A Taste for Variety*

Galit Ashkenazi-Golan† Dominik Karos‡ Ehud Lehrer§

April 13, 2024

Abstract

A decision maker repeatedly chooses one of a finite set of actions. In each period, the decision maker's payoff depends on 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, a stationary strategy can always 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.

Keywords: Repeated decision problem; intertemporal choice; time-inconsistent preferences; habit formation

JEL Classification: C61, C73, D01, D91

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.

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

†Department of Mathematics, London School of Economics

‡Center for Mathematical Economics, Bielefeld University

§School of Mathematical Sciences, Tel Aviv University and Durham University, Durham, UK.

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

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.

Classical economic theory assumes static preferences and discounted utility, as proposed by Samuelson (1937) and later motivated with an axiomatic foundation by Koopmans (1960). Since then, this approach has been challenged in various contexts. Arguably, the most developed one is choice under uncertainty. Based on the famous example of Allais (1953) dynamic consistency was challenged, and two major branches of the literature emerged: one focussed on behavioral aspects and challenged expected utility as a whole (e.g., Machina, 1989; Thaler, 1981); the other focussed 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. Yet, this paper does not cover dynamic inconsistencies that originate in uncertainty, so we refrain from providing an extensive overview of the literature here and refer to Etner et al. (2012).

In this paper, we exclusively cover complete information. In this case as well, several forms of time-inconsistent behavior are present: first, there is the classical present bias in which agents over-discount future payoffs (O'Donoghue and Rabin, 1999). As we will investigate how the past (rather than the future) affects current decisions, this is not the behavior we are interested in. Somewhat closer in spirit is the research on reference-dependent utility (Kőszegi and Rabin, 2006; O'Donoghue and Sprenger, 2018) if 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.

Our decision maker's preferences are more closely related to the idea of "habit formation" and to the model 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).

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.

The paper is organized as follows. In Section 2 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 3 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 4 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 5 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 6 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 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 $\vec{a} : \{1, \ldots, T\} \to A$, 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 $(\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_t)$, 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 $\vec{a}^{t-1} = (\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_{t-1})$.

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

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

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

Example 2.1. 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, $\vec{a} = (a, b, a, b, b, b, a, a, a, a, a, a, b, \ldots)$. 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 $\left(U_\gamma^T(\vec{a})\right)_{T \in \mathbb{N}\setminus\{0\}}$ 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 $\vec{a} \in A^\infty$

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

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

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

Thus, for each $v < \overline{V}_\gamma$ there is a history $\vec{a}$ whose average utility is at least v in infinitely many periods. Likewise, for each $v < \underline{V}_\gamma$ there is a history $\vec{a}$ whose average utility is at least v in all but finitely many periods.

Example 2.2. Recall the history $\vec{a}$ from Example 2.1. As t gets large, the average frequency of the action that is played in a block is approximated by

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

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

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

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

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

The average frequency of a at x (taken over the periods where a has been chosen) is still given in (1). The average frequency of b at $x > 0$ is given by

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

Thus, the average utility at x is approximated by

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

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

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

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

3 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 3.1. *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*

$$\lim_{t\to\infty} \varphi\left(a \middle| \vec{a}^t\right) = \frac{\gamma - |A^*| + \displaystyle\sum_{b \in A^*} \frac{u(a)}{u(b)}}{\gamma \displaystyle\sum_{b \in A^*} \frac{u(a)}{u(b)}} \tag{2}$$

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 $(1 - \gamma f_a)\, u(a) = (1 - \gamma f_b)\, u(b)$ for all $a, b \in A^*$. Solving for b and summing over all b we find that

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

Solving for f_a delivers (2). ■

The expression in (2) 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 2.1, 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 $\varphi\left(a|\vec{a}\right) = \lim_{t\to\infty} \varphi\left(a\middle|\vec{a}^{t-1}\right)$. If there is no risk of confusion, we will even write $\varphi(a) = \varphi\left(a|\vec{a}\right)$. The limit of the average utilities is then given by

$$\overline{V}_\gamma\left(\vec{a}\right) = \underline{V}_\gamma\left(\vec{a}\right) = \lim_{T\to\infty} U_\gamma^T\left(\vec{a}\right) = \sum_{a\in A} \varphi\left(a\right)\left(1-\gamma\varphi\left(a\right)\right)u(a). \tag{3}$$

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

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

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

Example 3.2. 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 $\varphi\left(a\right) = \frac{10\gamma-9}{11\gamma}$ and $\varphi\left(b\right) = \frac{\gamma+9}{11\gamma}$. 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 (3), it is given by

$$V_\gamma^* = \max_{x\in\Delta(A)} \sum_{a\in A} x_a\left(1-\gamma x_a\right)u(a), \tag{4}$$

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 (4) 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 3.3. *Let $\vec{a} \in A^\infty$ be the history evolving from an optimal stationary strategy. Then*

$$\varphi\left(a\right) = \frac{2\gamma - |A^*| + \sum_{b\in A^*} \frac{u(a)}{u(b)}}{2\gamma \sum_{b\in A^*} \frac{u(a)}{u(b)}} \tag{5}$$

for all $a \in A^*$, where $A^* \subseteq A$ is the set of actions with $\varphi(a) > 0$.

Proof. The first-order conditions of the maximization problem in (4) 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 3.1 one obtains (5). ∎

Example 3.4. 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 $\gamma \leq \frac{9}{20}$ 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,

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

In particular, this expression is strictly larger than the average utility of the greedy strategy in Example 3.2. □

A special case of optimal stationary histories emerges if A contains exactly two elements and $\gamma = 1$. In this case Proposition 3.3 immediately implies the following corollary.

Corollary 3.5. *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 3.3, yet Examples 2.2 and 3.4 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 2.2 achieves an average utility of $\frac{14}{3}\ln(2) \approx 3.23$, while the best stationary strategy in Example 3.4 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

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

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.

