Average Behaviour in Discrete-Time Imprecise Markov Chains: A Study of Weak Ergodicity

Natan T'Joens, Jasper De Bock

FLip, Ghent University, Belgium

Abstract

We study the limit behaviour of upper and lower bounds on expected time averages in imprecise Markov chains; a generalised type of Markov chain where the local dynamics, traditionally characterised by transition probabilities, are now represented by sets of 'plausible' transition probabilities. Our first main result is a necessary and sufficient condition under which these upper and lower bounds, called upper and lower expected time averages, will converge as time progresses towards infinity to limit values that do not depend on the process' initial state. Our condition is considerably weaker than that needed for ergodic behaviour; a similar notion which demands that marginal upper and lower expectations of functions at a single time instant converge to so-called limit— or steady state—upper and lower expectations. For this reason, we refer to our notion as 'weak ergodicity'. Our second main result shows that, as far as this weakly ergodic behaviour is concerned, one should not worry about which type of independence assumption to adopt—epistemic irrelevance, complete independence or repetition independence. The characterisation of weak ergodicity as well as the limit values of upper and lower expected time averages do not depend on such a choice. Notably, this type of robustness is not exhibited by the notion of ergodicity and the related inferences of limit upper and lower expectations. Finally, though limit upper and lower expectations are often used to provide approximate information about the limit behaviour of time averages, we show that such an approximation is sub-optimal and that it can be significantly improved by directly using upper and lower expected time averages.

Keywords: Imprecise Markov chain, Upper expectation, Upper transition operator, Expected time average, Weak Ergodicity, Epistemic irrelevance, Complete independence, Repetition independence

1. Introduction

Markov chains [17, 23] are probabilistic models that are used to describe the uncertain dynamics of a large variety of stochastic processes. One of the key results in the field is the point-wise ergodic theorem. It establishes a relation between the long-term time average $f_{av}(X_{1:k}) = \frac{1}{k}\sum_{i=1}^{k} f(X_i)$ of a real-valued function f and its limit expectation $E_{\infty}(f) = \lim_{k \to +\infty} E(f(X_k))$, which is guaranteed to exist if the Markov chain is ergodic.[1] For this reason, limit expectations and limit distributions have become central objects of interest. Of course, if one is interested in the long-term behaviour of time averages, one could also study the expected values $E(f_{av}(X_{1:k}))$ of these averages directly. This is not often done though, because, if the Markov chain is ergodic, the limit of these expected time averages coincides with the aforementioned limit expectations, which can straightforwardly be obtained by solving a linear eigenproblem [17]. However, for a Markov chain that is not ergodic,

---

[1]The term ergodicity has various meanings; sometimes it refers to properties of an invariant measure, sometimes it refers to properties such as irreducibility (with or without aperiodicity), regularity, ... Our usage of the term follows conventions introduced in earlier work [8, 15] on imprecise Markov chains; see Sections 2 and 4.

the limit expectation $\mathrm{E}_\infty(f)$ does not necessarily exist and can therefore not be used to provide us with information about the average behaviour of f. The expected time average $\mathrm{E}(f_{\mathrm{av}}(X_{1:k}))$—or its limit for $k \to +\infty$, if it exists—then serves as a seemingly suitable alternative; the figure on the right depicts a basic example where this is the case. So we see that even in the context of traditional "precise" Markov chains, expected time averages have the potential to be more informative about the long-term average of f compared to the limit expectation $\mathrm{E}_\infty(f)$.

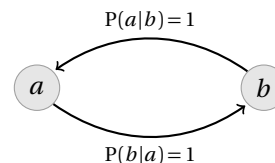In this work, we consider a generalisation of Markov chains, called imprecise Markov chains [8,

Figure 1: The graph above represents a cyclic Markov chain with two states a and b. Suppose that we want to have an idea about the average number of times that the process will be in state b. So, we are interested in the limit behaviour of $f_{\mathrm{av}}(X_{1:k})$ where $f(a) \coloneqq 0$ and $f(b) \coloneqq 1$. Though the expectation $\mathrm{E}_\infty(f)$ does not exist (if the initial distribution differs from the uniform distribution), the expectation $\mathrm{E}(f_{\mathrm{av}}(X_{1:k}))$ converges to 1/2 irrespectively of the process' initial state or distribution. This value is clearly representative for the long-term average of f.

10, 16], for which the study of long-term average behaviour becomes somewhat more complex. Imprecise Markov chains are sets of traditional "precise" probabilistic models, where the Markov property (history independence) and the time-homogeneity property apply to this set of precise models as a whole, but not necessarily to the individual models themselves. In fact, one distinguishes between three different types of imprecise Markov chains (IMC's): [2]

- *IMC under epistemic irrelevance:* the individual models do not (necessarily) satisfy the Markov property, nor the time-homogeneity property.

- *IMC under complete independence:* the individual models satisfy the Markov property, but not (necessarily) the time-homogeneity property.

- *IMC under repetition independence:* the individual models satisfy both the Markov property and the time-homogeneity property.

So an imprecise Markov chain under repetition independence only allows one to incorporate model uncertainty about the numerical values of the transition probabilities that make up a Markov chain, while an imprecise Markov chain under epistemic irrelevance also allows one to take into account uncertainty about the structural assumptions of being time-homogeneous and satisfying the Markov property. Regardless of the type of imprecise Markov chain that is used, one is typically interested in obtaining tight upper and lower bounds on inferences for the individual constituting models. The operators that represent these upper and lower bounds are respectively called upper and lower expectations and we will, for the time being, denote them by $\overline{\mathrm{E}}(\cdot)$ and $\underline{\mathrm{E}}(\cdot)$ respectively.

Just like ergodicity in traditional Markov chains, an imprecise Markov chain is said to be ergodic if the limit upper expectation $\overline{\mathrm{E}}_\infty(f) = \lim_{k \to +\infty} \overline{\mathrm{E}}(f(X_k))$ and the limit lower expectation $\underline{\mathrm{E}}_\infty(f) = \lim_{k \to +\infty} \underline{\mathrm{E}}(f(X_k))$ exist and do not depend on the process' initial state or distribution. There are necessary and sufficient conditions for when this is the case [15] as well as an imprecise variant of the point-wise ergodic theorem [8].[3] An important difference with traditional Markov chains, however, is that even if an imprecise Markov chain is ergodic and the limit upper expectation $\overline{\mathrm{E}}_\infty(f)$ and the limit lower expectation $\underline{\mathrm{E}}_\infty(f)$ exist, the upper and lower expected

---

[2]A fourth type of imprecise Markov chain that is often encountered in the literature, especially in the more general context where imprecise Markov chains are simply regarded as special credal networks, are IMC's under strong independence [1, 5, 6, 16]; convex hulls of IMC's under complete independence. However, as one of us argues in [7, Section 3] for the case of credal networks, we are of the opinion that such models lack a clear and sensible meaning. Moreover, the resulting upper and lower expectations—the inferences that we will be interested in here—are identical to those for IMC's under complete independence. We will therefore not consider them in our study.

[3]These results only hold for imprecise Markov chains under epistemic irrelevance and under complete independence.

time averages $\overline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ and $\underline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ may not converge to them—that is, to $\overline{\mathrm{E}}_\infty(f)$ and $\underline{\mathrm{E}}_\infty(f)$, respectively. Nevertheless, because they (i) give conservative bounds [21, Lemma 57], (ii) are fairly easy to compute [10] and (iii) satisfy a point-wise ergodic theorem [8], the inferences $\overline{\mathrm{E}}_\infty(f)$ and $\underline{\mathrm{E}}_\infty(f)$ are often used as descriptors of the long-term behaviour of imprecise Markov chains, even if one is actually interested in time averages. This comes at a cost though: as we will show in Section 4, both types of inferences can differ greatly, with limit upper and lower expectations sometimes providing far too conservative bounds.

Unfortunately, apart from some experiments in [21], little is known about the long-term behaviour of the upper and lower expected time averages $\overline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ and $\underline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$. The aim of this paper is to remedy this situation. Our main result is an accessibility condition that is necessary and sufficient for these upper and lower expected time averages to (each) converge to a limit value that does not depend on the process' initial state (or distribution). Remarkably, this condition is considerably weaker than the one required for ergodicity. This explains why we call this type of behaviour 'weak ergodicity' (or 'weakly ergodic behaviour'). Moreover, we also show that this notion of weak ergodicity does not depend on the adopted type of imprecise Markov chain; whether one considers an imprecise Markov chain under epistemic irrelevance, complete independence or repetition independence is not relevant for the weakly ergodic behaviour of the Markov chain. More precisely, given sufficient model parameters, both the accessibility condition that characterises weak ergodicity, as well as—if this condition is satisfied—the limit values of the inferences $\overline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ and $\underline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ are the same, no matter what kind of IMC we consider. Conventional ergodicity does not exhibit this kind of robustness; we illustrate this in Example 2. This provides yet another argument for why (limits of) upper and lower expected time averages—$\overline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$ and $\underline{\mathrm{E}}(f_{\mathrm{av}}(X_{1:k}))$—candidate as the objects of interest when looking at the long-term average behaviour of imprecise Markov chains.

The outline of the paper is as follows. We start by introducing "precise" Markov chains in Section 2 and subsequently generalise towards the case of imprecise Markov chains in Section 3. We then focus, as a first step, on average behaviour in imprecise Markov chains under epistemic irrelevance and complete independence, temporarily leaving imprecise Markov chains under repetition independence out of the picture. As mentioned before, we will study two types of inferences: (limits of) upper and lower expectations of a function evaluated at a single time instant, and (limits of) upper and lower expected time averages of a function. In Section 4, we give recursive expressions for how these inferences evolve through time, introduce the notions of ergodicity and weak ergodicity, and moreover illustrate, using two basic examples, that weak ergodicity has some considerable advantages over "conventional" ergodicity when it comes to characterising average behaviour. Section 5 introduces essential mathematical machinery needed in order to arrive to our results in Sections 6–8; we explain what it means for a map to be topical and introduce some graph-theoretic notions. In the subsequent section, Section 6, we derive a sufficient condition for weak ergodicity by borrowing an eigenproblem result from the theory of topical maps. Section 7 then shows that this condition can be replaced by a weaker one that is not only sufficient, but also necessary. Finally, we consider the case of imprecise Markov chains under repetition independence and relate their weak ergodicity to that of imprecise Markov chains under epistemic irrelevance or complete independence. This will be the subject of Section 8. As mentioned before, it will turn out that for all three types of imprecise Markov chains, weak ergodicity is characterised by the same condition and the limit upper (lower) expected time averages are all equal.

This paper extends upon an earlier conference paper [25]; we provide proofs for the results in [25], and extend our study of weak ergodicity to also include imprecise Markov chains under repetition independence. In order not to lose the reader's focus, we have chosen to relegate some of the more technical proofs to an appendix at the end of the paper. This is particularly true for the results in Sections 7 and 8, where the main text provides, for the most part, an informal argument that aims to provide intuition.

2. Markov Chains

We consider an infinite sequence $X_1 X_2 X_3 \cdots$ of uncertain states, where each state X_k at time $k \in \mathbb{N}$ takes values in some finite set $\mathscr{X}$, called the *state space*. Such a sequence $X_1 X_2 X_3 \cdots$ will be called a *(discrete-time) stochastic process*. For any $k, \ell \in \mathbb{N}$ such that $k \leq \ell$, we use $X_{k:\ell}$ to denote the finite subsequence $X_k \cdots X_\ell$ of states that takes values in $\mathscr{X}^{\ell-k+1}$. Moreover, for any $k, \ell \in \mathbb{N}$ such that $k \leq \ell$ and any $x_{k:\ell} \in \mathscr{X}^{\ell-k+1}$, we use $X_{k:\ell} = x_{k:\ell}$ to denote the event that $X_k = x_k \cdots X_\ell = x_\ell$. The uncertain dynamics of a stochastic process are then typically described by probabilities of the form $P(X_{k+1} = x_{k+1} | X_{1:k} = x_{1:k})$, for any $k \in \mathbb{N}$ and any $x_{1:k+1} \in \mathscr{X}^{k+1}$. They represent beliefs about which state the process will be in at time $k+1$ given that we know that it was in the states $x_1 \cdots x_k$ at time instances 1 through k. Additionally, our beliefs about the value of the initial state X_1 can be represented by probabilities $P(X_1 = x_1)$ for all $x_1 \in \mathscr{X}$. The local probability assessments $P(X_{k+1} = x_{k+1} | X_{1:k} = x_{1:k})$ and $P(X_1 = x_1)$ can now be combined to construct a global probability model P that describes the dynamics of the process on a more general level. This can be done in various ways; one of the most common ones being a measure-theoretic approach where countable additivity plays a central role. For our purposes however, we will only require finite additivity. Regardless, once you have such a global probability model P, it can then be used to define expectations and make inferences about the uncertain behaviour of the process.

For any set A, let us write $\mathscr{L}(A)$ to denote the set of all real-valued functions on A. Throughout, for any $B \subseteq A$, we use $\mathbb{I}_B$ to denote the *indicator* of B: the function in $\mathscr{L}(A)$ that takes the value 1 in B and 0 otherwise. We will only be concerned with (upper and lower) expectations of *finitary functions*: functions that depend on the state of the process at a finite number of time instances. So if f is finitary, we can write $f = g(X_{1:k})$ for some $k \in \mathbb{N}$ and some $g \in \mathscr{L}(\mathscr{X}^k)$. Note that finitary functions are bounded; this follows from their real-valuedness and the fact that $\mathscr{X}$ is finite. The expectation of a finitary function $f(X_{1:k})$ conditional on some event $X_{1:\ell} = x_{1:\ell}$, with $\ell < k$, simply reduces to a finite weighted sum:

$$E_P(f(X_{1:k}) | X_{1:\ell} = x_{1:\ell}) = \sum_{x_{\ell+1:k} \in \mathscr{X}^{k-\ell}} f(x_{1:k}) \prod_{i=\ell}^{k-1} P(X_{i+1} = x_{i+1} | X_{1:i} = x_{1:i}).$$

A particularly interesting case arises when studying stochastic processes that are described by a probability model P that satisfies

$$P(X_{k+1} = y | X_{1:k} = x_{1:k}) = P(X_{k+1} = y | X_k = x_k),$$

for all $k \in \mathbb{N}$, all $y \in \mathscr{X}$ and all $x_{1:k} \in \mathscr{X}^k$. This property, known as the *Markov property*, states that given the present state of the process the future behaviour of the process does not depend on its history. A process of this type is called a *Markov chain*. We moreover call it *(time) homogeneous* if additionally $P(X_{k+1} = y | X_k = x) = P(X_2 = y | X_1 = x)$, for all $k \in \mathbb{N}$ and all $x, y \in \mathscr{X}$. Hence, together with the assessments $P(X_1 = x_1)$, the dynamics of a homogeneous Markov chain are fully characterised by the probabilities $P(X_2 = y | X_1 = x)$. These probabilities are typically gathered in a *transition matrix* T; a row-stochastic $|\mathscr{X}| \times |\mathscr{X}|$ matrix T that is defined by $T(x, y) := P(X_2 = y | X_1 = x)$ for all $x, y \in \mathscr{X}$. This matrix representation T can be regarded as a linear operator from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$, defined for any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{X}$ by

$$Tf(x) := \sum_{y \in \mathscr{X}} f(y) P(X_2 = y | X_1 = x) = E_P(f(X_2) | X_1 = x).$$

Conveniently, for any $k \in \mathbb{N}$, we also have that

$$E_P(f(X_{k+1}) | X_k = x) = \sum_{y \in \mathscr{X}} f(y) P(X_{k+1} = y | X_k = x) = \sum_{y \in \mathscr{X}} f(y) P(X_2 = y | X_1 = x) = Tf(x).$$

4

More generally, it holds that $E_P(f(X_{k+\ell})|X_k = x) = T^\ell f(x)$ for all $k \in \mathbb{N}$, all $\ell \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and all $x \in \mathcal{X}$. Then, under some well-known accessibility conditions [15, Proposition 3], the expectation $T^\ell f(x)$ converges for increasing ℓ towards a constant $E_\infty(f)$ independently of the initial state x. If this is the case for all $f \in \mathcal{L}(\mathcal{X})$, the homogeneous Markov chain will have a steady-state distribution, represented by the limit expectation E_∞, and we call the Markov chain *ergodic*. The expectation E_∞ is in particular also useful if we are interested in the limit behaviour of expected time averages. Indeed, let $f_{av}(X_{1:k}) := \frac{1}{k} \sum_{i=1}^{k} f(X_i)$ be the time average of some function $f \in \mathcal{L}(\mathcal{X})$ evaluated at the time instances 1 through k. Then, according to [21, Theorem 38], the limit of the expected average $\lim_{k \to +\infty} E_P(f_{av}(X_{1:k}))$ coincides with the limit expectation $E_\infty(f)$. One of the aims of this paper is to explore to which extent this remains true for imprecise Markov chains.

