overline{V}_\gamma - 2\varepsilon$. As $\varepsilon > 0$ was arbitrary, we have $\lim_{T\to\infty} v_\gamma^T \geq \overline{V}_\gamma$.

Assume that there is $c > 0$ such that $\lim_{T\to\infty} v_\gamma^T \geq \overline{V}_\gamma + 4c$. Let T^0 be such that $v_\gamma^{T'} \geq \lim_{T\to\infty} v_\gamma^T - c$ for all $T' \geq T^0$. There is $T_1 \geq T^0$ such that

$$v_\gamma^{T'} \geq \lim_{T\to\infty} v_\gamma^T - c \geq \overline{V}_\gamma + 3c = \sup_{\vec{a}\in A^\infty} \limsup_T U_\gamma^T\left(\vec{a}\right) + 3c \tag{32}$$

for all $T' \geq T_1$. For each $\vec{a}$ there is $T_2\left(\vec{a}\right) \geq T_1$ such that $\limsup_T U_\gamma^T\left(\vec{a}\right) \geq U^{T'}\left(\vec{a}\right) - c$ for all $T' \geq T_2\left(\vec{a}\right)$. In particular,

$$\sup_{\vec{a}\in A^\infty} \limsup_T U_\gamma^T\left(\vec{a}\right) + 3c \geq \sup_{\vec{a}\in A^\infty} U_\gamma^{T_2(\vec{a})}\left(\vec{a}\right) + 2c \geq \lim_{T\to\infty} v_\gamma^T + c, \tag{33}$$

where the last inequality holds since $T_2\left(\vec{\alpha}\right) \geq T_1 \geq T_0$ for all $\vec{a} \in A^\infty$. From (32) and (33) we obtain $v_\gamma^{T'} \geq \lim_{T\to\infty} v_\gamma^T + c$ for all $T' \geq T_1$, which is impossible as $\left(v_\gamma^T\right)$ converges by Proposition 6.1. □

6.2. Stationary strategies cannot achieve $\overline{V}_\gamma$

The next main result shows that (generically) $\overline{V}_\gamma$ cannot be achieved by any stationary strategy, or, to be more precise, that the optimal stationary strategy achieves an average payoff that is strictly less than $\overline{V}_\gamma$. In the proof we start from a stationary history $\vec{a}$ and then iteratively construct a sequence of histories by only switching two actions in each step. We begin with the following lemma, which provides a sufficient condition for increasing all future payoffs by swapping the positions of two actions a and b within some history $\vec{a}$. Let t be the last occurrence of a before b. Then the switch of a and b increases the average payoff if

$$\varphi(a \mid \vec{a}^{t-1})u(a) - \varphi(b \mid \vec{a}^{t-1})u(b) > 0.$$

The intuition here is that actions with high base utility, when shifted backward, can be played with a relatively small fatigue cost.

Lemma 6.3. *Let $a, b \in A$ and let $\vec{a}, \vec{b} \in A^\infty$ be such that there are $s > t$ with $\vec{a}_t = \vec{b}_s = a$, $\vec{a}_s = \vec{b}_t = b$, $\vec{a}_{t'} = \vec{b}_{t'}$ for all $t' \neq t, s$, and $\vec{a}_{t'} \neq a$ for all $t' = t, \ldots, s-1$. Then*

$$U_\gamma^T(\vec{b}) - U_\gamma^T(\vec{a}) \geq \frac{\gamma(s-t)}{(s-1)T}\left(\varphi\left(a\middle|\vec{a}^{t-1}\right)u(a) - \varphi\left(b\middle|\vec{a}^{t-1}\right)u(b)\right) \tag{34}$$

for all $T \geq s$.

As the proof of Lemma 6.3 is fairly straightforward, it is deferred to the Appendix. Equation (34) provides us with a lower bound for the increase in payoff that results from switching two actions. In order to prove the next theorem, we start with a stationary history and then successively swap actions, such that in each step the premise of Lemma 6.3 is satisfied. We show that the increase in payoff (which applies to all sufficiently late periods) is significant.

Theorem 6.4. *Let A be a finite set of actions. Then $\overline{V}_\gamma > V_\gamma^*$ if and only if for the optimal stationary frequency $\varphi \in \Delta(A)$ there are two actions $a, b \in A$ with $\varphi(a), \varphi(b) > 0$ and $u(a) \neq u(b)$.*

Proof. Necessity is clear. We show sufficiency. Let $\vec{a}$ be an optimal stationary strategy, write $\varphi(a)$ for $\varphi\left(a\middle|\vec{a}\right)$, and denote by A^* the set of actions $a \in A$ with $\varphi(a) > 0$. By Proposition 4.3,

$$\varphi(a)u(a) = \frac{2\gamma - |A^*| + \sum_{b\in A^*}\frac{u(a)}{u(b)}}{2\gamma\sum_{b\in A^*}\frac{1}{u(b)}} = \frac{u(a)}{2\gamma} - \frac{|A| - 2\gamma}{2\gamma\sum_{b\in A^*}\frac{1}{u(b)}},$$

which implies that $\varphi(a) \geq \varphi(b)$ if and only if $u(a) \geq u(b)$. Moreover, we have

$$\varphi(a)u(a) - \varphi(b)u(b) = \frac{u(a) - u(b)}{2\gamma} \tag{35}$$

for all $a, b \in A^*$. As V_γ^* and $\overline{V}_\gamma$ depend continuously on γ and $\{u(a)\}_{a\in A}$, and $\varphi(\cdot) \in \mathbb{Q}^A$ if $\gamma, u(a) \in \mathbb{Q}$ for all $a \in A$ by Proposition 4.3, we can assume without loss of generality that $\varphi(\cdot) \in \mathbb{Q}^A$. Thus, there are integers $m_a \in \mathbb{N}$ for all $a \in A$ such that $\varphi(a) = \frac{m_a}{m}$, where $m = \sum_{a\in A^*} m_a$. Again, without loss of generality, we can assume that $\vec{a}$ is the infinite repetition of a sequence of length m in which each action $a \in A^*$ is played exactly m_a times. Let $\underline{a} \in \arg\min_{a\in A^*} u(a)$ and $\overline{a} \in \arg\max_{a\in A^*} u(a)$. By the premise of the theorem, $u\left(\underline{a}\right) < u\left(\overline{a}\right)$, so that $m_{\overline{a}} > m_{\underline{a}}$.

The proof of the following claim is deferred to the appendix.

Claim 1: *For all $t \geq 2$ and all $a \in A^*$ it holds that*

$$\varphi(a) - \frac{m}{t-1} \leq \varphi\left(a\middle|\vec{a}^{t-1}\right) \leq \varphi(a) + \frac{m}{t-1}. \quad \square$$

Define now the following constants

$$\delta = \frac{u\left(\overline{a}\right) - u\left(\underline{a}\right)}{4\gamma},$$
$$q' = \frac{\varphi\left(\underline{a}\right)\left(u\left(\underline{a}\right) + u\left(\overline{a}\right)\right)}{\delta + \varphi\left(\underline{a}\right)\left(u\left(\underline{a}\right) + u\left(\underline{b}\right)\right)},$$
$$q = \max\left(\frac{3}{4}, q'\right),$$
$$\eta = \frac{1-q}{64}\varphi\left(\underline{a}\right)\gamma\delta,$$

and observe that $\delta > 0$, so that $q < 1$ and $\eta > 0$. Let

$$\varepsilon \le \min\left(\frac{\delta}{u\left(\underline{a}\right) + u\left(\overline{a}\right), \frac{1}{2}\eta}\right)$$

and let

$$T_1 \ge \frac{2m - 1 + (m+1)\left(\delta + \varphi\left(\underline{a}\right)\left(u\left(\underline{a}\right) + u\left(\overline{a}\right)\right)\right)}{\delta}$$

be a multiple of m and be such that $\left|\varphi\left(a\middle|\vec{a}^t\right) - \varphi(a)\right| \le \varepsilon$ for all $a \in A$ and all $t \ge T_1$. Finally, let

$$T^* \ge \max\left(2T_1, T_1 + 4m, T_1 + \frac{4\left(\varphi\left(\underline{a}\right) + 2m\right)}{(1-q)\varphi\left(\underline{a}\right)}\right)$$

be such that $\left|U_\gamma^T\left(\vec{a}\right) - V_\gamma^*\right| \le \eta$ for all $T \ge T^*$. We show that there is an infinite sequence of T's with $v^T \ge V^* + \eta$ for all $T \ge T^*$. For this purpose we will construct for any T in this sequence a history $\vec{b}$ with $U_\gamma^T\left(\vec{b}\right) \ge V^* + \eta$. By Proposition 6.2, this is sufficient to prove the theorem.

