Supplemental Material for "Reinforcement with Fading Memories"

Kuang Xu and Se-Young Yun

July, 2019

Appendix.

A. Technical Preliminaries

PROPOSITION 6 **(Doob's inequality).** *(cf. Section 12.6, Grimmett and Stirzaker (2001)) Let $\{X(t)\}_{t\geq 0}$ be a discrete- or continuous-time non-negative submartingale. Fix $T\geq 0$. We have that*

$$\mathbb{P}\left(\sup_{0\leq t\leq T} X(t)\geq c\right)\leq \frac{\mathbb{E}\left(X(T)\right)}{c},\quad \forall c>0.$$

PROPOSITION 7 **(Gronwall's lemma).** *(cf. Section 1.3, Ames and Pachpatte (1997)) Let $f,b:\mathbb{R}_+\to\mathbb{R}_+$ be continuous and non-negative functions, and let $a:\mathbb{R}_+\to\mathbb{R}_+$ be a continuous, positive and non-decreasing function. If $f(t)\leq a(t)+\int_0^t b(s)f(s)ds$, for all $t\in[0,T]$, then*

$$f(T)\leq a(T)\exp\left(\int_0^T b(t)\,dt\right).$$

B. Proofs of Propositions

B.1. Proof of Proposition 1 The proof of Proposition 1 is largely based on maximal inequalities for continuous-time martingales, and the main steps follow the approach developed in Kurtz (1978). Define $X^m(t)\in\mathbb{R}_+^{\mathcal{K}}$, where

$$X_k^m(t)=\overline{Q}_k^m(t)-\left(\overline{Q}_k^m(0)+\int_0^t G_k(\overline{Q}^m(s),\overline{C}^m(s))ds\right),\quad t\in\mathbb{R}_+,k\in\mathcal{K}. \tag{123}$$

The proof proceeds in two stages: we first show, via Gronwall's lemma (Proposition 7 in Supplemental Material Section A), that bounding the deviation of $\overline{Q}_k^m(\cdot)$ from $V_k^m(\cdot)$ can be reduced to bounding the magnitude of $X_k^m(\cdot)$. We then show that this can be accomplished by writing $X_k^m(\cdot)$ as a sum of two continuous-time martingales. From the definition of $\{V^m(t)\}_{t\geq 0}$, we have that

$$\begin{aligned}
&\left\|\overline{Q}^m(t)-V^m(t)\right\|\\
=&\sum_{k\in\mathcal{K}}\left|\overline{Q}_k^m(t)-q_k^0-\int_0^t G_k(V^m(s),\overline{C}^m(s))ds\right|\\
\leq&\sum_{k\in\mathcal{K}}\left|\overline{Q}_k^m(0)-q_k^0\right|+\sum_{k\in\mathcal{K}}\left|X_k^m(t)+\int_0^t\left(G_k(\overline{Q}^m(s),\overline{C}^m(s))-G_k(V^m(s),\overline{C}^m(s))\right)ds\right|\\
\leq&\left(\left\|\overline{Q}^m(0)-q^0\right\|+\|X^m(t)\|\right)+\int_0^t\sum_{k\in\mathcal{K}}\left|G_k(\overline{Q}^m(s),\overline{C}^m(s))-G_k(V^m(s),\overline{C}^m(s))\right|ds\\
\overset{(a)}{\leq}&\left(\left\|\overline{Q}^m(0)-q^0\right\|+\|X^m(t)\|\right)+\int_0^t\sum_{k\in\mathcal{K}}\left|\overline{Q}_k^m(s)-V_k^m(s)\right|ds\\
=&\left(\left\|\overline{Q}^m(0)-q^0\right\|+\|X^m(t)\|\right)+\int_0^t\left\|\overline{Q}^m(s)-V^m(s)\right\|ds.
\end{aligned}\tag{124}$$

For step (a), we observe that for all $c,k\in\mathcal{K}$, $G_k(\cdot,c)$ is a 1-Lipschitz continuous function. We now apply Gronwall's lemma (Proposition 7 in Supplemental Material Section A) to Eq. (124), with $a(t)=\left\|\overline{Q}^m(0)-q^0\right\|+\|X^m(t)\|$, $b(t)=1$, and $f(t)=\left\|\overline{Q}^m(t)-V^m(t)\right\|$, and obtain that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T}\left\|\overline{Q}^m(t)-V^m(t)\right\|>\epsilon\right)\\
\le&\mathbb{P}\left(\sup_{0\le t\le T}\left(\left\|\overline{Q}^m(0)-q^0\right\|+\|X^m(t)\|\right)e^t>\epsilon\right)\\
\le&\mathbb{P}\left(\left\|\overline{Q}^m(0)-q^0\right\|>\epsilon e^{-T}/2\right)+\mathbb{P}\left(\sup_{0\le t\le T}\|X^m(t)\|>\epsilon e^{-T}/2\right).
\end{aligned}\tag{125}$$

Because we have assumed that $\lim_{m\to\infty}\mathbb{P}\left(\left\|\overline{Q}^m(0)-q^0\right\|>\epsilon\right)=0$ for all $\epsilon>0$, to prove Lemma 1, it therefore suffices to show that

$$\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\|X^m(t)\|>\epsilon\right)=0,\quad\forall\epsilon>0.\tag{126}$$

We now prove Eq. (126) by expressing $X_k^m(\cdot)$ as the sum of two continuous-time martingales corresponding to the arrivals and departures of rewards, respectively. Fix $t\in\mathbb{R}_+$ and $k\in\mathcal{K}$, and denote by $A_k^m(t)$ and $D_k^m(t)$ the amount of rewards associated with action k that have arrived and departed, respectively, during the interval $[0,t]$. In particular, we can write

$$Q_k^m(t)=Q_k^m(0)+A_k^m(t)-D_k^m(t),\quad t\in\mathbb{R}_+.\tag{127}$$

We have that

$$\begin{aligned}
&|X_k^m(t)|\\
=&\left|\frac{1}{m}(A_k^m(mt)-D_k^m(mt))-\int_0^t G_k(\overline{Q}^m(s),\overline{C}^m(s))ds\right|\\
=&\left|\frac{1}{m}\left(A_k^m(mt)-\int_0^{mt}\lambda_k\mathbb{I}(C^m(s)=k)ds\right)-\frac{1}{m}\left(D_k^m(mt)-\int_0^{mt}Q_k^m(s)m^{-1}ds\right)\right|\\
\le&\left|\frac{1}{m}\left(A_k^m(mt)-\int_0^{mt}\lambda_k\mathbb{I}(C^m(s)=k)ds\right)\right|+\left|\frac{1}{m}\left(D_k^m(mt)-\int_0^{mt}Q_k^m(s)m^{-1}ds\right)\right|.
\end{aligned}\tag{128}$$

For the remainder of the proof, we will focus on showing that, for all $\epsilon>0$,

$$\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{1}{m}\left(A_k^m(mt)-\int_0^{mt}\lambda_k\mathbb{I}(C^m(s)=k)ds\right)\right|>\epsilon\right)=0,\quad\text{and}\tag{129}$$

$$\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{1}{m}\left(D_k^m(mt)-\int_0^{mt}Q_k^m(s)m^{-1}ds\right)\right|>\epsilon\right)=0.\tag{130}$$

In light of Eq. (128), the above two equations imply the validity of Eq. (126), which proves Proposition 1. The proofs for both Eqs. (129) and (130) hinge upon the following technical result, which states that the sample path of a certain time-inhomogenous Poisson process stays close to its mean, with high probability. Note that a similar result for uniform-rate Poisson processes can be found in Lemma 7.6 of Massey and Whitt (1998). The proof of the lemma is given in Supplemental Material Section C.6, and uses a similar line of argument as that of Theorem 2.2 in Kurtz (1978).

LEMMA 8. *Fix $l\in\mathbb{Z}^{K+1}$ and $T\in\mathbb{R}_+$. Let $\{N(t)\}_{t\ge0}$ be the counting process where $N(t)$ is the number of times in $[0,t]$ that the process $W^m(\cdot)$ jumps from state w to $w+l$, for some $w\in$*

$\mathbb{Z}^{K+1}$. *Denote by* $\psi:\mathbb{Z}^{K+1}\to\mathbb{R}_+$ *the corresponding rate function of* $N(\cdot)$*, so that the instantaneous transition rate of* $N(\cdot)$ *at time* t *is equal to* $\psi(W^m(t))$*. For all* $\epsilon,\phi>0$*, we have that*

$$\mathbb{P}\left(\sup_{0\le t\le T}\left|N(t)-\int_0^t\psi(W^m(s))ds\right|\ge\epsilon\right)\le 2e^{-\phi T\cdot h(\epsilon/\phi T)}+\mathbb{P}\left(\sup_{0\le t\le T}\psi(W^m(t))\ge\phi\right), \tag{131}$$

where $h(x)=(1+x)\log(1+x)-x$.

We now prove Eqs. (129) and (130). In the context of Lemma 8, $A_k(\cdot)$ is the counting process with $l=e_k$, where e_k is the vector $(K+1)\times 1$ vector whose kth coordinate is 1 and all other coordinates zero, corresponding to an arrival to $Q_k^m(\cdot)$. The rate of $A_k(\cdot)$ at time t is equal to $\lambda_k\mathbb{I}(C^m(t)=k)$, which is bounded from above by λ_k for all $t\in\mathbb{R}_+$. By applying Lemma 8, with $\psi(W^m(t))=\lambda_k\mathbb{I}(C^m(t)=k)$ and $\phi=\lambda_k$, we have that $\mathbb{P}\left(\sup_{0\le t\le T}\psi(W^m(t))>\phi\right)=0$, and for all $\epsilon>0$,

$$\begin{aligned}
&\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{1}{m}\left(A_k^m(mt)-\int_0^{mt}\lambda_k\mathbb{I}(C^m(s)=k)ds\right)\right|\ge\epsilon\right)\\
=&\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le mT}\left|A_k^m(t)-\int_0^t\lambda_k\mathbb{I}(C^m(s)=k)ds\right|\ge m\epsilon\right)\\
\le&\lim_{m\to\infty}2\exp\left(-\lambda_k mT\cdot h\left(\frac{\epsilon}{\lambda_k T}\right)\right)=0.
\end{aligned}\tag{132}$$