4 Stationary strategies achieve $\underline{V}_\gamma$

Finding a strategy such that the evolving history $\vec{a}$ that achieves $\underline{V}_\gamma(\vec{a}) = \underline{V}_\gamma$ is essentially a dynamic program on a countable state space. Unfortunately, these problems typically lack a tractable structure, so there are no general results that could be helpful in the current context. Thus, we will have to develop some tools to obtain our result in Subsection 4.1. In Subsection 4.2 we shall then investigate how the optimal frequencies change as the parameter γ varies.

4.1 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 $(\vec{a}_{t_1+1}, \vec{a}_{t_1+2}, \ldots, \vec{a}_{t_2})$. The average utility within this block is given by

$$W_\gamma = W_\gamma\left(\vec{a}, t_1, t_2\right) = \frac{1}{t_2 - t_1} \sum_{s=t_1+1}^{t_2} \left(1 - \gamma\varphi\left(a_s \mid \vec{a}^{s-1}\right)\right) u(a_s). \tag{6}$$

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

$$p(a) = p\left(a; \vec{a}, t_1, t_2\right) = \frac{1}{t_2 - t_1} \sum_{s=t_1+1}^{t_2} \mathbb{1}_{\vec{a}_s = a}. \tag{7}$$

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

$$\widetilde{U}_\gamma = \widetilde{U}_\gamma\left(\vec{a}, t_1, t_2\right) = \sum_{a\in A} p(a)\left(1 - \gamma\varphi\left(a \middle| \vec{a}^{t_1}\right)\right) u(a). \tag{8}$$

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.

Lemma 4.1. *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*

$$\left|W_\gamma - \widetilde{U}_\gamma\right| \leq 2\frac{t_2 - t_1}{t_1}\gamma \sum_{a\in A} u(a). \tag{9}$$

Proof. By the definition of W_γ, we have

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

$$= \sum_{a \in A} u(a)p(a) - \frac{1}{t_2 - t_1} \sum_{s=t_1+1}^{t_2} \gamma\varphi\left(a_s \middle| \vec{a}^{s-1}\right) u(a_s). \tag{10}$$

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

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

where in the last step we use that $t_1 \leq s - 1 \leq t_2$ and $\varphi\left(a\middle|\vec{a}^{t_1}\right) \leq 1$. From (8) and (10) we now obtain

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

as required. ■

With Lemma 4.1 we can now prove our first main result, namely that there is a stationary strategy such that the evolving history $\vec{a}$ satisfies $\underline{V}_\gamma = \underline{V}_\gamma(\vec{a}) = V^*_\gamma$. The idea of the proof is to suppose by contradiction that there is a non-stationary history $\vec{a}$ with $\underline{V}_\gamma(\vec{a}) \geq V^*_\gamma + c$ for some strictly positive constant c. This infinite history is split up into blocks that all satisfy the conditions of Lemma 4.1, i.e., that are not too long. Denote by $\varphi^k(a) = \varphi\left(a \middle| \vec{a}^{t_k}\right)$ the frequency of a at the beginning of the k-th block and for each block k consider the number

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

We show that these numbers would behave awkwardly if $\underline{V}_\gamma$ were bounded away from V_γ^*. Specifically, denote by H_K the weighted average of $x_1, \ldots, x_K$, where the weight of x_k equals the relative length of the k-th block within the first K-blocks. We then conclude that $\limsup_K H_K$ is bounded away from $\limsup_K x_K$, which is impossible.

Theorem 4.2. *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 $\underline{V}\left(\vec{a}\right) = 4c + V_\gamma^*$ for some constant $c > 0$. Thus,

$$U_\gamma^T\left(\vec{a}\right) \geq V_\gamma^* + 3c \tag{11}$$

for all sufficiently large T. Let T_1 be such that (11) 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 (6), (7), and (8). 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 4.1,

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

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 (11) and the definition of T_1 that

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

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

$$\begin{aligned} V_\gamma^* + 3c &\leq U_\gamma^{t_K}\left(\vec{a}\right) \\ &= \frac{t_1}{t_K} W_\gamma^1 + \sum_{k=2}^{K} \frac{t_k - t_{k-1}}{t_K} W_\gamma^k \end{aligned}$$

$$\begin{aligned}
&= \frac{T_1}{\lceil (1+\alpha)^{K-1}T_1\rceil} W_\gamma^1 + \sum_{k=2}^{K} \frac{\lceil (1+\alpha)^{k-1}T_1\rceil - \lceil (1+\alpha)^{k-2}T_1\rceil}{\lceil (1+\alpha)^{K-1}T_1\rceil} W_\gamma^k \\
&\leq \frac{T_1}{(1+\alpha)^{K-1}T_1} W_\gamma^1 + \sum_{k=2}^{K} \frac{\lceil (1+\alpha)^{k-2}T_1 + \alpha(1+\alpha)^{k-2}T_1\rceil - \lceil (1+\alpha)^{k-2}T_1\rceil}{\lceil (1+\alpha)^{K-1}T_1\rceil} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\lceil (1+\alpha)^{k-2}T_1\rceil + \lceil \alpha(1+\alpha)^{k-2}T_1\rceil - \lceil (1+\alpha)^{k-2}T_1\rceil}{\lceil (1+\alpha)^{K-1}T_1\rceil} W_\gamma^k \\
&= \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\lceil \alpha(1+\alpha)^{k-2}T_1\rceil}{\lceil (1+\alpha)^{K-1}T_1\rceil} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\lceil \alpha(1+\alpha)^{k-2}\rceil \cdot T_1}{(1+\alpha)^{K-1}T_1} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\alpha(1+\alpha)^{k-2}+1}{(1+\alpha)^{K-1}} W_\gamma^k \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{\alpha(1+\alpha)^{k-2}+1}{(1+\alpha)^{K-1}} \left(\tilde{U}_\gamma^k + 2\left(\alpha + \frac{1}{(1+\alpha)^{k-2}T_1}\right)\gamma \sum_{a\in A} u(a)\right) \\
&\leq \frac{1}{(1+\alpha)^{K-1}} W_\gamma^1 + \sum_{k=2}^{K} \frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}} \widetilde{U}_\gamma^k \\
&\qquad + 2\gamma \sum_{a\in A} u(a) \sum_{k=2}^{K} \frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}} \left(\alpha + \frac{1}{(1+\alpha)^{k-2}}\right), \qquad (13)
\end{aligned}$$