3. Imprecise Markov Chains

If the basic probabilities $P(X_{k+1}|X_{1:k} = x_{1:k})$ that describe a stochastic process are imprecise, in the sense that we only have partial information about them, then we can still model the process' dynamics by considering a set $\mathcal{T}_{x_{1:k}}$ of such probabilities, for all $k \in \mathbb{N}$ and all $x_{1:k} \in \mathcal{X}^k$. This set $\mathcal{T}_{x_{1:k}}$ is then interpreted as the set of all probability mass functions $P(X_{k+1}|X_{1:k} = x_{1:k})$ that we deem "plausible". We here consider the special case where the sets $\mathcal{T}_{x_{1:k}}$ satisfy a Markov property, meaning that $\mathcal{T}_{x_{1:k}} = \mathcal{T}_{x_k}$ for all $k \in \mathbb{N}$ and all $x_{1:k} \in \mathcal{X}^k$. Similarly to the precise case, the sets $\mathcal{T}_x$, for all $x \in \mathcal{X}$, can be gathered into a single object: the set $\mathcal{T}$ of all row stochastic $|\mathcal{X}| \times |\mathcal{X}|$ matrices T such that, for all $x \in \mathcal{X}$, the probability mass function $T(x, \cdot)$ is an element of $\mathcal{T}_x$. A set $\mathcal{T}$ of transition matrices defined in this way is called *separately specified* [16, Definition 11.6]; this property asserts that, for any two transition matrices $T_1, T_2 \in \mathcal{T}$ and any subset $A \subseteq \mathcal{X}$, there is a third transition matrix $T_3 \in \mathcal{T}$ such that $T_3(x, \cdot) = T_1(x, \cdot)$ for all $x \in A$ and $T_3(y, \cdot) = T_2(y, \cdot)$ for all $y \in \mathcal{X} \setminus A$. For any such set $\mathcal{T}$, the corresponding *imprecise Markov chain under epistemic irrelevance* $\mathscr{P}_{\mathcal{T}}^{ei}$ [9, 10] is the set of all (precise) probability models P such that $P(X_{k+1}|X_{1:k} = x_{1:k}) \in \mathcal{T}_{x_k}$ for all $k \in \mathbb{N}$ and all $x_{1:k} \in \mathcal{X}^k$. The values of the probabilities $P(X_1 = x_1)$ will be of no importance to us, because we will focus solely on (upper and lower) expectations conditional on the value of the initial state X_1.

Clearly, an imprecise Markov chain $\mathscr{P}_{\mathcal{T}}^{ei}$ also contains non-homogeneous, and even non-Markovian processes. So the Markov property does in this case not apply to the individual probability assessments, but rather to the sets $\mathcal{T}_{x_{1:k}}$. The model $\mathscr{P}_{\mathcal{T}}^{ei}$ is therefore a generalisation of a traditional Markov chain where we allow for model uncertainty about, on the one hand, the mass functions $P(X_{k+1}|X_{1:k} = x_{1:k})$ and, on the other hand, about structural assumptions such as the Markov and time-homogeneity property. In order to make inferences that are robust with respect to this model uncertainty, we will use *upper* and *lower expectations* [2, 11, 27]. These operators are respectively defined as the tightest upper and lower bound on the expectation E_P associated with any probability model P in $\mathscr{P}_{\mathcal{T}}^{ei}$:

$$\overline{E}_{\mathcal{T}}^{ei}(f|A) := \sup_{P \in \mathscr{P}_{\mathcal{T}}^{ei}} E_P(f|A) \quad \text{and} \quad \underline{E}_{\mathcal{T}}^{ei}(f|A) := \inf_{P \in \mathscr{P}_{\mathcal{T}}^{ei}} E_P(f|A),$$

for any finitary function f and any event A of the form $X_{1:k} = x_{1:k}$. The operators $\overline{E}_{\mathcal{T}}^{ei}$ and $\underline{E}_{\mathcal{T}}^{ei}$ are related by conjugacy, meaning that $\underline{E}_{\mathcal{T}}^{ei}(\cdot|\cdot) = -\overline{E}_{\mathcal{T}}^{ei}(-\cdot|\cdot)$, which allows us to focus on only one of them; upper expectations in our case. The lower expectation $\underline{E}_{\mathcal{T}}^{ei}(f|A)$ of a finitary function f can then simply be obtained by considering the upper expectation $-\overline{E}_{\mathcal{T}}^{ei}(-f|A)$. Moreover, note that *upper* and *lower probabilities* can simply be regarded as special cases of upper and lower expectations: for any event B such that $\mathbb{I}_B$ is finitary and any event A of the form $X_{1:k} = x_{1:k}$, we can define them as respectively $\overline{P}_{\mathcal{T}}^{ei}(B|A) := \overline{E}_{\mathcal{T}}^{ei}(\mathbb{I}_B|A)$ and $\underline{P}_{\mathcal{T}}^{ei}(B|A) := \underline{E}_{\mathcal{T}}^{ei}(\mathbb{I}_B|A)$.

Apart from epistemic irrelevance, there are also other types of independence assumptions for imprecise Markov chains that impose more stringent conditions on the individual composing probability models. For a given set $\mathcal{T}$, the imprecise Markov chain under *complete independence* $\mathscr{P}_{\mathcal{T}}^{ci}$ is the subset of $\mathscr{P}_{\mathcal{T}}^{ei}$ that contains

all—possibly non-homogeneous—Markov chains in $\mathscr{P}_{\mathscr{T}}^{\text{ei}}$ [6, 21, 22]. The models $\mathscr{P}_{\mathscr{T}}^{\text{ci}}$, also known as 'Markov set-chains' [14], were the first types of imprecise Markov chains to be thoroughly studied. They can be motivated in a rather straightforward way, using a 'sensitivity analysis interpretation'; the set $\mathscr{T}$ is then regarded as a result of our ignorance about some "true" transition matrix T_k that may depend on the time k. A third type of imprecise Markov chain that we will associate with $\mathscr{T}$ is the corresponding imprecise Markov chain under *repetition independence* $\mathscr{P}_{\mathscr{T}}^{\text{ri}}$, which is the subset of $\mathscr{P}_{\mathscr{T}}^{\text{ei}}$ containing all homogeneous Markov chains [4, 21]. Similarly as for $\mathscr{P}_{\mathscr{T}}^{\text{ci}}$, the model $\mathscr{P}_{\mathscr{T}}^{\text{ri}}$ can be motivated using a sensitivity analysis interpretation, where the unknown matrix T_k is now assumed to be fixed in time. Observe that the models $\mathscr{P}_{\mathscr{T}}^{\text{ci}}$ and $\mathscr{P}_{\mathscr{T}}^{\text{ri}}$ do not allow us to incorporate uncertainty about the Markov assumption, a feature that only imprecise Markov chains under epistemic irrelevance have. Moreover, though imprecise Markov chains under epistemic irrelevance can also be justified starting from a sensitivity analysis interpretation—the underlying 'true' probability model is in that case not assumed to be Markov—they are especially suitable when we regard the sets $\mathscr{T}_x$ as arising from the—subjective—beliefs of a subject that is uncertain about the process' next state value, like Walley does [27]. The fact that these sets $\mathscr{T}_x$ satisfy a Markov property, then simply means that our subject's beliefs are solely based on the current state x of the process; we refer to [7, 9] for further details.

Similarly to how we defined upper and lower expectations for imprecise Markov chains under epistemic irrelevance, we can define the upper expectations $\overline{\text{E}}_{\mathscr{T}}^{\text{ci}}$ and $\overline{\text{E}}_{\mathscr{T}}^{\text{ri}}$ (and the lower expectations $\underline{\text{E}}_{\mathscr{T}}^{\text{ci}}$ and $\underline{\text{E}}_{\mathscr{T}}^{\text{ri}}$) as the tightest upper (and lower) bounds on the expectations corresponding to the models in $\mathscr{P}_{\mathscr{T}}^{\text{ci}}$ and $\mathscr{P}_{\mathscr{T}}^{\text{ri}}$, respectively. Upper and lower probabilities can also be defined in the same way as before; as upper and lower expectations of indicators. Furthermore, note that, since $\mathscr{P}_{\mathscr{T}}^{\text{ri}} \subseteq \mathscr{P}_{\mathscr{T}}^{\text{ci}} \subseteq \mathscr{P}_{\mathscr{T}}^{\text{ei}}$, we have that

$$\overline{\text{E}}_{\mathscr{T}}^{\text{ri}}(f|A) \leq \overline{\text{E}}_{\mathscr{T}}^{\text{ci}}(f|A) \leq \overline{\text{E}}_{\mathscr{T}}^{\text{ei}}(f|A),$$

for any finitary function f and any event A of the form $X_{1:k} = x_{1:k}$. Henceforth, we let $\mathscr{T}$ be some generic set of transition matrices that is separately specified.

In this paper, we will be specifically concerned with two types of inferences: the conditional upper (and lower) expectation of a function $f \in \mathscr{L}(\mathscr{X})$ evaluated at a single time instant k, and the conditional upper (and lower) expectation of the time average $f_{\text{av}}(X_{1:k})$ of a function $f \in \mathscr{L}(\mathscr{X})$, given that we start in some $x \in \mathscr{X}$. For imprecise Markov chains under epistemic irrelevance and under complete independence, both of these inferences coincide [21, Theorem 51 & Theorem 52]. For any $f \in \mathscr{L}(\mathscr{X})$, any $x \in \mathscr{X}$ and any $k \in \mathbb{N}$, we will denote them by

$$\overline{\text{E}}_k(f|x) = \overline{\text{E}}_{\mathscr{T}}^{\text{ei}}(f(X_k)|X_1 = x) = \overline{\text{E}}_{\mathscr{T}}^{\text{ci}}(f(X_k)|X_1 = x)$$
$$\text{and } \overline{\text{E}}_{\text{av},k}(f|x) = \overline{\text{E}}_{\mathscr{T}}^{\text{ei}}(f_{\text{av}}(X_{1:k})|X_1 = x) = \overline{\text{E}}_{\mathscr{T}}^{\text{ci}}(f_{\text{av}}(X_{1:k})|X_1 = x),$$

where the dependency on $\mathscr{T}$ is implicit. The corresponding lower expectations can be obtained through conjugacy: $\underline{\text{E}}_k(f|x) = -\overline{\text{E}}_k(-f|x)$ and $\underline{\text{E}}_{\text{av},k}(f|x) = -\overline{\text{E}}_{\text{av},k}(-f|x)$ for all $f \in \mathscr{L}(\mathscr{X})$, all $x \in \mathscr{X}$ and all $k \in \mathbb{N}$. As we will discuss shortly, the behaviour (or evolution) of both of these inferences can be recursively expressed in terms of a single so-called upper transition operator $\overline{T}$. These relations will form the starting point for our further study of the limit behaviour of these inferences. However, similar expressions seem not to exist for the upper expectations $\overline{\text{E}}_{\mathscr{T}}^{\text{ri}}(f(X_k)|X_1 = x)$ and $\overline{\text{E}}_{\mathscr{T}}^{\text{ri}}(f_{\text{av}}(X_{1:k})|X_1 = x)$ corresponding to an imprecise Markov chain $\mathscr{P}_{\mathscr{T}}^{\text{ri}}$ under repetition independence. As a consequence, such inferences demand a somewhat different approach. For the moment, we therefore omit them from our discussion. We will come back to them in Section 8.

4. Transition Operators, Ergodicity and Weak Ergodicity

Inferences of the form $\overline{\text{E}}_k(f|x)$—and, more specifically, upper (and lower) probabilities of events of the form $X_k \in A$, with $A \subseteq \mathscr{X}$—were among the first ones to be thoroughly studied in imprecise Markov chains [10, 14, 24].

Recent work on the topic [8, 10, 15] is crucially based on the observation that $\overline{\mathrm{E}}_{k+1}(f|x)$ can be elegantly rewritten as the k-th iteration of the map $\overline{T}\colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ defined by

$$\overline{T}h(x) := \sup_{T \in \mathcal{T}} Th(x) = \sup_{T(x,\cdot) \in \mathcal{T}_x} \sum_{y \in \mathcal{X}} T(x,y)h(y),$$

for all $x \in \mathcal{X}$ and all $h \in \mathcal{L}(\mathcal{X})$. Concretely, $\overline{\mathrm{E}}_k(f|x) = [\overline{T}^{k-1}f](x)$ for all $x \in \mathcal{X}$ and all $k \in \mathbb{N}$ [10, Theorem 3.1]. The map $\overline{T}$ therefore plays a similar role as the transition matrix T in traditional Markov chains, which is why it is called the *upper transition operator* corresponding to the set $\mathcal{T}$. Moreover, observe that (the value of) the upper transition operator $\overline{T}$, for any $h \in \mathcal{L}(\mathcal{X})$, can always be approximated arbitrarily closely by an element of $\mathcal{T}$:

S1. $(\forall \epsilon > 0)(\forall h \in \mathcal{L}(\mathcal{X}))(\exists T \in \mathcal{T})\, \overline{T}h - \epsilon \le Th \le \overline{T}h.$

This property is a direct consequence of the fact that $\mathcal{T}$ is separately specified—see for example [19, Lemma 1] where we assume the set $\mathcal{T}$ to be closed—and will be used later on in Section 8.

In an analogous way, inferences of the form $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ can be obtained as the k-th iteration of the map $\overline{T}_f\colon \mathcal{L}(\mathcal{X}) \to \mathcal{L}(\mathcal{X})$ defined by $\overline{T}_f h := f + \overline{T}h$ for all $h \in \mathcal{L}(\mathcal{X})$. In particular, if we let $\tilde{m}_{f,1} := f = \overline{T}_f(0)$ and

$$\tilde{m}_{f,k} := f + \overline{T}\tilde{m}_{f,k-1} = \overline{T}_f \tilde{m}_{f,k-1} \text{ for all } k > 1, \tag{1}$$

then it follows from [21, Lemma 41] that $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \frac{1}{k}\tilde{m}_{f,k}(x)$ for all $x \in \mathcal{X}$ and all $k \in \mathbb{N}$. Applying Equation (1) repeatedly, we find that for all $x \in \mathcal{X}$ and all $k \in \mathbb{N}$:

$$\overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \tfrac{1}{k}\tilde{m}_{f,k}(x) = \tfrac{1}{k}[\overline{T}_f^{k-1}\tilde{m}_{f,1}](x) = \tfrac{1}{k}[\overline{T}_f^{k}(0)](x). \tag{2}$$

The same formula can also be obtained as a special case of the results in [26].