The proof of Eq. (130) uses essentially the same idea, but the argument needs to be more delicate due to the fact that the rate of the counting process $D_k^m(t)$, which is equal to $Q_k^m(t)\mu=Q_k^m(t)m^{-1}$, is not bounded over the state space. Therefore, we first derive an upper bound on the tail probabilities of $Q_k^m(t)$, as follows. Let $\psi(\cdot)$ be the rate function for $D_k(\cdot)$ as in Lemma 8, and $\phi=2q_k^0+(1+\epsilon)\lambda_k T$. Fixing $\epsilon>0$, we have that, for all $r\in\mathbb{Z}_+$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le mT}\psi(W^m(t))\ge\phi\right)=\mathbb{P}\left(\sup_{0\le t\le mT}Q_k^m(t)m^{-1}\ge\phi\right)\\
\le&\mathbb{P}(Q_k^m(0)\ge 2mq_k^0)+\max_{r=1,\ldots,2mq_k^0}\mathbb{P}\left(\sup_{0\le t\le mT}Q_k^m(t)\ge m\phi\,\middle|\,Q^m(0)=r\right)\\
=&\mathbb{P}(Q_k^m(0)\ge 2mq_k^0)+\max_{r=1,\ldots,2mq_k^0}\mathbb{P}\left(\sup_{0\le t\le mT}Q_k^m(t)-Q_k^m\ge m\phi-r\,\middle|\,Q^m(0)=r\right)\\
\le&\mathbb{P}(Q_k^m(0)\ge 2mq_k^0)+\max_{r=1,\ldots,2mq_k^0}\mathbb{P}\left(\sup_{0\le t\le mT}A_k^m(t)\ge m\phi-r\,\middle|\,Q^m(0)=r\right)\\
\overset{(a)}{=}&\mathbb{P}(Q_k^m(0)\ge 2mq_k^0)+\mathbb{P}\left(\sup_{0\le t\le mT}A_k^m(t)\ge(1+\epsilon)m\lambda_k T\right)\\
\overset{(b)}{\le}&\mathbb{P}(Q_k^m(0)\ge 2mq_k^0)+\mathbb{P}\left(\sup_{0\le t\le mT}\left(A_k^m(t)-\int_0^t\lambda_k\mathbb{I}(C^m(s)=k)ds\right)\ge m[(1+\epsilon)\lambda_k T-\lambda_k T]\right)\\
\overset{(c)}{\le}&\mathbb{P}(m^{-1}Q_k^m(0)\ge 2q_k^0)+2\exp\left(-\lambda_k mT\cdot h(\epsilon)\right),
\end{aligned}\tag{133}$$

where step (a) follows from the fact that $A_k^m(\cdot)$ is independent of $Q_k^m(0)$, and hence the maximum in the second term is attained by setting $r=2mq_k^0$. Step (b) follows from $\lambda_k\mathbb{I}(C^m(s)=k)\le\lambda_k$ for all $t\in\mathbb{R}_+$, and (c) from the inequality in Eq. (132), by replacing ϵ with $\epsilon\lambda_k T$.

We are now ready to establish Eq. (130):

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{1}{m}\left(D_k^m(mt)-\int_0^{mt}Q_k^m(s)m^{-1}ds\right)\right|>\epsilon\right)\\
=&\mathbb{P}\left(\sup_{0\le t\le mT}\left|D_k^m(t)-\int_0^t Q_k^m(s)m^{-1}ds\right|>m\epsilon\right)
\end{aligned}$$

$$\begin{aligned}
&\overset{(a)}{\leq} 2\exp\left(-\phi m T\cdot h\left(\frac{\epsilon}{\phi T}\right)\right)+\mathbb{P}\left(\sup_{0\leq t\leq mT}\psi(W^m(t))\geq\phi\right)\\
&\overset{(b)}{\leq} 2\exp\left(-\phi m T\cdot h\left(\frac{\epsilon}{\phi T}\right)\right)+2\exp\left(-\lambda_k m T\cdot h(\epsilon)\right)+\mathbb{P}(m^{-1}Q_k^m(0)\geq 2q_k^0),
\end{aligned}\tag{134}$$

where step (a) follows from Lemma 8, and (b) from Eq. (133). Note that ϕ,ϵ and T are positive constants, and by our assumption, $\lim_{m\to\infty}\mathbb{P}(|m^{-1}Q_k^m(0)-q_k^0|>\delta)=0$ for all $\delta>0$. Therefore, Eq. (130) follows by taking the limit in the above inequality as $m\to\infty$. This completes the proof of Proposition 1.

B.2. Proof of Proposition 2 For the purpose of this proof, we will use an alternative representation of the fluid solutions using integral equations. Define the drift function, $F:\mathbb{R}_+^{\mathcal{K}}\to\mathbb{R}_+^{\mathcal{K}}$:

$$F_k(q)=\lambda_k p_k(q)-q_k,\quad q\in\mathcal{K},\tag{135}$$

where $p(\cdot)$ is defined in Eq. (14). It can be verified from the definition of $p(\cdot)$ that there exists $l_F\in\mathbb{R}_+$ such that $F_k(\cdot)$ is l_F-Lipschitz continuous for all $k\in\mathcal{K}$. Fix $q^0\in\mathbb{R}_+^{\mathcal{K}}$, let $\{q(t)\}_{t\in\mathbb{R}_+}$, $q(t)\in\mathbb{R}_+^{\mathcal{K}}$, be a solution to the following integral equation:

$$q_k(t)=q_k^0+\int_0^t F_k(q(s))ds,\quad k\in\mathcal{K}.\tag{136}$$

Similar to Lemma 1, it is not difficult to show that the function $q(\cdot)$ defined in Eq. (136) exists, is unique, and coincides with the fluid solution with initial condition q^0. We have that

$$\begin{aligned}
&\|V^m(t)-q(t)\|\\
=&\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s))-F_k(q(s))ds\right|\\
\leq&\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s))-F_k(V^m(s))ds\right|+\int_0^t\|F(V^m(s))-F(q(s))\|\,ds\\
\leq&\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s))-F_k(V^m(s))ds\right|+l_F\int_0^t\|V^m(s)-q(s)\|\,ds,
\end{aligned}\tag{137}$$

where the last inequalities come from the fact that $F_k(\cdot)$ is l_F-Lipschitz continuous for all $k\in\mathcal{K}$. In a manner analogous to Eq. (124) from the proof of Proposition 1, by applying Gronwall's lemma (Proposition 7 in Supplemental Material Section A), we have that, for all $\epsilon>0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\leq t\leq T}\|V^m(t)-q(t)\|>\epsilon\right)\\
\leq&\mathbb{P}\left(\sup_{0\leq t\leq T}\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s))-F_k(V^m(s))ds\right|>\epsilon e^{-l_F T}\right).
\end{aligned}\tag{138}$$

Therefore, in order to establish Proposition 2, it suffices to show that

$$\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\leq t\leq T}\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s))-F_k(V^m(s))ds\right|>\epsilon\right)=0,\quad\forall\epsilon>0.\tag{139}$$

In what follows, we will show (139) by using the discrete-time embedded process of $C^m(\cdot)$ and analyzing the system dynamics at the times when new choices are chosen. We begin by introducing

some notation. Fixing $i \in \mathbb{Z}_+$, we denote by S_i the ith update point, i.e., time of the ith update in $C^m(\cdot)$, with $S_0 \triangleq 0$, and

$$\tau_i^m = m^{-1}(S_i - S_{i-1}), \quad i \in \mathbb{N}. \tag{140}$$

Note that $\{\tau_i\}_{i\in\mathbb{N}}$ are independent exponential random variables with mean $(m\beta_m)^{-1}$. Let $\{\overline{Q}^m[i]\}_{i\in\mathbb{N}}$ be the discrete-time process, where $\overline{Q}^m[i]$ corresponds to the value of $\overline{Q}^m(\cdot)$ immediately following the ith update point:

$$\overline{Q}_k^m[i] = m^{-1}Q_k^m(S_i) = \overline{Q}_k^m(S_i/m), \quad k \in \mathcal{K}, i \in \mathbb{N}. \tag{141}$$

and $\{Z_k^m[i]\}_{i\in\mathbb{N}}$ be the process of indicator variables:

$$Z_k^m[i] = \mathbb{I}(C^m(S_i) = k), \quad k \in \mathcal{K}, i \in \mathbb{N}. \tag{142}$$

That is, $Z_k^m[i] = 1$ if action k is selected on the ith update point. Finally, let $\{\widehat{Q}^m(t)\}_{t\geq 0}$ be a right-continuous piece-wise constant process which coincides with $\overline{Q}^m(\cdot)$ at the points $\{S_i/m\}_{i\in\mathbb{N}}$:

$$\widehat{Q}^m(t) = \overline{Q}^m[i], \quad \forall t \in [S_i/m,\, S_{i+1}/m). \tag{143}$$

Fix $k \in \mathcal{K}$, and denote by I_t the number of updates in $C^m(\cdot)$ by time t, i.e.,

$$I_t = \max\{i : S_i \leq t\}. \tag{144}$$

By the triangle inequality, we have that

$$\begin{aligned}
&\lambda_k^{-1} \sup_{0\leq t\leq T} \left| \int_0^t G_k(V^m(s), \overline{C}^m(s)) - F_k(V^m(s))ds \right| \\
=&\lambda_k^{-1} \sup_{0\leq t\leq T} \left| \int_0^t \lambda_k \left(\mathbb{I}(\overline{C}^m(s) = k) - p_k(V^m(s)) \right) ds \right| \\
\leq& \sup_{0\leq t\leq T} \left[\left| \int_0^t \left(\mathbb{I}(\overline{C}^m(s) = k) - p_k(\widehat{Q}^m(s)) \right) ds \right| + \left| \int_0^t \left(p_k(\widehat{Q}^m(s)) - p_k(V^m(s)) \right) ds \right| \right] \\
=& \sup_{0\leq t\leq T} \left[\left| \sum_{i=1}^{I_{mt}+1} \tau_i^m \left(Z_k^m[i] - p_k(\overline{Q}^m[i]) \right) \right| + \left| \int_0^t \left(p_k(\widehat{Q}^m(s)) - p_k(V^m(s)) \right) ds \right| \right] \\
\leq& \sup_{0\leq t\leq T} \left| \sum_{i=1}^{I_{mt}+1} \tau_i^m \left(Z_k^m[i] - p_k(\overline{Q}^m[i]) \right) \right| \\
&\quad + \sup_{0\leq t\leq T} \left| \int_0^t \left(p_k(\widehat{Q}^m(s)) - p_k(V^m(s)) \right) ds \right|.
\end{aligned} \tag{145}$$

We now derive upper bounds on tail probabilities for each term on the right-hand side of Eq. (145). For the first term, define

$$\xi_n^m = \sum_{i=1}^{n} \tau_i^m \left(Z_k^m[i] - p_k(\overline{Q}^m[i]) \right), \quad n \in \mathbb{N}. \tag{146}$$

Fix $m \in \mathbb{N}$ and $\epsilon > 0$. Let

$$J_\epsilon^m = (1+\epsilon)m\beta_m T. \tag{147}$$

(To avoid the excessive use of floors and ceilings, we assume that J_ϵ^m is a positive integer. The results extend easily to the general case.) Define the events

$$\mathcal{A}_\epsilon^m = \{I_{mT} < J_\epsilon^m\}, \quad \text{and} \quad \mathcal{B}_\epsilon^m = \left\{ \max_{0\leq n\leq J_\epsilon^m} |\xi_n^m| \leq \epsilon \right\}. \tag{148}$$

We have that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T}\left|\sum_{i=1}^{I_{mt}+1}\tau_i^m\left(Z_k^m[i]-p_k(\overline{Q}^m[i])\right)\right|\ge\epsilon\right)\\
=&\mathbb{P}\left(\max_{0\le n\le I_{mT}}|\xi_n^m|\ge\epsilon\right)\\
\le&1-\mathbb{P}(\mathcal{A}_\epsilon^m\cap\mathcal{B}_\epsilon^m)\\
\le&(1-\mathbb{P}(\mathcal{A}_\epsilon^m))+(1-\mathbb{P}(\mathcal{B}_\epsilon^m)).
\end{aligned}\tag{149}$$

With the above equation in mind, we now proceed to demonstrate that both $)1-\mathbb{P}(\mathcal{A}_\epsilon^m))$ and $(1-\mathbb{P}(\mathcal{B}_\epsilon^m))$ converge to zero in the limit as $m\to\infty$. Note that I_{mT} is a Poisson random variable with mean $m\beta_m T$. Using elementary tail bounds on the Poisson distribution, we have that

$$1-\mathbb{P}(\mathcal{A}_\epsilon^m)=\mathbb{P}(I_{mT}\ge(1+\epsilon)m\beta_m T)\le\frac{1+\epsilon}{\epsilon^2}(m\beta_m T)^{-1},\tag{150}$$

which converges to zero as $m\to\infty$.