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

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

By (13) we, hence, have

$$V_\gamma^* + 2c \leq W_\gamma^1 + \sum_{k=2}^{K} \frac{(1+\alpha)^{k-2}\alpha+1}{(1+\alpha)^{K-1}} \widetilde{U}_\gamma^k. \qquad (14)$$

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

$$\langle x, y\rangle := \sum_{a\in A} x_a y_a u(a) \qquad \text{and} \qquad \|x\|^2 := \langle x, x\rangle. \tag{15}$$

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

$$\begin{aligned}
\varphi^k(a) - \varphi^{k-1}(a) &= \frac{t_{k-1}}{t_k}\varphi^{k-1}(a) + \frac{t_k - t_{k-1}}{t_k} p^{k-1}(a) - \varphi^{k-1}(a) \\
&= \frac{t_k - t_{k-1}}{t_k}\left(p^{k-1}(a) - \varphi^{k-1}(a)\right) \\
&\geq \frac{(1+\alpha)^k - (1+\alpha)^{k-1} - 1}{(1+\alpha)^k} d^k(a) \\
&\geq \beta d^k(a) - \frac{1}{(1+\alpha)^k}.
\end{aligned} \tag{16}$$

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 (4)

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

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 (14) that for all $0 < \alpha < \bar{\alpha}$ and $K \geq K(\alpha)$

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

Hence,

$$\begin{aligned}
2c &\leq \sum_{k=1}^K \frac{(1+\alpha)^{k-1}\alpha}{(1+\alpha)^K}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle = \sum_{k=1}^K \frac{\alpha}{(1+\alpha)^{K-k+1}}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle \\
&= \frac{1}{1+\alpha}\sum_{k=1}^K \frac{\alpha}{(1+\alpha)^{K-k}}\langle d^k, \mathbf{1} - \gamma\varphi^k\rangle.
\end{aligned} \tag{17}$$

For any K and α define

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

Note that $\sum_{k=1}^{K} \frac{\alpha}{(1+\alpha)^{K-k+1}} \leq 1$, 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,

$$H(K,\alpha) - H(K-1,\alpha) = -\beta H(K-1,\alpha) + \beta \left\|\mathbf{1} - \gamma\varphi^K\right\|^2. \tag{18}$$

Define

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

Then

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

Thus,

$$H(K,\alpha) - H(K-1,\alpha) \leq \varepsilon_{K,\alpha} - 2\beta\gamma \sum_{k=1}^{K-1} \frac{\alpha}{(1+\alpha)^{K-k}} \langle d^k, \mathbf{1} - \gamma\varphi^k\rangle$$

$$+\beta^2\gamma^2\sum_{k=1}^{K-1}\frac{\alpha}{(1+\alpha)^{K-k}}\|d^k\|^2+\frac{(K-2)\alpha}{(1+\alpha)^{K+1}}$$

and together with (17) and (18), we get

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

or equivalently,

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

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

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

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

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

for all $\alpha \leq \alpha^*$ and all $K \geq K^*$. Consequently, due to (18), for $\alpha \leq \alpha^*$ we have,

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

We reached a contradiction. ■

4.2 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 4.3. *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 (4). Then*

$$\sum_{i=1}^{k} \frac{d}{d\gamma} x_\gamma(a_i) \leq 0 \tag{19}$$

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

Proof. First, observe that if $x_\gamma(a_i) = 0$ for some i, $x_\gamma(a_j) = 0$ for all $j \geq i$. Indeed, if $u(a_i) > u(a_j)$, this is immediately clear. If $u(a_i) = u(a_j)$, then let $y_\gamma(a_i) = x_\gamma(a_j)$, $y_\gamma(a_j) = x_\gamma(a_i)$ and $y_\gamma(a) = x_\gamma(a)$ for all $a \neq a_i, a_j$. Then y_γ is a solution of (4), contradicting uniqueness. This means that $\sum_{i=1}^{k^*} \frac{d}{d\gamma} x^\gamma(a_i) \leq 0$, with equality if $\lambda_{k^*+1} > 0$, and $\sum_{i=1}^{k} \frac{d}{d\gamma} x^\gamma(a_i) = 0$ for all $k \geq k^* + 1$.

It remains to prove the claim for $k < k^*$. Let $A^* = \{a \in A : x^\gamma(a) > 0\} = \{a_1, \ldots, a_{k^*}\}$. The Lagrangian of maximization problem (4) is

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

with first-order conditions

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

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

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

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

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

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

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

Taking the derivative of both sides with respect to γ gives

$$\frac{d}{d\gamma} x^\gamma(a_k) = \frac{1}{2\gamma^2}\left(\frac{1}{k^*}\sum_{i=1}^{k^*} u(a_i) - u(a_k)\right) + \frac{1}{u(a_k)}\sum_{i=1}^{k^*} \frac{d}{d\gamma} x^\gamma(a_i) u(a_i). \tag{20}$$

Suppose first that $\sum_{i=1}^{k^*} \frac{d}{d\gamma} x^\gamma(a_i) u(a_i) > 0$. Then the right-hand side of (20) is increasing in k, as $u(a_k)$ is decreasing in k. Thus, in particular,

$$\frac{d}{d\gamma} x^\gamma(a_1) \le \frac{d}{d\gamma} x^\gamma(a_2) \le \cdots \le \frac{d}{d\gamma} x^\gamma(a_{k^*}). \tag{21}$$

Suppose that (19) does not hold. Then there is $k < k^*$ such that $\sum_{i=1}^{k} \frac{d}{d\gamma} x^\gamma(a_i) > 0$. Thus, by (21), we must have $\sum_{i=1}^{k^*} \frac{d}{d\gamma} x^\gamma(a_i) \ge \sum_{i=1}^{k} \frac{d}{d\gamma} x^\gamma(a_i) > 0$, which is impossible. Suppose next that $\sum_{i=1}^{k^*} \frac{d}{d\gamma} x^\gamma(a_i) u(a_i) \le 0$. Then, for all $\ell \le k^*$,

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

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

As $\sum_{a \in A} x_\gamma(a) = 1$ for all $\gamma \in (0,1]$, an immediate consequence of Lemma 4.3 is that the aggregated weight on the poor actions is increasing as γ increases. So, for comparing any two value γ, γ' we obtain the following corollary.