So, let $T \ge T^*$ be a multiple of m. We construct $\vec{b}$ by iteratively switching actions $\underline{a}$ and $\overline{a}$ between $T_1 + 1$ and T. Specifically, let $\vec{c} \in \mathcal{A}^\infty$ be a history that has been reached after some switches, and let $T_1 + 1 \le s \le T$ be the first occurrence of $\underline{a}$ such that the period of the last previous occurrence of $\overline{a}$, denoted by $t < s$, satisfies

$$\varphi\left(\overline{a}\middle|\vec{c}^{t-1}\right)u\left(\overline{a}\right) - \varphi\left(\underline{a}\middle|\vec{c}^{t-1}\right)u\left(\underline{a}\right) \ge \delta. \tag{36}$$

If such $t, s \le T$ do not exist, let $\vec{b} = \vec{c}$. Otherwise, note that (36) does not depend on s, so that the minimality of s implies that there are no instances of $\underline{a}$ between $t+1$ and $s-1$. Let $\vec{d}$ be such that $\vec{d}_t = \underline{a}$, $\vec{d}_s = \overline{a}$, and $\vec{d}_{t'} = \vec{c}_{t'}$ for all $t' \ne t, s$. By Lemma 6.3,

$$U_\gamma^T\left(\vec{d}\right) - U_\gamma^T\left(\vec{c}\right) \ge \frac{\gamma(s-t)}{(s-1)T}\delta.$$

Thus, we say that the switch of $\overline{a}$ and $\underline{a}$ in $\vec{c}$ is *beneficial*. Repeat the procedure with $\vec{d}$ and continue as long as possible. Note that all beneficial switches will shift occurrences of $\underline{a}$ towards the beginning, i.e., T_1, and occurrences of $\overline{a}$ towards the end, i.e., T. Thus, for each finite $T \ge T^*$ there is a finite number of beneficial switches, so $\vec{b}$ is well-defined. Moreover, for any $T_1 + 1 \le t \le T - m$ it holds that the sequence $\left(\vec{b}_{t+1}, \ldots, \vec{b}_{t+m}\right)$ contains exactly $m_{\underline{a}} + m_{\overline{a}}$ periods in which either $\underline{a}$ or $\overline{a}$ are being chosen.

The proof of the next claim can also be found in the appendix.

Claim 2: *In history $\vec{b}$, the last occurrence of $\underline{a}$ until T is at some $s \le T_1 + \left(T - T_1\right)q$.* □

So, in history $\vec{b}$ there are no occurrences of $\underline{a}$ between $T_1 + q\left(T - T_1\right)$ and T. Let t^* be the smallest integer such that $t^* \ge T_1 + \frac{1+q}{2}\left(T - T_1\right)$ and let k be the number of occurrences of $\underline{a}$ in $\vec{a}$ between $t^* + 1$ and T. Then, with the bounds in Claim 1, and since $T \ge T^* \ge T_1 + \frac{4(\varphi(\underline{a}) + m)}{(1-q)\varphi(\underline{a})}$ by construction,

$$\begin{aligned}
k &= \varphi\left(\underline{a} \mid \vec{a}^T\right)T - \varphi\left(\underline{a} \mid \vec{a}^{t^*}\right)t^* \\
&\ge \varphi\left(\underline{a}\right)T - \varphi\left(\underline{a}\right)t^* - 2m \\
&\ge \varphi\left(\underline{a}\right)\left(T - \left(T_1 + \frac{1+q}{2}\left(T - T_1\right) + 1\right)\right) - 2m \\
&= \varphi\left(\underline{a}\right)\left(T - T_1\right)\frac{1-q}{2} - \varphi\left(\underline{a}\right) - m \\
&= \varphi\left(\underline{a}\right)\left(T - T_1\right)\frac{1-q}{4} + \varphi\left(\underline{a}\right)\left(T - T_1\right)\frac{1-q}{4} - \varphi\left(\underline{a}\right) - 2m
\end{aligned}$$

$$\geq \varphi\left(\underline{a}\right)\left(T-T_1\right)\frac{1-q}{4}+\varphi\left(\underline{a}\right)\left(T_1+\frac{4\left(\varphi\left(\underline{a}\right)+2m\right)}{(1-q)\varphi\left(\underline{a}\right)}-T_1\right)\frac{1-q}{4}-\varphi\left(\underline{a}\right)-2m$$
$$=\varphi\left(\underline{a}\right)\left(T-T_1\right)\frac{1-q}{4}. \tag{37}$$

Let $s^1,\dots,s^k$ be the times of all occurrences of $\underline{a}$ in $\vec{a}$ with $T_1+\frac{1+q}{2}\left(T-T_1\right)\leq s_1\leq\cdots\leq s_k\leq T$. Let $t_1\leq\cdots\leq t_k$ be the last k occurrences of $\underline{a}$ in $\vec{b}$, and recall that $t_\ell\leq T_1+q\left(T-T_1\right)$ for all $\ell=1,\dots,k$. Thus, $s^\ell-t^\ell\geq\frac{1+q}{2}\left(T-T_1\right)$ for all $\ell=1,\dots,k$. Define histories $\vec{b}(0),\dots,\vec{b}(k)$ as follows. Let $\vec{b}(0)=\vec{b}$ and for all $\ell=1,\dots,k$, let $\vec{b}_{t^\ell}(\ell)=\vec{b}_{s^\ell}(\ell-1)$, $\vec{b}_{s^\ell}(\ell)=\vec{b}_{t^\ell}(\ell-1)$, and $\vec{b}_t(\ell)=\vec{b}_t(\ell-1)$ for all $t\neq t^\ell,s^\ell$. Using Lemma 6.3 and the fact that $\frac{T-T_1}{T}\geq\frac{1}{2}$ by construction, we therefore have

$$\begin{aligned}U_\gamma^T\left(\vec{b}(\ell)\right)-U_\gamma^T\left(\vec{b}(\ell+1)\right)&\geq\gamma\frac{1}{T}\frac{s^\ell-t^\ell}{s^\ell-1}\left(\varphi(\overline{a}|\vec{b}^{t^\ell-1})u\left(\overline{a}\right)-\varphi(\underline{a}|\vec{b}^{t^\ell-1})u\left(\underline{a}\right)\right)\\&\geq\frac{\gamma}{T}\frac{(1+q)\left(T-T_1\right)}{2T}\delta\\&\geq\frac{\gamma}{4T}\delta\end{aligned}$$

for all $\ell=0,\dots,k-1$. Observe that the iterative procedure that we used to construct $\vec{b}$ must have passed through these histories and, in particular, through $\vec{b}(k)$. Thus, with the lower bound in (37) for k we have

$$\begin{aligned}U_\gamma^T\left(\vec{b}\right)-U_\gamma^T\left(\vec{a}\right)&\geq U_\gamma^T\left(\vec{b}\right)-U_\gamma^T\left(\vec{b}(k)\right)\\&=\sum_{\ell=0}^{k-1}\left(U_\gamma^T\left(\vec{b}(\ell)\right)-U_\gamma^T\left(\vec{b}(\ell+1)\right)\right)\\&\geq k\frac{\gamma}{4T}\delta\\&\geq\varphi\left(\underline{a}\right)\frac{1-q}{4}\frac{T-T_1}{T}\frac{\gamma}{4}\delta\\&\geq\frac{1-q}{32}\gamma\delta\varphi\left(\overline{a}\right)\\&=2\eta.\end{aligned}$$

Thus,

$$v_\gamma^T\geq U_\gamma^T\left(\vec{b}\right)\geq U_\gamma^T\left(\vec{a}\right)+2\eta\geq V_\gamma^*-\eta+2\eta=V_\gamma^*+\eta$$

for all sufficiently large T that are multiples of m. In particular, $\overline{V}_\gamma=\lim_T v_\gamma^T\geq V_\gamma^*+\eta$, which completes the proof of the theorem.

The construction in the proof of Theorem 6.4 also sheds some light on why the optimal limit superior cannot be achieved by a stationary strategy. As the decision maker who optimizes it only cares about the average payoff at some periods, she can first refrain for a long time from playing the good actions, only to play it later with a smaller fatigue punishment. The limit superior does not depend on all these periods where her average payoff is low, but only takes into account the few (but still infinitely many) periods that she optimizes.

6.3. A strategy that achieves $\overline{V}_\gamma$

Taking advantage of Proposition 6.2, we now construct a strategy that guarantees a payoff close to $\overline{V}_\gamma$ over an infinite set of periods. Let $\{T_k\}_k$ be a sequence of times such that

$$\frac{\sum_{j<k}T_j}{T_k}<\frac{1}{k^3}.$$

For every k, let $\vec{a}(T_k)\in A^{T_k}$ be a history of length T_k that achieves $v_\gamma^{T_k}$. That is, $v_\gamma^{T_k}=U_\gamma^{T_k}\left(\vec{a}(T_k)\right)$. We construct an infinite sequence $\vec{a}$ by concatenating these finite sequences one after another. Let,

$$\vec{a}=(\vec{a}(T_1),\vec{a}(T_2),\dots). \tag{38}$$

The next proposition shows that $\vec{a}$ achieves $\overline{V}_\gamma$ as its limit superior. Specifically, it demonstrates that the supremum in the definition of $\overline{V}_\gamma$ (see Equation (1)) can be replaced by a maximum.