We next turn to the value of $\mathbb{P}(\mathcal{B}_\epsilon^m)$. Recall from the definition of $Z_k^m[i]$ that

$$\mathbb{E}(Z_k^m[i]\,|\,\overline{Q}^m[i])=p_k(\overline{Q}^m[i]).\tag{151}$$

It is therefore not difficult to verify that $\{\xi_n^m\}_{n\in\mathbb{N}}$ is a martingale, and our objective would be to derive an upper bound on its maximum upward excursion over $\{0,\ldots,J_\epsilon^m\}$. Unfortunately, we cannot apply the Azuma-Hoeffding inequality, because the ith increment of $\{\xi_n^m\}_{n\in\mathbb{N}}$ involves the term τ_i^m, which does not admit a bounded support. Instead, we will use the following upper bound on the moment generating function of ξ_n^m, whose proof is based on Doob's inequality and is given in Supplemental Material Section C.7.

LEMMA 9. *Fix $n\in\mathbb{N}$ and $\theta\in(0,m\beta_m)$. We have that*

$$\mathbb{E}\left(\exp\left(\theta\xi_n^m\right)\right)\le\exp\left(\frac{\theta^2 n}{4m\beta_m\left(m\beta_m-\theta\right)}\right).\tag{152}$$

We are now ready to establish an upper bound on the quantity $1-\mathbb{P}(\mathcal{B}_\epsilon^m)$. Recall that $J_\epsilon^m=(1+\epsilon)m\beta_m T$. Fix $\eta\in(0,\min\{T,1\}/2)$, and let $\theta_0=\eta\frac{\epsilon}{1+\epsilon}T^{-1}m\beta_m$. In particular, $\theta_0\in(0,m\beta_m/2)$. We have that

$$\begin{aligned}
1-\mathbb{P}(\mathcal{B}_\epsilon^m)=&\mathbb{P}\left(\max_{0\le n\le J_\epsilon^m}\xi_n^m\ge\epsilon\right)\\
\overset{(a)}{\le}&\inf_{\theta>0}\exp\left(-\theta\epsilon\right)\mathbb{E}\left(\exp\left(\theta\xi_{J_\epsilon^m}^m\right)\right)\\
\overset{(b)}{\le}&\exp\left(-\theta_0\epsilon+\frac{J_\epsilon^m\theta_0^2/4}{m\beta_m\left(m\beta_m-\theta_0\right)}\right)\\
\overset{(c)}{\le}&\exp\left(-\theta_0\epsilon+\frac{J_\epsilon^m\theta_0^2}{2(m\beta_m)^2}\right)\\
=&\exp\left(-\theta_0\left(\epsilon-\frac{(1+\epsilon)T\theta_0}{2m\beta_m}\right)\right)\\
\overset{(d)}{\le}&\exp\left(-\theta_0\left(\epsilon-\epsilon/2\right)\right)\\
=&\exp\left(-\eta\frac{\epsilon^2}{2+2\epsilon}T^{-1}m\beta_m\right).
\end{aligned}\tag{153}$$

Step (a) follows from Doob's inequality and the fact that, for any $\theta>0$, the sequence $\{\exp\left(\theta\xi_n^m\right)\}_{n\in\mathbb{N}}$ is a submartingale, and (b) from Lemma 9 with $n=J_\epsilon^m$. Step (c) is due to $\theta_0<m\beta_m/2$, and hence $m\beta_m-\theta_0>m\beta_m/2$. Finally, step (d) follows from the definition of θ_0 and that $\eta<1$.

With a derivation analogous to that of Eq. (153), we have that

$$\mathbb{P}\left(\min_{0\le n\le J_\epsilon^m}\xi_n^m\le-\epsilon\right)\le\exp\left(-\eta\frac{\epsilon^2}{2+2\epsilon}T^{-1}m\beta_m\right). \tag{154}$$

Therefore, combining Eq. (149) with Eqs. (150), (153) and (154), we have that

$$\begin{aligned}&\mathbb{P}\left(\sup_{0\le t\le T}\left|\sum_{i=1}^{I_{mt}+1}\tau_i^m\left(Z_k^m[i]-p_k(\overline{Q}^m[i])\right)\right|\ge\epsilon\right)\\ \le&(1-\mathbb{P}(\mathcal{A}_\epsilon^m))+(1-\mathbb{P}(\mathcal{B}_\epsilon^m))\\ \le&\frac{1+\epsilon}{T\epsilon^2}(m\beta_m)^{-1}+2\exp\left(-\eta\frac{\epsilon^2}{2+2\epsilon}T^{-1}m\beta_m\right).\end{aligned} \tag{155}$$

Since $m\beta_m\to\infty$ as $m\to\infty$, the above equation further implies that

$$\begin{aligned}&\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\sum_{i=1}^{I_{mt}+1}\tau_i^m\left(Z_k^m[i]-p_k(\overline{Q}^m[i])\right)\right|\ge\epsilon\right)\\ \le&\lim_{m\to\infty}\left(\frac{1+\epsilon}{T\epsilon^2}(m\beta_m)^{-1}+2\exp\left(-\eta\frac{\epsilon^2}{2+2\epsilon}T^{-1}m\beta_m\right)\right)\\ =&0\end{aligned} \tag{156}$$

We now bound the second term in Eq. (145). It is not difficult to verify, from Eq. (14), that there exists $l_p\in\mathbb{R}_+$ such that $p_k(\cdot)$ is l_p-Lipschitz continuous for all $k\in\mathcal{K}$. We have that

$$\begin{aligned}&\mathbb{P}\left(\sup_{0\le t\le T}\left|\int_0^t\left(p_k(\widehat{Q}^m(s))-p_k(V^m(s))\right)ds\right|\ge\epsilon\right)\\ \le&\mathbb{P}\left(\sup_{0\le t\le T}\int_0^t l_p\left|\widehat{Q}_k^m(s)-V_k^m(s)\right|ds\ge\epsilon\right)\\ \le&\mathbb{P}\left(\int_0^T l_p\left|\widehat{Q}_k^m(s)-V_k^m(s)\right|ds\ge\epsilon\right).\end{aligned} \tag{157}$$

It therefore suffices to show that, if $m\beta_m\to\infty$ as $m\to\infty$, then

$$\lim_{m\to\infty}\mathbb{P}\left(\int_0^T\left|\widehat{Q}_k^m(s)-V_k^m(s)\right|ds\ge\epsilon\right)=0,\quad\forall\epsilon>0. \tag{158}$$

To this end, we have that

$$\begin{aligned}&\int_0^T\left|\widehat{Q}_k^m(s)-V_k^m(s)\right|ds\\ =&\sum_{i=0}^{I_{mT}}\int_{S_i/m}^{S_{i+1}/m}\left|\widehat{Q}_k^m(t)-V_k^m(s)\right|ds\\ \le&\sum_{i=0}^{I_{mT}}\int_{S_i/m}^{S_{i+1}/m}\left|\widehat{Q}_k^m(t)-V_k^m(S_i/m)\right|+\left|V_k^m(s)-V_k^m(S_i/m)\right|ds\\ \overset{(a)}{=}&\sum_{i=0}^{I_{mT}}\tau_{i+1}^m\left|\overline{Q}_k^m[i]-V_k^m(S_i/m)\right|+\sum_{i=0}^{I_{mT}}\int_{S_i/m}^{S_{i+1}/m}|V_k^m(s)-V_k^m(S_i/m)|ds\\ \le&T\sup_{0\le s\le T}\left|\overline{Q}_k^m(s)-V_k^m(s)\right|+\sum_{i=0}^{I_{mT}}\tau_{i+1}^m\left(\tau_{i+1}^m\sup_{0\le s\le T}\left|G_k(V_k^m(s),\overline{C}^m(s))\right|\right)\end{aligned}$$

$$
\begin{aligned}
&= T \sup_{0\le s\le T}\left|\overline{Q}_k^m(s) - V_k^m(s)\right| + \left(\sup_{0\le s\le T}\left|G_k(V_k^m(s),\overline{C}^m(s))\right|\right)\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2\\
&\overset{(b)}{\le} T \sup_{0\le s\le T}\left|\overline{Q}_k^m(s) - V_k^m(s)\right| + \left(\lambda_k + \sup_{0\le s\le T}|V_k^m(s)|\right)\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2\\
&\overset{(c)}{\le} T \sup_{0\le s\le T}\left|\overline{Q}_k^m(s) - V_k^m(s)\right| + [q_0 + \lambda_k(T+1)]\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2. \qquad (159)
\end{aligned}
$$

In step (a) we have invoked the property that $\widehat{Q}_k^m(\cdot)$ is piece-wise constant. Step (b) follows from the definition of $G_k(\cdot,\cdot)$ in Eq. (35), and (c) from the fact that $|V_k^m(t)| \le q^0 + \lambda_k t$ for all $t \in \mathbb{R}_+$ (Eq. (38)). It remains to derive an upper bound on the tail probabilities of the term $\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2$, which is isolated in the form of the following lemma. The proof involves an elementary application of Markov's inequality, and is given in Supplemental Material Section C.8.