Corollary 4.4. *Let $A = \{a_1, \ldots, a_m\}$ with $u(a_1) \ge u(a_2) \ge \cdots \ge u(a_m)$, let $0 < \gamma < \gamma' \le 1$, 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 $\varphi(\cdot|\vec{b})$.*

5 Stationary strategies do not achieve $\overline{V}_\gamma$

In Examples 2.2 and 3.4 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 construct a sequence $\left(v^T_\gamma\right)_{T\in\mathbb{N}}$ with $\lim_{T\to\infty} v^T_\gamma = \overline{V}_\gamma$, and then show that $\lim_{T\to\infty} v^T_\gamma$ is bounded away from V^*_γ.

5.1 Approximating $\overline{V}_\gamma$

For every $T \geq 1$ define

$$v^T_\gamma = \max_{\vec{a}\in A^\infty} U^T_\gamma\left(\vec{a}\right).$$

That is, v^T_γ denotes the maximal average payoff that can be obtained from a history of length T. We show that the sequence $\left(v^T_\gamma\right)_{T\in\mathbb{N}}$ converges. The idea of the proof is to show that for a history $\vec{a}$ of 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. 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 4.2. These blocks are then "stretched" by some factor $\delta > 1$ such that $S = \delta T$, and the within-block frequencies and average payoffs of the k-block of $\vec{a}$ and $\vec{b}$ are close.

Proposition 5.1. *The sequence* $\left(v^T_\gamma\right)_{T\in\mathbb{N}}$ *converges.*

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

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

$$1 \geq \alpha = \frac{\varepsilon}{16\gamma\sum_a u(a)} - \frac{2}{t_1} \geq \frac{1+2\varepsilon^2}{\varepsilon^2 t_1} - \frac{2}{t_1} = \frac{1}{\varepsilon^2 t_1} > \frac{1}{t_1}. \tag{22}$$

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 $k = 2, \ldots, K-1$ let $t_k = r^k$, and let $t_K = T^* > t_{K-1}$. Let $\vec{a} \in A^\infty$ be such that $U_\gamma^{T^*}(\vec{a}) = v_\gamma^{T^*}$. As in the proof of Theorem 4.2, let the k-th block of $\vec{a}$ be the finite sequence $\left(\vec{a}_{t_k+1}, \ldots, \vec{a}_{t_{k+1}}\right)$. For $k = 0, \ldots, K-1$, denote the average payoff, the frequency, and the approximation by $W_\gamma^k = W_\gamma(\vec{a}, t_k, t_{k+1})$, $p^k(a) = p(a; \vec{a}, t_k, t_{k+1})$, and $\widetilde{U}_\gamma^k = \widetilde{U}_\gamma(\vec{a}, t_k, t_{k+1})$, respectively, as in Equations (6), (7), and (8). 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 $k = 1, \ldots, K$. Thus, by Lemma 4.1 and the definition of α,

$$\left|W^k - \widetilde{U}_\gamma^k\right| \leq 2\left(\alpha + \frac{1}{t_1}\right)\gamma \sum_{a\in A} u(a) \leq 2\frac{\varepsilon}{16\gamma \sum_a u(a)}\gamma \sum_{a\in A} u(a) = \frac{1}{8}\varepsilon \tag{23}$$

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

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

$$\vec{b}_s = \begin{cases} \arg\min_{a\in A} \varphi\left(a \mid b^{s-1}\right) - \varphi\left(a \mid a^{t_1}\right), & \text{if } s = t_1+1, \ldots, s^1, \\ \arg\min_{a\in A} \frac{1}{s-s^k}\sum_{s'=s^k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a} - p^k(a), & \text{if } s^k+1 \leq s \leq s^{k+1}, \text{ where } k \geq 1. \end{cases} \tag{24}$$

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 $k = 0, \ldots, K-1$ denote the average payoff, the frequency, and the approximation in $\vec{b}$ by $Y_\gamma^k = W_\gamma(\vec{b}, t_k, t_{k+1})$, $q^k(a) = p(a; \vec{b}, t_k, t_{k+1})$, and $\widetilde{V}_\gamma^k = \widetilde{U}_\gamma(\vec{b}, t_k, t_{k+1})$, respectively, as in Equations (6), (7), and (8). As $\frac{s^{k+1}-s^k}{s^k} \leq \alpha + \frac{2}{t_1}$ for all $k \geq 1$, Lemma 4.1 together with the definition of α give

$$\left|Y^k - \widetilde{V}_\gamma^k\right| \leq 2\left(\alpha + \frac{2}{t_1}\right)\gamma \sum_{a\in A} u(a) = 2\frac{\varepsilon}{16\gamma \sum_a u(a)}\gamma \sum_{a\in A} u(a) = \frac{1}{8}\varepsilon. \tag{25}$$