Proposition 6.5. *Let $\vec{a}$ be defined as in (38). Then $\overline{V}_\gamma(\vec{a})=\overline{V}_\gamma$.*

The proof appears in the Appendix.

7. Discounting

In the context of a taste for variety, discounting can have two meanings. The first is the classical discounting of future payoffs. This means that we value future positive payoffs less than present ones. For example, we would rather have a delicious meal today than a week from now. The second is a discount on the effect of past uses of actions. As before, this means that the more we experience something, the less we enjoy it. For example, if we eat the same meal every day, we will eventually get tired of it. However, if we had a delicious meal yesterday, we would prefer the same meal today less than if we had it only a year ago.

This section focuses on discounting the effects of the past. We will analyze optimal behavior both in the case where future payoffs are discounted and in the case where the decision maker maximizes the limit inferior of average payoffs as before. To take the meaning of a discounted past into account, we define for a discount factor $\lambda \in (0,1)$ the *discounted frequency* of a in the history $\vec{a}^{t-1}$ as

$$\varphi^{\lambda}\left(a\middle|\vec{a}^{t-1}\right)=\begin{cases}\frac{1-\lambda}{1-\lambda^{t-1}}\sum_{s=1}^{t-1}\lambda^{t-s-1}\mathbb{1}_{\vec{a}_s=a}, & \text{if } t\geq 2,\\ 0, & \text{if } t=1.\end{cases}$$

7.1. Discounting future payoffs and past frequencies

A decision maker who discounts both the past and future payoffs derives from $\vec{a}$ the payoff

$$U_{\gamma}^{\lambda,\delta}\left(\vec{a}\right)=(1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}\left(1-\gamma\varphi^{\lambda}\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(a_t\right),$$

where $\delta > 0$ is the future discount factor.

Let

$$V_{\gamma}^{\lambda,\delta}=\max_{\vec{a}} U_{\gamma}^{\lambda,\delta}\left(\vec{a}\right).$$

The maximum exists since $U_{\gamma}^{\lambda,\delta}(\cdot)$ is a continuous function defined on the compact set consisting of all infinite histories.[3]

When choosing an action, the decision maker considers two distinct implications for overall utility. First, the action directly affects the immediate utility in the current period. Second, it influences future payoffs through the fatigue factor associated with repeated use of the same action. This decision process is analogous to that in a standard Markov Decision Problem and is therefore fully consistent with the principle of dynamic consistency. In other words, the decision can be made *ex ante* by planning all future actions in advance. Then, at any point in time, the decision made at the outset remains optimal.

The following theorem states that if the discount on frequency is sufficiently large, then for a sufficiently patient decision maker, the best achievable payoff approaches the optimal stationary payoff, V_{γ}^{*}. Specifically, for every $\varepsilon > 0$, there exists a lower bound λ_0 such that when $\lambda > \lambda_0$ and δ is sufficiently large, $V_{\gamma}^{\lambda,\delta}$ does not exceed V_{γ}^{*} by much. This situation exemplifies another instance where cyclical consumption is optimal.

The proof technique resembles that of Theorem 5.2; the main idea centers on defining a potential function H as in (41). However, here there is no need to partition the periods into blocks. Instead, the discount factor δ allows for a definition that enables two expressions for H, forming the key step of the proof in Equation (43). After establishing this equation, the proof proceeds by analyzing its two sides to obtain the desired result.

Theorem 7.1. *There is a function $\delta(\lambda) < 1$ s.t. for every $\varepsilon > 0$ there is λ_0 satisfying*

$$V_{\gamma}^{*}\leq V_{\gamma}^{\lambda,\delta}<V_{\gamma}^{*}+\varepsilon,$$

for every $\lambda > \lambda_0$ and $\delta > \delta(\lambda)$.

Proof. As γ is fixed, for simplicity we drop it from the notation, that is, $V^{\lambda,\delta}=V_{\gamma}^{\lambda,\delta}$ and $V^{*}=V_{\gamma}^{*}$. Clearly, $V^{\lambda,\delta}\geq V^{*}$ for sufficiently large λ and δ. This is so, because $U^{\lambda,\delta}\left(\vec{a}\right)=V^{*}$ for a stationary history $\vec{a}$ that achieves V^{*}. We show the inverse direction. Let $\vec{a}$ be an arbitrary sequence and let $\varepsilon > 0$. We show that for sufficiently large λ and δ,

$$U^{\lambda,\delta}\left(\vec{a}\right)<V^{*}+\varepsilon. \tag{39}$$

[3] The set of all histories is the infinite product A^{∞}, i.e., the Cartesian product of A with itself countably many times. When endowed with the product topology, this set is compact by Tychonoff's Theorem. Moreover, the function $U_{\gamma}^{\lambda,\delta}(\cdot)$ is continuous on this space. Intuitively, two histories that coincide over a sufficiently long initial segment yield similar payoffs. This stands in sharp contrast to the functions $\underline{V}_{\gamma}$ and $\overline{V}_{\gamma}$ (see Footnote 2). Although they are defined on the same compact set, these functions are not continuous: their values are determined by the "tail" of the action sequences, whereas $U_{\gamma}^{\lambda,\delta}(\cdot)$ depends essentially on finite, though sufficiently long, prefixes of the histories.

To simplify the proof, we use the notation $\varphi^{t-1,\lambda}(a) = \varphi^{\lambda}\left(a \middle| \vec{a}^{t-1}\right)$. Note that for every $t \geq 2$, $\sum_a \varphi^{t-1,\lambda}(a) = 1$. We let $\varphi^{t-1,\lambda}$ be the probability distribution that assigns probability $\varphi^{t-1,\lambda}(a)$ to a. Denote also $\eta^t = (1-\delta)\delta^{t-1}$ and $\beta^t = \frac{1-\lambda}{1-\lambda^t}$. Finally, let $\mathbf{1}^t$ stand for the unit A-dimensional vector assigning 1 to a when $a_t = a$, and 0 to all other members of A. Using these notations and the inner product introduced in (16), we have

$$U^{\lambda,\delta}(\vec{a}) = \sum_{t=1}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle. \tag{40}$$

Denote

$$H := H(\vec{a}) = \sum_{t=1}^{\infty} \frac{\eta^t}{\beta^t} \left\| \mathbf{1} - \gamma\varphi^{t,\lambda} \right\|^2. \tag{41}$$

One easily verifies that $\varphi^{t,\lambda} = \varphi^{t-1,\lambda} + \beta^t\left(\mathbf{1}^t - \varphi^{t-1,\lambda}\right)$. Hence, with $\epsilon_1 = (1-\delta)\left\|\mathbf{1} - \gamma\varphi^{1,\lambda}\right\|$, one obtains

$$\begin{aligned}
H &= \sum_{t=1}^{\infty} \frac{\eta^t}{\beta^t} \left\| \mathbf{1} - \gamma\varphi^{t,\lambda} \right\|^2 \\
&= \epsilon_1 + \sum_{t=2}^{\infty} \frac{\eta^t}{\beta^t} \left\| \left(\mathbf{1} - \gamma\left(\varphi^{t-1,\lambda}\right)\right) - \gamma\beta^t\left(\mathbf{1}^t - \varphi^{t-1,\lambda}\right) \right\|^2 \\
&= \epsilon_1 + \sum_{t=2}^{\infty} \frac{\eta^t}{\beta^t} \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 - 2\sum_{t=2}^{\infty} \frac{\eta^t}{\beta^t} \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \gamma\beta^t\left(\mathbf{1}^t - \varphi^{t-1,\lambda}\right) \rangle \\
&\qquad + \gamma^2 \sum_{t=2}^{\infty} \frac{\eta^t}{\beta^t} \left\| \beta^t\left(\mathbf{1}^t - \varphi^{t-1,\lambda}\right) \right\|^2 \\
&= \epsilon_1 + \sum_{t=2}^{\infty} \left(\frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}}\right) \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 + \sum_{t=2}^{\infty} \frac{\eta^{t-1}}{\beta^{t-1}} \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 \\
&\qquad - 2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle + 2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \varphi^{t-1,\lambda} \rangle \\
&\qquad + \gamma^2 \sum_{t=2}^{\infty} \eta^t\beta^t \left\| \mathbf{1}^t - \gamma\varphi^{t-1,\lambda} \right\|^2.
\end{aligned} \tag{42}$$

Since $\sum_{t=2}^{\infty} \frac{\eta^{t-1}}{\beta^{t-1}} \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 = H$, we can observe that H appears on both sides of equation (42). This allows us, after rearranging equation (42), to derive the key equation of the proof:

$$\begin{aligned}
2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle &= \epsilon_1 + \sum_{t=2}^{\infty} \left(\frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}}\right) \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 + 2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \varphi^{t-1,\lambda} \rangle \\
&\qquad + \gamma^2 \sum_{t=2}^{\infty} \eta^t\beta^t \left\| \mathbf{1}^t - \gamma\varphi^{t-1,\lambda} \right\|^2.
\end{aligned} \tag{43}$$

The remainder of the proof is devoted to analyzing both sides of Equation (43). We start with the left-hand side. Note that, by the definition of $U^{\lambda,\delta}$,

$$2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle = 2\gamma U^{\lambda,\delta}(\vec{a}) - 2\gamma(1-\delta)u(\vec{a}_1).$$

Thus, there is δ_0 such that for all $\delta > \delta_0$ we have

$$2\gamma U^{\lambda,\delta}(\vec{a}) \leq 2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle + \frac{\varepsilon\gamma}{3}. \tag{44}$$

This completes the analysis of the left-hand side of (43), and we turn to the expressions on the right-hand side of it.