These expressions for $\overline{\mathrm{E}}_k(f|x)$ and $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ in terms of the respective operators $\overline{T}$ and $\overline{T}_f$ are particularly useful when we aim to characterise the limit behaviour of these inferences. As will be elaborated on in the next section, there are conditions on $\overline{T}$ that are necessary and sufficient for $\overline{\mathrm{E}}_k(f|x)$ to converge to a limit value that does not depend on the process' initial state $x \in \mathcal{X}$. If this is the case for all $f \in \mathcal{L}(\mathcal{X})$, the imprecise Markov chain (under epistemic irrelevance or complete independence) is called *ergodic* [8, 15] and we then denote the constant limit value by $\overline{\mathrm{E}}_\infty(f) := \lim_{k \to +\infty} \overline{\mathrm{E}}_k(f|x)$. By analogy, we call an imprecise Markov chain (under epistemic irrelevance or complete independence) *weakly ergodic* if, for all $f \in \mathcal{L}(\mathcal{X})$, $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ exists and does not depend on the initial state x. For a weakly ergodic imprecise Markov chain, we denote the common limit value by $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) := \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$. In contrast with conventional ergodicity, weak ergodicity and other properties of the long-term behaviour of $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ are almost entirely unexplored. The main result of the first part of this paper, which focuses on imprecise Markov chains under epistemic irrelevance and complete independence, is a necessary and sufficient condition for weak ergodicity. As we will see, this condition is weaker than that needed for conventional ergodicity, hence our choice of terminology. The following example shows that this difference already becomes apparent in the precise case.

Example 1. *Recall the situation sketched in Figure 1, where $\mathcal{X} = \{a,b\}$ and where $\mathcal{T}$ consists of a single matrix $T = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Fix any function $f = \left[\begin{smallmatrix} f_a \\ f_b \end{smallmatrix}\right] \in \mathcal{L}(\mathcal{X})$. Clearly, $\overline{T}$ is not ergodic because $\overline{T}^{(2\ell+1)}f = T^{(2\ell+1)}f = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]f = \left[\begin{smallmatrix} f_b \\ f_a \end{smallmatrix}\right]$ and $\overline{T}^{(2\ell)}f = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]f = \left[\begin{smallmatrix} f_a \\ f_b \end{smallmatrix}\right]$ for all $\ell \in \mathbb{N}_0$. $\overline{T}$ is weakly ergodic though, because*

$$\overline{T}_f^{(2\ell)}(0) = \ell\left[\begin{smallmatrix} f_a+f_b \\ f_a+f_b \end{smallmatrix}\right] \text{ and } \overline{T}_f^{(2\ell+1)}(0) = f + \overline{T}\,\overline{T}_f^{(2\ell)}(0) = f + \ell\left[\begin{smallmatrix} f_a+f_b \\ f_a+f_b \end{smallmatrix}\right],$$

for all $\ell \in \mathbb{N}_0$, which implies that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) := \lim_{k \to +\infty} \overline{T}_f^{k}(0)/k = (f_a + f_b)/2$ exists. $\Diamond$

Notably, even if an imprecise Markov chain is ergodic (and hence also weakly ergodic) and therefore both $\overline{\mathrm{E}}_\infty(f)$ and $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ exist, these inferences will not necessarily coincide. This was first observed in an experimental setting [21, Section 7.6], but the differences that were observed there were marginal. The following example shows that these differences can in fact be very substantial.

Example 2. *Let $\mathscr{X} = \{a, b\}$, let $\mathscr{T}_a$ be the set of all probability mass functions on $\mathscr{X}$ and let $\mathscr{T}_b$ be the set that consists of the single probability mass function that puts all mass in a; see Figure 2. Then, for any $f = \begin{bmatrix} f_a \\ f_b \end{bmatrix} \in \mathscr{L}(\mathscr{X})$, we have that*

$$\overline{T}f(x) = \begin{cases} \max f & \text{if } x = a; \\ f_a & \text{if } x = b, \end{cases} \quad \text{and} \quad \overline{T}^2 f(x) = \begin{cases} \max \overline{T} f = \max f & \text{if } x = a; \\ \overline{T}f(a) = \max f & \text{if } x = b. \end{cases}$$

It follows that $\overline{T}^k f = \max f$ for all $k \geq 2$, so the limit upper expectation $\overline{\mathrm{E}}_\infty(f)$ exists and is equal to $\max f$ for all $f \in \mathscr{L}(\mathscr{X})$. In particular, we have that $\overline{\mathrm{E}}_\infty(\mathbb{I}_b) = 1$. On the other hand, we find that $\overline{T}_{\mathbb{I}_b}^{(2\ell)}(0) = \ell$ and $\overline{T}_{\mathbb{I}_b}^{(2\ell+1)}(0) = \mathbb{I}_b + \overline{T}\,\overline{T}_{\mathbb{I}_b}^{(2\ell)}(0) = \begin{bmatrix} \ell \\ \ell+1 \end{bmatrix}$ for all $\ell \in \mathbb{N}_0$. This implies that the upper expectation $\overline{\mathrm{E}}_{\mathrm{av},\infty}(\mathbb{I}_b) := \lim_{k\to+\infty} \overline{T}_{\mathbb{I}_b}^k(0)/k$ exists and is equal to $1/2$. This value differs significantly from the limit upper expectation $\overline{\mathrm{E}}_\infty(\mathbb{I}_b) = 1$.

In fact, this result could have been expected simply by taking a closer look at the dynamics that correspond to $\mathscr{T}$. Indeed, it follows directly from $\mathscr{T}$ that, if the system is in state b at some instant, then it will surely be in a at the next time instant. Hence, the system can only reside in state b for maximally half of the time, resulting in an upper expected average that converges to $1/2$. These underlying dynamics have little effect on the limit upper expectation $\overline{\mathrm{E}}_\infty(\mathbb{I}_b)$ though, because it is only concerned with the upper expectation of $\mathbb{I}_b$ evaluated at a single time instant.

Finally, we want to draw attention to the fact that the upper expectation $\overline{\mathrm{E}}_{\mathrm{av},\infty}(\mathbb{I}_b) = 1/2$ is actually reached by a compatible homogeneous Markov chain. Specifically, it is reached by the Markov chain from Example 1, which is indeed compatible because its transition matrix $T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is in $\mathscr{T}$. This already illustrates what will be established later on in Section 8: when we are interested in the limit behaviour of the inferences $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$, we can simply treat them as upper envelopes of the expectations $\mathrm{E}_{\mathrm{P}}(f_{\mathrm{av}}(X_{1:k})|X_1 = x)$ that correspond to the compatible homogeneous Markov chains P. This is not the case for the limit behaviour of the inferences $\overline{\mathrm{E}}_k(f|x)$ though; for instance, in the current example, where $\overline{\mathrm{E}}_\infty(\mathbb{I}_b) = 1$, the expectation $\mathrm{E}_{\mathrm{P}}(\mathbb{I}_b(X_k)|X_1 = b)$, for any $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$, is lower or equal than $1/2$ for k even. This is left as an exercise for the reader—Hint: for any $T = \begin{bmatrix} p & 1-p \\ 1 & 0 \end{bmatrix} \in \mathscr{T}$, find $c_1, c_2 \in \mathbb{R}$ such that $T^2 \mathbb{I}_a = c_1 + c_2 \mathbb{I}_a$, and then use this observation to find an expression for $T^{2\ell} \mathbb{I}_a$ and $T^{2\ell+1} \mathbb{I}_b = (1-p)T^{2\ell}\mathbb{I}_a$ for all $\ell \in \mathbb{N}$. ◊
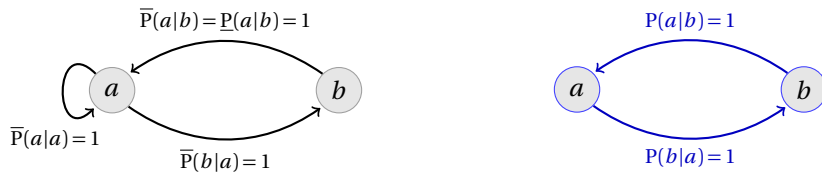

Figure 2: The graph on the left hand side illustrates, in terms of upper and lower transition probabilities, how the state of the process can change from one time instant to the next. The blue graph on the right hand side depicts the compatible precise Markov chain P for which the upper bound $\overline{\mathrm{E}}_{\mathrm{av},\infty}(\mathbb{I}_b) = 1/2$ is reached.

Although we have used sets $\mathscr{T}$ of transition matrices to define imprecise Markov chains, it should at this point be clear that, if we are interested in the inferences $\overline{\mathrm{E}}_k(f|x)$ and $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ and their limit values, then it suffices to specify $\overline{T}$. In fact, we will temporarily forget about $\mathscr{T}$ and simply assume that $\overline{T}$ is a general upper transition operator on $\mathscr{L}(\mathscr{X})$. That is, we assume $\overline{T}$ to be any operator from $\mathscr{L}(\mathscr{X})$ to $\mathscr{L}(\mathscr{X})$ that satisfies

U1. $\overline{T}h \leq \max h$ [upper bounds];

U2. $\overline{T}(h+g) \leq \overline{T}h + \overline{T}g$ [sub-additivity];

U3. $\overline{T}(\lambda h) = \lambda \overline{T}h$ [non-negative homogeneity],

for all $h, g \in \mathscr{L}(\mathscr{X})$ and all real $\lambda \geq 0$ [8, 15, 18]. This can be done without loss of generality because it is well-established that any operator $\overline{T}$ that is defined as an upper envelope of a set $\mathscr{T}$ of transition matrices—as we did in Section 4—always satisfies U1–U3 [27, Theorem 2.6.3]. Then again, any upper transition operator can uniquely be represented by a closed, convex set of transition matrices that is seperately specified [27, Theorem 3.3.3], so there is no gain in generality either. Apart from the axioms above, our results and proofs will also rely on the following three properties that are implied by U1–U3 [27, Section 2.6.1]:

U4. $\min h \leq \overline{T}h \leq \max h$ [boundedness];

U5. $\overline{T}(\mu + h) = \mu + \overline{T}h$ [constant additivity];

U6. if $h \leq g$ then $\overline{T}h \leq \overline{T}g$ [monotonicity];

U7. $\overline{T}h - \overline{T}g \leq \overline{T}(h-g)$ [mixed sub-additivity],

for all $h, g \in \mathscr{L}(\mathscr{X})$ and all real μ. Henceforth, we will simply say that the upper transition operator $\overline{T}$ is ergodic if $[\overline{T}^k f]$ converges to a constant for all $f \in \mathscr{L}(\mathscr{X})$ and, analogously, we will say that it is weakly ergodic if $\frac{1}{k}[\overline{T}_f^k(0)]$ converges to a constant for all $f \in \mathscr{L}(\mathscr{X})$. So ergodicity and weak ergodicity of $\overline{T}$ is equivalent to the respective notions for an imprecise Markov chain under epistemic irrelevance or complete independence with upper transition operator $\overline{T}$.

5. Accessibility Relations and Topical Maps

To characterise ergodicity and weak ergodicity, we will make use of some well-known graph-theoretic concepts, suitably adapted to the imprecise Markov chain setting; we recall the following from [10] and [15]. The *upper accessibility graph* $\mathscr{G}(\overline{T})$ corresponding to $\overline{T}$ is defined as the directed graph with vertices $x_1 \cdots x_n \in \mathscr{X}$, where $n := |\mathscr{X}|$, with an edge from x_i to x_j if $\overline{T}\mathbb{I}_{x_j}(x_i) > 0$. For any two vertices x_i and x_j, we say that x_j is *accessible* from x_i, denoted by $x_i \to x_j$, if $x_i = x_j$ or if there is a directed path from x_i to x_j, which means that there is a sequence $x_i = x_0', x_1', \cdots, x_k' = x_j$ of vertices, with $k \in \mathbb{N}$, such that there is an edge from $x_{\ell-1}'$ to x_ℓ' for all $\ell \in \{1, \cdots, k\}$. The following result shows that such a directed path exists if and only if there is some $k \in \mathbb{N}$ such that the k-step upper probability $\overline{T}^k \mathbb{I}_{x_j}(x_i)$ to transition from x_i to x_j is positive.

Lemma 1. [15, Proposition 4] *For any two vertices x and y, there is a directed path of length $k \in \mathbb{N}$ from x to y if and only if $\overline{T}^k \mathbb{I}_y(x) > 0$.*

We say that two vertices x_i and x_j *communicate* and write $x_i \longleftrightarrow x_j$ if both $x_i \to x_j$ and $x_j \to x_i$. The relation $\to$ is a preorder (reflexive and transitive), and therefore, $\longleftrightarrow$ is an equivalence relation (reflexive, symmetric and transitive) for which the equivalence classes are called *communication classes*. Moreover, it is well-known that these communication classes then form a partition $\mathscr{C}$ of $\mathscr{X}$. Sometimes, we will allow ourselves a slight abuse of terminology, and call any non-empty set $\mathscr{S} \subseteq \mathscr{X}$ a *class* in $\mathscr{G}(\overline{T})$. We call the graph $\mathscr{G}(\overline{T})$ *strongly connected* if any two vertices x_i and x_j in $\mathscr{G}(\overline{T})$ communicate, or equivalently, if $\mathscr{X}$ itself is a communication class.

We also extend the domain of the relation $\to$ to include all communication classes by saying that $A \to B$, for any two $A, B \in \mathscr{C}$, if $x \to y$ for at least one (and hence—since A and B are communication classes—all) $x \in A$ and $y \in B$. Then it can easily be seen that $\to$ induces a partial order (reflexive, antisymmetric and transitive) on the set $\mathscr{C}$ [Lemma 23, Appendix A]. $\overline{T}$ (or $\mathscr{G}(\overline{T})$) is then said to have a *top class* $\mathscr{R}$ if there is a communication class $\mathscr{R}$ that dominates every other communication class in this partial order, or equivalently [Lemma 25, Appendix A], since $\mathscr{X}$—and hence also $\mathscr{C}$—is finite, if $\mathscr{R}$ is the only maximal (undominated) communication class of this

ordering. One can also show [Lemma 27, Appendix A] that the top class $\mathscr{R}$ exists if and only if the set of all vertices that can be reached from anywhere in $\mathscr{G}(\overline{T})$ is non-empty, in which case the top class is equal to this set:

$$\mathscr{G}(\overline{T}) \text{ has a top class } \mathscr{R} \Longleftrightarrow \mathscr{R} = \{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\} \neq \emptyset. \tag{3}$$

Finally, we also say that a class $\mathscr{S}$ is *closed* if $x \nrightarrow y$ for all $x \in \mathscr{S}$ and all $y \in \mathscr{S}^c$. Since the notions of closedness and maximality coincide for communication classes [Lemma 24, Appendix A], it follows that the top class $\mathscr{R}$, if it exists, is the only closed communication class in $\mathscr{G}(\overline{T})$.

Having a top class is necessary for $\overline{T}$ to be ergodic, but it is not sufficient. Sufficiency additionally requires that the top class $\mathscr{R}$ is regular and absorbing [15, Proposition 3]. These properties are defined, for any closed class $\mathscr{S}$, as

A1. $(\forall x \in \mathscr{S})(\exists k^* \in \mathbb{N})(\forall k \geq k^*) \min \overline{T}^k \mathbb{I}_x > 0$ [Regularity];

A2. $(\forall x \in \mathscr{S}^c)(\exists k \in \mathbb{N}) \overline{T}^k \mathbb{I}_{\mathscr{S}^c}(x) < 1$ [Absorbing].

We will say that $\overline{T}$ is *top class regular* (TCR) if it has a top class that is regular, and analogously for *top class absorbing* (TCA).[4] Top class regularity represents aperiodic behaviour: it demands that there is some time instant $k^* \in \mathbb{N}$ such that all of the elements in the top class $\mathscr{R}$ are accessible from each other in k steps, for any $k \geq k^*$. In the case of traditional Markov chains, top class regularity suffices as a necessary and sufficient condition for ergodicity [10, 17]. However, in the imprecise case, we need the additional condition of being top class absorbing, which ensures that the top class will eventually be reached. It requires that, if the process starts from any state $x \in \mathscr{R}^c$, the lower probability that it will ever transition to $\mathscr{R}$ is strictly positive; see [10] for a more detailed discussion. This notion of an absorbing (top) class, however, is not to be confused with what is called an absorbing state in standard literature on Markov chains. The latter simply is a state such that, once entered, it can never be left anymore [17, Section 3.1]; in our terminology, this is the same as a closed class that consists of a single state.