By construction, $\varphi\left(.\middle|\vec{a}^{t_1}\right) = \varphi\left(.\middle|\vec{b}^{t_1}\right)$. By (24), 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 $\varepsilon > 0$

$$\left|\varphi\left(a\middle|\vec{a}^{t_1}\right) - \varphi\left(a\middle|\vec{b}^{s}\right)\right| \leq \frac{|A|-1}{s} \leq \frac{|A|-1}{t_1} \leq \varepsilon^3 \frac{|A|-1}{|A| + 8\gamma\sum_a u(a)} \leq \varepsilon^2 \tag{26}$$

for all $s \leq s^1$. Let $k \geq 1$ and $s^k + 1 \leq s \leq s^{k+1}$. By (24), action a is only being played at s if $\frac{1}{s-s^k}\sum_{s'=s^k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a} \leq p^k(a)$. Thus, $p^k(a) \leq \frac{1}{s-s^k} + \frac{1}{s-s^k}\sum_{s'=s^k+1}^{s} \mathbb{1}_{\vec{b}_{s'}=a}$. In particular, for $s = s^{k+1}$ it holds that

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

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

$$\begin{aligned} \left|p^k(a) - q^k(a)\right| &\leq \frac{|A|-1}{s^{k+1}-s^k} \leq \frac{|A|-1}{t_{k+1}-t_k} \\ &\leq \frac{|A|-1}{\alpha t_k - 1} \leq \frac{|A|-1}{\alpha\,(1+\alpha)^k\, t_1 - 1} \\ &\leq \frac{|A|-1}{\frac{1}{\varepsilon^2}\,(1+\alpha)^k - k} \leq \varepsilon^2 \frac{|A|-1}{(1+\alpha)^k - \varepsilon^2} \\ &\leq \frac{\varepsilon}{8\,|A|\,u(a)} \end{aligned} \tag{27}$$

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

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

Further, by using (27) we find for $k \geq 2$ that

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

where $x_1, x_2 \leq 1$ 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,

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

for all $k \geq 1$, so that

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

for all $k \geq 0$. (The case $k = 0$ follows from (26).) Hence,

$$\begin{aligned}\left|\widetilde{U}_\gamma^k-\widetilde{V}_\gamma^k\right| &= \left|\sum_{a\in A}p^k(a)\left(1-\varphi\left(a\middle|\vec{a}^{t_k}\right)\right)u(a)-\sum_{a\in A}q^k(a)\left(1-\varphi^k\left(a\middle|\vec{b}^{s^k}\right)\right)u(a)\right|\\ &= \left|\sum_{a\in A}\left(p^k(a)-q^k(a)\right)\left(1-\varphi\left(a\middle|\vec{a}^{t_k}\right)\right)u(a)-\sum_{a\in A}q^k(a)\left(\varphi\left(a\middle|\vec{a}^{t_k}\right)-\varphi\left(a\middle|\vec{b}^{s^k}\right)\right)u(a)\right|\\ &\leq \sum_{a\in A}\left|p^k(a)-q^k(a)\right|u(a)-\sum_{a\in A}\left|\varphi\left(a\middle|\vec{a}^{t_k}\right)-\varphi\left(a\middle|\vec{b}^{s^k}\right)\right|u(a)\\ &\leq \frac{1}{4}\varepsilon \end{aligned} \tag{28}$$

for all $k \geq 1$. From (23), (25), an (28) 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

$$\begin{aligned}\left|U_\gamma^T\left(\vec{b}\right)-U_\gamma^S\left(\vec{a}\right)\right| &\leq \sum_{k=0}^{K-1}\left|\frac{t_{k+1}-t_k}{T}W_\gamma^k-\frac{s^{k+1}-s^k}{S}Y_\gamma^k\right|\\ &\leq \sum_{k=0}^{K-1}\left|\frac{\delta t_{k+1}-\delta t_k}{\delta T}W_\gamma^k-\frac{s^{k+1}-s^k}{\delta T}Y_\gamma^k\right|\end{aligned}$$

$$\leq \sum_{k=0}^{K-1} \left(\left| \frac{\delta t_{k+1} - \delta t_k}{\delta T} \left(W_\gamma^k - Y_\gamma^k \right) \right| + \left| \frac{1}{\delta T} Y_\gamma^k \right| \right)$$
$$\leq \sum_{k=0}^{K-1} \frac{(1+\alpha)^{k+1} t_1 - (1+\alpha)^k t_1 + 1}{(1+\alpha)^{K-1} t_1} \left| W_\gamma^k - Y_\gamma^k \right| + \frac{K}{\delta (1+\alpha)^{K-1} t_1} \sum_{a \in A} u(a)$$
$$\leq \frac{\varepsilon}{2} \sum_{k=0}^{K-1} \left(\alpha (1+\alpha)^{k-K+1} + \frac{1}{(1+\alpha)^{K-1} t_1} \right) + \frac{K}{(1+\alpha)^{K-1} t_1} \sum_{a \in A} u(a)$$
$$\leq \frac{\varepsilon}{2} \left(1 + \frac{K}{(1+\alpha)^K t_1} \right) + \frac{K}{(1+\alpha)^{K-1} t_1} \sum_{a \in A} u(a).$$

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,

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

Thus,

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

as required. ∎

After we have established that $\lim_{T \to \infty} v_\gamma^T$ is well-defined, we shall now show that the sequence converges to $\overline{V}_\gamma$.

Proposition 5.2. *It holds that* $\lim_{T \to \infty} v_\gamma^T = \overline{V}_\gamma$.

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

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

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

$$v_\gamma^{T'} \geq \lim_{T \to \infty} v_\gamma^T - c \geq \overline{V}_\gamma + 3c = \sup_{\vec{a} \in A^\infty} \limsup_T U_\gamma^T (\vec{a}) + 3c \tag{29}$$

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

$$\sup_{\vec{a}\in A^\infty} \limsup_T U_\gamma^T(\vec{a}) + 3c \geq \sup_{\vec{a}\in A^\infty} U_\gamma^{T_2(\vec{a})}(\vec{a}) + 2c \geq \lim_{T\to\infty} v_\gamma^T + c, \tag{30}$$

where the last inequality holds since $T_2(\vec{\alpha}) \geq T_1 \geq T_0$ for all $\vec{a} \in A^\infty$. From (29) and (30) 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 5.1. ■

5.2 Establishing suboptimality of V_γ^*

The next main result shows that $\overline{V}_\gamma$ 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 show that the effect of these switches is significant, i.e., that for the final history $\vec{b}$ it holds that $\overline{V}_\gamma(\vec{b}) > V_\gamma^*(\vec{a}) + \eta$ for some constant $\eta > 0$. We start with the following lemma that captures the effect of such pairwise switches.