LEMMA 10. *Suppose that $m\beta_m \to \infty$ as $m \to \infty$. We have that*

$$
\lim_{m\to\infty}\mathbb{P}\left(\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2 \ge \epsilon\right) = 0. \quad \forall \epsilon > 0. \qquad (160)
$$

We are now ready to prove Eq. (158). By Eq. (159), we have that

$$
\begin{aligned}
&\lim_{m\to\infty}\mathbb{P}\left(\int_0^T\left|\widehat{Q}_k^m(s) - v_k^m(s)\right| ds \ge \epsilon\right)\\
&\le \lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le s\le T}\left|\overline{Q}_k^m(s) - V_k^m(s)\right| \ge \frac{\epsilon}{2T}\right) + \lim_{m\to\infty}\mathbb{P}\left(\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2 \ge \frac{\epsilon}{2[q_0+\lambda_k(T+1)]}\right)\\
&= 0. \qquad (161)
\end{aligned}
$$

for all $\epsilon > 0$, where the first inequality follows from a union bound, and the last step from Proposition 1 and Lemma 10. Substituting Eqs. (156) and (161) into Eq. (145), we have that

$$
\begin{aligned}
&\lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\sum_{k\in\mathcal{K}}\left|\int_0^t G_k(V^m(s),\overline{C}^m(s)) - F_k(V^m(s))ds\right| > \epsilon\right)\\
&= \lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\sum_{i=1}^{I_{mt}+1}\tau_i^m\left(Z_k^m[i] - p_k(\overline{Q}^m[i])\right)\right| \ge \epsilon/\lambda_k\right)\\
&\qquad + \lim_{m\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\int_0^t\left(p_k(\widehat{Q}^m(s)) - p_k(V^m(s))\right)ds\right| \ge \epsilon/\lambda_k\right) = 0. \qquad (162)
\end{aligned}
$$

This establishes Eq. (139), which, in light of Eq. (138) completes the proof of Proposition 2.

B.3. Proof of Proposition 3 *Proof.* We begin by showing the following, strengthened version of Lemma 4, which states that a similar stochastic dominance property holds for $Q^m(\infty)$ even when conditioning on $C^m(\cdot)$ being of a specific value.

LEMMA 11. *Let $W^m = (Q^m, C^m)$ be a random vector drawn from the steady-state distribution, $W^m(\infty)$. There exist constants , m_0 and $\gamma > 0$, such that for all $m \ge m_0$,*

$$
Q_i^m \,\big|\, \{C^m = k\} \preceq \gamma U_{\lambda_1}^m, \quad \forall i, k \in \mathcal{K}. \qquad (163)
$$

Proof. Let $\mathcal{Q}^m = \{x/m : x \in \mathbb{Z}_+^{\mathcal{K}}\}$. We have that, for all $k \in \mathcal{K}$, $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}(C^m[n] = k) &= \sum_{u \in \mathcal{Q}^m} \frac{u_k \vee \alpha_0}{\sum_{i\in\mathcal{K}}(u_i \vee \alpha_0)} \mathbb{P}\left(\overline{Q}^m[n] = u\right) \\ &\geq \sum_{u \in \mathcal{Q}^m} \frac{\alpha_0}{K\alpha_0 + \sum_{i\in\mathcal{K}} u_i} \mathbb{P}\left(\overline{Q}^m[n] = u\right) \end{aligned} \tag{164}$$

By Eq. (103) of Lemma 7, the above equation implies that

$$\mathbb{P}(C^m = k) = \lim_{n\to\infty} \mathbb{P}(C^m[n] = k) \geq \sum_{u \in \mathcal{Q}^m} \frac{\alpha_0}{K\alpha_0 + \sum_{i\in\mathcal{K}} u_i} \mathbb{P}\left(\overline{Q}^m[\infty] = u\right). \tag{165}$$

where the last step follows from the fact that $\frac{\alpha_0}{K\alpha_0+\sum_{i\in\mathcal{K}} u_i}$ is always bounded from above by $1/K$.

By Eq. (84), we have that $U^m_{\lambda_1}/m \to \lambda_1$ almost surely as $m \to \infty$. Therefore, there exist m_0 and $y > 0$, such that

$$\mathbb{P}\left(\frac{1}{m} U^m_{\lambda_1} \geq \lambda_1 + y\right) \leq \frac{1}{2K^2}, \quad \forall m \geq m_0. \tag{166}$$

Combining Eq. (102) in Lemma 7 with Eqs. (165) and (166), we have that, for all $m \geq m^0$,

$$\begin{aligned}&\mathbb{P}(C^m = k) \\ &\overset{(a)}{\geq} \sum_{u \in \mathcal{Q}^m} \frac{\alpha_0}{K\alpha_0 + \sum_{i\in\mathcal{K}} u_i} \mathbb{P}\left(\overline{Q}^m[\infty] = u\right) \\ &\geq \mathbb{P}\left(\max_{i\in\mathcal{K}} \overline{Q}^m_i[\infty] \leq \lambda_1 + y\right) \frac{\alpha_0}{K\alpha_0 + K(\lambda_1 + y)} + \mathbb{P}\left(\max_{i\in\mathcal{K}} \overline{Q}^m_i[\infty] > \lambda_1 + y\right) \cdot 0 \\ &\overset{(b)}{\geq} \left(1 - K\mathbb{P}\left(\frac{1}{m} U^m_{\lambda_1} \leq \lambda_1 + y\right)\right) \frac{\alpha_0}{K\alpha_0 + K(\lambda_1 + y)} \\ &\overset{(c)}{\geq} \frac{\alpha_0\,(1 - 1/2K)}{K\alpha_0 + K(\lambda_1 + y)} \\ &= \gamma^{-1}, \end{aligned} \tag{167}$$

where $\gamma \overset{\triangle}{=} \frac{K\alpha_0 + K(\lambda_1+y)}{\alpha_0(1-1/2K)}$. Step (a) follows from Eq. (165), (b) from Lemma 4 and a union bound, and (c) from Eq. (166).

Fix $x \in \mathbb{Z}_+$. We have that, for all $m \geq m_0$,

$$\begin{aligned}\mathbb{P}\left(Q^m_i \geq x \,\middle|\, C^m = k\right) &= \frac{\mathbb{P}\left(Q^m_i \geq x,\, C^m = k\right)}{\mathbb{P}(C^m = k)} \leq \frac{\mathbb{P}\left(Q^m_i \geq x\right)}{\mathbb{P}(C^m = k)} \\ &\overset{(a)}{\leq} \frac{\mathbb{P}\left(U^m_{\lambda_1} \geq x\right)}{\mathbb{P}(C^m = k)} \\ &\overset{(b)}{\leq} \gamma \mathbb{P}\left(U^m_{\lambda_1} \geq x\right), \end{aligned} \tag{168}$$

for all $i, k \in \mathcal{K}$, where step (a) follows from Lemma 4, and (b) from Eq. (167). Since the above inquality holds for all $x \in \mathbb{Z}_+$, this completes the proof of Lemma 11. □

We now prove the convergence in Eq. (106). Recall that the first update point, S^m_1, is exponentially distributed with mean $1/\beta_m$, and independent from $W^m(0)$. Define the event $\mathcal{E}^m = \{S^m_1 \geq h(m)\}$. We have that

$$\mathbb{P}(\mathcal{E}^m) = \mathbb{P}(S^m_1 \geq h(m)) = \exp(-\beta_m h(m)) \to 1, \quad \text{as } m \to \infty. \tag{169}$$

Fix $i, k \in \mathcal{K}$. Recall from Eq. (84) that $U^m_{\lambda_1}/m$ converges to λ_1 almost surely as $m \to \infty$. This implies that there exists $\upsilon > 0$, independent of m, such that

$$\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}^m_i[0] > \upsilon \,\middle|\, C^m[0] = k\right) \overset{(a)}{\leq} \limsup_{m\to\infty} \mathbb{P}\left(U^m_{\lambda_1}/m > \gamma^{-1}\upsilon\right) = 0, \tag{170}$$

where step (a) follows from Lemma 11.

Fix $x \in \mathbb{R}_+$. We have that

$$\begin{aligned}
&\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m[1] \geq x \,\middle|\, C^m[0]=k\right)\\
\leq &\limsup_{m\to\infty} \left[\mathbb{P}\left(\overline{Q}_i^m[1] \geq x \,\middle|\, C^m[0]=k,\, \mathcal{E}^m\right) + (1-\mathbb{P}(\mathcal{E}^m \,|\, C^m[0]=k))\right]\\
\overset{(a)}{=} &\limsup_{m\to\infty} \left[\mathbb{P}\left(\overline{Q}_i^m[1] \geq x \,\middle|\, C^m[0]=k,\, \mathcal{E}^m\right) + (1-\mathbb{P}(\mathcal{E}^m))\right]\\
\overset{(b)}{=} &\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m[1] \geq x \,\middle|\, C^m[0]=k,\, \mathcal{E}^m\right)
\end{aligned} \tag{171}$$

where step (a) follows from the independence between $\mathcal{E}^m$ and $C^m[0]$, and (b) from Eq. (169).

We now bound the term on the right-hand side of Eq. (171). Denote by $B(n,p)$ a binomial random variable with n trials and a success probability of p per trial, and by $U_\lambda^m(t)$ the number of jobs in system at time t in an initially empty $M/M/\infty$ queue with arrival rate λ and departure rate $1/m$. We have that

$$\begin{aligned}
&\mathbb{P}\left(\overline{Q}_i^m[1] \geq x \,\middle|\, C^m[0]=k,\, \mathcal{E}^m\right)\\
\overset{(a)}{=}&\mathbb{P}\left(\frac{1}{m}\left(U^m_{\lambda_k}(S_1^m)\mathbb{I}(i=k) + B\left(Q_i^m(0), e^{-S_1^m/m}\right)\right) \geq x \,\middle|\, C^m[0]=k, \mathcal{E}^m\right)\\
\overset{(b)}{\leq}&\mathbb{P}\left(\frac{1}{m}\left(U^m_{\lambda_k}(S_1^m)\mathbb{I}(i=k) + B\left(Q_i^m(0), e^{-h(m)/m}\right)\right) \geq x \,\middle|\, C^m[0]=k\right)\\
\overset{(c)}{\leq}&\mathbb{P}\left(\frac{1}{m}\left(U^m_{\lambda_k}\mathbb{I}(i=k) + B\left(Q_i^m(0), e^{-h(m)/m}\right)\right) \geq x \,\middle|\, C^m[0]=k\right)\\
\leq&\mathbb{P}\left(\frac{1}{m}\left(U^m_{\lambda_k}\mathbb{I}(i=k) + B\left(\upsilon m, e^{-h(m)/m}\right)\right) \geq x \,\middle|\, C^m[0]=k, \overline{Q}_i^m(0) \leq \upsilon\right) + \mathbb{P}\left(\overline{Q}_i^m(0) > \upsilon \,\middle|\, C^m[0]=k\right)\\
=&\mathbb{P}\left(\frac{1}{m}\left(U^m_{\lambda_k}\mathbb{I}(i=k) + B\left(\upsilon m, e^{-h(m)/m}\right)\right) \geq x\right) + \mathbb{P}\left(\overline{Q}_i^m(0) > \upsilon \,\middle|\, C^m[0]=k\right).
\end{aligned} \tag{172}$$

For step (a), note that each unit of reward initially present at time $t=0$ has probability of $\exp(-S_1^m/m)$ or remaining in the system by $t=S_1^m$. Therefore, the rewards in site i at time S_1^m satisfy the following decomposition:

$$Q_i^m[1] \overset{d}{=} U^m_{\lambda_k}(S_1^m)\mathbb{I}(i=k) + B\left(Q_i^m(0), e^{-S_1^m/m}\right). \tag{173}$$

The first term corresponds to the units of rewards at $t=S_1^m$ that had arrived during the interval $(0, S_1^m)$, and hence is non-zero only if $i=k$. The second term corresponds to those individuals initially present at $t=0$ who remained in the system by $t=S_1^m$. Step (b) follows from the definition of $\mathcal{E}^m$, and (c) from the well-known fact that the number of jobs in system in an initially empty $M/M/\infty$ queue at any time is always stochastically dominated by its steady-state distribution.