As $\epsilon_1 = \frac{\eta^1}{\beta^1} \left\| \mathbf{1} - \gamma\varphi^{1,\lambda} \right\|^2 = 1-\delta$, there is δ_1 such that $\epsilon_1 < \varepsilon\gamma/3$ for all $\delta > \delta_1$. We show that the first sum on the right-hand side of (43) is, for λ and δ sufficiently close to 1, bounded by the same constant. For this purpose note first that $\left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 \leq 1$ uniformly. Next, observe that

$$\sum_{t=2}^{\infty} \left| \frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}} \right| \leq \frac{1-\delta}{1-\lambda} \sum_{t=1}^{\infty} \left| \delta^t\left(1-\lambda^{t+1}\right) - \delta^{t-1}\left(1-\lambda^t\right) \right|$$

$$= \frac{1-\delta}{1-\lambda}\sum_{t=1}^{\infty}\delta^{t-1}\left|\delta\left(1-\lambda^{t+1}\right)-\left(1-\lambda^{t}\right)\right|. \tag{45}$$

Let T be the largest integer such that $\delta(1-\lambda^{t+1}) \geq (1-\lambda^t)$. (One easily checks that T is well defined.) The right-hand side of (45) is then bounded from above by

$$\begin{aligned}
&\frac{1-\delta}{1-\lambda}\sum_{t=1}^{T}\delta^{t-1}\left(\delta\left(1-\lambda^{t+1}\right)-\left(1-\lambda^{t}\right)\right)+\frac{1-\delta}{1-\lambda}\sum_{t=T+1}^{\infty}\delta^{t-1}\left(\left(1-\lambda^{t}\right)-\delta\left(1-\lambda^{t+1}\right)\right)\\
&\leq \frac{1-\delta}{1-\lambda}\sum_{t=1}^{T}\delta^{t-1}\left(\left(1-\lambda^{t+1}\right)-\left(1-\lambda^{t}\right)\right)+\frac{1-\delta}{1-\lambda}\sum_{t=T+1}^{\infty}\delta^{t-1}\left(\left(1-\lambda^{t+1}\right)-\delta\left(1-\lambda^{t+1}\right)\right)\\
&\leq (1-\delta)\sum_{t=1}^{T}\delta^{t-1}\lambda^{t}+\frac{1-\delta}{1-\lambda}\sum_{t=T+1}^{\infty}\delta^{t-1}\left(1-\lambda^{t+1}-\delta\left(1-\lambda^{t+1}\right)\right)\\
&\leq \lambda\frac{1-\delta}{1-\lambda\delta}+\frac{(1-\delta)^2}{1-\lambda}\sum_{t=T+1}^{\infty}\delta^{t-1}\left(1-\lambda^{t+1}\right)\leq \frac{1-\delta}{1-\lambda\delta}+\frac{(1-\delta)^2}{1-\lambda}\sum_{t=T+1}^{\infty}\delta^{t-1}\\
&\leq \frac{1-\delta}{1-\lambda\delta}+\frac{1-\delta}{1-\lambda}\\
&\leq 2\frac{1-\delta}{1-\lambda}.
\end{aligned} \tag{46}$$

For any λ, there exists a function $\delta_2(\lambda)$ such that, when $\delta > \delta_2(\lambda)$, the inequality $2\frac{1-\delta_1(\lambda)}{1-\lambda} < \frac{\varepsilon\gamma}{3}$ holds. Consequently, using (45) and (46), we find

$$\begin{aligned}
\sum_{t=2}^{\infty}\left(\frac{\eta^t}{\beta^t}-\frac{\eta^{t-1}}{\beta^{t-1}}\right)\left\|1-\gamma\varphi^{t-1,\lambda}\right\|^2 &\leq \frac{1-\delta}{1-\lambda}\sum_{t=1}^{\infty}\delta^{t-1}\left|\delta\left(1-\lambda^{t+1}\right)-\left(1-\lambda^{t}\right)\right|\leq 2\frac{1-\delta}{1-\lambda}\\
&< \frac{\varepsilon\gamma}{3}.
\end{aligned} \tag{47}$$

So far for the first term on the right-hand side of (43).

As for the second sum on the right-hand side of (43), recall that for every t the vector $\varphi^{t-1,\lambda}$ is a distribution over A, so that by the definition of V^* we have, for every t, $\left\langle \mathbf{1}-\varphi^{t-1,\lambda},\varphi^{t-1,\lambda}\right\rangle \leq V^*$. Thus,

$$2\gamma\sum_{t=2}^{\infty}\eta^t\left\langle \mathbf{1}-\gamma\varphi^{t-1,\lambda},\varphi^{t-1,\lambda}\right\rangle \leq 2\gamma V^*\sum_{t=2}^{\infty}\eta^t = 2\gamma V^*(1-\delta)\sum_{t=2}^{\infty}\delta^{t-1}\leq 2\gamma V^*. \tag{48}$$

For the last sum on the right-hand side of (43), first note that

$$\sum_{t=2}^{\infty}\eta^t\beta^t = \sum_{t=2}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}\leq \sum_{t=1}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}. \tag{49}$$

For all $\lambda < 1$ there is t^* such that $\frac{1-\lambda}{1-\lambda^t} \leq 1-\lambda+\frac{\varepsilon\gamma}{6}$ for all $t \geq t^*$. Indeed, note that $\frac{1-\lambda}{1-\lambda^t} = \left(\sum_{s=0}^{t-1}\lambda^s\right)^{-1} \longrightarrow 1-\lambda$ as $t\to\infty$. Moreover, for each t^* there $\delta' < 1$ such that $\sum_{s=1}^{t^*}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t} \leq \frac{\varepsilon\gamma}{6}$ for all $\delta > \delta'$. Hence, for each $\lambda < 1$ there is $\delta_3(\lambda)$ such that by (49) one obtains,

$$\begin{aligned}
\gamma^2\sum_{t=2}^{\infty}\eta^t\beta^t\left\|\mathbf{1}^t-\gamma\varphi^{t-1,\lambda}\right\|^2 &\leq \sum_{t=2}^{\infty}\eta^t\beta^t\\
&\leq \sum_{t=1}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}\\
&= \sum_{s=1}^{t^*}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}+\sum_{t=t^*+1}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}\\
&\leq \frac{\varepsilon\gamma}{6}+(1-\delta)\left(1-\lambda+\frac{\varepsilon\gamma}{6}\right)\sum_{t=t^*+1}^{\infty}\delta^t\\
&\leq 1-\lambda+\varepsilon\gamma/3\\
&\leq \frac{2\varepsilon\gamma}{3},
\end{aligned} \tag{50}$$

for all $\lambda \geq 1-\frac{\varepsilon\gamma}{3}$ and $\delta \geq \delta_3(\lambda)$.

We now combine (44), which bounds the left-hand side of (43), with (47), (48), and (50), which bound the various terms on the right-hand side, to obtain

$$2\gamma U^{\lambda,\delta}\left(\vec{a}\right) \le 2\gamma \sum_{t=2}^{\infty} \eta^t \left\langle \mathbf{1} - \gamma \varphi^{t-1,\lambda}, \mathbf{1}^t \right\rangle + \frac{\epsilon\gamma}{3} \le \frac{\epsilon\gamma}{3} + \frac{\epsilon\gamma}{3} + 2\gamma V^* + \frac{2\epsilon\gamma}{3} + \frac{\epsilon\gamma}{3}$$
$$= \frac{5\epsilon\gamma}{3} + 2\gamma V^*.$$

Dividing by 2γ yields $U^{\lambda,\delta}\left(\vec{a}\right) \le V^* + \epsilon$, as required. □

An immediate consequence of Theorem 7.1 is that a sufficiently patient decision maker can not achieve significantly more than what a stationary strategy would deliver. More precisely, a discounting decision maker can achieve V^*_γ by playing a stationary history. But for each $\epsilon > 0$ there are λ and δ such that a decision maker who discounts the past with factor at least λ and future payoffs with factor at least δ will not achieve a discounted payoff of more than $V^*_\gamma + \epsilon$.