From a practical point of view, an important feature of both of these accessibility conditions is that they can be easily checked in practice, as is shown in [15, Section 5]. Strictly speaking, though, the method for checking A2 that is presented in [15, Proposition 6], only applies to regular top classes. However, a closer look at the proof—of [15, Proposition 6]—shows that it does not rely on the regularity of the top class, and that the method can therefore be applied to any top class. Hence, the condition of (TCA)—the central condition of this paper that will turn out to be necessary and sufficient for weak ergodicity—can be easily verified in practice by first checking the existence of a top class $\mathscr{R}$—for instance, using (3)—and then checking A2 using the method described in [15, Proposition 6]. A more explicit treatment of this subject, however, would lead us too far, and we therefore leave it to this informal argument.

The characterisation of ergodicity using (TCR) and (TCA) was strongly inspired by the observation that upper transition operators are part of a specific collection of order-preserving maps, called *topical maps*. These are maps $F : \mathbb{R}^n \to \mathbb{R}^n$ that satisfy, for all $h, g \in \mathbb{R}^n$ and all $\mu \in \mathbb{R}$,

T1. $F(\mu + h) = \mu + Fh$ [constant additivity];

T2. if $h \leq g$ then $F(h) \leq F(g)$ [monotonicity].

To show this, we identify $\mathscr{L}(\mathscr{X})$ with the finite-dimensional linear space $\mathbb{R}^n$, with $n = |\mathscr{X}|$; this is clearly possible because both are isomorphic. That every upper transition operator is topical now follows trivially from U5 and U6. What is perhaps less obvious, but can be derived in an equally trivial way, is that the operator $\overline{T}_f$ is also topical. This allows us to apply results for topical maps to $\overline{T}_f$ in order to find necessary and sufficient conditions for weak ergodicity.

---

[4] The definitions for (TCR) and (TCA) in [15] differ slightly from ours; they are equivalent though, as can be seen from Lemmas 27 and 28.

6. A Sufficient Condition for Weak Ergodicity

As a first step, we aim to find sufficient conditions for the existence of $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$. To that end, recall from Section 4 that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ exists if—and only if—the limit $\lim_{k\to+\infty} \overline{T}_f^k(0)/k$ exists, and in that case $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) = \lim_{k\to+\infty} \overline{T}_f^k(0)/k$. Then, since $\overline{T}_f$ is topical, the following lemma implies that it is also equal to $\lim_{k\to+\infty} \overline{T}_f^k h/k$ for any $h \in \mathscr{L}(\mathscr{X})$.

Lemma 2. [13, Lemma 3.1] *Consider any topical map $F\colon \mathbb{R}^n \to \mathbb{R}^n$. If the limit $\lim_{k\to+\infty} F^k h/k$ exists for some $h \in \mathbb{R}^n$, then the limit exists for all $h \in \mathbb{R}^n$ and they are all equal.*

Hence, if $\lim_{k\to+\infty} \overline{T}_f^k h/k$ converges to a constant vector μ for some $h \in \mathscr{L}(\mathscr{X})$, then $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ exists and is equal to μ. This condition is clearly satisfied if the map $\overline{T}_f$ has an (additive) eigenvector $h \in \mathscr{L}(\mathscr{X})$, meaning that $\overline{T}_f^k h = h + k\mu$ for some $\mu \in \mathbb{R}$ and all $k \in \mathbb{N}_0$. In that case, we have that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) = \mu$, where μ is called the eigenvalue corresponding to h. Moreover, there can then only be one such eigenvalue μ.

Corollary 3. *Consider any topical map $F\colon \mathbb{R}^n \to \mathbb{R}^n$. If F has an (additive) eigenvalue μ, then it is the only eigenvalue of F.*

Proof. Suppose that F has two eigenvalues μ_1 and μ_2, and let h_1 and h_2 be the corresponding eigenvectors. Then we have that $F^k h_1 = h_1 + k\mu_1$ for all $k \in \mathbb{N}_0$, which immediately implies that $\lim_{k\to+\infty} F^k h_1/k = \mu_1$. In a similar way, we obtain that $\lim_{k\to+\infty} F^k h_2/k = \mu_2$. Then, due to Lemma 2, we have that $\mu_1 = \mu_2$. $\quad\square$

To find conditions that guarantee the existence of an eigenvector of $\overline{T}_f$, we will make use of results from [12] and [13]. There, accessibility graphs are defined in a slightly different way: for any topical map $F\colon \mathbb{R}^n \to \mathbb{R}^n$, they let $\mathscr{G}'(F)$ be the graph with vertices $v_1, \cdots, v_n$ and an edge from v_i to v_j if $\lim_{\alpha\to+\infty}[F(\alpha \mathbb{I}_{v_j})](v_i) = +\infty$. Subsequently, for such a graph $\mathscr{G}'(F)$, the accessibility relation $\cdot \to \cdot$ and corresponding notions (e.g. 'strongly connected', 'top class', …) are defined as in Section 5. If we identify the vertices $v_1, \cdots, v_n$ in $\mathscr{G}'(\overline{T})$ and $\mathscr{G}'(\overline{T}_f)$ with the different states $x_1, \cdots, x_n$ in $\mathscr{X}$, this can in particular be done for the topical maps $\overline{T}$ and $\overline{T}_f$. The following results show that the resulting graphs coincide with the one defined in Section 5.

Lemma 4. *For any two vertices x and y in $\mathscr{G}'(\overline{T})$, there is an edge from x to y in $\mathscr{G}'(\overline{T})$ if and only if there is an edge from x to y in $\mathscr{G}(\overline{T})$.*

Proof. Consider any two vertices x and y in the graph $\mathscr{G}'(\overline{T})$. By definition, there is an edge from x to y if $\lim_{\alpha\to+\infty}[\overline{T}(\alpha \mathbb{I}_y)](x) = +\infty$. Due to U3, this is equivalent to the condition that $\lim_{\alpha\to+\infty} \alpha[\overline{T}\mathbb{I}_y](x) = +\infty$. Since moreover $0 \le \overline{T}\mathbb{I}_y \le 1$ by U4, this condition reduces to $\overline{T}\mathbb{I}_y(x) > 0$. $\quad\square$

Corollary 5. *The graphs $\mathscr{G}'(\overline{T}_f)$, $\mathscr{G}'(\overline{T})$ and $\mathscr{G}(\overline{T})$ are identical.*

Proof. Lemma 4 implies that $\mathscr{G}'(\overline{T})$ and $\mathscr{G}(\overline{T})$ are identical. Moreover, that $\mathscr{G}'(\overline{T}_f)$ is equal to $\mathscr{G}'(\overline{T})$, follows straightforwardly from the definition of $\overline{T}_f$. $\quad\square$

In principle, we could use this result to directly obtain the desired condition for the existence of an eigenvector from [12, Theorem 2]. However, [12, Theorem 2] is given in a multiplicative framework and would need to be reformulated in an additive framework in order to be applicable to the map $\overline{T}_f$; see [12, Section 2.1]. This can be achieved with a bijective transformation, but we prefer to not do so because it would require too much extra terminology and notation. Instead, we will derive an additive variant of [12, Theorem 2] directly from [12, Theorem 9] and [12, Theorem 10].

The first result establishes that the existence of an eigenvector is equivalent to the fact that trajectories are bounded with respect to the Hilbert semi-norm $\|\cdot\|_{\mathrm{H}}$, defined by $\|h\|_{\mathrm{H}} := \max h - \min h$ for all $h \in \mathbb{R}^n$.

Theorem 6. [12, Theorem 9] *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a topical map. Then F has an eigenvector in $\mathbb{R}^n$ if and only if $\{\|F^k h\|_{\mathrm{H}} \colon k \in \mathbb{N}\}$ is bounded for some (and hence all) $h \in \mathbb{R}^n$.*

That the boundedness of a single trajectory indeed implies the boundedness of all trajectories follows from the non-expansiveness of a topical map with respect to the Hilbert semi-norm [12]. The second result that we need uses the notion of a *super-eigenspace*, defined for any topical map F and any $\mu \in \mathbb{R}$ as the set $S^{\mu}(F) \coloneqq \{h \in \mathbb{R}^n \colon F h \le h + \mu\}$.

Theorem 7. [12, Theorem 10] *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a topical map such that the associated graph $\mathscr{G}'(F)$ is strongly connected. Then all of the super-eigenspaces are bounded in the Hilbert semi-norm.*

Together, these theorems imply that any topical map $F\colon \mathbb{R}^n \to \mathbb{R}^n$ for which the graph $\mathscr{G}'(F)$ is strongly connected, has an eigenvector. The connection between both is provided by the fact that trajectories cannot leave an eigenspace. The following result formalises this.

Theorem 8. *Let $F\colon \mathbb{R}^n \to \mathbb{R}^n$ be a topical map such that the associated graph $\mathscr{G}'(F)$ is strongly connected. Then F has an eigenvector in $\mathbb{R}^n$.*

Proof. Consider any $h \in \mathbb{R}^n$ and any $\mu \in \mathbb{R}$ such that $\max(F h - h) \le \mu$. Then $F h \le h + \mu$, so $h \in S^{\mu}(F)$. Now notice that $F(F h) \le F(h + \mu) = F h + \mu$ because of T1 and T2, which implies that also $F h \in S^{\mu}(F)$. In the same way, we can also deduce that $F^2 h \in S^{\mu}(F)$ and, by repeating this argument, that the whole trajectory corresponding to h remains in $S^{\mu}(F)$. This trajectory is bounded because of Theorem 7, which by Theorem 6 guarantees the existence of an eigenvector. $\qquad\square$

In particular, if $\mathscr{G}'(\overline{T}_f)$ is strongly connected then $\overline{T}_f$ has an eigenvector, which on its turn implies the existence of $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ as explained earlier. If we combine this observation with Corollary 5, we obtain the following result.

Proposition 9. *$\overline{T}$ is weakly ergodic if the associated graph $\mathscr{G}(\overline{T})$ is strongly connected. In that case, for any $f \in \mathscr{L}(\mathscr{X})$, the limit value $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ is equal to the unique (additive) eigenvalue of $\overline{T}_f$.*

Proof. Suppose that $\mathscr{G}(\overline{T})$ is strongly connected. Then, by Corollary 5, $\mathscr{G}'(\overline{T}_f)$ is also strongly connected. Hence, for any $f \in \mathscr{L}(\mathscr{X})$, since $\overline{T}_f$ is a topical map, Theorem 8 guarantees the existence of an eigenvector of $\overline{T}_f$. Let μ be the corresponding eigenvalue. By Corollary 3, this eigenvalue μ is the only, and therefore unique, eigenvalue corresponding to $\overline{T}_f$. As explained in the beginning of this section, it now follows from Lemma 2 that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ exists and is equal to μ, so we indeed find that $\overline{T}$ is weakly ergodic. $\qquad\square$

In the remainder of this paper, we will use the fact that $\overline{T}$ is an upper transition operator—so not just any topical map—to strengthen this result. In particular, we will show that the condition of being strongly connected can be replaced by a weaker one: being top class absorbing. Nonetheless, the result above can already be useful in practice because checking whether a graph is strongly connected can be done rather efficiently; in any case more efficiently than checking for (TCA) since that requires us to first check the existence of a top class anyway. Hence, when interested in weakly ergodic behaviour, and when the dimensions of the considered model are large, one may prefer to verify this before checking the weaker condition of being top class absorbing.[5]

---

[5]Note that one might want to do so even in case the considered imprecise Markov chain fails to satisfy (TCR)—and therefore fails to be ergodic. This is illustrated by Example 1 which depicts a situation where it is immediately clear that (TCR) is not satisfied—since $\mathscr{G}(\overline{T})$ is cyclic—but where the graph $\mathscr{G}(\overline{T})$ is nevertheless strongly connected.

7. A Necessary and Sufficient Condition for Weak Ergodicity

In order to gain some intuition about how to obtain a more general sufficient condition for weak ergodicity, consider the case where $\overline{T}$ has a closed (or, equivalently, maximal) communication class $\mathscr{S}$ and the process' initial state x is in $\mathscr{S}$. Since $\mathscr{S}$ is closed, the process surely remains in $\mathscr{S}$ and hence, it is to be expected that the time average of f will not be affected by the dynamics of the process outside $\mathscr{S}$. Moreover, the communication class $\mathscr{S}$ is a strongly connected component, so one would expect that, due to Proposition 9, the upper expected time average $\overline{E}_{av,k}(f|x)$ converges to a constant that does not depend on the state $x \in \mathscr{S}$. Our intuition is formalised by the following proposition. Its proof, as well as those of the other statements in this section (apart from Theorem 14), can be found in the Appendix.

Proposition 10. *For any closed communication class $\mathscr{S}$ of $\overline{T}$, any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{S}$, the inference $\overline{E}_{av,k}(f|x)$ is equal to $\overline{E}_{av,k}(f\mathbb{I}_{\mathscr{S}}|x)$ and converges to a limit value as k recedes to infinity. This limit value is furthermore the same for all $x \in \mathscr{S}$.*

As a next step, we want to extend the domain of convergence of $\overline{E}_{av,k}(f|x)$ to all states $x \in \mathscr{X}$. To do so, we will impose the additional property of being top class absorbing (TCA), which, as explained in Section 5, demands that there is a strictly positive (lower) probability to reach the top class $\mathscr{R}$ in a finite time period. Once in $\mathscr{R}$, the process can never escape $\mathscr{R}$ though. One would therefore expect that as time progresses—as more of these finite time periods go by—this lower probability increases, implying that the process will eventually be in $\mathscr{R}$ with practical certainty. Furthermore, if the process transitions from $x \in \mathscr{R}^c$ to a state $y \in \mathscr{R}$, then Proposition 10 guarantees that $\overline{E}_{av,k}(f|y)$ converges to a limit and that this limit value does not depend on the state y. Finally, since the average is taken over a growing time interval, the initial finite number of time steps that it took for the process to transition from x to y will not influence the time average of f in the limit. This leads us to suspect that $\overline{E}_{av,k}(f|x)$ converges to the same limit as $\overline{E}_{av,k}(f|y)$. Since this argument applies to any $x \in \mathscr{R}^c$, we are led to believe that $\overline{T}$ is weakly ergodic. The following result confirms this.

Proposition 11. *$\overline{T}$ is weakly ergodic if it satisfies (TCA).*

Conversely, suppose that $\overline{T}$ does not satisfy (TCA). Then there are two possibilities: either there is no top class or there is a top class but it is not absorbing. If there is no top class, then it can be easily deduced that there are at least two closed communication classes $\mathscr{S}_1$ and $\mathscr{S}_2$. As discussed earlier, the process cannot leave the classes $\mathscr{S}_1$ and $\mathscr{S}_2$ once it has reached them. So if it starts in one of these communication classes, the process' dynamics outside this class are irrelevant for the behaviour of the resulting time average. In particular, if we let f be the function that takes the constant value c_1 in $\mathscr{S}_1$ and c_2 in $\mathscr{S}_2$, with $c_1 \neq c_2$, then we would expect that $\overline{E}_{av,k}(f|x) = c_1$ and $\overline{E}_{av,k}(f|y) = c_2$ for all $k \in \mathbb{N}_0$, any $x \in \mathscr{S}_1$ and any $y \in \mathscr{S}_2$. In fact, this can easily be formalised by means of Proposition 10. Hence, $\overline{E}_{av,\infty}(f|x) = c_1 \neq c_2 = \overline{E}_{av,\infty}(f|y)$, so the upper transition operator $\overline{T}$ cannot be weakly ergodic. In other words, if $\overline{T}$ is weakly ergodic, there must be a top class.

Proposition 12. *$\overline{T}$ has a top class if it is weakly ergodic.*

Finally, suppose that there is a top class $\mathscr{R}$, but that it is not absorbing. This implies that there is an $x \in \mathscr{R}^c$ and a compatible precise model such that the process is guaranteed to remain in $\mathscr{R}^c$ given that it started in x.[6] If we now let $f = \mathbb{I}_{\mathscr{R}^c}$, then conditional on the fact that $X_0 = x$, the expected time average of f corresponding to this precise model is equal to 1. Furthermore, since $f \leq 1$, no other process can yield a higher expected time average. The upper expected time average $\overline{E}_{av,k}(f|x)$ is therefore equal to 1 for all $k \in \mathbb{N}$. However, using Proposition 10, we can also show that $\overline{E}_{av,k}(f|y) = 0$ for any $y \in \mathscr{R}$ and all $k \in \mathbb{N}$. Hence, $\overline{E}_{av,\infty}(f|x) = 1 \neq 0 = \overline{E}_{av,\infty}(f|y)$, which precludes $\overline{T}$ from being weakly ergodic.