Because $h(m)/m \to \infty$ as $m\to\infty$, we have that $\lim_{m\to\infty}\mathbb{E}\left(\frac{1}{m}B\left(\upsilon m, e^{-h(m)/m}\right)\right)=0$. Applying Markov's inequality, we obtain that

$$\frac{1}{m}B\left(\upsilon m, e^{-h(m)/m}\right) \overset{P}{\to} 0, \quad \text{as } m\to\infty, \tag{174}$$

where $\overset{P}{\to}$ denotes convergence in probability. Recall from Eq. (84) that, almost surely,

$$\frac{1}{m}U^m_{\lambda_k}\mathbb{I}(i=k) \to \lambda_k\mathbb{I}(i=k) = q^*_{k,i}. \tag{175}$$

Fix $\epsilon > 0$, and substitute Eq. (172) into Eq. (171). We have that

$$\begin{aligned}
&\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m[1] \geq q_{k,i}^* + \epsilon \,\middle|\, C^m[0] = k\right) \\
=&\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m[1] \geq q_{k,i}^* + \epsilon \,\middle|\, C^m[0] = k, \mathcal{E}^m\right) \\
\leq&\limsup_{m\to\infty} \mathbb{P}\left(\frac{1}{m}\left(U_{\lambda_k}^m \mathbb{I}(i=k) + B\left(\upsilon m, e^{-h(m)/m}\right)\right) \geq q_{k,i}^* + \epsilon\right) \\
&+\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m(0) > \upsilon \,\middle|\, C^m[0] = k\right) \\
\overset{(a)}{=}&\limsup_{m\to\infty} \mathbb{P}\left(\frac{1}{m}\left(U_{\lambda_k}^m \mathbb{I}(i=k) + B\left(\upsilon m, e^{-h(m)/m}\right)\right) \geq q_{k,i}^* + \epsilon\right) \\
\overset{(b)}{=}&0,
\end{aligned} \tag{176}$$

where step (a) follows from Eq. (170), and (b) from Eqs. (174) and (175). Using the same line of arguments as that in Eqs. (171) through (176), we can show that

$$\limsup_{m\to\infty} \mathbb{P}\left(\overline{Q}_i^m[1] \leq q_{k,i}^* - \epsilon \,\middle|\, C^m[0] = k\right) = 0, \tag{177}$$

which, along with Eq. (176), yields that

$$\limsup_{m\to\infty} \mathbb{P}\left(\left|\overline{Q}_i^m[1] - q_{k,i}^*\right| > \epsilon \,\middle|\, C^m[0] = k\right) = 0. \tag{178}$$

Since the above equation holds for all $i, k \in \mathcal{K}$, this proves Eq. (106) in Proposition 3.

We now turn to Eq. (107). Fix $i, k \in \mathcal{K}$. Using essentially identical arguments as those for Eq. (178), we can show that

$$\limsup_{m\to\infty} \mathbb{P}\left(\left|m^{-1} Q_i^m(h(m)) - q_{k,i}^*\right| > \epsilon \,\middle|\, C^m[0] = k\right) = 0, \quad \forall \epsilon > 0. \tag{179}$$

We have that,

$$\begin{aligned}
Q_i^m(h(m))|\{C^m[0] = k\} &\overset{(a)}{\preceq} U_{\lambda_1}^m(h(m)) + Q_i^m(0)|\{C^m[0] = k\} \\
&\overset{(b)}{\preceq} U_{\lambda_1}^m(h(m)) + \gamma U_{\lambda_1}^m \\
&\preceq (\gamma+1) U_{\lambda_1}^m,
\end{aligned} \tag{180}$$

Step (a) follows from a decomposition similar to Eq. (173), by the writing the recallable rewards at time $h(m)$ as those who arrived after $t = 0$, which is dominated by $U_{\lambda_1}^m(h(m))$, and those who were in the system at $t = 0$, which is dominated by $Q_i^m(0)|\{C^m[0] = k\}$. Step (b) follows from Lemma 4. Since $U_{\lambda_1}^m$ is a Poisson distribution with mean $m\lambda_1$, it is not difficult show that there exists a random variable $Y \in \mathbb{R}_+$, such that

$$m^{-1} Q_i^m(h(m))|\{C^m[0] = k\} \overset{(a)}{\preceq} \frac{1}{m} U_{\lambda_1}^m(\gamma+1) \preceq Y, \quad \forall m \in \mathbb{N}. \tag{181}$$

Combining Eqs. (106) and (181), the dominated convergence theorem implies that, for all $i, k \in \mathcal{K}$,

$$\lim_{m\to\infty} \mathbb{E}\left(\left|m^{-1} Q_i^m(h(m)) - q_{k,i}^*\right| \,\middle|\, C^m[0] = k\right) = 0. \tag{182}$$

This shows Eq. (107), and thus completes the proof of Proposition 3. □

B.4. Proof of Proposition 5 *Proof.* Fix $\eta < 1$. Under the polynomial reward-matching model, the fluid solution satisfies

$$\dot{q}_k(t) = \lambda_k \frac{(q_k \vee \alpha_0)^\eta}{\sum_{i\in\mathcal{K}}(q_i \vee \alpha_0)^\eta} - q_k(t), \quad \forall k \in \mathcal{K}, \tag{183}$$

setting the left-hand side to 0, we have that a state q is an invariant state of the fluid solutions if

$$\frac{q_k}{(q_k \vee \alpha_0)^\eta} = \lambda_k \frac{1}{\sum_{i\in\mathcal{K}}(q_i \vee \alpha_0)^\eta}, \quad \forall k \in \mathcal{K}. \tag{184}$$

We first show that the above equations admit a unique solution, q^I. That is, the fluid solutions admit a unique invariant state. Suppose, for the sake of contradiction, that there exist two distinct invariant states q^I and $\tilde{q}^I$. Let

$$Z = \sum_{i\in\mathcal{K}} (q_i^I \vee \alpha_0)^\eta \tag{185}$$

and $\tilde{Z} = \sum_{i\in\mathcal{K}}(\tilde{q}_i^I \vee \alpha_0)^\eta$ denote the denominators on the right-hand side of Eq. (184) under q^I and $\tilde{q}^I$, respectively. From Eq. (184), by considering separately two cases depending on whether q_k^I is smaller than α_0, we have that the invariant state satisfies

$$q_k^I = \begin{cases} \left(\frac{\lambda_k}{Z}\right)^{\frac{1}{1-\eta}} & (\geq \alpha_0), \quad \text{if} \quad \lambda_k \geq Z\alpha_0^{1-\eta}, \\ \lambda_k \frac{\alpha_0^\eta}{Z} & (< \alpha_0), \quad \text{if} \quad \lambda_k < Z\alpha_0^{1-\eta}, \end{cases} \tag{186}$$

which indicates that if $Z = \tilde{Z}$, then $q^I = \tilde{q}^I$. Therefore, in order for q^I and $\tilde{q}^I$ to be distinct, we must have that $\tilde{Z} \neq Z$. Without loss of generality, let us assume that $\tilde{Z} > Z$. Because $\eta < 1$, by Eq. (186), we have that q_k^I is a monotonically decreasing function of Z, for all k. We thus have that $\tilde{q}_k^I < q_k^I$ for all $k \in \mathcal{K}$. This leads to a contradiction, since when $\tilde{q}_k^I < q_k^I$ for all $k \in \mathcal{K}$, we will necessarily have that $\tilde{Z}$ is strictly less than Z. This proves that the solution to Eq. (184) must be unique.

We now find the unique invariant state q^I, and for now we assume that such q^I exists. Note that when $\eta < 1$, q_k^I is a monotonically increasing function of λ_k for $\lambda_k \geq 0$. Eq. (186) implies that $q_i \geq q_j$ if and only if $\lambda_i \geq \lambda_j$, which further implies that $q_1^I \geq \cdots \geq q_K^I$. Since q_k^I is non-increasing in k, we may define i^* as the unique index such that

$$q_{i^*}^I \geq \alpha_0, \quad \text{and} \quad q_{i^*+1}^I < \alpha_0, \tag{187}$$

where we define $i^* = 0$ if $q_1^I < \alpha_0$, and $i^* = K$ if $q_K^I \geq \alpha_0$.

We now consider different values of α_0. Suppose that $\alpha_0 \geq \lambda_1/K$. It is not difficult to verify that in this case $i^* = 0$, $q_k^I < \alpha_0$ for all $k \in \mathcal{K}$, and $Z = \sum_{i\in\mathcal{K}}(q_i^I \vee \alpha_0)^\eta = K\alpha_0^\eta$. It follows from Eq. (186) that we must have $\lambda_1 < Z\alpha^{1-\eta} = K\alpha_0$, or equivalently, $\alpha_0 > \lambda_1/K$, and that

$$p_k^I = \lambda_k/K, \quad \forall k \in \mathcal{K}. \tag{188}$$

This proves Item 1 in the proposition.

Consider next the other extreme where α_0 is so small that $i^* = K$ and $q_k^I \geq \alpha_0$ for all k. By Eq. (186), this is to say that

$$\lambda_K \geq Z\alpha_0^{1-\eta}. \tag{189}$$

In this case, we have that $Z = \sum_{i\in\mathcal{K}} \left(\frac{\lambda_i}{Z}\right)^{\frac{\eta}{1-\eta}}$, which leads to, after rearrangement,

$$Z = \left(\sum_{i\in\mathcal{K}} \lambda_i^{\frac{\eta}{1-\eta}}\right)^{1-\eta}. \tag{190}$$

Substituting the value of Z from Eq. (190) into (189), we obtain the condition on α_0:

$$\alpha_0 \le \lambda_K \frac{\lambda_K^{\frac{\eta}{1-\eta}}}{\sum_{i\in\mathcal{K}} \lambda_i^{\frac{\eta}{1-\eta}}}. \tag{191}$$

The expression of q_k^I in Eq. (118) is obtained by substituting Eq. (190) into the top line of Eq. (186). This proves Item 2 of the proposition.