7.2. An undiscounted evaluation of payoffs while frequencies are discounted

We now consider a hybrid model where the frequency is discounted while the future payoffs are not. As before, λ is the discount on the frequency of past actions. We fix the fatigue factor γ and for convenience, we omit it from the notations. Define

$$U^{\lambda,T}_\gamma\left(\vec{a}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(1 - \gamma\varphi^\lambda\left(\vec{a}_t \middle| \vec{a}^{t-1}\right)\right) u\left(\vec{a}_t\right)$$

and let

$$\underline{V}^\lambda_\gamma\left(\vec{a}\right) = \liminf_{T\to\infty} U^{\lambda,T}_\gamma\left(\vec{a}\right) \qquad \text{and} \qquad \underline{V}^\lambda_\gamma = \sup\left\{\underline{V}^\lambda_\gamma\left(\vec{a}\right) \middle| \vec{a} \in A^\infty\right\}$$

$\underline{V}^\lambda_\gamma$ represents the best achievable limit (inferior) of the finite averages when the frequency of past actions is discounted by λ. As a consequence of Theorem 7.1, we show that for sufficiently large λ, this limit cannot exceed V^* by a significant amount. This implies that the best achievable payoff with stationary strategies is nearly optimal in the current model.

Proposition 7.2. *For every $\epsilon > 0$ there is $\lambda_0 < 1$ such that for every $\lambda > \lambda_0$,*

$$V^*_\gamma \le \underline{V}^\lambda_\gamma < V^*_\gamma + \epsilon. \tag{51}$$

The proof is deferred to the Appendix.

8. Summary

We have analyzed a dynamic decision problem where a decision maker's utility from a particular action decreases as the action is used more frequently. This phenomenon is captured through a fatigue factor, quantifying the extent to which utility diminishes with repeated use of the same action. The optimal strategy involves a stationary approach in behaviorally significant scenarios, such as when utility is evaluated using the limit inferior or when discounting is applied. In other words, periodic consumption emerges as the most effective strategy.

The decision problem is a Markov Decision Process with countably many states. We introduce a novel analytical technique based on potential functions to address its complexity. These potential functions provide a powerful tool for deriving the key results, allowing us to characterize the optimal strategies and establish conditions under which periodic consumption is optimal.

Our approach not only simplifies the analysis but also offers insights into how the decision maker can balance the trade-offs between immediate rewards and long-term utility, especially under conditions of diminishing returns when an action is repeatedly applied. By employing potential functions, we provide a systematic method for identifying optimal policies in complex decision-making scenarios characterized by fatigue effects.

Declaration of competing interest

Declarations of interest: none.

Appendix A

The proof of Lemma 5.1. By the definition of W_γ, we have

$$W_\gamma = \frac{1}{t_2 - t_1}\sum_{s=t_1+1}^{t_2}\left(1 - \gamma\varphi\left(\vec{a}_s \middle| \vec{a}^{s-1}\right)\right) u(\vec{a}_s)$$

$$
\begin{aligned}
&= \frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\sum_{a\in A}\mathbb{1}_{\vec{a}_s=a}u(a) - \frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\gamma\varphi\left(\vec{a}_s\middle|\vec{a}^{s-1}\right)u(\vec{a}_s)\\
&= \sum_{a\in A}u(a)\left(\frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\mathbb{1}_{\vec{a}_s=a}\right) - \frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\gamma\varphi\left(\vec{a}_s\middle|\vec{a}^{s-1}\right)u(\vec{a}_s)\\
&= \sum_{a\in A}u(a)p(a) - \frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\gamma\varphi\left(\vec{a}_s\middle|\vec{a}^{s-1}\right)u(\vec{a}_s). \qquad (52)
\end{aligned}
$$

Furthermore, for every $a\in A$ and $t_1+1\le s\le t_2$,

$$
\begin{aligned}
\left|\varphi\left(a\middle|\vec{a}^{s-1}\right)-\varphi\left(a\middle|\vec{a}^{t_1}\right)\right| &= \left|\frac{1}{s-1}\sum_{r=1}^{s-1}\mathbb{1}_{\vec{a}_r=a} - \frac{1}{t_1}\sum_{r=1}^{t_1}\mathbb{1}_{\vec{a}_r=a}\right|\\
&= \left|\frac{1}{t_1}\frac{t_1}{s-1}\sum_{r=1}^{t_1}\mathbb{1}_{\vec{a}_r=a} - \frac{1}{t_1}\sum_{r=1}^{t_1}\mathbb{1}_{\vec{a}_r=a} + \frac{1}{s-1}\sum_{r=t_1+1}^{s-1}\mathbb{1}_{\vec{a}_r=a}\right|\\
&\le \left|\frac{1}{t_1}\left(\frac{t_1}{s-1}-1\right)\sum_{r=1}^{t_1}\mathbb{1}_{\vec{a}_r=a}\right| + \frac{1}{s-1}\sum_{r=t_1+1}^{s-1}1\\
&= \frac{s-1-t_1}{s-1}\varphi\left(a\middle|\vec{a}^{t_1}\right) + \frac{s-1-t_1}{s-1}\\
&< 2\frac{t_2-t_1}{t_1},
\end{aligned}
$$

where in the last step we use that $t_1\le s-1\le t_2$ and $\varphi\left(a\middle|\vec{a}^{t_1}\right)\le 1$. From (10) and (52) we now obtain

$$
\begin{aligned}
\left|W_\gamma-\widetilde{U}_\gamma\right| &= \gamma\left|\sum_{a\in A}p(a)\varphi\left(a\middle|\vec{a}^{t_1}\right)u(a) - \frac{1}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\varphi\left(\vec{a}_s\middle|\vec{a}^{s-1}\right)u(\vec{a}_s)\right|\\
&\le \gamma\sum_{a\in A}\frac{u(a)}{t_2-t_1}\left|\varphi\left(a\middle|\vec{a}^{t_1}\right)\sum_{s=t_1+1}^{t_2}\mathbb{1}_{\vec{a}_s=a} - \sum_{s=t_1+1}^{t_2}\mathbb{1}_{\vec{a}_s=a}\varphi\left(a\middle|\vec{a}^{s-1}\right)\right|\\
&\le \gamma\sum_{a\in A}\frac{u(a)}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\mathbb{1}_{\vec{a}_s=a}\left|\varphi\left(a\middle|\vec{a}^{t_1}\right)-\varphi\left(a\middle|\vec{a}^{s-1}\right)\right|\\
&< \gamma\sum_{a\in A}\frac{u(a)}{t_2-t_1}\sum_{s=t_1+1}^{t_2}\mathbb{1}_{\vec{a}_s=a}2\frac{t_2-t_1}{t_1}\\
&= 2\frac{t_2-t_1}{t_1}\gamma\sum_{a\in A}u(a)p(a)\\
&\le 2\frac{t_2-t_1}{t_1}\gamma\sum_{a\in A}u(a)
\end{aligned}
$$

as required. □

The proof of Lemma 5.3. First, observe that if $x_\gamma\left(a_i\right)=0$ for some i, $x_\gamma\left(a_j\right)=0$ for all $j\ge i$. Indeed, if $u\left(a_i\right)>u\left(a_j\right)$, this is immediately clear. If $u\left(a_i\right)=u\left(a_j\right)$, then let $y_\gamma\left(a_i\right)=x_\gamma\left(a_j\right)$, $y_\gamma\left(a_j\right)=x_\gamma\left(a_i\right)$ and $y_\gamma(a)=x_\gamma(a)$ for all $a\ne a_i,a_j$. Then y_γ is a solution of (6), contradicting uniqueness. This means that $\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)\le 0$, with equality if $\lambda_{k^*+1}>0$, and $\sum_{i=1}^{k}\frac{d}{d\gamma}x^\gamma\left(a_i\right)=0$ for all $k\ge k^*+1$.

It remains to prove the claim for $k<k^*$. Let $A^*=\{a\in A:\ x^\gamma(a)>0\}=\left\{a_1,\dots,a_{k^*}\right\}$. The Lagrangian of maximization problem (6) is

$$
\sum_{i=1}^{m}x\left(a_i\right)\left(1-x\left(a_i\right)\right)u\left(a_i\right)+\lambda_i x\left(a_i\right)-\mu\left(\sum_{i=1}^{m}x\left(a_i\right)-1\right),
$$

with first-order conditions

$$
u\left(a_i\right)\left(1-2\gamma x\left(a_i\right)\right)+\lambda_i-\mu=0.
$$

for $i=1,\dots,m$. For all i we either have $x\left(a_i\right)=0$ or $\lambda_i=0$; in the latter case

$$u\left(a_i\right)\left(1-2\gamma x\left(a_i\right)\right)=\mu.$$

Summing over all $a_i \in A^*$, solving for μ and substituting in we find that

$$u(a)(1-2\gamma x(a))=\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)\left(1-2\gamma x\left(a_i\right)\right)=\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)-2\gamma\sum_{i=1}^{k^*}x\left(a_i\right)u\left(a_i\right).$$

So, x_γ satisfies for all $k=1,\dots,k^*$,

$$x_\gamma\left(a_k\right)=\frac{1}{2\gamma u\left(a_k\right)}\left(u\left(a_k\right)-\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)\right)+\frac{1}{u\left(a_k\right)}\sum_{i=1}^{k^*}x^\gamma\left(a_i\right)u\left(a_i\right).$$