---

[6]For the sake of this intuitive explanation, we have assumed the set $\mathscr{T}$ to be closed. If this is not the case, we can actually not guarantee the existence of such a precise compatible model. Our proof of Proposition 13 uses a more involved—yet less intuitive—argument that does not require $\mathscr{T}$ to be closed.

Proposition 13. $\overline{T}$ *satisfies (TCA) if it is weakly ergodic and has a top class.*

Together with Propositions 11 and 12, this allows us to conclude that (TCA) is a necessary and sufficient condition for weak ergodicity.

Theorem 14. $\overline{T}$ *is weakly ergodic if and only if it satisfies (TCA).*

Proof. That (TCA) is a sufficient condition follows from Proposition 11. Necessity follows from Proposition 12 together with Proposition 13. □

8. Weak Ergodicity for Imprecise Markov Chains Under Repetition Independence

So far, we have ignored imprecise Markov chains under repetition independence in our analysis of weakly ergodic behaviour. Within the field of imprecise probability, these imprecise Markov chains are less studied because (i) they can incorporate fewer types of model uncertainty and (ii) they seem difficult to handle computationally, in the sense that there are (almost) no methods that are able to efficiently solve inference problems for such models; see [21, Section 5.7] for more details. On the other hand, their relevance should not be underestimated; they model the practical situation where we believe there is a single and fixed transition matrix T, but only have partial knowledge about the numerical values that make up this matrix.

The concepts of ergodicity and weak ergodicity can be defined in a similar way as for imprecise Markov chains under epistemic irrelevance and imprecise Markov chains under complete independence. For any $f \in \mathscr{L}(\mathscr{X})$, any $x \in \mathscr{X}$ and any $k \in \mathbb{N}$, we let

$$\overline{\mathrm{E}}_k^{\mathrm{ri}}(f|x) := \overline{\mathrm{E}}_{\mathscr{T}}^{\mathrm{ri}}(f(X_k)|X_1 = x) \quad \text{and} \quad \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) := \overline{\mathrm{E}}_{\mathscr{T}}^{\mathrm{ri}}(f_{\mathrm{av}}(X_{1:k})|X_1 = x).$$

Then we say that the imprecise Markov chain $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is ergodic if, for all $f \in \mathscr{L}(\mathscr{X})$, the upper expectation $\overline{\mathrm{E}}_k^{\mathrm{ri}}(f|x)$ converges to a value $\overline{\mathrm{E}}_\infty^{\mathrm{ri}}(f) := \lim_{k \to +\infty} \overline{\mathrm{E}}_k^{\mathrm{ri}}(f|x)$ that does not depend on $x \in \mathscr{X}$. Analogously, we call $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ weakly ergodic if, for all $f \in \mathscr{L}(\mathscr{X})$, the upper expectation $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ converges to a value $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) := \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ that does not depend on $x \in \mathscr{X}$. Unfortunately, however, and unlike what we saw for the inferences $\overline{\mathrm{E}}_k(f|x)$ and $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ in imprecise Markov chains under epistemic irrelevance and complete independence, there are no known recursive expressions that describe the behaviour of the inferences $\overline{\mathrm{E}}_k^{\mathrm{ri}}(f|x)$ and $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ in terms of a single (topical) map $F \colon \mathscr{L}(\mathscr{X}) \to \mathscr{L}(\mathscr{X})$. Hence, in order to study weak ergodicity (and conventional ergodicity) for these models, we will have to rely on a different approach than the one we have used before. Moreover, because of the absence of any recursive expressions, it is a priori not certain whether the inferences $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ will only depend on $\mathscr{T}$ through the upper transition operator $\overline{T}$. So, contrary to what we did in Sections 4 to 7, we cannot simply forget about the set $\mathscr{T}$ here. Instead, we will regard the set $\mathscr{T}$—which is always assumed to be separately specified—as the primary object that determines the values of the inferences $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$.

Before we begin our analysis, recall that the model $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ consists of all homogeneous (precise) Markov chains P that are compatible with $\mathscr{T}$, in the sense that $\mathrm{P}(X_2|X_1 = x) \in \mathscr{T}_x$ for all $x \in \mathscr{X}$—where we already took into account the homogeneity and the Markov property for each P. Since each homogeneous Markov chain P has a transition matrix T, we could alternatively write that $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ if and only if $T(x, \cdot) \in \mathscr{T}_x$ for all $x \in \mathscr{X}$ or, since $\mathscr{T}$ is separately specified, if and only if $T \in \mathscr{T}$. Furthermore, note that each $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is itself an imprecise Markov chain, where the upper transition operator is now linear and characterised by the transition matrix T of P. The independence assumption—epistemic irrelevance, complete independence or repetition independence—obviously does not matter here; they are all equivalent. Hence, the recursive expressions (1) and (2), and, more generally, all our results for imprecise Markov chains under epistemic irrelevance or complete independence, also hold for a homogeneous Markov chain $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$. The transition matrix T corresponding to P then takes the role of a particular upper transition operator $\overline{T}$. In the remainder, we will assume that the reader has taken notice of this fact and

we will simply apply our results from the previous sections in this more specific case without further ado. On top of that, we will often use the basic relation that $T^k h \leq \overline{T}^k h$ for all $T \in \mathcal{T}$, all $h \in \mathcal{L}(\mathcal{X})$ and all $k \in \mathbb{N}_0$. This intuitive inequality can be easily derived from the definition of $\overline{T}$ and the monotonicity [U6] of $\overline{T}$ and T.

For notational convenience, we use T_f and $\mathrm{E}_{\mathrm{av},k}^\mathrm{P}$ to denote the objects $\overline{T}_f$ and $\overline{\mathrm{E}}_{\mathrm{av},k}$—or, equivalently, $\overline{\mathrm{E}}_{\mathrm{av},k}^\mathrm{ri}$—that correspond to any homogeneous Markov chain $\mathrm{P} \in \mathscr{P}_\mathcal{T}^\mathrm{ri}$; so we let $T_f h := f + T h$, with T the transition matrix of P, and $\mathrm{E}_{\mathrm{av},k}^\mathrm{P}(f_{\mathrm{av}} | x) := \mathrm{E}_\mathrm{P}(f_{\mathrm{av}}(X_{1:k}) | X_1 = x)$ for all $f, h \in \mathcal{L}(\mathcal{X})$, all $x \in \mathcal{X}$ and all $k \in \mathbb{N}$. Taking into account the previous considerations, we can then write that, for all $f \in \mathcal{L}(\mathcal{X})$, all $x \in \mathcal{X}$ and all $k \in \mathbb{N}$,

$$\overline{\mathrm{E}}_{\mathrm{av},k}^\mathrm{ri}(f | x) = \sup_{\mathrm{P} \in \mathscr{P}_\mathcal{T}^\mathrm{ri}} \mathrm{E}_{\mathrm{av},k}^\mathrm{P}(f | x) = \sup_{T \in \mathcal{T}} \tfrac{1}{k}[T_f^k(0)](x), \tag{4}$$

where the last equality holds because of Equation (2) and the fact that $\mathrm{P} \in \mathscr{P}_\mathcal{T}^\mathrm{ri}$ if and only if $T \in \mathcal{T}$ (for the transition matrix T correponding to P). The expression above will now serve as the main starting point in our further study of $\overline{\mathrm{E}}_{\mathrm{av},k}^\mathrm{ri}(f | x)$.

As a first step, we show that an imprecise Markov chain $\mathscr{P}_\mathcal{T}^\mathrm{ri}$ is weakly ergodic if the upper transition operator $\overline{T}$ corresponding to $\mathcal{T}$ has a graph $\mathscr{G}(\overline{T})$ that is strongly connected. Moreover, in that case, we also have that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^\mathrm{ri}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $f \in \mathcal{L}(\mathcal{X})$. The result can be obtained in a rather straightforward fashion from the following lemma, which essentially states that, if $\overline{T}_f$ has an eigenvalue, there is a homogeneous Markov chain $\mathrm{P} \in \mathscr{P}_\mathcal{T}^\mathrm{ri}$ that behaves 'approximately' weakly ergodic. Its proof makes use of the supremum norm $\|\cdot\|_\infty$, defined by $\|h\|_\infty := \max_{x \in \mathcal{X}} |h(x)|$ for all $h \in \mathcal{L}(\mathcal{X})$.

Lemma 15. *Consider any $f \in \mathcal{L}(\mathcal{X})$. If the map $\overline{T}_f$ has an eigenvalue μ, then, for any $\epsilon > 0$, there is some $T \in \mathcal{T}$ such that*

$$\mu - \epsilon \leq \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)] \leq \limsup_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)] \leq \mu.$$

Proof. Suppose that the map $\overline{T}_f$ has an eigenvalue μ. Then $\overline{T}_f$ has at least one eigenvector $h \in \mathcal{L}(\mathcal{X})$ for which it holds that $\overline{T}_f^k h = h + k\mu$ for all $k \in \mathbb{N}$. Fix any $\epsilon > 0$. Then, since $\mathcal{T}$ is assumed to be separately specified, we are allowed to use S1, which implies that there is some $T \in \mathcal{T}$ such that $\overline{T}_f h - \epsilon = f + \overline{T} h - \epsilon \leq f + T h = T_f h \leq \overline{T}_f h$. The function h is an eigenvector of $\overline{T}_f$ with eigenvalue μ, so it follows that $h + \mu - \epsilon \leq T_f h \leq h + \mu$. Since T_f is monotone [T2] and satisfies constant additivity [T1]—because T is monotone [U6] and satisfies constant additivity [U5]—this implies that

$$h + 2\mu - 2\epsilon \leq T_f h + \mu - \epsilon = T_f(h + \mu - \epsilon) \leq T_f^2 h \leq T_f(h + \mu) = T_f h + \mu \leq h + 2\mu.$$

Using the same properties of T_f, this on its turn implies that

$$h + 3\mu - 3\epsilon \leq T_f^2 h + \mu - \epsilon = T_f(T_f h + \mu - \epsilon) = T_f^2(h + \mu - \epsilon)$$
$$\leq T_f^3 h \leq T_f^2(h + \mu) = T_f(T_f h + \mu) = T_f^2 h + \mu \leq h + 3\mu.$$

Repeating the argument above allows us to conclude that $h + k\mu - k\epsilon \leq T_f^k h \leq h + k\mu$ for all $k \in \mathbb{N}$, and therefore, that $\|h + k\mu - T_f^k h\|_\infty \leq k\epsilon$ for all $k \in \mathbb{N}$. Furthermore, T_f is topical [since, as we already mentioned above, T_f satisfies T1 and T2], so T_f is non-expansive with respect to the supremum norm [13, Proposition 1.1]. This implies that $\|T_f^k h - T_f^k(0)\|_\infty \leq \|T_f^{k-1} h - T_f^{k-1}(0)\|_\infty \leq \cdots \leq \|h - 0\|_\infty = \|h\|_\infty$ for all $k \in \mathbb{N}$. Then, recalling that $\|h + k\mu - T_f^k h\|_\infty \leq k\epsilon$ for all $k \in \mathbb{N}$, we can use the triangle inequality to infer that $\|h + k\mu - T_f^k(0)\|_\infty \leq k\epsilon + \|h\|_\infty$ or, equivalently, that $\|\tfrac{1}{k} h + \mu - \tfrac{1}{k} T_f^k(0)\|_\infty \leq \epsilon + \tfrac{1}{k}\|h\|_\infty$ for all $k \in \mathbb{N}$. Hence, we have that $(\tfrac{1}{k} h + \mu) - (\epsilon + \tfrac{1}{k}\|h\|_\infty) \leq$

$\frac{1}{k} T_f^k(0)$ for all $k \in \mathbb{N}$, which implies that

$$\mu - \epsilon \leq \liminf_{k \to +\infty} \tfrac{1}{k} T_f^k(0). \tag{5}$$

On the other hand, we have that $\limsup_{k \to +\infty} \frac{1}{k} T_f^k(0) \leq \limsup_{k \to +\infty} \frac{1}{k} \overline{T}_f^k(0)$ because $T_f^k(0) \leq \overline{T}_f^k(0)$ for all $k \in \mathbb{N}$. Furthermore, since h is an eigenvector of $\overline{T}_f$ with eigenvalue μ, we have that $\lim_{k \to +\infty} \frac{1}{k} \overline{T}_f^k h = \mu$, which by Lemma 2 implies that $\lim_{k \to +\infty} \frac{1}{k} \overline{T}_f^k(0) = \mu$. Hence, we have that $\limsup_{k \to +\infty} \frac{1}{k} T_f^k(0) \leq \mu$, which, together with Equation (5) and the fact that obviously $\liminf_{k \to +\infty} \frac{1}{k} T_f^k(0) \leq \limsup_{k \to +\infty} \frac{1}{k} T_f^k(0)$, implies the desired statement. $\qquad \square$

Proposition 16. *If $\mathscr{G}(\overline{T})$ is strongly connected, then $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic and we have that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$.*

Proof. Fix any $f \in \mathscr{L}(\mathscr{X})$. Since $\mathscr{G}(\overline{T})$ is strongly connected, Proposition 9 implies that $\overline{T}$ is weakly ergodic and that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) = \mu$ where μ is the unique eigenvalue of the map $\overline{T}_f$. Then it follows from Lemma 15 that, for any $\epsilon > 0$, there is a $T \in \mathscr{T}$ such that $\mu - \epsilon \leq \liminf_{k \to +\infty} \frac{1}{k}[T_f^k(0)]$. So, for any $x \in \mathscr{X}$, we have that

$$\mu \leq \sup_{T \in \mathscr{T}} \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)](x) \leq \liminf_{k \to +\infty} \sup_{T \in \mathscr{T}} \tfrac{1}{k}[T_f^k(0)](x) = \liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x),$$

where we used Equation (4) in the last step. On the other hand, since $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \leq \overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ for all $k \in \mathbb{N}$ (because $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}} \subseteq \mathscr{P}_{\mathscr{T}}^{\mathrm{ci}} \subseteq \mathscr{P}_{\mathscr{T}}^{\mathrm{ei}}$), we also have that

$$\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \leq \liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f) = \mu.$$

Hence, we conclude that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ exists and is equal to $\mu = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $x \in \mathscr{X}$ and therefore, that $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic and that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$. $\qquad \square$

Proposition 16 provides a sufficient condition—having a graph $\mathscr{G}(\overline{T})$ that is strongly connected—for an imprecise Markov chain $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ to be weakly ergodic. However, as was the case for the imprecise Markov chains $\mathscr{P}_{\mathscr{T}}^{\mathrm{ei}}$ and $\mathscr{P}_{\mathscr{T}}^{\mathrm{ci}}$, this condition can actually be weakened to the condition that $\overline{T}$ should be top class absorbing (TCA). It will moreover turn out that this weaker accessibility condition is not only sufficient, but also necessary.