Finally, fix α_0 such that $\lambda_K \frac{\lambda_K^{\frac{\eta}{1-\eta}}}{\sum_{i\in\mathcal{K}} \lambda_i^{\frac{\eta}{1-\eta}}} < \alpha_0 \le \lambda_1/K$. In this case, we have that $1 \le i^* \le K-1$. By Eq. (186), we have that for all $k \ge i^*+1$,

$$\begin{aligned} q_k^I =& \lambda_k \frac{\alpha_0^\eta}{Z} \\ =& \frac{\lambda_{i^*}}{Z} \cdot \frac{\lambda_k}{\lambda_{i^*}} \alpha^\eta \\ =& \left(q_{i^*}^I\right)^{1-\eta} \frac{\lambda_k}{\lambda_{i^*}} \alpha_0^\eta, \end{aligned} \tag{192}$$

where the last equality follows from the fact that $q_{i^*}^I \ge \alpha_0$, and hence $q_{i^*}^I = \left(\frac{\lambda_{i^*}}{Z}\right)^{\frac{1}{1-\eta}}$. This yields

$$q_k^I = \begin{cases} q_{i^*}^I \left(\frac{\lambda_k}{\lambda_{i^*}}\right)^{\frac{1}{1-\eta}}, & k = 1, \dots, i^*-1, \\ \left(q_{i^*}^I\right)^{1-\eta} \frac{\lambda_k}{\lambda_{i^*}} \alpha_0^\eta, & k = i^*+1, \dots, K. \end{cases}, \tag{193}$$

which proves Eq. (122) in Item 3. It remains to identify the values of i^* and $q_{i^*}^I$. By Eq. (186), in order to have $q_{i^*}^I \ge \alpha_0$, it is necessary and sufficient to have

$$\begin{aligned} \lambda_{i^*} \ge& \left(\sum_{i\in\mathcal{K}} (q_i^I \vee \alpha_0)^\eta\right) \alpha_0^{1-\eta} \\ \overset{(a)}{=}& \left((K-i^*)\alpha_0^\eta + \sum_{i=1}^{i^*} (q_i^I)^\eta\right) \alpha_0^{1-\eta} \\ \overset{(b)}{=}& \left((K-i^*)\alpha_0^\eta + \sum_{i=1}^{i^*} \left(\frac{\lambda_i}{\lambda_{i^*}}\right)^{\frac{\eta}{1-\eta}} \left(q_{i^*}^I\right)^\eta\right) \alpha_0^{1-\eta} \\ \overset{(c)}{\ge}& \left((K-i^*) + \sum_{i=1}^{i^*} \left(\frac{\lambda_i}{\lambda_{i^*}}\right)^{\frac{\eta}{1-\eta}}\right) \alpha_0 \\ =& g(i^*), \end{aligned} \tag{194}$$

where the equalities (a) and (b) are based on Eq. (193), and inequality (c) uses the fact that $q_{i^*}^I \ge \alpha_0$. Analogously, in order to have $q_{i^*}^I < \alpha_0$, it is necessary and sufficient to have

$$\begin{aligned} \lambda_{i^*+1} <& \left(\sum_{i\in\mathcal{K}} (q_i^I \vee \alpha_0)^\eta\right) \alpha_0^{1-\eta} \\ =& \left((K-i^*)\alpha_0^\eta + \sum_{i=1}^{i^*} (q_i^I)^\eta\right) \alpha_0^{1-\eta} \\ =& \left((K-i^*)\alpha_0^\eta + \sum_{i=1}^{i^*} \left(\frac{\lambda_i}{\lambda_{i^*}}\right)^{\frac{\eta}{1-\eta}} \left(q_{i^*}^I\right)^\eta\right) \alpha_0^{1-\eta} \end{aligned}$$

$$\begin{aligned}&\overset{(a)}{<}\left((K-i^*-1)\alpha_0^\eta+\sum_{i=1}^{i^*+1}\left(\frac{\lambda_i}{\lambda_{i^*+1}}\right)^{\frac{\eta}{1-\eta}}\alpha_0^\eta\right)\alpha_0^{1-\eta}\\&=g(i^*+1),\end{aligned}\tag{195}$$

where inequality (a) is derived from the bottom line of Eq. (193):

$$\left(q_{i^*}^I\right)^\eta=\left(\frac{q_{i^*+1}^I}{\alpha_0^\eta}\right)^{\frac{\eta}{1-\eta}}\left(\frac{\lambda_{i^*}}{\lambda_{i^*+1}}\right)^{\frac{\eta}{1-\eta}}>\alpha_0^\eta\left(\frac{\lambda_{i^*}}{\lambda_{i^*+1}}\right)^{\frac{\eta}{1-\eta}}.$$

Eqs. (194) and (195) thus give a charaterization of i^* in terms of the problem primitives, and establish Eq. (120). Note that the proceeding analysis has already shown that i^* is unique and lies in $\{1,\dots,K-1\}$, although the uniqueness can also be derived easily by noticing that $g(\cdot)$ is a non-decreasing function (since $\left(\frac{\lambda_j}{\lambda_i}\right)^{\frac{\eta}{1-\eta}}\geq 1$ whenever $j\leq i$) and the λ_i's are non-increasing in i.

Finally, we show that $q_{i^*}^I$ exists and is given by the unique solution to Eq. (121). Substituting the expressions for the q_k^I's from Eq. (193) into Eq. (184) leads to Eq. (121). In particular, $q_{i^*}^I$ is the solution to the following equation:

$$\left(q_{i^*}^I\right)^{1-\eta}=\frac{\lambda_{i^*}}{(K-i^*)\alpha_0^\eta+\left(q_{i^*}^I\right)^\eta\sum_{j=1}^{i^*}\left(\frac{\lambda_j}{\lambda_{i^*}}\right)^{\frac{\eta}{1-\eta}}}.\tag{196}$$

To see that such a solution exists and is unique for any fixed $i^*\in\{1,\dots,K-1\}$, note that since $\eta<1$, the left-hand side of the above equation is a strictly increasing function in $q_{i^*}^I$, which grows from 0 to ∞ as $q_{i^*}^I$ varies from 0 to ∞; the right-hand side, on the other hand, is a strictly decreasing function in $q_{i^*}^I$, which, as $q_{i^*}^I$ varies from 0 to ∞, decreases from $\frac{\lambda_{i^*}}{(K-i^*)\alpha_0^\eta}$ to 0. Together, they imply that Eq. (196) must admit a unique solution in $\mathbb{R}_+$. This completes the proof of Proposition 5. □

C. Proofs for Lemmas

C.1. Proof of Lemma 1 *Proof.* The existence and uniqueness of the fluid solution follow from Picard's existence theorem (Section 2, Chapter 1, Coddington and Levinson (1955)) by verifying that the right-hand side of Eq. (15) is uniformly Lipschitz-continuous in $q(t)$ over $\mathbb{R}_+^{\mathcal{K}}$, and (trivially) continuous in t. To show the fluid solution's continuous dependence on initial condition, note that because of the Lipschitz continuity of $p_k(\cdot)$, there exists a constant $l>0$, such that for initial conditions $x,y\in\mathbb{R}_+^{\mathcal{K}}$ and $t\in\mathbb{R}_+$, we have that

$$\begin{aligned}\|q(x,t)-q(y,t)\|\leq&\|x-y\|+\int_0^t\|(p(q(x,s))-q(x,s))-(p(q(y,s))-q(y,s))\|ds\\\leq&\|x-y\|+l\int_0^t\|q(x,s)-q(y,s)\|ds\\\leq&\|x-y\|e^{lt},\end{aligned}\tag{197}$$

where the last inequality follows from Gronwall's lemma (Proposition 7). Therefore, $\lim_{x\to y}q(x,t)=q(y,t)$ for all $y\in\mathbb{R}_+^{\mathcal{K}}$. This completes the proof of the lemma. □

C.2. Proof of Lemma 4 *Proof.* We will use a simple coupling argument as follows. Fixing $k\in\mathcal{K}$, the evolution of the process $U_{\lambda_k}^m(\cdot)$ corresponds to that of $Q_k^m(\cdot)$ if action k were selected for all $t\in\mathbb{R}_+$, and it follows, given the same initial condition, that $Q_k^m(t)$ is stochastically dominated by $U_k^m(t)$ for all $t\in\mathbb{R}_+$. Since $U_k^m(\cdot)$ is positive recurrent for all $k\in\mathcal{K}$, we know that $Q^m(\cdot)$ is also positive recurrent, which in turn implies the positive recurrence of $\overline{W}^m(\cdot)$, because the evolution

of $C^m(\cdot)$ is derived by sampling from $\mathcal{K}$ based solely on the value of $Q^m(\cdot)$. Lemma 4 follows form the above-mentioned stochastic dominance, the fact that under any finite initial condition, $U_k^m(t)$ converges in distribution to $U_{\lambda_k}^m$ as $t\to\infty$, and the observation that $U_{\lambda'}^m \preceq U_{\lambda}^m$ whenever $\lambda\geq\lambda'\geq 0$. □

C.3. Proof of Lemma 5 *Proof.* To show Eq. (87), observe that

$$\begin{aligned}
&\lim_{x\to\infty}\sup_{m\in\mathbb{N}}(1-\pi_W^m(\overline{\mathcal{Q}}_x))\\
&=\lim_{x\to\infty}\sup_{m\in\mathbb{N}}\mathbb{P}\left(\max_{k\in\mathcal{K}}\overline{Q}_k^m(\infty)>x\right)\overset{(a)}{\leq}\lim_{x\to\infty}\sup_{m\in\mathbb{N}}K\mathbb{P}(U_{\lambda_1}^m/m>x)\\
&=\lim_{x\to\infty}\sup_{m\in\mathbb{N}}Ke^{-m\lambda_1}\sum_{i=mx}^{\infty}\frac{(m\lambda_1)^i}{i!}\overset{(b)}{\leq}\lim_{x\to\infty}\sup_{m\in\mathbb{N}}K\sum_{i=mx}^{\infty}\left(\frac{em\lambda_1}{i}\right)^i\\
&\leq\lim_{x\to\infty}\sup_{m\in\mathbb{N}}K\sum_{i=mx}^{\infty}\left(\frac{em\lambda_1}{mx}\right)^i\leq\lim_{x\to\infty}K\sum_{i=x}^{\infty}\left(\frac{e\lambda_1}{x}\right)^i\\
&\leq\lim_{x\to\infty}K2^{-(x-1)}\\
&=0,
\end{aligned}\tag{198}$$

where step (a) follows from Lemma 4 and the union bound, and (b) from the elementary inequality $i!\geq(i/e)^i$. Eq. (198) thus shows that for all ϵ, $\inf_{m\in\mathbb{N}}\pi_W^m(\overline{\mathcal{Q}}_x)>1-\epsilon$ for all sufficiently large x, which proves Eq. (87). □