Lemma 5.3. *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*

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

for all $T \geq s$.

Proof. By the conditions on $\vec{a}$ and $\vec{b}$,

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

and

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

Thus,

$$\begin{aligned} T\left(U_\gamma^T(\vec{b}) - U_\gamma^T(\vec{a})\right) &= \left(1 - \gamma\varphi\left(b\middle|\vec{b}^{t-1}\right)\right)u(b) + \left(1 - \gamma\varphi\left(a\middle|\vec{b}^{s-1}\right)\right)u(a) \\ &\quad - \left(1 - \gamma\varphi\left(a\middle|\vec{a}^{t-1}\right)\right)u(a) - \left(1 - \gamma\varphi\left(b\middle|\vec{a}^{s-1}\right)\right)u(b) \end{aligned}$$

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

$$= \gamma\left(\left(\frac{1}{s-1}\sum_{r=1}^{s-1}\mathbb{1}_{\vec{a}_r=b} - \frac{1}{t-1}\sum_{r=1}^{t-1}\mathbb{1}_{\vec{b}_r=b}\right)u(b)\right.$$

$$\left. + \left(\frac{1}{t-1}\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=a} - \frac{1}{s-1}\sum_{r=1}^{s-1}\mathbb{1}_{\vec{b}_r=a}\right)u(a)\right)$$

$$\geq \frac{\gamma}{(s-1)(t-1)}\left((t-s)\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=b}u(b) + (s-t)\sum_{r=1}^{t-1}\mathbb{1}_{\vec{a}_r=a}u(a)\right)$$

$$= \frac{\gamma\,(s-t)}{s-1}\left(\varphi\left(a \mid \vec{a}^{t-1}\right)u(a) - \varphi\left(b \mid \vec{a}^{t-1}\right)u(b)\right)$$

as required. ■

Lemma 5.3 provides a sufficient condition to increase all future payoffs by swapping the position of an action $b \in A$ with its next previous occurrence of $a \in A$ within $\vec{a}$. Such a switch increases the average payoff if

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

Intuitively, actions with high basic utility will be shifted towards the back, so that they can be played with small punishment.

Theorem 5.4. *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 the sufficiency. Let $\vec{a}$ be an optimal stationary strategy, write $\varphi(a)$ for $\varphi\left(a|\vec{a}\right)$, and denote by A^* be the set of actions $a \in A$ with $\varphi\left(a\right) > 0$. By Proposition 3.3,

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

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

$$\varphi(a)u(a) - \varphi(b)u(b) = \frac{u(a) - u(b)}{2\gamma} \tag{31}$$

for all $a, b \in A^*$. As V_γ^* and $\overline{V}_\gamma$ depend continuously on γ and $\{u(a)\}_{a\in A}$, and $\varphi\left(\cdot\right) \in \mathbb{Q}^A$ if $\gamma, u(a) \in \mathbb{Q}$ for all $a \in A$ by Proposition 3.3, we can assume without loss of generality that $\varphi\left(\cdot\right) \in \mathbb{Q}^A$. Thus, there are integers $m_a \in \mathbb{N}$ for all $a \in A$ such that $\varphi\left(a\right) = \frac{m_a}{m}$, 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, $u(\underline{a}) < u(\overline{a})$, so that $m_{\overline{a}} > m_{\underline{a}}$.

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

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

Proof. 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. ■

Define the following constants

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

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

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

and let

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

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

Finally, let

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

be such that $\left|U_\gamma^T\left(\vec{a}\right) - V_\gamma^*\right| \leq \eta$ for all $T \geq T^*$. We show that there is an infinite sequence of T's with $v^T \geq V^* + \eta$ for all $T \geq T^*$. For this purpose we will construct for any T in this sequence a history $\vec{b}$ with $U_\gamma^T\left(\vec{b}\right) \geq V^* + \eta$. By Proposition 5.2, this is sufficient to prove the theorem.

So, let $T \geq T^*$ 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 $T_1 + 1 \leq s \leq T$ be the first occurrence of $\underline{a}$ such that the period of the last previous occurrence of $\overline{a}$, denoted by $t < s$, satisfies

$$\varphi\left(\overline{a}\middle|\vec{c}^{t-1}\right)u\left(\overline{a}\right) - \varphi\left(\underline{a}\middle|\vec{c}^{t-1}\right)u\left(\underline{a}\right) \geq \delta. \tag{32}$$

If such $t, s \leq T$ do not exist, let $\vec{b} = \vec{c}$. Otherwise, note that (32) 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 $t' \neq t, s$. By Lemma 5.3,

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

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 $T \geq T^*$ there is a finite number of beneficial switches, so $\vec{b}$ is well-defined. Moreover, for any $T_1 + 1 \leq t \leq T - m$ 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.