Taking the derivative of both sides with respect to γ gives

$$\frac{d}{d\gamma}x^\gamma\left(a_k\right)=\frac{1}{2\gamma^2}\left(\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)-u\left(a_k\right)\right)+\frac{1}{u\left(a_k\right)}\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)u\left(a_i\right). \tag{53}$$

Suppose first that $\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)u\left(a_i\right)>0$. Then the right-hand side of (53) is increasing in k, as $u\left(a_k\right)$ is decreasing in k. Thus, in particular,

$$\frac{d}{d\gamma}x^\gamma\left(a_1\right)\le\frac{d}{d\gamma}x^\gamma\left(a_2\right)\le\cdots\le\frac{d}{d\gamma}x^\gamma\left(a_{k^*}\right). \tag{54}$$

Suppose that (24) does not hold. Then there is $k<k^*$ such that $\sum_{i=1}^{k}\frac{d}{d\gamma}x^\gamma\left(a_i\right)>0$. Thus, by (54), we must have $\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)\ge\sum_{i=1}^{k}\frac{d}{d\gamma}x^\gamma\left(a_i\right)>0$, which is impossible. Suppose next that $\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)u\left(a_i\right)\le 0$. Then, for all $\ell\le k^*$,

$$\begin{aligned}\sum_{k=1}^{\ell}\frac{d}{d\gamma}x^\gamma\left(a_k\right)&=\sum_{k=1}^{\ell}\frac{1}{2\gamma^2}\left(\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)-u\left(a_k\right)\right)+\sum_{k=1}^{\ell}\frac{1}{u\left(a_k\right)}\sum_{i=1}^{k^*}\frac{d}{d\gamma}x^\gamma\left(a_i\right)u\left(a_i\right)\\&\le\frac{1}{2\gamma^2}\left(\frac{\ell}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)-\sum_{k=1}^{\ell}u\left(a_k\right)\right)\\&\le 0,\end{aligned}$$

where the last inequality holds because $\frac{1}{k^*}\sum_{i=1}^{k^*}u\left(a_i\right)\le\frac{1}{\ell}\sum_{k=1}^{\ell}u\left(a_k\right)$ for all $\ell\le k^*$. □

The proof of Lemma 6.3. By the conditions on $\vec{a}$ and $\vec{b}$,

$$\sum_{r=1}^{s-1}\mathbb{1}_{\vec{a}_r=b}\ge\sum_{r=1}^{t-1}\mathbb{1}_{\vec{b}_r=b}=\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=b}$$

and

$$\sum_{r=1}^{s-1}\mathbb{1}_{\vec{b}_r=a}=\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=a}.$$

Thus,

$$\begin{aligned}T\left(U_\gamma^T(\vec{b})-U_\gamma^T(\vec{a})\right)&=\left(1-\gamma\varphi\left(b\middle|\vec{b}^{t-1}\right)\right)u(b)+\left(1-\gamma\varphi\left(a\middle|\vec{b}^{s-1}\right)\right)u(a)\\&\qquad-\left(1-\gamma\varphi\left(a\middle|\vec{a}^{t-1}\right)\right)u(a)-\left(1-\gamma\varphi\left(b\middle|\vec{a}^{s-1}\right)\right)u(b)\\&=\gamma\left(\left(\varphi\left(b\middle|\vec{a}^{s-1}\right)-\varphi\left(b\middle|\vec{b}^{t-1}\right)\right)u(b)+\left(\varphi\left(a\middle|\vec{a}^{t-1}\right)-\varphi\left(a\middle|\vec{b}^{s-1}\right)\right)u(a)\right)\\&=\gamma\left(\left(\frac{1}{s-1}\sum_{r=1}^{s-1}\mathbb{1}_{\vec{a}_r=b}-\frac{1}{t-1}\sum_{r=1}^{t-1}\mathbb{1}_{\vec{b}_r=b}\right)u(b)\right.\\&\qquad\left.+\left(\frac{1}{t-1}\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=a}-\frac{1}{s-1}\sum_{r=1}^{s-1}\mathbb{1}_{\vec{b}_r=a}\right)u(a)\right)\\&\ge\frac{\gamma}{(s-1)(t-1)}\left((t-s)\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=b}u(b)+(s-t)\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=a}u(a)\right)\\&=\frac{\gamma(s-t)}{s-1}\left(\varphi\left(a\mid\vec{a}^{t-1}\right)u(a)-\varphi\left(b\mid\vec{a}^{t-1}\right)u(b)\right)\end{aligned}$$

as required. □

The proof of Claim 1. Let $t \geq 1$. There is $k \in \mathbb{N}$ such that $km \leq t \leq (k+1)m$. At t, a was chosen at least km_a times, but no more than $km_a + (t - km)$ times. Thus,

$$\varphi\left(a \mid \vec{a}^t\right) \geq \frac{km_a}{t} = \frac{km}{t}\varphi(a) = \varphi(a) - \frac{t-km}{t} \geq \varphi(a) - \frac{(k+1)m-km}{t} = \varphi(a) - \frac{m}{t}$$

and

$$\varphi\left(a \mid \vec{a}^t\right) \leq \frac{km_a + t - km}{t} = \frac{km}{t}\varphi(a) + \frac{t-km}{t} \leq \varphi(a) + \frac{(k+1)m-km}{t} = \varphi(a) + \frac{m}{t}.$$

Shifting from t to $t-1$ completes the proof. □

The proof of Claim 2. Suppose first that in history $\vec{b}$, there is between $T_1 + 1$ and T no occurrence of $\overline{a}$ before $\underline{a}$. Let s be the last occurrence of $\bar{a}$ and let s^* be the largest multiple of m with $s^* \leq s \leq T$. As each occurrence of $\overline{a}$ between $T_1 + 1$ and s^* has been replaced by an occurrence of $\underline{a}$ that originally occurred after s^*, it holds that $\varphi\left(\underline{a}\right)(T - s^*) \geq \varphi\left(\overline{a}\right)\left(s^* - T_1\right)$. Thus,

$$2\varphi\left(\underline{a}\right)\left(s^* - T_1\right) \leq \left(\varphi\left(\underline{a}\right) + \varphi\left(\overline{a}\right)\right)\left(s^* - T_1\right) \leq \varphi\left(\underline{a}\right)\left(T - T_1\right),$$

which means that $s^* \leq \frac{1}{2}\left(T_1 + T\right)$. Hence, since by construction $T \geq T^* \geq T_1 + 4m$, we find that

$$s \leq s^* + m \leq \frac{1}{2}\left(T_1 + T\right) + \frac{1}{4}\left(T - T_1\right) \leq T_1 + q\left(T - T_1\right),$$

as required.

Suppose next that after all beneficial switches have been made, there is at least one occurrence of $\overline{a}$ before the last occurrence of $\underline{a}$. Let t be the period of said occurrence of $\overline{a}$. Then, since by the definition of $\vec{b}$ the switch of $\overline{a}$ with the last occurrence of $\underline{a}$ is not beneficial, we have

$$\delta \geq \varphi\left(\overline{a}\middle|\vec{b}^{t-1}\right)u\left(\overline{a}\right) - \varphi\left(\underline{a}\middle|\vec{b}^{t-1}\right)u\left(\underline{a}\right). \tag{55}$$

As there is only one occurrence of $\underline{a}$ in $\vec{b}$ between t and T, we have that $(t-1)\varphi\left(\underline{a}\middle|\vec{b}^{t-1}\right) = T\varphi\left(\underline{a}\right) - 1$, so that

$$\varphi\left(\underline{a}\middle|\vec{b}^{t-1}\right) = \frac{T}{t-1}\varphi\left(\underline{a}\right) - \frac{1}{t-1}. \tag{56}$$

Similarly,

$$(t-1)\,\varphi\left(\overline{a}\middle|\vec{b}^{t-1}\right) = (t-1)\varphi\left(\overline{a}\middle|\vec{a}^{t-1}\right) - \left(\left(T\varphi\left(\underline{a}\right) - 1\right) - (t-1)\varphi\left(\underline{a}\middle|\vec{a}^{t-1}\right)\right),$$

where the expression in the round brackets describes the number of occurrences of $\overline{a}$ that originally lay between $T_1 + 1$ and t but have been switched away for some $\underline{a}$ that originally occurred after t. Using the bounds that we derived in Claim 1, we find that

$$\begin{aligned}(t-1)\,\varphi\left(\overline{a}\middle|\vec{b}^{t-1}\right) &\geq (t-1)\left(\varphi\left(\overline{a}\right) - \frac{m}{t-1}\right) - \left(T\varphi\left(\underline{a}\right) - 1 - (t-1)\left(\varphi\left(\underline{a}\right) - \frac{m}{t-1}\right)\right)\\ &= (t-1)\varphi\left(\overline{b}\right) - (T-(t-1))\,\varphi\left(\underline{a}\right) - (2m-1)\end{aligned}$$