C.4. Proof of Lemma 6 *Proof.* We first show the following uniform convergence property:

$$\lim_{m\to\infty}\sup_{x\in S\cap\mathcal{Q}^m}\mathbb{P}\left(\|q(x,t)-\overline{Q}^m(x,t)\|>\delta\right)=0,\quad\forall\delta>0.\tag{199}$$

where $\overline{Q}^m(x,\cdot)$ denotes a process $\overline{Q}^m(\cdot)$ initialized with $\mathbb{P}(\overline{Q}^m(0)=x)=1$. Suppose, for the sake of contradiction, that there exist $\delta>0$ and a sequence $\{x_i\}_{i\in\mathbb{N}}$, $x_i\in S\cap\mathcal{Q}^m$, such that

$$\limsup_{i\to\infty}\mathbb{P}\left(\|q(x_i,t)-\overline{Q}^m(x_i,t)\|>\delta\right)>0.\tag{200}$$

Because S is compact, there exists a sub-sequence $\{x_{i_j}\}_{j\in\mathbb{N}}\subset\{x_i\}_{i\in\mathbb{N}}$ and $x^*\in S$ such that $x_{i_j}\to x^*$ as $j\to\infty$. We have that

$$\limsup_{j\to\infty}\mathbb{P}\left(\|q(x_{i_j},t)-\overline{Q}^m(x_{i_j},t)\|>\delta\right)\overset{(a)}{=}\limsup_{j\to\infty}\mathbb{P}\left(\|q(x^*,t)-\overline{Q}^m(x_{i_j},t)\|>\delta\right)>0,\tag{201}$$

where step (a) follows from the fact that, for all $t\in\mathbb{R}_+$, $q(x,t)$ is continuous with respect to x (Lemma 1). This leads to a contradiction with Theorem 5, and hence proves Eq. (199).

C.5. Proof of Lemma 7 *Proof.* It is not difficult verify that $\{W^m[n]\}_{n\in\mathbb{Z}_+}$ is a time-homogeneous, aperiodic, and irreducible Markov chain. Because the continuous-time process, $\{W^m(t)\}_{t\in\mathbb{R}_+}$, is positive recurrent, so is its discrete-time counterpart, $\{W^m[n]\}_{n\in\mathbb{N}}$, and $W^m[n]$ converges to its steady-state distribution, $W^m[\infty]$, as $n\to\infty$. Eq. (102) follows from the same argument as that of Lemma 4. We now show Eq. (103). Denote by $N^m(t)$ the index of the last update point by time t:

$$N^m(t)=\sup\{n:S_n^m\leq t\},\tag{202}$$

and by $T^m(t)$ its value

$$T^m(t)=S_{N^m(t)}^m.\tag{203}$$

Recall that the update points are generated according to a Poisson process, and it not difficult to show that, almost surely, $N^m(t) \to \infty$ and $T^m(t) \to \infty$, as $t \to \infty$. We thus have that, for all $k \in \mathcal{K}$,

$$\mathbb{P}(C^m[\infty] = k) = \lim_{n\to\infty} \mathbb{P}(C^m[n] = k) = \lim_{t\to\infty} \mathbb{P}(C^m(T^m(t)) = k) = \mathbb{P}(C^m(\infty) = k). \tag{204}$$

The same argument applies for $Q^m[\infty]$ versus $Q^m(\cdot)$. This completes the proof of Lemma 7. □

C.6. Proof of Lemma 8

Proof. Define

$$\gamma(t) = N(t) - \int_0^t \psi(W^m(s))ds, \quad t \in \mathbb{R}_+. \tag{205}$$

Let $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$ be the natural filtration associated with $W^m(\cdot)$. It is not difficult to show, by the definition of $N(\cdot)$, that $\{\gamma(t)\}_{t\in\mathbb{R}_+}$ is martingale with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$. Define the stopping time

$$T_s = \inf\{t : \psi(W^m(t)) \geq \phi\}. \tag{206}$$

Let $\{\tilde{N}(t)\}_{t\in\mathbb{N}}$ be a counting process defined as:

$$\tilde{N}(t) = N(t \wedge T_s), \quad t \in \mathbb{R}_+. \tag{207}$$

That is, $\tilde{N}(t)$ coincides with $N(t)$ up until $t = T_s$, and stays constant afterwards. Let

$$\tilde{\psi}(t) \triangleq \psi(W^m(t))\mathbb{I}(t \leq T_s). \tag{208}$$

Then, it is not difficult to show that $\tilde{N}(\cdot)$ is a counting process whose instantaneous rate at time t is $\tilde{\psi}(t)$, and the process

$$\tilde{\gamma}(t) = \tilde{N}(t) - \int_0^t \tilde{\psi}(s)ds, \quad t \in \mathbb{R}_+,$$

is a martingale with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$. From the definition of $\tilde{\gamma}(\cdot)$ and $\gamma(\cdot)$, we have that, for all $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0\leq t\leq T} |\gamma(t)| \geq \epsilon\right) \\
=&\mathbb{P}\left(\sup_{0\leq t\leq T} |\gamma(t)| \geq \epsilon, T_s > T\right) + \mathbb{P}\left(\sup_{0\leq t\leq T} |\gamma(t)| \geq \epsilon, T_s \leq T\right) \\
\leq&\mathbb{P}\left(\sup_{0\leq t\leq T} |\tilde{\gamma}(t)| \geq \epsilon\right) + \mathbb{P}\left(T_s \leq T\right) \\
=&\mathbb{P}\left(\sup_{0\leq t\leq T} |\tilde{\gamma}(t)| \geq \epsilon\right) + \mathbb{P}\left(\sup_{0\leq t\leq T} \psi(W^m(t)) \geq \phi\right).
\end{aligned} \tag{209}$$

To complete the proof, therefore, it suffices to show that

$$\mathbb{P}\left(\sup_{0\leq t\leq T} |\tilde{\gamma}(t)| \geq \epsilon\right) \leq 2e^{-\phi T \cdot h(\epsilon/\phi T)}. \tag{210}$$

We now show Eq. (210) using Doob's inequality (Proposition 6 in Supplemental Material Section A), by following a line of arguments similar to that used in the proof of Theorem 2.2 of Kurtz (1978). First, we introduce a representation of the Markov process $W^m(\cdot)$ using Poisson processes. Let $\{\Xi_y(\cdot)\}_{y\in\mathbb{Z}^{|\mathcal{K}|+1}}$ a family of mutually independent unit-rate Poisson counting processes, indexed by $\mathbb{Z}^{|\mathcal{K}|+1}$. For every $y \in \Xi_y$, let $\psi_y(\cdot)$ be the rate function of $W^m(\cdot)$ for the jump with value y, i.e.,

$\psi_y(w)$ is the instantaneous rate at which $W^m(\cdot)$ jumps to state $w+y$ when in state $w \in \mathbb{Z}_+^{K+1}$. Then, the process $W^m(\cdot)$ can be expressed as a solution to the following integral equation:

$$W^m(t) = W^m(0) + \sum_{y \in \mathbb{Z}_+^{K+1}} y \, \Xi_y \left(\int_0^t \psi_y(W^m(s)) ds \right), \quad t \in \mathbb{R}_+. \tag{211}$$

In this representation, $\Xi_y \left(\int_0^t \psi_y(W^m(s)) ds \right)$ counts the number of jumps of value y over the interval $[0,t]$. We thus have that, by setting $y = l$,

$$N(t) = \Xi_l \left(\int_0^t \psi(W^m(s)) ds \right), \quad t \in \mathbb{R}_+, \tag{212}$$

and

$$\tilde{\gamma}(t) = \tilde{N}(t) - \int_0^t \tilde{\psi}(s) ds = \Xi_l \left(\int_0^t \tilde{\psi}(s) ds \right) - \int_0^t \tilde{\psi}(s) ds, \quad t \in \mathbb{R}_+. \tag{213}$$

Fix $\theta > 0$. Since $\tilde{\gamma}(\cdot)$ is a martingale and $\exp(\cdot)$ a positive convex function, $\{\exp(\theta \tilde{\gamma}(t))\}_{t \in \mathbb{R}_+}$ is a submartingale. From Eq. (213), we have that

$$\mathbb{E}\left(\exp\left(\tilde{\gamma}(T)\right)\right) = \mathbb{E}\left(\exp\left(\Xi_l(\tau_T) - \tau_T\right)\right), \tag{214}$$

where $\tau_t = \int_0^t \tilde{\psi}(s) ds$. Note that τ_T is a stopping time with respect to $\Xi_l(\cdot)$. Since by definition $\tilde{\psi}(t) \le \phi$ for all t, we have that $\tau_T \le \phi T$. Applying the optional sampling theorem for submartingales indexed by partially ordered sets (cf. Washburn and Willsky (1981)) to $\{\exp(\theta \tilde{\gamma}(t))\}_{t \in \mathbb{R}_+}$, we have that

$$\mathbb{E}\left(\exp\left(\theta \tilde{\gamma}(T)\right)\right) \le \mathbb{E}\left(\exp\left(\theta \Xi_l(\phi T) - \theta \phi T\right)\right) \le \exp\left((e^\theta - \theta - 1)\phi T\right), \tag{215}$$

where the last inequality follows from the fact that $\Xi_l(\phi T)$ is a Poisson random variable with mean ϕT whose moment generating function is given by $\mathbb{E}(\exp(a \Xi_l(\phi T))) = \exp(\phi T(e^a - 1))$, $a \in \mathbb{R}$. Analogously, we can show that

$$\mathbb{E}\left(\exp\left(-\theta \tilde{\gamma}(T)\right)\right) \le \exp\left(\left(e^{-\theta} + \theta - 1\right)\phi T\right) \le \exp\left(\left(e^\theta - \theta - 1\right)\phi T\right), \tag{216}$$

where the last inequality follows from the fact that $e^\theta - e^{-\theta} \ge 2\theta$ for all $\theta \ge 0$.