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

Proof. 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 $\varphi\left(\underline{a}\right)\left(T - s^*\right) \geq \varphi\left(\overline{a}\right)\left(s^* - T_1\right)$. 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

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

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

$$\delta \geq \varphi\left(\overline{a} \middle| \vec{b}^{t-1}\right) u(\overline{a}) - \varphi\left(\underline{a} \middle| \vec{b}^{t-1}\right) u(\underline{a}). \tag{33}$$

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

$$\varphi\left(\underline{a}\middle|\vec{b}^{t-1}\right) = \frac{T}{t-1}\varphi(\underline{a}) - \frac{1}{t-1}. \tag{34}$$

Similarly,

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

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

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

Therefore

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

This, together with (31), (33), and (34) shows that

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

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

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 $\left(\vec{b}_{t+1},\ldots,\vec{b}_{t+m}\right)$ 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 of the claim. ■

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

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

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

$$=\varphi\left(\underline{a}\right)\left(T-T_1\right)\frac{1-q}{4}. \tag{35}$$

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 $\vec{b}_{t^\ell}\left(\ell\right)=\vec{b}_{s^\ell}\left(\ell-1\right)$, $\vec{b}_{s^\ell}\left(\ell\right)=\vec{b}_{t^\ell}\left(\ell-1\right)$, and $\vec{b}_t\left(\ell\right)=\vec{b}_t\left(\ell-1\right)$ for all $t\neq t^\ell,s^\ell$. Using Lemma 5.3 and the fact that $\frac{T-T_1}{T}\geq\frac{1}{2}$ by construction, we therefore have

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

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 (35) for k we have

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

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, $\overline{V}_\gamma=\lim_T v_\gamma^T\geq V_\gamma^*+\eta$. ■

6 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.

To take the second meaning 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

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

The fatigue parameter γ is fixed throughout, and we do not append it in notations. The utility derived from $\vec{a}$ is defined as,

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

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

Let

$$V^{\lambda,\delta} = \max_{\vec{a}} U^{\lambda,\delta}\left(\vec{a}\right).$$

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

Theorem 6.1. *There is a function $\delta(\lambda) < 1$ s.t. for every $\varepsilon > 0$ there is λ_0 satisfying*

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

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

The theorem states that the stationary strategy is optimal for increasingly patient decision makers. This is another case where cyclical consumption is optimal.

Proof. 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,

$$U^{\lambda,\delta}(\vec{a}) < V^* + \varepsilon. \tag{36}$$

To simplify the proof, we use the notation $\varphi^{t-1,\lambda}(a) = \varphi^{\lambda}\big(a \big| \vec{a}^{t-1}\big)$. 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 (15), we have

$$U^{\lambda,\delta}(\vec{a}) = \sum_{t=1}^{\infty} \eta^t \big\langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \big\rangle. \tag{37}$$

Denote

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

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

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

Since $\sum_{t=2}^{\infty} \frac{\eta^{t-1}}{\beta^{t-1}} \left\| \mathbf{1} - \gamma \varphi^{t-1,\lambda} \right\|^2 = H$, we obtain after rearranging the following key equation:

$$2\gamma \sum_{t=2}^{\infty} \eta^t \left\langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \right\rangle = \epsilon_1 + \sum_{t=2}^{\infty} \left(\frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}} \right) \left\| \mathbf{1} - \gamma\varphi^{t-1,\lambda} \right\|^2 + 2\gamma \sum_{t=2}^{\infty} \eta^t \left\langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \varphi^{t-1,\lambda} \right\rangle$$
$$+ \gamma^2 \sum_{t=2}^{\infty} \eta^t \beta^t \left\| \mathbf{1}^t - \gamma\varphi^{t-1,\lambda} \right\|^2. \tag{38}$$

Note that, by the definition of $U^{\lambda,\delta}$,

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

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

$$2\gamma U^{\lambda,\delta}(\vec{a}) \leq 2\gamma \sum_{t=2}^{\infty} \eta^t \left\langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \right\rangle + \frac{\varepsilon\gamma}{3}. \tag{39}$$

We turn to the expressions on the right-hand side of (38). 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 next show that the first sum on the right-hand side of (38) 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

$$\sum_{t=2}^{\infty} \left| \frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}} \right| \leq \frac{1-\delta}{1-\lambda} \sum_{t=1}^{\infty} \left| \delta^t \left(1 - \lambda^{t+1}\right) - \delta^{t-1} \left(1 - \lambda^t\right) \right|$$
$$= \frac{1-\delta}{1-\lambda} \sum_{t=1}^{\infty} \delta^{t-1} \left| \delta \left(1 - \lambda^{t+1}\right) - \left(1 - \lambda^t\right) \right|. \tag{40}$$

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 (40) is then bounded from above by

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

$$\leq \frac{1-\delta}{1-\lambda\delta} + \frac{1-\delta}{1-\lambda}$$
$$\leq 2\frac{1-\delta}{1-\lambda}. \tag{41}$$

As for any λ there is a function $\delta_2(\lambda)$ such that when $\delta > \delta_2(\lambda)$, it holds that $2\frac{1-\delta_1(\lambda)}{1-\lambda} < \varepsilon\gamma/3$, we find with (40) and (41)

$$\sum_{t=2}^{\infty}\left(\frac{\eta^t}{\beta^t} - \frac{\eta^{t-1}}{\beta^{t-1}}\right)\left\|1-\gamma\varphi^{t-1,\lambda}\right\|^2 \leq \frac{1-\delta}{1-\lambda}\sum_{t=1}^{\infty}\delta^{t-1}\left|\delta\left(1-\lambda^{t+1}\right)-\left(1-\lambda^t\right)\right| \leq 2\frac{1-\delta}{1-\lambda}$$
$$< \frac{\varepsilon\gamma}{3} \tag{42}$$

as required.

As for the second sum on the right-hand side of (38), 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, $\langle \mathbf{1}-\varphi^{t-1,\lambda}, \varphi^{t-1,\lambda}\rangle \leq V^*$. Thus,

$$2\gamma\sum_{t=2}^{\infty}\eta^t\langle \mathbf{1}-\gamma\varphi^{t-1,\lambda}, \varphi^{t-1,\lambda}\rangle \leq 2\gamma V^*\sum_{t=2}^{\infty}\eta^t = 2\gamma V^*(1-\delta)\sum_{t=2}^{\infty}\delta^{t-1} \leq 2\gamma V^*. \tag{43}$$

For the last sum on the right-hand side of (38), first note that

$$\sum_{t=2}^{\infty}\eta^t\beta^t = \sum_{t=2}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t} \leq \sum_{t=1}^{\infty}(1-\delta)\delta^{t-1}\frac{1-\lambda}{1-\lambda^t}. \tag{44}$$

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 (44)