Both the sufficiency and necessity of (TCA) can be made intuitive using arguments that are somewhat similar to the ones we have used in Section 7, where we discussed the sufficiency and necessity of (TCA) for imprecise Markov chains under epistemic irrelevance and complete independence. Before we explain why (TCA) is sufficient here as well, first observe that, for any closed class $\mathscr{S}$ in the graph $\mathscr{G}(\overline{T})$—associated with $\mathscr{T}$ through $\overline{T}$—the class $\mathscr{S}$ is also closed in the graph $\mathscr{G}(T)$, for all $T \in \mathscr{T}$:

Lemma 17. *If a class $\mathscr{S}$ is closed in $\mathscr{G}(\overline{T})$, then, for all $T \in \mathscr{T}$, the class $\mathscr{S}$ is also closed in $\mathscr{G}(T)$.*

Proof. Let $\mathscr{S}$ be any closed class in $\mathscr{G}(\overline{T})$ and consider any $x \in \mathscr{S}$. By definition, we have that $x \nrightarrow y$ in $\mathscr{G}(\overline{T})$ for any $y \in \mathscr{S}^c$, which by Lemma 1 implies that $\overline{T}^k \mathbb{I}_y(x) \leq 0$ for all $k \in \mathbb{N}$. Furthermore, for any $T \in \mathscr{T}$ and all $k \in \mathbb{N}$, we have that $T^k \mathbb{I}_y(x) \leq \overline{T}^k \mathbb{I}_y(x)$, so it follows that $T^k \mathbb{I}_y(x) \leq 0$. Hence, by Lemma 1, $x \nrightarrow y$ in $\mathscr{G}(T)$. This holds for all $x \in \mathscr{S}$ and all $y \in \mathscr{S}^c$, so $\mathscr{S}$ is also closed in $\mathscr{G}(T)$ for any $T \in \mathscr{T}$. $\qquad \square$

Now suppose that the upper transition operator $\overline{T}$ corresponding to $\mathscr{T}$ satisfies (TCA) with top class $\mathscr{R}$, and that the process' initial state x is in $\mathscr{R}$. Because of the lemma above, $\mathscr{R}$ is also closed in $\mathscr{G}(T)$ for any $T \in \mathscr{T}$. Hence, according to any (homogeneous) Markov chain P with transition matrix T, the process' state can never leave $\mathscr{R}$. So it is to be expected that the inference $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x)$ will not be influenced by the behaviour of the Markov

chain P outside of $\mathscr{R}$. Since this is true for any $T \in \mathscr{T}$, and since $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ is simply an upper envelope of $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x)$ over all compatible Markov chains P, it seems that, in our study of the inference $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$, we can limit ourselves to that part of $\mathscr{T}$ and $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ that describes the process' dynamics for states in $\mathscr{R}$. Moreover, it can be shown (see Lemma 45 in Appendix E) that the upper transition operator $\overline{T}'$ and the graph $\mathscr{G}(\overline{T}')$ associated with such a restricted version $\mathscr{T}'$ of $\mathscr{T}$, are itself versions of the original $\overline{T}$ and $\mathscr{G}(\overline{T})$ restricted to the states in $\mathscr{R}$. Since $\mathscr{R}$ is a strongly connected component in $\mathscr{G}(\overline{T})$, the graph $\mathscr{G}(\overline{T}')$ will therefore be strongly connected as a whole and we are led to believe, due to Proposition 16, that the inference $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ will converge to a constant that does not depend on the state $x \in \mathscr{R}$. So, one would expect that, if we limit ourselves to initial states in $\mathscr{R}$, the condition of (TCA) is sufficient for weakly ergodic behaviour.

In order to see why this can also be expected if we allow the process to start in $\mathscr{R}^c$, first consider the following straightforward lemma.

Lemma 18. *If $\overline{T}$ satisfies (TCA) with top class $\mathscr{R}$, then, for all $T \in \mathscr{T}$, the class $\mathscr{R}$ is an absorbing closed class in $\mathscr{G}(T)$.*

Proof. Since $\mathscr{R}$ is the top class in $\mathscr{G}(\overline{T})$, it is closed in $\mathscr{G}(\overline{T})$ [Lemma 25, Appendix A] and therefore, by Lemma 17, it is also closed in $\mathscr{G}(T)$ for all $T \in \mathscr{T}$. So it remains to prove that $\mathscr{R}$ is moreover absorbing in $\mathscr{G}(T)$ for all $T \in \mathscr{T}$. To do so, recall that $\mathscr{R}$ is an absorbing (and closed) class in $\mathscr{G}(\overline{T})$, meaning that $\overline{T}^{k_x} \mathbb{I}_{\mathscr{R}^c}(x) < 1$ for any $x \in \mathscr{R}^c$ and some $k_x \in \mathbb{N}$. Then, for any $T \in \mathscr{T}$, since $T^{k_x} \mathbb{I}_{\mathscr{R}^c} \leq \overline{T}^{k_x} \mathbb{I}_{\mathscr{R}^c}$, we also have that $T^{k_x} \mathbb{I}_{\mathscr{R}^c}(x) < 1$ which implies that $\mathscr{R}$ is indeed an absorbing closed class in $\mathscr{G}(T)$. $\qquad\square$

So if we consider any Markov chain $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ and let $T \in \mathscr{T}$ be the corresponding transition matrix, the class $\mathscr{R}$ is closed and absorbing in $\mathscr{G}(T)$. We can then reason in a similar way as in the paragraph that preceded Proposition 11; due to the fact that $\mathscr{R}$ is absorbing and closed, it is to be expected that, according to the model P, the process is eventually in $\mathscr{R}$ with practical certainty. Unlike before, however, $\mathscr{R}$ is not necessarily a communication class in $\mathscr{G}(T)$, so we cannot assert that P behaves in a weakly ergodic way for initial states in $\mathscr{R}$. In other words, for any $x \in \mathscr{R}$, the expected time average $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x)$ may not converge or, if it does converge, it may possibly depend on the initial state $x \in \mathscr{R}$. Nonetheless, it can be inferred from our previous discussion and Lemma 15 that the model $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ can be chosen in such a way that its behaviour for initial states $x \in \mathscr{R}$ is 'approximately' weakly ergodic, in the sense that $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x)$ will eventually lie at ϵ-distance from a constant μ that does not depend on $x \in \mathscr{R}$. Recalling that $\mathscr{R}$ is closed and absorbing—for all $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ and therefore, also for any specific choice of P—the same can be expected for initial states $y \in \mathscr{R}^c$. As a result, we would then get that $\mu - \epsilon \leq \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ for any $x \in \mathscr{X}$ and for k large enough. Moreover, it can easily be shown that the constant μ is the limit value of the inference $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ (see, for instance, Proposition 9), so μ is also an upper bound for the limit values of $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$. Both observations taken together, one would expect that the inference $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ converges to the constant μ for all $x \in \mathscr{X}$. Our next result confirms this.

Proposition 19. *If $\overline{T}$ satisfies (TCA), then $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic and $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$.*

Conversely, consider any $\mathscr{T}$ such that the corresponding $\overline{T}$ does not satisfy (TCA). Then either $\mathscr{G}(\overline{T})$ has no top class, in which case it must necessarily have two (or more) different closed communication classes $\mathscr{S}_1$ and $\mathscr{S}_2$, or $\mathscr{G}(\overline{T})$ has a top class but it is not absorbing. Consider the first case. Since $\mathscr{S}_1$ and $\mathscr{S}_2$ are both closed in $\mathscr{G}(\overline{T})$, they are also both closed (but not necessarily a communication class) in $\mathscr{G}(T)$ for any $T \in \mathscr{T}$, because of Lemma 17. So, according to any Markov chain $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ with transition matrix T, the process surely remains in $\mathscr{S}_i$ with $i \in \{1, 2\}$ once it has reached $\mathscr{S}_i$. Hence, if we let $c_1, c_2 \in \mathbb{R}$ such that $c_1 \neq c_2$ and assume that $f(x) = c_i$ for all $x \in \mathscr{S}_i$—which is possible because $\mathscr{S}_1$ and $\mathscr{S}_2$ are two different communication classes in $\mathscr{G}(\overline{T})$ and are therefore disjoint—we would expect that $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x)$ is simply equal to c_i for all $k \in \mathbb{N}$ and all $x \in \mathscr{S}_i$. Since this is the case for all $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$, we are inclined to conclude that $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) = c_i$ for all $k \in \mathbb{N}$ and all $x \in \mathscr{S}_i$, and therefore that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x_1) = c_1 \neq c_2 = \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x_2)$ for any $x_1 \in \mathscr{S}_1$ and any $x_2 \in \mathscr{S}_2$. This would preclude

$\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ from being weakly ergodic. As a result, if we assume that $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic, then $\mathscr{G}(\overline{T})$ must have a top class.

Proposition 20. *If $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic, then the graph $\mathscr{G}(\overline{T})$ has a top class.*

Finally, consider the case that $\mathscr{G}(\overline{T})$ has a top class $\mathscr{R}$ that is not absorbing and recall the discussion that lead to Proposition 13. There, we relied on the fact that, since $\mathscr{R}$ is not absorbing, there is some precise model $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ei}}$ for which the process is guaranteed to remain in $\mathscr{R}^c$ given that it started in some state $x \in \mathscr{R}^c$.[7] As a matter of fact, this compatible model P can always be chosen in such a way that it is a homogeneous Markov chain, and therefore such that $\mathrm{P} \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$. Hence, if we again let f be the indicator $\mathbb{I}_{\mathscr{R}^c}$, then we would obtain that $\mathrm{E}_{\mathrm{av},k}^{\mathrm{P}}(f|x) = 1$. Since no other $\mathrm{P}' \in \mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ can then yield a higher expected time average for this f, we would also have that $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) = 1$ for all $k \in \mathbb{N}$. On the other hand, since $\mathscr{R}$ is closed and $f(y) = 0$ for any $y \in \mathscr{R}$, it is to be expected—due to the same reasons as we have come to explain in the paragraph above, where both $\mathscr{S}_1$ and $\mathscr{S}_2$ were closed—that $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|y) = 0$ for all $k \in \mathbb{N}$. So in conclusion, we would have that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) = 1 \neq 0 = \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|y)$, again precluding $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ from being weakly ergodic.

Proposition 21. *If $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic and the graph $\mathscr{G}(\overline{T})$ has a top class, then $\overline{T}$ satisfies (TCA).*

It now only remains to combine Propositions 19, 20 and 21 to establish our second main result.

Theorem 22. *An imprecise Markov chain under repetition independence $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic if and only if the upper transition operator $\overline{T}$ corresponding with the separately specified set $\mathscr{T}$ satisfies (TCA). Furthermore, in that case, we have that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$.*

Proof. Sufficiency follows from Proposition 19; necessity follows from Proposition 20 and Proposition 21. The last statement—that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ for all $f \in \mathscr{L}(\mathscr{X})$ in the case that $\overline{T}$ satisfies (TCA)—also follows from Proposition 19. $\qquad\square$

9. Conclusion

The most important conclusion of our study of upper and lower expected time averages is the following (see Theorems 14 and 22): The condition of being top class absorbing is necessary and sufficient for weakly ergodic behaviour of an imprecise Markov chain, irrespectively of the imposed independence assumption. In that case, upper (and lower) expected time averages converge to limit values that are constant, not only for all possible initial states (or distributions) of the process, but also for all possible types of independence assumptions.

Now, if we compare our notion of weak ergodicity with that of conventional ergodicity—which guarantees the existence of a limit upper and lower expectation—we believe that weak ergodicity, and the associated (limits of) upper and lower expected time averages, should become the new objects of interest when it comes to characterising long-term average behaviour. Our conviction is based on the following three arguments:

i. Weak ergodicity requires less stringent conditions to be satisfied than conventional ergodicity, which additionally requires top class regularity. We illustrated this difference in Example 1, where we considered a(n imprecise) Markov chain that satisfies (TCA) but not (TCR).

ii. The inferences $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ are able to provide us with more information about how time averages might behave, compared to limit expectations $\overline{\mathrm{E}}_{\infty}(f)$. To see why, recall Example 2, where the inference $\overline{\mathrm{E}}_{\mathrm{av},\infty}(\mathbb{I}_b) =$

---

[7]Once more, this is only valid under the assumption that $\mathscr{T}$ is closed. The formal proof of Proposition 21, however, does not rely on $\mathscr{T}$ being closed.

1/2 significantly differed from $\overline{\mathrm{E}}_\infty(\mathbb{I}_b) = 1$. Clearly, the former was more representative for the limit behaviour of the time average of $\mathbb{I}_b$. As a consequence of [21, Lemma 57], a similar statement holds for general functions. In particular, it implies that $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f) \leq \overline{\mathrm{E}}_\infty(f)$ for any function $f \in \mathscr{L}(\mathscr{X})$. Since both inferences are upper bounds, $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ is therefore at least as (and sometimes much more) informative as $\overline{\mathrm{E}}_\infty(f)$.

iii. The characterisation of weak ergodicity, as well as the limit values $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$—or, equivalently, $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f)$—of upper (and lower) expected time averages, do not depend on the type of independence assumption that we impose on the imprecise Markov chain. As we have illustrated in Example 2, conventional ergodicity does not exhibit this kind of robustness. We perceive this as an advantage in favor of weak ergodicity. On the one hand, it provides a clear practical benefit because one should not spend time and/or money in searching for the appropriate independence assumption; it simply does not matter. On the other hand, it also opens doors for further theoretical research on this topic, because the limit values $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ are now approximated by two different objects: the upper expectations $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ and the upper expectations $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$. It is, for instance, not unreasonable to think that this feature might very well be a crucial step in developing efficient algorithms for the computation of $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$.

That said, there is also one important feature that limit upper and lower expectations have, but that we did not consider yet for upper and lower expected time averages: an (imprecise) point-wise ergodic theorem [8, Theorem 32]. For the limit upper and lower expectations of an ergodic imprecise Markov chain, this result states that

$$\underline{\mathrm{E}}_\infty(f) \leq \liminf_{k \to +\infty} f_{\mathrm{av}}(X_{1:k}) \leq \limsup_{k \to +\infty} f_{\mathrm{av}}(X_{1:k}) \leq \overline{\mathrm{E}}_\infty(f),$$

with lower probability one. In order for limit upper and lower expected time averages to be the undisputed quantities of interest when studying long-term time averages, a similar result would need to be obtained for weak ergodicity, where the role of $\overline{\mathrm{E}}_\infty(f)$ and $\underline{\mathrm{E}}_\infty(f) := -\overline{\mathrm{E}}_\infty(-f)$ is taken over by $\overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$ and $\underline{\mathrm{E}}_{\mathrm{av},\infty}(f) := -\overline{\mathrm{E}}_{\mathrm{av},\infty}(-f)$, respectively. If such a result would hold, it would provide us with (strictly almost sure) bounds on the limit values attained by time averages that are not only more informative than the current ones, but also guaranteed to exist under weaker conditions and equal for all types of independence assumptions. In fact, we are happy to report that we already established such a result. However, the proof necessitates a mathematical framework that strongly differs from the one used here, and we therefore intend to present it in future work.

Another topic that we would like to consider in the future, is the convergence of the inferences $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ and $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x)$ in general, without imposing that their limit values should be constant for all states x in $\mathscr{X}$. We suspect that this kind of convergence will require no conditions at all.

Acknowledgments

References

[1] A. Antonucci, C.P. de Campos, and M. Zaffalon. Probabilistic graphical models. In T. Augustin, F.P.A. Coolen, G. de Cooman, and M.C.M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 207–229. John Wiley & Sons, Chichester, 2014.

[2] T. Augustin, F.P.A. Coolen, G. de Cooman, and M.C.M. Troffaes. *Introduction to Imprecise Probabilities.* John Wiley & Sons, Chichester, 2014.

[3] G. Birkhoff. *Lattice Theory,* volume 25 of *American Mathematical Society colloquium publications.* American Mathematical Society, Providence, Rhode Island, 1940. Third edition.

[4] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5(2):165–181, 2000.

[5] F.G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199–233, 2000.

[6] F.G. Cozman. Sets of probability distributions, independence, and convexity. *Synthese*, 186(2):577–600, 2012.

[7] J. De Bock. Credal networks under epistemic irrelevance. *International Journal of Approximate Reasoning*, 85:107–138, 2017.

[8] G. de Cooman, J. De Bock, and S. Lopatatzidis. Imprecise stochastic processes in discrete time: global models, imprecise Markov chains, and ergodic theorems. *International Journal of Approximate Reasoning*, 76:18–46, 2016.

[9] G. de Cooman, F. Hermans, A. Antonucci, and M. Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9):1029–1052, 2010.

[10] G. de Cooman, F. Hermans, and E. Quaeghebeur. Imprecise Markov chains and their limit behaviour. *Probability in the Engineering and Informational Sciences*, 23(4):597–635, 2009.