Note that for any $\theta > 0$, both $\{\exp(\theta \gamma(t))\}_{t \ge 0}$ and $\{\exp(-\theta \gamma(t))\}_{t \ge 0}$ are non-negative submartingales. Using Doob's inequality and Eqs. (215) and Eq. (216), we have that, for all $\epsilon, \theta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{0 \le t \le T} \tilde{\gamma}(t) \ge \epsilon \right) + \mathbb{P}\left(\sup_{0 \le t \le T} -\tilde{\gamma}(t) \ge \epsilon \right) \\
=&\mathbb{P}\left(\exp(\theta \tilde{\gamma}(T)) \ge \exp(\theta \epsilon)\right) + \mathbb{P}\left(\exp(-\theta \tilde{\gamma}(T)) \ge \exp(\theta \epsilon)\right) \\
\le&\frac{\mathbb{E}\left(\exp\left(\theta \tilde{\gamma}(T)\right)\right)}{\exp(\theta \epsilon)} + \frac{\mathbb{E}\left(\exp\left(-\theta \tilde{\gamma}(T)\right)\right)}{\exp(\theta \epsilon)} \\
\le&2 \exp\left(\phi T \left(e^\theta - \theta - 1\right) - \theta \epsilon\right).
\end{aligned} \tag{217}$$

By setting $\theta = \log(1 + \epsilon/\phi T)$ in Eq. (217), we conclude that

$$\mathbb{P}\left(\sup_{0 \le t \le T} |\tilde{\gamma}(t)| \ge \epsilon \right) \le 2 \exp(-\phi T \cdot h(\epsilon/\phi T)),$$

where $h(x) \triangleq (1+x)\log(1+x) - x$. This completes the proof. □

C.7. Proof of Lemma 9 *Proof.* We show the result by induction. For the base case, we extend the definition of $\{\xi_i^m\}_{i\in\mathbb{N}}$ by letting $\xi_0^m \triangleq 0$, and it is not difficult to see that the inequality holds when $n=0$.

Fix $i\in\{0,\dots,n-1\}$, and suppose that

$$\mathbb{E}\left(\exp\left(\theta\xi_i^m\right)\right) \le \exp\left(\frac{i\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right). \tag{218}$$

In light of the base case, it then suffices to show that the above equation implies

$$\mathbb{E}\left(\exp\left(\theta\xi_{i+1}^m\right)\right) \le \exp\left(\frac{(i+1)\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right). \tag{219}$$

Let $\{\mathcal{C}_l\}_{l\in\mathbb{Z}_+}$ be the natural filtration induced by $\{\tau_l^m, Z_k^m[l], \overline{Q}^m[l]\}_{l\in\mathbb{N}}$, with $\mathcal{C}_0 \triangleq \emptyset$. We have that

$$\begin{aligned}\mathbb{E}\left(\exp\left(\theta\xi_{i+1}^m\right)\,\middle|\,\mathcal{C}_i\right) =& \mathbb{E}\left(\exp\left(\theta\xi_i^m\right)\exp\left(\theta\tau_{i+1}^m\left(Z_k^m[i+1]-p_k(\overline{Q}^m[i+1])\right)\right)\middle|\,\mathcal{C}_i\right)\\ \overset{(a)}{=}& \exp\left(\theta\xi_i^m\right)\mathbb{E}\left(\exp\left(\theta\tau_{i+1}^m\left(Z_k^m[i+1]-p_k(\overline{Q}^m[i+1])\right)\right)\middle|\,\mathcal{C}_i\right),\end{aligned} \tag{220}$$

where step (a) follows from ξ_i^m being $\mathcal{C}_i$-measurable. We now develop an upper bound for the second term on the right-hand side of Eq. (220), as follows.

$$\begin{aligned}&\mathbb{E}\left(\exp\left(\theta\tau_{i+1}^m\left(Z_k^m[i+1]-p_k(\overline{Q}^m[i+1])\right)\right)\middle|\,\mathcal{C}_i\right)\\ =&\mathbb{E}\left(\mathbb{E}\left(\exp\left(\theta\tau_{i+1}^m\left(Z_k^m[i+1]-p_k(\overline{Q}^m[i+1])\right)\right)\middle|\,\overline{Q}^m[i+1]\right)\middle|\,\mathcal{C}_i\right)\\ \overset{(a)}{=}&\mathbb{E}\left(\mathbb{E}\left(p_k(\overline{Q}^m[i+1])\exp\left(\theta\left(1-p_k(\overline{Q}^m[i+1])\right)\tau_{i+1}^m\right)\right.\right.\\ &\left.\left.+\left(1-p_k(\overline{Q}^m[i+1])\right)\exp\left(-\theta p_k(\overline{Q}^m[i+1])\tau_{i+1}^m\right)\middle|\,\overline{Q}^m[i+1]\right)\middle|\,\mathcal{C}_i\right)\\ \overset{(b)}{=}&\mathbb{E}\left(\frac{p_k(\overline{Q}^m[i+1])m\beta_m}{m\beta_m-\theta\left(1-p_k(\overline{Q}^m[i+1])\right)}+\frac{\left(1-p_k(\overline{Q}^m[i+1])\right)m\beta_m}{m\beta_m+\theta p_k(\overline{Q}^m[i+1])}\middle|\,\mathcal{C}_i\right)\\ =&\mathbb{E}\left(\frac{m\beta_m}{m\beta_m+\theta p_k(\overline{Q}^m[i+1])}\cdot\frac{m\beta_m+2\theta p_k(\overline{Q}^m[i+1])-\theta}{m\beta_m+\theta p_k(\overline{Q}^m[i+1])-\theta}\middle|\,\mathcal{C}_i\right)\\ =&\mathbb{E}\left(1+\frac{\theta^2 p_k(\overline{Q}^m[i+1])\left(1-p_k(\overline{Q}^m[i+1])\right)}{\left(m\beta_m+\theta p_k(\overline{Q}^m[i+1])\right)\left(m\beta_m+\theta p_k(\overline{Q}^m[i+1])-\theta\right)}\middle|\,\mathcal{C}_i\right)\\ \overset{(c)}{\le}&\mathbb{E}\left(\exp\left(\frac{\theta^2 p_k(\overline{Q}^m[i+1])\left(1-p_k(\overline{Q}^m[i+1])\right)}{\left(m\beta_m+\theta p_k(\overline{Q}^m[i+1])\right)\left(m\beta_m+\theta p_k(\overline{Q}^m[i+1])-\theta\right)}\right)\middle|\,\mathcal{C}_i\right)\\ \overset{(d)}{\le}&\exp\left(\frac{\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right).\end{aligned} \tag{221}$$

Step (a) follows from the fact that, for a given value of $\overline{Q}^m[i+1]$, $Z_k^m[i+1]$ is a Bernoulli random variable with $\mathbb{P}\left(Z_k^m[i+1]=1\right)=p_k(\overline{Q}^m[i+1])$, and is independent from τ_{i+1}^m. For step (b), note that τ_{i+1} is an exponential random variable with mean $(m\beta_m)^{-1}$, independent from $\mathcal{C}_i$ and $\overline{Q}^m[i+1]$. Its moment generating function is given by $\mathbb{E}\left(e^{h\tau_{i+1}}\right)=\frac{\beta_m}{\beta_m-h}$ for all $h\le\beta_m$, where h, in our case, corresponds to $\theta(1-p_k(\overline{Q}^m[i+1]))$ and $\theta p_k(\overline{Q}^m[i+1])$, for the two terms respectively. Step (c) stems from the fact that $1+x\le e^x$ for all $x\in\mathbb{R}_+$. Finally, for step (d) we have used the fact that $p_k(\overline{Q}^m[i+1])\in[0,1]$ by definition, and that $\theta<m\beta_m$; the exponent is hence bounded from above by setting $p_k(\overline{Q}^m[i+1])$ to $1/2$ and 0 in the numerator and denominator, respectively.

Substituting Eq. (221) into Eq. (220), and invoking the induction hypothesis of Eq. (218), we have that

$$\begin{aligned}\mathbb{E}\left(\exp\left(\theta\xi_{i+1}^m\right)\right) =& \mathbb{E}\left(\mathbb{E}\left(\exp\left(\theta\xi_{i+1}^m\right)\,\middle|\,\mathcal{C}_i\right)\right)\\ =& \mathbb{E}\left(\exp\left(\theta\xi_i^m\right)\mathbb{E}\left(\exp\left(\theta\tau_{i+1}^m\left(Z_k^m[i+1]-p_k(\overline{Q}^m[i+1])\right)\right)\middle|\,\mathcal{C}_i\right)\right)\\ \le& \exp\left(\frac{i\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right)\exp\left(\frac{\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right)\\ \le& \exp\left(\frac{(i+1)\theta^2/4}{m\beta_m\left(m\beta_m-\theta\right)}\right).\end{aligned} \tag{222}$$

This proves our claim. □

C.8. Proof of Lemma 10 *Proof.* Fix $\epsilon>0$ and $m\in\mathbb{N}$. We have that

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2\ge\epsilon\right) \le& \mathbb{P}\left(\sum_{i=1}^{I_{mT}+1}(\tau_i^m)^2\ge\epsilon\,,\,I_{mT}<(1+\epsilon)m\beta_mT\right)+\mathbb{P}\left(I_{mT}\ge(1+\epsilon)m\beta_mT\right)\\ \overset{(a)}{\le}& \mathbb{P}\left(\sum_{i=1}^{(1+\epsilon)m\beta_mT}(\tau_i^m)^2\ge\epsilon\right)+\frac{1+\epsilon}{T\epsilon^2}(m\beta_m)^{-1}\\ \overset{(b)}{\le}& \frac{(1+\epsilon)m\beta_mT\cdot\mathbb{E}((\tau_1^m)^2)}{\epsilon}+\frac{1+\epsilon}{T\epsilon^2}(m\beta_m)^{-1}\\ \overset{(c)}{=}& 2(1+\epsilon)T(\epsilon m\beta_m)^{-1}+\frac{1+\epsilon}{T\epsilon^2}(m\beta_m)^{-1},\end{aligned} \tag{223}$$

where step (a) follows from Eq. (150), (b) from the Markov's inequality, and (c) from τ_i^m being an exponentially distributed random variable with mean $(m\beta_m)^{-1}$ and hence $\mathbb{E}((\tau_1^m)^2)=2(m\beta_m)^{-2}$. Because $m\beta_m\to\infty$ as $m\to\infty$, the claim follows. □

D. Comparison to Perfect Memory We discuss in this appendix what could happen if there were no memory decay in our model, i.e., if $\mu=0$, and why it is different from the limiting regime considered in this paper, with $\mu\to0$.

When $\mu=0$, all recallable rewards remain in the system indefinitely. If we view the recallable rewards sampled at the update points as a discrete-time process, and in addition set the exploration parameter α to 0, then our model becomes essentially the same as the choice process analyzed by Beggs (2005), where it is shown that the probability of choosing the best action converges to one as $t\to\infty$ under Luce's rule, as long as each action is associated with some strictly positive initial rewards. If α is a positive constant, because there is no memory decay, the effect of α disappears as soon as the rewards for all actions exceed α, and the same conclusion should hold.

Therefore, one would expect that when $\mu=0$ and $\alpha\ge0$, the choice probability under the reward-matching rule will concentrate on the best action as $t\to\infty$, regardless of the update rate, β. This is however different from the conclusion of Theorem 1, which shows the existence of two distinct limiting steady-state probabilities, one of which does not exhibit concentration on the best action. These observations thus suggest that our scaling regime do capture unique effects of imperfect memory. This is perhaps not too surprising in hindsight: if we had set μ to zero, any positive update rate β would become, by definition, significantly greater than μ, and hence the second (memory-deficient) regime in Theorems 1 and 2 could not have appeared when $\mu=0$.