Therefore

$$\varphi\left(\overline{a}\middle|\vec{b}^{t-1}\right) \geq \varphi\left(\overline{a}\right) - \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right) - \frac{2m-1}{t-1}.$$

This, together with (35), (55), and (56) shows that

$$\begin{aligned}\delta &\geq \left(\varphi\left(\overline{a}\right) - \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right) - \frac{2m-1}{t-1}\right)u\left(\overline{a}\right) - \left(\frac{T}{t-1}\varphi\left(\underline{a}\right) - \frac{1}{t-1}\right)u\left(\underline{a}\right)\\ &= \varphi\left(\overline{a}\right)u\left(\overline{a}\right) - \varphi\left(\underline{a}\right)u\left(\underline{a}\right) - \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right) + u\left(\underline{a}\right)\right) - \frac{2m-1}{t-1}u\left(\overline{a}\right)\\ &= \frac{u\left(\overline{a}\right) - u\left(\underline{a}\right)}{2\gamma} - \varphi\left(\underline{a}\right)u\left(\underline{a}\right) - \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right) + u\left(\underline{a}\right)\right) - \frac{2m-1}{t-1}u\left(\overline{a}\right)\\ &= 2\delta - \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right) + u\left(\underline{a}\right)\right) - \frac{2m-1}{t-1}u\left(\overline{a}\right).\end{aligned}$$

Thus,

$$\delta \leq \frac{T-(t-1)}{t-1}\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right) + u\left(\underline{a}\right)\right) + \frac{2m-1}{t-1}u\left(\overline{a}\right)$$

and solving for t delivers

$$\begin{aligned} t &\leq T \frac{\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} + \frac{2m-1}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} + 1 \\ &= Tq' + \frac{2m-1}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} + 1. \end{aligned}$$

Let s be the period of the last occurrence of $\underline{a}$ in $\vec{b}$. Then $s \leq t+m$. Indeed, the sequence $\left(\vec{b}_{t+1},\dots,\vec{b}_{t+m}\right)$ contains at least $m_{\underline{a}}+m_{\overline{a}}$ periods in which either $\underline{a}$ or $\overline{a}$ is chosen, and at the first such period $\underline{a}$ is chosen by construction. Therefore,

$$\begin{aligned} s &\leq Tq' + \frac{2m-1}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} + 1 + m \\ &= \left(T-T_1\right)q' + T_1 - \left(1-q'\right)T_1 + \frac{2m-1}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} + 1 + m \\ &= \left(T-T_1\right)q' + T_1 + \frac{-\delta T_1 + 2m - 1 + (m+1)\left(\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)\right)}{\delta+\varphi\left(\underline{a}\right)\left(u\left(\overline{a}\right)+u\left(\underline{a}\right)\right)} \\ &\leq \left(T-T_1\right)q + T_1, \end{aligned}$$

where in the last step we use the lower bound for T_1 and $q \geq q'$. This concludes the proof. □

The proof of Proposition 6.5. Denote $S_k = \sum_{j\leq k} T_j$. Recall that by the choice of the sequence $\{T_k\}$, it holds that $\frac{S_{k-1}}{T_k} < \frac{1}{k^3}$. We show that $\left|U_\gamma^{S_{k+1}}\left(\vec{a}\right) - U_\gamma^{T_{k+1}}(\vec{a}(T_{k+1}))\right|$ is small and start with analyzing $U_\gamma^{S_{k+1}}\left(\vec{a}\right)$. Note that $S_{k+1} > S_k + (k+1)^3 S_k > (k+1)S_k$.

$$\begin{aligned} U_\gamma^{S_{k+1}}\left(\vec{a}\right) &= \frac{1}{S_{k+1}} \sum_{t=1}^{S_{k+1}} \left(1-\gamma\varphi\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(\vec{a}_t\right) \\ &= \frac{1}{S_{k+1}} \left(\sum_{t=1}^{(k+1)S_k} \left(1-\gamma\varphi\left(a_t\middle|\vec{a}^{t-1}\right)\right)u\left(a_t\right) + \sum_{t=(k+1)S_k+1}^{S_{k+1}} \left(1-\gamma\varphi\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(\vec{a}_t\right)\right). \end{aligned} \tag{57}$$

Thus,

$$\begin{aligned} \left|U_\gamma^{S_{k+1}}\left(\vec{a}\right) - U_\gamma^{T_{k+1}}(\vec{a}(T_{k+1}))\right| \leq{}& \frac{(k+1)S_k}{S_{k+1}} \max_{a\in A}[u(a)] \\ &+ \left|\frac{1}{S_{k+1}} \sum_{t=(k+1)S_k+1}^{S_{k+1}} \left(1-\gamma\varphi\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(\vec{a}_t\right)\right. \\ &\left. - \frac{1}{T_{k+1}} \sum_{t=1}^{T_{k+1}} \left(1-\gamma\varphi\left(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}\right)\right)u\left(\vec{a}(T_{k+1})_t\right)\right|. \end{aligned} \tag{58}$$

Since $T_{k+1} > (k+1)^3 S_k > k^2 S_k \geq k S_k$, the right term in the previous equation is

$$\begin{aligned} &\frac{1}{T_{k+1}} \sum_{t=1}^{T_{k+1}} (1-\gamma\varphi\left(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}\right))\,u\left(\vec{a}(T_{k+1})_t\right) \\ &= \frac{1}{T_{k+1}} \left(\sum_{t=1}^{kS_k} (1-\gamma\varphi(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}))\,u\left(\vec{a}(T_{k+1})_t\right)\right. \\ &\qquad\left. + \sum_{t=kS_k+1}^{T_{k+1}} (1-\gamma\varphi(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}))\,u\left(\vec{a}(T_{k+1})_t\right)\right) \\ &\leq \frac{kS_k}{T_{k+1}} \max_{a\in A}[u(a)] + \frac{T_{k+1}-kS_k}{T_{k+1}} \frac{1}{T_{k+1}-kS_k} \sum_{t=kS_k+1}^{T_{k+1}} (1-\gamma\varphi(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}))\,u\left(\vec{a}(T_{k+1})_t\right) \\ &\leq \frac{1}{k} \max_{a\in A}[u(a)] + \left(1-\frac{1}{k}\right) \frac{1}{T_{k+1}-kS_k} \sum_{t=kS_k+1}^{T_{k+1}} (1-\gamma\varphi(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}))\,u\left(\vec{a}(T_{k+1})_t\right) \end{aligned} \tag{59}$$

where the last inequality additionally uses that

$$\max_{a\in A}[u(a)] \geq \frac{1}{T_{k+1}-kS_k}\sum_{t=kS_k+1}^{T_{k+1}} \left(1-\gamma\varphi(\vec{a}(T_{k+1})_t \middle| \vec{a}(T_{k+1})^{t-1}\right) u\left(\vec{a}(T_{k+1})_t\right).$$

Since $\frac{kS_k}{S_{k+1}}\max_{a\in A}[u(a)] < \frac{1}{k}$ for k large enough, we obtain from (58) and (59) that

$$\begin{aligned}
&\left|U_\gamma^{S_{k+1}}\left(\vec{a}\right) - U_\gamma^{T_{k+1}}\left(\vec{a}(T_{k+1})\right)\right| \\
&\leq \frac{2}{k}\max_{a\in A}[u(a)] + \frac{1}{k} + \left|\frac{1}{S_{k+1}}\sum_{t=(k+1)S_k+1}^{S_{k+1}}\left(1-\gamma\varphi\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(\vec{a}_t\right)\right. \\
&\qquad \left. - \left(1-\frac{1}{k}\right)\frac{1}{T_{k+1}-kS_k}\sum_{t=kS_k+1}^{T_{k+1}}\left(1-\gamma\varphi\left(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}\right)\right)u\left(\vec{a}(T_{k+1})_t\right)\right|. \qquad (60)
\end{aligned}$$

Since $S_{k+1}-(k+1)S_k = S_{K+1}-S_k-kS_k = T_{k+1}-kS_k$, the two summations in the previous equation are of equal length. Moreover, they involve the same actions, that is, $a(T_{k+1})_t = a_{t+S_k}$ for every $t = kS_k+1,\ldots,T_{k+1}$. The only difference lies in the fatigue element preceding each utility term. However, since

$$\left|\varphi(\vec{a}(T_{k+1})_t \mid \vec{a}(T_{k+1})^{t-1}) - \varphi(\vec{a}_{t+S_k} \mid \vec{a}^{t+S_k-1})\right| \leq \frac{S_k}{t-1} \leq \frac{S_k}{kS_k} = \frac{1}{k},$$

for every $t = kS_k+1,\ldots,T_{k+1}$, it holds that

$$\begin{aligned}
&\left|\sum_{t=(k+1)S_k+1}^{S_{k+1}}\left(1-\gamma\varphi\left(\vec{a}_t\middle|\vec{a}^{t-1}\right)\right)u\left(\vec{a}_t\right) - \sum_{t=kS_k+1}^{T_{k+1}}\left(1-\gamma\varphi\left(\vec{a}(T_{k+1})_t\middle|\vec{a}(T_{k+1})^{t-1}\right)\right)u\left(\vec{a}(T_{k+1})_t\right)\right| \\
&\leq \left(T_{k+1}-kS_k\right)\frac{1}{k}\max_{a\in A}[u(a)]
\end{aligned}$$