[11] G. de Cooman and M.C.M. Troffaes. *Lower Previsions*. Wiley, Chichester, 2014.

[12] S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Transactions of the American Mathematical Society*, 356(12):4931–4950, 2004.

[13] J. Gunawardena. From max-plus algebra to nonexpansive mappings: a nonlinear theory for discrete event systems. *Theoretical Computer Science*, 293(1):141–167, 2003.

[14] D.J. Hartfiel. *Markov set-chains*. Lecture Notes in Mathematics. Springer, Berlin, 1998.

[15] F. Hermans and G. de Cooman. Characterisation of Ergodic Upper Transition Operators. *International Journal of Approximate Reasoning*, 53(4):573–583, 2012.

[16] F. Hermans and D. Škulj. Stochastic Processes. In T. Augustin, F.P.A. Coolen, G. de Cooman, and M.C.M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 258–278. John Wiley & Sons, Chichester, 2014.

[17] J.G. Kemeny and J.L. Snell. *Finite Markov chains*. Undergraduate Text in Mathematics. Springer-Verlag, New York, 1976.

[18] T. Krak, J. De Bock, and A. Siebes. Imprecise continuous-time Markov chains. *International Journal of Approximate Reasoning*, 88:452–528, 2017.

[19] T. Krak, N. T'Joens, and J. De Bock. Hitting times and probabilities for imprecise Markov chains. Extended version of [20], arXiv:1905.08781.

[20] T. Krak, N. T'Joens, and J. De Bock. Hitting times and probabilities for imprecise Markov chains. In *International Symposium on Imprecise Probabilities : Theories and Applications, ISIPTA 2019, Proceedings*, volume 103, pages 265–275. PMLR, 2019.

[21] S. Lopatatzidis. *Robust Modelling and Optimisation in Stochastic Processes using Imprecise Probabilities, with an Application to Queueing Theory*. PhD thesis, Ghent University, 2017.

[22] T. Seidenfeld. Conditional independence, imprecise probabilities, null-events, and graph-theoretic models. Seymour Geisser Distinguished Lecture, School of Statistics, University of Minnesota, Minneapolis, MN, 2007.

[23] A.N. Shiryaev. *Probability*. Graduate Texts in Mathematics. Springer, New York, second edition, 1996.

[24] D. Škulj. Discrete-time Markov chains with interval probabilities. *International Journal of Approximate Reasoning*, 50(8):1314–1329, 2009.

[25] N. T'Joens and J. De Bock. Limit behaviour of upper and lower expected time averages in discrete-time imprecise Markov chains. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU2020, Proceedings*, volume 1237 of *Communications in Computer and Information Science*, pages 224–238, 2020.

[26] N. T'Joens, T. Krak, J. De Bock, and G. de Cooman. A recursive algorithm for computing inferences in imprecise Markov chains. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2019, Proceedings*, volume 11726 of *Lecture Notes in Artificial Intelligence*, pages 455–465, 2019.

[27] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

Appendix A. Accessibility relations

Lemma 23. *The relation → induces a partial order on the partition $\mathscr{C}$ of all communication classes.*

Proof. That → is reflexive and transitive on $\mathscr{C}$ follows immediately from the reflexivity and transitivity of the relation → on the singletons. To see that it is also antisymmetric, consider any two sets A and B in $\mathscr{C}$. Then if $A \to B$ and $B \to A$, it should be clear that any two vertices in $A \cup B$ communicate and therefore that $A \cup B$ is a communication class. Since A and B are two sets in the partition $\mathscr{C}$ of communication classes, $A \cup B$ can only be a communication class as well if $A = B$. Hence, the relation → induces a partial order on the partition $\mathscr{C}$ of all communication classes. □

The following lemma establishes our claim about the equivalence—for communication classes—between the notions of maximality and closedness.

Lemma 24. *Consider any communication class $\mathscr{S} \in \mathscr{C}$. Then $\mathscr{S}$ is closed if and only if $\mathscr{S}$ is maximal.*

Proof. Suppose that $\mathscr{S}$ is closed; so $x \not\to y$ for all $x \in \mathscr{S}$ and all $y \in \mathscr{S}^c$. Consider any $\mathscr{S}' \in \mathscr{C}$ such that $\mathscr{S}' \neq \mathscr{S}$, and note that $\mathscr{S} \cap \mathscr{S}' = \emptyset$ because $\mathscr{C}$ is a partition. Then, since $\mathscr{S}$ is closed, there is no $x \in \mathscr{S}$ and $y \in \mathscr{S}'$ such

that $x \to y$, which implies that $\mathscr{S}'$ is not accessible from (or does not dominate) $\mathscr{S}$. Because this is the case for all $\mathscr{S}' \in \mathscr{C}$ such that $\mathscr{S}' \neq \mathscr{S}$, we conclude that $\mathscr{S}$ is maximal. Conversely, suppose that $\mathscr{S}$ is maximal, and consider any $x \in \mathscr{S}$ and any $y \in \mathscr{S}^c$. Let $\mathscr{S}' \in \mathscr{C}$ be such that $y \in \mathscr{S}'$; there is exactly one such $\mathscr{S}'$ because $\mathscr{C}$ forms a partition of $\mathscr{X}$. Moreover, $\mathscr{S}' \neq \mathscr{S}$ because $y \in \mathscr{S}' \cap \mathscr{S}^c$. Then, since $\mathscr{S}$ is maximal, we have that $\mathscr{S} \not\to \mathscr{S}'$ and therefore, in particular, $x \not\to y$. This is true for all $x \in \mathscr{S}$ and all $y \in \mathscr{S}^c$, so we conclude that $\mathscr{S}$ is closed. $\qquad\square$

The following result is well-known in order theory. However, since we could not immediately find an appropriate reference, we have chosen to provide a proof of our own.

Lemma 25. *A communication class $\mathscr{S}$ is the top class if and only if $\mathscr{S}$ is the only maximal—or, equivalently, closed—communication class.*

Proof. First note that maximality and closedness can indeed be used interchangeably here; see Lemma 24. To see that the direct implication holds, suppose that $\mathscr{S}$ is the top class. Then observe that $\mathscr{S}$ is maximal because otherwise there would be some $C \in \mathscr{C} \setminus \{\mathscr{S}\}$ such that $\mathscr{S} \to C$ and (since $\mathscr{S}$ is the top class) $C \to \mathscr{S}$, and therefore, by the antisymmetry of $\to$, that $\mathscr{S} = C$, contradicting that $C \in \mathscr{C} \setminus \{\mathscr{S}\}$. It is also the only maximal communication class because each $C \in \mathscr{C}$ is dominated by $\mathscr{S}$.

Now suppose that $\mathscr{S}$ is the only maximal communication class. To prove the converse implication, we will rely on the following observation: for any finite sequence $C_1, \cdots, C_n$ of different communication classes such that $C_i \to C_j$ and $C_i \neq \mathscr{S}$ for all $i, j \in \{1, \cdots, n\}$ such that $i \leq j$, there is a $C_{n+1} \in \mathscr{C}$ such that $C_i \neq C_{n+1}$ and $C_i \to C_{n+1}$ for all $i \in \{1, \cdots, n\}$. Indeed, $C_n \neq \mathscr{S}$, which implies that C_n is not maximal—$\mathscr{S}$ is the only maximal communication class—and therefore that there is some $C_{n+1} \in \mathscr{C}$ such that $C_n \to C_{n+1}$ (and obviously $C_n \neq C_{n+1}$). Since $C_i \to C_n$ for any $i \in \{1, \cdots, n\}$, the transitivity of $\to$ implies that also $C_i \to C_{n+1}$. Moreover, C_{n+1} differs from any C_i with $i \in \{1, \cdots, n-1\}$, because we would otherwise have that $C_i \to C_n$ and $C_n \to C_{n+1} = C_i$, and therefore, by the antisymmetry of $\to$, that $C_i = C_n$. This would contradict our assumptions about $C_1, \cdots, C_n$ and since we already established that $C_n \neq C_{n+1}$, we indeed conclude that $C_i \neq C_{n+1}$ and $C_i \to C_{n+1}$ for all $i \in \{1, \cdots, n\}$.

Now fix any C_1 such that $C_1 \neq \mathscr{S}$. Then we can use the rule above to show that $C_1 \to \mathscr{S}$ and therefore—since $C_1 \neq \mathscr{S}$ is arbitrary—that $\mathscr{S}$ is the top class. Indeed, since $C_1 \neq \mathscr{S}$, it follows from this rule that there is a $C_2 \in \mathscr{C}$ such that $C_1 \neq C_2$ and $C_1 \to C_2$. If also $C_2 \neq \mathscr{S}$, there is a third $C_3 \in \mathscr{C}$ such that $C_i \neq C_3$ and $C_i \to C_3$ for all $i \in \{1, 2\}$. If also $C_3 \neq \mathscr{S}$, then there is a fourth $C_4 \in \mathscr{C}$ such that $C_i \neq C_4$ and $C_i \to C_4$ for all $i \in \{1, \cdots, 3\}$, and so on, always continuing to extend this sequence in the same way. Then, since $\mathscr{X}$—and therefore also $\mathscr{C}$—is finite, and since $\mathscr{S} \in \mathscr{C}$, we will eventually find that, at some point, the next element C_n of this sequence is such that $C_n = \mathscr{S}$. Then, due to the fact that $C_i \to C_n$ for all $i \in \{1, \cdots, n-1\}$, we have in particular that $C_n = \mathscr{S}$ is accessible from C_1. $\qquad\square$

Corollary 26. *If $\overline{T}$ does not have a top class, then it has at least two (disjoint) closed communication classes.*

Proof. Once more, let $\mathscr{C}$ be the partition of all communication classes in $\mathscr{X}$. Since we know from Lemma 23 that $\to$ induces a partial order relation on $\mathscr{C}$ and because $\mathscr{C}$ is finite, there is at least one maximal communication class [3, Theorem 3]. The case that there is exactly one is impossible, because by Lemma 25 that would mean that there is a top class. Hence, there are at least two maximal—or, by Lemma 24, closed—communication classes in $\mathscr{X}$. Furthermore, both communication classes are necessarily disjoint because $\mathscr{C}$ is a partition of $\mathscr{X}$. $\qquad\square$

Lemma 27. *$\overline{T}$ has a top class $\mathscr{R}$ if and only if $\mathscr{R}' := \{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\} \neq \emptyset$ and, in that case, $\mathscr{R} = \mathscr{R}'$.*

Proof. Suppose that $\overline{T}$ has a top class $\mathscr{R}$. Then it follows from the fact that $\mathscr{R}$ dominates (or, in other words, is accessible from) each other communication class $\mathscr{S} \in \mathscr{C}$, that any $x \in \mathscr{X}$ is in $\mathscr{R}$ if and only if $y \to x$ for all

$y \in \mathscr{X}$. Indeed, on the one hand, if x is in $\mathscr{R}$, then for any $y \in \mathscr{X}$, if we let $\mathscr{Y}$ be the unique communication class that contains y, we find that $y \to x$ because $\mathscr{Y} \to \mathscr{R}$. And on the other hand, if $y \to x$ for all $y \in \mathscr{X}$, then for any communication class $\mathscr{Y}$, if we let y be any element of $\mathscr{Y}$, we find that $\mathscr{Y} \to \mathscr{R}$ because $y \to x$. Hence, since $\mathscr{R}$ is implicitly assumed to be non-empty if it exists, we have that $\mathscr{R}' = \mathscr{R} \neq \emptyset$.

Conversely, suppose that $\mathscr{R}' = \{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\} \neq \emptyset$. Then first observe that $\mathscr{R}'$ is a communication class. Indeed, on the one hand, for any $x, z \in \mathscr{R}'$, the definition of $\mathscr{R}'$ trivially implies that $x \longleftrightarrow z$. And on the other hand, for any $x \in \mathscr{R}'$ and $z \in \mathscr{X}$ such that $x \longleftrightarrow z$, we know that $z \in \mathscr{R}'$ because, for all $y \in \mathscr{X}$, we have that $y \to x \to z$ and hence $y \to z$. So $\mathscr{R}'$ is a communication class. Since for any other communication class $\mathscr{S} \in \mathscr{C}$, we have that $y \to x$ for any $x \in \mathscr{R}'$ and any $y \in \mathscr{S}$, it follows that $\mathscr{S} \to \mathscr{R}'$ for all $\mathscr{S} \in \mathscr{C}$. Hence, $\mathscr{R}'$ is the (non-empty) top class $\mathscr{R}$. $\qquad \square$

Lemma 28. $\overline{T}$ *satisfies (TCR) if and only if* $\mathscr{R}' := \{x \in \mathscr{X} : (\exists k^* \in \mathbb{N})(\forall k \geq k^*) \min \overline{T}^k \mathbb{I}_x > 0\} \neq \emptyset$ *and, in that case,* $\mathscr{R}'$ *is the top class* $\mathscr{R}$ *of* $\overline{T}$.

Proof. Suppose that $\overline{T}$ satisfies (TCR) and let $\mathscr{R} \neq \emptyset$ be the corresponding top class. Then, since $\mathscr{R}$ is regular [A1], we clearly have that $\mathscr{R} \subseteq \mathscr{R}'$. Moreover, since $\min \overline{T}^{k_x} \mathbb{I}_x > 0$ for any $x \in \mathscr{R}'$ and some $k_x \in \mathbb{N}$, it follows from Lemma 1 that any $x \in \mathscr{R}'$ is accessible from anywhere in $\mathscr{G}(\overline{T})$. Hence, due to Lemma 27, we also have that $\mathscr{R}' \subseteq \mathscr{R} = \{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\}$. As a conclusion, we have that $\mathscr{R}' = \mathscr{R} \neq \emptyset$.

Conversely, suppose that $\mathscr{R}' = \{x \in \mathscr{X} : (\exists k^* \in \mathbb{N})(\forall k \geq k^*) \min \overline{T}^k \mathbb{I}_x > 0\} \neq \emptyset$. Again, since $\min \overline{T}^{k_x} \mathbb{I}_x > 0$ for any $x \in \mathscr{R}'$ and some $k_x \in \mathbb{N}$, Lemma 1 implies that any $x \in \mathscr{R}'$ is accessible from anywhere in $\mathscr{G}(\overline{T})$. So, we have that $\mathscr{R}' \subseteq \{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\}$, which by Lemma 27 and the fact that $\mathscr{R}' \neq \emptyset$, and therefore $\{x \in \mathscr{X} : y \to x \text{ for all } y \in \mathscr{X}\} \neq \emptyset$, implies that the top class $\mathscr{R}$ exists, that it is non-empty, and that $\mathscr{R}' \subseteq \mathscr{R}$. To show that also $\mathscr{R} \subseteq \mathscr{R}'$, consider any $x \in \mathscr{R}$ and any $y \in \mathscr{R}'$. Due to Lemma 27, x is accessible from anywhere in $\mathscr{G}(\overline{T})$ and hence definitely from y, so there is a directed path from y to x. Let $k \in \mathbb{N}$ be the length of this path. Furthermore, since $y \in \mathscr{R}'$, there is some $k^* \in \mathbb{N}$ such that $\min \overline{T}^{k'} \mathbb{I}_y > 0$ for all $k' \geq k^*$, and therefore that $\overline{T}^{k'} \mathbb{I}_y(z) > 0$ for all $k' \geq k^*$ and all $z \in \mathscr{X}$. Fix any such $k' \geq k^*$ and any such $z \in \mathscr{X}$. Then, according to Lemma 1, there is a directed path of length k' from z to y. Hence, recalling that there is a directed path of length k from y to x, we infer that there is a directed path of length $k + k'$ from z to x, and therefore, again by Lemma 1, that $\overline{T}^{k+k'} \mathbb{I}_x(z) > 0$. Since this holds for any $k' \geq k^*$ and any $z \in \mathscr{X}$, we have that $\min \overline{T}^{k+k'} \mathbb{I}_x > 0$ for all $k' \geq k^*$, or equivalently, that $\min \overline{T}^{\ell} \mathbb{I}_x > 0$ for all $\ell \geq k + k^*$. As a result, x is an element of $\mathscr{R}'$. Since $x \in \mathscr{R}$ was chosen arbitrarily, it follows that $\mathscr{R} \subseteq \mathscr{R}'$. $\qquad \square$

Appendix B. Proof of Proposition 10

In the following, we will often use the fact that, since $\overline{T}$ is an upper transition operator, the iterates of $\overline{T}$ will also be upper transition operators. This can easily be derived using the properties U1–U7 and an induction argument in k. For an illustration of how to do so, we refer to [19, Lemma 23].