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

$$\leq \frac{2\varepsilon\gamma}{3}, \tag{45}$$

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

Hence, from (38), (42), (43), and (45) we obtain

$$\begin{aligned} 2\gamma U^{\lambda,\delta}(\vec{a}) &\leq 2\gamma \sum_{t=2}^{\infty} \eta^t \langle \mathbf{1} - \gamma\varphi^{t-1,\lambda}, \mathbf{1}^t \rangle + \frac{\varepsilon\gamma}{3} \leq \frac{\varepsilon\gamma}{3} + \frac{\varepsilon\gamma}{3} + 2\gamma V^* + \frac{2\varepsilon\gamma}{3} + \frac{\varepsilon\gamma}{3} \\ &= \frac{5\varepsilon\gamma}{3} + 2\gamma V^*. \end{aligned}$$

Dividing by 2γ yields $U^{\lambda,\delta}(\vec{a}) \leq V^* + \varepsilon$, as required. ■

7 Summary

We have discussed a dynamic decision problem in which the decision maker's utility derived from a certain action diminishes with the frequency of its use. In the interesting cases where utility is measured by the limit inferior or discounting is applied, the optimal outcome is achieved by a stationary strategy, meaning that periodic consumption is optimal.

References

Allais, M. (1953). "Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axiomes de l'Ecole Americaine". In: *Econometrica* 21, pp. 503–546.

Baucells, M., M. Weber, and F. Welfens (2011). "Reference Point Formation and Updating". In: *Management Science* 57, pp. 506–519.

Becker, G. S. and K. M. Murphy (1988). "A Theory of Rational Addiction". In: *Journal of Political Economy* 96.4, pp. 675–700.

Boldrin, M., L. J. Christiano, and J. D. Fisher (2001). "Habit Persistence, Asset Returns, and the Business Cycle". In: *American Economic Review* 91, pp. 149–166.

Bornstein, R. F., A. R. Kale, and K. R. Cornell (1990). "Boredom as Limiting Condition on the Mere Exposure Effect". In: *Journal of Personality and Social Psychology* 58, pp. 791–800.

Constantinides, G. M. (1990). "Habit Formation: A Resolution to the Equity Premium Puzzle". In: *Journal of Political Economy* 98, pp. 519–543.

Crisp, R. J., R. R. C. Hutter, and B. Young (2008). "When Mere Exposure Leads to Less Liking: The Incremental Threat Effect in Intergroup Contexts". In: *British Journal of Psychology* 100, pp. 133–149.

Etner, J., M. Jeleva, and J.-M. Tallon (2012). "Decision Theory Under Ambiguity". In: *Journal of Economic Surveys* 26, pp. 234–270.

Gilboa, I. and D. Schmeidler (1995). "Case-based Decision Theory". In: *The quarterly Journal of economics* 110.3, pp. 605–639.

Gonzales, C., J. F. Lerch, and C. Lebiere (2003). "Instance-based Learning in Dynamic Decision Making". In: *Cognitive Science* 27, pp. 501–635.

He, Y., J. S. Dyer, and J. C. Butler (2013). "On the Axiomatization of the Satiation and Habit Formation Utility Models". In: *Operations Research* 61, pp. 1399–1410.

Kaiser, K. and R. Schwabe (2012). "Preference for Variety". In: *The B.E. Journal of Theoretical Economics* 12, Article 3.

Koopmans, T. C. (1960). "Stationary Ordinal Utility and Impatience". In: *Econometrica: Journal of the Econometric Society*, pp. 287–309.

Kőszegi, B. and M. Rabin (2006). "A Model of Reference-dependent Preferences". In: *The Quarterly Journal of Economics* 121.4, pp. 1133–1165.

Liao, H.-I., S.-L. Yeh, and S. Shimojo (2011). "Novelty vs. Familiarity Principles in Preference Decisions: Task-Context of Past Experience Matters". In: *Frontiers in Psychology* 2, Article 43.

Machina, M. J. (1989). "Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty". In: *Journal of Economic Literature* 27, pp. 1622–1668.

Mehra, R. and E. C. Prescott (1985). "The Equity Premium: A Puzzle". In: *Journal of Monetary Economics* 15, pp. 145–161.

O'Donoghue, T. and M. Rabin (1999). "Doing It Now or Later". In: *American Economic Review* 89, pp. 103–124.

O'Donoghue, T. and C. Sprenger (2018). "Reference-Dependent Preferences". In: *Handbook of Behavioral Economics: Applications and Foundations I.* Ed. by B. D. bernheim, S. DellaVigna, and D. Laibson. Elsevier, pp. 1–77.

Rozen, K. (2010). "Foundations of Intrinsic Habit Formation". In: *Econometrica* 78.4, pp. 1341–1373.

Rustichini, A. and P. Siconolfi (2014). "Dynamic Theory of Preferences: Habit formation and Taste for Variety". In: *Journal of Mathematical Economics* 55, pp. 55–68.

Samuelson, P. A. (1937). "A Note on Measurement of Utility". In: *The Review of Economic Studies* 4, pp. 155–161.

Stewart, N., N. Chater, and G. D. A. Brown (2006). "Decision by Sampling". In: *Cognitive Psychology* 53, pp. 1–26.

Thaler, R. (1981). "Some Empirical Evidence on Dynamic Inconsistency". In: *Economics Letters* 8, pp. 201–207.

Williams, S. M. (1987). “Repeated Exposure and the Attractiveness of Synthetic Speech: An Inverted U-Relationship”. In: *Current Psychological Research & Reviews* 6, pp. 148–154.

Zajonc, R. B. (1968). “Attitudinal Effects of Mere Exposure”. In: *Journal of Personality and Social Psychology, Monograph Supplement* 9, pp. 1–27.

Zajonc, R. B., P. Shaver, T. Carol, and V. K. David (1972). “Exposure, Satiation, and Stimulus Discriminability”. In: *Journal of Personality and Social Psychology* 21, pp. 270–280.