As moreover,

$$\max\left(\frac{1}{S_{k+1}}, \left(1-\frac{1}{k}\right)\frac{1}{T_{k+1}-kS_k}\right) < \frac{1}{T_{k+1}-kS_k},$$

we obtain from (60) that

$$|U_\gamma^{S_{k+1}}\left(\vec{a}\right) - v_\gamma^{T_{k+1}}| = |U_\gamma^{S_{k+1}}\left(\vec{a}\right) - U_\gamma^{T_{k+1}}(\vec{a}(T_{k+1}))| < \frac{3}{k}\max_{a\in A}[u(a)] + \frac{1}{k}.$$

This shows that $\overline{V}_\gamma\left(\vec{a}\right) \geq \lim_k U_\gamma^{S_{k+1}}\left(\vec{a}\right) = \lim_k v_\gamma^{T_k} = \overline{V}_\gamma$. The last equality is due to Proposition 6.2. This completes the proof. □

The proof of Proposition 7.2. For convenience, we keep γ fixed and drop it from the notation. The inequality $V^* \leq \underline{V}^\lambda$ is clear. For the strict inequality, assume, to the contrary, that there exists $\varepsilon > 0$ such that $\underline{V}^\lambda > V^* + 3\varepsilon$ for an infinite sequence of λ values converging to 1. This assumption implies that, for each such λ, there exists an infinite history $\vec{a}(\lambda)$ such that

$$\underline{V}^{\lambda,T}\left(\vec{a}(\lambda)\right) > V^* + 3\varepsilon.$$

Therefore, we conclude that there exists a time $T_0(\lambda)$ such that $U^{\lambda,T}\left(\vec{a}(\lambda)\right) > V^* + 3\varepsilon$ for all $T \geq T_0(\lambda)$.

We now resort to Theorem 7.1 and apply it to ε. It states that there is λ, sufficiently close to 1, such that for any δ big enough,

$$V^{\lambda,\delta} < V^* + \varepsilon. \qquad (61)$$

From now on we fix this λ and denote $\vec{a} = \vec{a}(\lambda)$.

Recall that

$$U^{\lambda,\delta}\left(\vec{a}\right) = (1-\delta)\sum_{t=1}^{\infty}\delta^{t-1}\left(1-\gamma\varphi^\lambda\left(a_t\middle|\vec{a}^{t-1}\right)\right)u\left(a_t\right),$$

which is a discounted sum. As such (see Lehrer and Sorin (1992)), it can be expressed as a convex combination of the finite averages $U^{\lambda,T}(\vec{a})$. Specifically, there exist weights $w_T(\delta) \geq 0$ for $T = 1, 2, \ldots$, that sum to 1 such that

$$U^{\lambda,\delta}\left(\vec{a}\right) = \sum_T w_T(\delta)U^{\lambda,T}(\vec{a}). \qquad (62)$$

Moreover, for each T, we have $w_T(\delta) \to 0$ as $\delta \to 1$.

Let δ_0 be such that for any $\delta > \delta_0$,

$$\sum_{T<T_0(\lambda)} w_T(\delta)U^{\lambda,T}(\vec{a}) < \varepsilon,$$

and the total sum of the weights, excluding this prefix, is large, that is,

$$\sum_{T \geq T_0(\lambda)} w_T(\delta) > \frac{V^* + 2\varepsilon}{V^* + 3\varepsilon}.$$

Together with (62), this implies that

$$U^{\lambda,\delta}\left(\vec{a}\right) > \sum_{T \geq T_0(\lambda)} w_T(\delta) U^{\lambda,T}(\vec{a}) - \varepsilon > \left(\frac{V^* + 2\varepsilon}{V^* + 3\varepsilon}\right)(V^* + 3\varepsilon) - \varepsilon = V^* + \varepsilon.$$

Since $V^{\lambda,\delta} \geq U^{\lambda,\delta}\left(\vec{a}\right)$, this contradicts (61). We conclude that for every $\varepsilon > 0$ there exists $\lambda_0 < 1$ such that for every $\lambda > \lambda_0$, it holds that $\underline{V}^{\lambda} < V^* + \varepsilon$, as desired. □

Data availability

No data was used for the research described in the article.

References

Allais, M., 1953. Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine. Econometrica 21, 503–546.
Baucells, M., Weber, M., Welfens, F., 2011. Reference point formation and updating. Manag. Sci. 57, 506–519.
Becker, G.S., Murphy, K.M., 1988. A theory of rational addiction. J. Polit. Econ. 96 (4), 675–700.
Boldrin, M., Christiano, L.J., Fisher, J.D., 2001. Habit persistence, asset returns, and the business cycle. Am. Econ. Rev. 91, 149–166.
Bornstein, R.F., Kale, A.R., Cornell, K.R., 1990. Boredom as limiting condition on the mere exposure effect. J. Pers. Soc. Psychol. 58, 791–800.
Cerreia-Vioglio, S., Corrao, R., Lanzani, G., 2023. Nonlinear Fixed Points and Stationarity: Economic Applications. Available at SSRN 5107497.
Constantinides, G.M., 1990. Habit formation: a resolution to the equity premium puzzle. J. Polit. Econ. 98, 519–543.
Crisp, R.J., Hutter, R.R.C., Young, B., 2008. When mere exposure leads to less liking: the incremental threat effect in intergroup contexts. Br. J. Psychol. 100, 133–149.
Etner, J., Jeleva, M., Tallon, J.-M., 2012. Decision theory under ambiguity. J. Econ. Surv. 26, 234–270.
Gilboa, I., Schmeidler, D., 1995. Case-based decision theory. Q. J. Econ. 110 (3), 605–639.
Gonzales, C., Lerch, J.F., Lebiere, C., 2003. Instance-based learning in dynamic decision making. Cogn. Sci. 27, 501–635.
He, Y., Dyer, J.S., Butler, J.C., 2013. On the axiomatization of the satiation and habit formation utility models. Oper. Res. 61, 1399–1410.
Kaiser, K., Schwabe, R., 2012. Preference for variety. B. E. J. Theor. Econ. 12, 3.
Koopmans, T.C., 1960. Stationary ordinal utility and impatience. Econometrica, 287–309.
Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.
Lehrer, E., Sorin, S., 1992. A Tauberian theorem for stochastic games. Int. J. Game Theory 21 (1), 39–52.
Liao, H.-I., Yeh, S.-L., Shimojo, S., 2011. Novelty vs. Familiarity principles in preference decisions: task-context of past experience matters. Front. Psychol. 2, 43.
Machina, M.J., 1989. Dynamic consistency and non-expected utility models of choice under uncertainty. J. Econ. Lit. 27, 1622–1668.
Mehra, R., Prescott, E.C., 1985. The equity premium: a puzzle. J. Monet. Econ. 15, 145–161.
Al-Najjar, N.I., Shmaya, E., 2019. Recursive utility and parameter uncertainty. J. Econ. Theory 181, 274–288.
O'Donoghue, T., Sprenger, C., 2018. Reference-dependent preferences. In: Bernheim, B.D., DellaVigna, S., Laibson, D. (Eds.), Handbook of Behavioral Economics: Applications and Foundations I. Elsevier, pp. 1–77.
Rozen, K., 2010. Foundations of intrinsic habit formation. Econometrica 78 (4), 1341–1373.
Rustichini, A., Siconolfi, P., 2014. Dynamic theory of preferences: habit formation and taste for variety. J. Math. Econ. 55, 55–68.
Samuelson, P.A., 1937. A note on measurement of utility. Rev. Econ. Stud. 4, 155–161.
Sorin, S., 2002. A First Course on Zero-Sum Repeated Games, vol. 37. Springer Science & Business Media.
Stanca, L.M., 2024. Restricted Dynamic Consistency. Available at SSRN 4724184.
Stewart, N., Chater, N., Brown, G.D.A., 2006. Decision by sampling. Cogn. Psychol. 53, 1–26.
Thaler, R., 1981. Some empirical evidence on dynamic inconsistency. Econ. Lett. 8, 201–207.
Williams, S.M., 1987. Repeated exposure and the attractiveness of synthetic speech: an inverted U-relationship. Curr. Psychol. Res. Rev. 6, 148–154.
Zajonc, R.B., 1968. Attitudinal effects of mere exposure. J. Pers. Soc. Psychol. 9, 1–27.
Zajonc, R.B., Shaver, P., Carol, T., David, V.K., 1972. Exposure, satiation, and stimulus discriminability. J. Pers. Soc. Psychol. 21, 270–280.
Ziliotto, B., 2016. A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. Math. Oper. Res. 41 (4), 1522–1534.