Lemma 29. *If* $\overline{T}$ *is an upper transition operator then, for any* $k \in \mathbb{N}$, $\overline{T}^k$ *is an upper transition operator as well.*

The properties U2–U7 therefore also apply to $\overline{T}^k$:

U2′. $\overline{T}^k(h + g) \leq \overline{T}^k h + \overline{T}^k g$ [sub-additivity];

U3′. $\overline{T}^k(\lambda h) = \lambda \overline{T}^k h$ [non-negative homogeneity];

U4′. $\min h \leq \overline{T}^k h \leq \max h$ [boundedness];

U5′. $\overline{T}^k(\mu + h) = \mu + \overline{T}^k h$ [constant additivity];

22

U6′. if $h \le g$ then $\overline{T}^k h \le \overline{T}^k g$ [monotonicity];

U7′. $\overline{T}^k h - \overline{T}^k g \le \overline{T}^k(h-g)$ [mixed sub-additivity],

for all $k \in \mathbb{N}_0$, all $h, g \in \mathcal{L}(\mathcal{X})$, all real μ and all real $\lambda \ge 0$.

Many of the results in this appendix will make use of the graph-theoretic concepts and notations that were defined in Section 5. Unless mentioned otherwise, we will always implicitly assume that they correspond to the graph $\mathcal{G}(\overline{T})$ of $\overline{T}$. Note however that, due to Corollary 5, we could also equivalently consider the graphs $\mathcal{G}'(\overline{T})$ or $\mathcal{G}'(\overline{T}_f)$.

Lemma 30. *For any $\overline{T}$ with a closed class $\mathcal{S}$, we have that $\overline{T}^k \mathbb{I}_{\mathcal{S}^c}(x) = 0$ for all $x \in \mathcal{S}$ and all $k \in \mathbb{N}$.*

Proof. Consider any $x \in \mathcal{S}$. Then, since $\mathcal{S}$ is closed, we have that $x \nrightarrow y$ for any $y \in \mathcal{S}^c$, which by Lemma 1 implies that $\overline{T}^k \mathbb{I}_y(x) \le 0$ for all $k \in \mathbb{N}$. Hence,

$$0 \le \overline{T}^k \mathbb{I}_{\mathcal{S}^c}(x) = \left[\overline{T}^k \left(\sum_{y \in \mathcal{S}^c} \mathbb{I}_y \right) \right](x) \le \sum_{y \in \mathcal{S}^c} \overline{T}^k \mathbb{I}_y(x) \le 0 \text{ for all } k \in \mathbb{N},$$

where the first step uses U4′ and the third uses U2′. $\qquad\square$

Lemma 31. *For any $f \in \mathcal{L}(\mathcal{X})$ and any $\overline{T}$ with a closed class $\mathcal{S}$, we have that $\overline{T}_f h(x) = \overline{T}_f(h\mathbb{I}_{\mathcal{S}})(x)$ for all $h \in \mathcal{L}(\mathcal{X})$ and all $x \in \mathcal{S}$.*

Proof. Fix any $h \in \mathcal{L}(\mathcal{X})$ and any $x \in \mathcal{S}$. By sub-additivity [U2], we have that $\overline{T} h(x) \le \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) + \overline{T}(h\mathbb{I}_{\mathcal{S}^c})(x)$. Since $h\mathbb{I}_{\mathcal{S}^c} \le \|h\|_\infty \mathbb{I}_{\mathcal{S}^c}$, monotonicity [U6] therefore implies that

$$\overline{T} h(x) \le \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) + \overline{T}(\|h\|_\infty \mathbb{I}_{\mathcal{S}^c})(x) = \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) + \|h\|_\infty \overline{T}\mathbb{I}_{\mathcal{S}^c}(x)$$
$$= \overline{T}(h\mathbb{I}_{\mathcal{S}})(x),$$

where the first equality follows from non-negative homogeneity [U3] and the second from Lemma 30. Hence, we obtain that $\overline{T}_f h(x) \le \overline{T}_f(h\mathbb{I}_{\mathcal{S}})(x)$. To prove the converse inequality, observe that

$$\overline{T} h(x) \ge \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) - \overline{T}(-h\mathbb{I}_{\mathcal{S}^c})(x) \ge \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) - \overline{T}(\|h\|_\infty \mathbb{I}_{\mathcal{S}^c})(x)$$
$$= \overline{T}(h\mathbb{I}_{\mathcal{S}})(x) - \|h\|_\infty \overline{T}\mathbb{I}_{\mathcal{S}^c}(x)$$
$$= \overline{T}(h\mathbb{I}_{\mathcal{S}})(x),$$

where the first step follows from U7, the second follows from $-h\mathbb{I}_{\mathcal{S}^c} \le \|h\|_\infty \mathbb{I}_{\mathcal{S}^c}$ and monotonicity [U6], the third follows from non-negative homogeneity [U3] and the last from Lemma 30. So, we have that $\overline{T} h(x) \ge \overline{T}(h\mathbb{I}_{\mathcal{S}})(x)$ and therefore also that $\overline{T}_f h(x) \ge \overline{T}_f(h\mathbb{I}_{\mathcal{S}})(x)$. Hence, $\overline{T}_f h(x) = \overline{T}_f(h\mathbb{I}_{\mathcal{S}})(x)$ for all $h \in \mathcal{L}(\mathcal{X})$ and all $x \in \mathcal{S}$. $\qquad\square$

To prove Proposition 10, we will use the following notations that allow us to confine the dynamics of the process to a closed class. Let $\overline{T}$ be any upper transition operator and let $\mathcal{S}$ be any non-empty subset of $\mathcal{X}$. For any $h \in \mathcal{L}(\mathcal{X})$, let $h|_{\mathcal{S}} \in \mathcal{L}(\mathcal{S})$ denote the restriction of h to the domain $\mathcal{S}$. Additionally, for any $h \in \mathcal{L}(\mathcal{S})$, we let $h^\uparrow \in \mathcal{L}(\mathcal{X})$ denote the zero-extension of h into $\mathcal{L}(\mathcal{X})$, which takes the value $h(x)$ for $x \in \mathcal{S}$ and 0 elsewhere. Then note that $(h|_{\mathcal{S}})^\uparrow = h\mathbb{I}_{\mathcal{S}}$ for any $h \in \mathcal{L}(\mathcal{X})$ and $(g^\uparrow)|_{\mathcal{S}} = g$ for any $g \in \mathcal{L}(\mathcal{S})$. Let $\overline{T}_{f,\mathcal{S}} \colon \mathcal{L}(\mathcal{S}) \to \mathcal{L}(\mathcal{S})$ be defined by $\overline{T}_{f,\mathcal{S}} h := (\overline{T}_f h^\uparrow)|_{\mathcal{S}}$ for all $h \in \mathcal{L}(\mathcal{S})$.

Lemma 32. *For any $f \in \mathcal{L}(\mathcal{X})$ and any $\overline{T}$ with a closed class $\mathcal{S}$, we have that $(\overline{T}_f^k h)|_{\mathcal{S}} = \overline{T}_{f,\mathcal{S}}^k(h|_{\mathcal{S}})$ for all $h \in \mathcal{L}(\mathcal{X})$ and all $k \in \mathbb{N}$.*

Proof. We use an induction argument in $k \in \mathbb{N}$. That the statement holds for $k = 1$ follows immediately from Lemma 31. Indeed, for any $h \in \mathscr{L}(\mathscr{X})$, Lemma 31 says that $\overline{T}_f h(x) = \overline{T}_f(h \mathbb{I}_{\mathscr{S}})(x)$ for all $x \in \mathscr{S}$ or, equivalently, that $(\overline{T}_f h)|_{\mathscr{S}} = (\overline{T}_f(h \mathbb{I}_{\mathscr{S}}))|_{\mathscr{S}}$. This implies, by the definition of $\overline{T}_{f,\mathscr{S}}$ and the fact that $h \mathbb{I}_{\mathscr{S}} = (h|_{\mathscr{S}})^{\uparrow}$, that $(\overline{T}_f h)|_{\mathscr{S}} = \overline{T}_{f,\mathscr{S}}(h|_{\mathscr{S}})$ for any $h \in \mathscr{L}(\mathscr{X})$, which provides an induction base.

Now assume that the statement holds for all $i \in \{1, \cdots, k\}$, with $k \in \mathbb{N}$. Then, for any $h \in \mathscr{L}(\mathscr{X})$, we have that

$$(\overline{T}_f^{k+1} h)|_{\mathscr{S}} = (\overline{T}_f(\overline{T}_f^k h))|_{\mathscr{S}} = \overline{T}_{f,\mathscr{S}}((\overline{T}_f^k h)|_{\mathscr{S}}) = \overline{T}_{f,\mathscr{S}}(\overline{T}_{f,\mathscr{S}}^k(h|_{\mathscr{S}})) = \overline{T}_{f,\mathscr{S}}^{k+1}(h|_{\mathscr{S}}),$$

where the second equality follows from the fact that the statement holds for $i = 1$ and the third equality follows from the assumption that the statement holds for $i = k$. Combined with the induction base, this concludes the proof. $\qquad \square$

Lemma 33. *For any $f \in \mathscr{L}(\mathscr{X})$ and any $\overline{T}$ with a closed class $\mathscr{S}$, the map $\overline{T}_{f,\mathscr{S}}$ is topical.*

Proof. To prove T1, consider any $\mu \in \mathbb{R}$ and any $h \in \mathscr{L}(\mathscr{S})$. Since $\overline{T}_f$ satisfies T1, we have that $\overline{T}_f(\mu + h^{\uparrow}) = \mu + \overline{T}_f(h^{\uparrow})$ and therefore, also that $(\overline{T}_f(\mu + h^{\uparrow}))|_{\mathscr{S}} = \mu + (\overline{T}_f h^{\uparrow})|_{\mathscr{S}} = \mu + \overline{T}_{f,\mathscr{S}} h$. Moreover, by Lemma 32, we have that $(\overline{T}_f(\mu + h^{\uparrow}))|_{\mathscr{S}} = \overline{T}_{f,\mathscr{S}}((\mu + h^{\uparrow})|_{\mathscr{S}}) = \overline{T}_{f,\mathscr{S}}(\mu + h)$, implying that T1 holds. Finally, that monotonicity [T2] holds for $\overline{T}_{f,\mathscr{S}}$ follows directly from its definition and the fact $\overline{T}_f$ is monotone. $\qquad \square$

Lemma 34. *For any $f \in \mathscr{L}(\mathscr{X})$ and any $\overline{T}$ with a closed communication class $\mathscr{S}$, the map $\overline{T}_{f,\mathscr{S}}$ has exactly one (additive) eigenvalue.*

Proof. Consider any two states x and y in $\mathscr{S}$. Then, by the definition of $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$, there is an edge from x to y in $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$ if $\lim_{\alpha \to +\infty} \overline{T}_{f,\mathscr{S}}(\alpha \mathbb{I}_y)(x) = +\infty$. Moreover, by the definition of $\overline{T}_{f,\mathscr{S}}$, we have that $\overline{T}_{f,\mathscr{S}}(\alpha \mathbb{I}_y) = (\overline{T}_f(\alpha \mathbb{I}_y)^{\uparrow})|_{\mathscr{S}} = (\overline{T}_f(\alpha \mathbb{I}_y))|_{\mathscr{S}}$ for all $\alpha \in \mathbb{R}$, where we used $\mathbb{I}_y$ to denote the indicator of y in both $\mathscr{L}(\mathscr{S})$ and $\mathscr{L}(\mathscr{X})$ depending on the domain of the considered map. Hence, there is an edge from x to y in the graph $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$ if and only if $\lim_{\alpha \to +\infty} \overline{T}_f(\alpha \mathbb{I}_y)(x) = +\infty$ or, equivalently, if and only if there is an edge from x to y in the graph $\mathscr{G}'(\overline{T}_f)$. So $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$ is identical to the restriction of the graph $\mathscr{G}'(\overline{T}_f)$ to the vertices in $\mathscr{S}$. Now, x and y are two states in the closed communication class $\mathscr{S}$ of $\mathscr{G}(\overline{T})$, so we have that $x \to y$ in $\mathscr{G}(\overline{T})$. Moreover, the directed path from x to y remains within the closed class $\mathscr{S}$, because $x \not\to z$ for any $x \in \mathscr{S}$ and any $z \in \mathscr{S}^c$. Then, since $\mathscr{G}(\overline{T})$ is identical to $\mathscr{G}'(\overline{T}_f)$ because of Corollary 5, and since $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$ is the restriction of $\mathscr{G}'(\overline{T}_f)$ to $\mathscr{S}$, we find that $x \to y$ in $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$. Since this holds for any two vertices in $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$, it follows that $\mathscr{G}'(\overline{T}_{f,\mathscr{S}})$ is strongly connected. Because $\overline{T}_{f,\mathscr{S}}$ is topical by Lemma 33, Theorem 8 then guarantees the existence of an (additive) eigenvector $h \in \mathscr{L}(\mathscr{S})$. Let $\mu \in \mathbb{R}$ be the corresponding eigenvalue. By Corollary 3, this is the only eigenvalue of $\overline{T}_{f,\mathscr{S}}$. $\qquad \square$

Lemma 35. *For any $\overline{T}$ with a closed communication class $\mathscr{S}$, we have that*

$$\lim_{k \to +\infty} \tfrac{1}{k}[\overline{T}_f^k(0)](x) = \mu,$$

for all $f \in \mathscr{L}(\mathscr{X})$ and all $x \in \mathscr{S}$, where μ is the unique eigenvalue of the map $\overline{T}_{f,\mathscr{S}}$.

Proof. Consider any $\overline{T}$ with a closed communication class $\mathscr{S}$ and fix any $f \in \mathscr{L}(\mathscr{X})$. Lemma 34 guarantees that $\overline{T}_{f,\mathscr{S}}$ has a unique eigenvalue $\mu \in \mathbb{R}$. Let $h \in \mathscr{L}(\mathscr{S})$ be a corresponding (not necessarily unique) eigenvector. Then, in a similar way as argued below Lemma 2, we deduce that $\lim_{k \to +\infty} \overline{T}_{f,\mathscr{S}}^k(h)/k = \mu$. Since $\overline{T}_{f,\mathscr{S}}$ is topical due to Lemma 33, Lemma 2 then also implies that $\lim_{k \to +\infty} \overline{T}_{f,\mathscr{S}}^k(0|_{\mathscr{S}})/k = \mu$, with 0 the zero vector in $\mathscr{L}(\mathscr{X})$. Moreover, for all $x \in \mathscr{S}$ and all $k \in \mathbb{N}_0$, due to Lemma 32, we have that $\frac{1}{k}[\overline{T}_f^k(0)](x) = \frac{1}{k}[\overline{T}_{f,\mathscr{S}}^k(0|_{\mathscr{S}})](x)$. This allows us to conclude that $\lim_{k \to +\infty} \frac{1}{k}[\overline{T}_f^k(0)](x) = \mu$ for all $x \in \mathscr{S}$, where μ is the unique eigenvalue of the map $\overline{T}_{f,\mathscr{S}}$. $\qquad \square$

Lemma 36. *For any $\overline{T}$ with a closed class $\mathscr{S}$, we have that $[\overline{T}_f^k(0)]\mathbb{I}_{\mathscr{S}} = [\overline{T}_g^k(0)]\mathbb{I}_{\mathscr{S}}$ for any two $f, g \in \mathscr{L}(\mathscr{X})$ such that $g\mathbb{I}_{\mathscr{S}} = f\mathbb{I}_{\mathscr{S}}$ and all $k \in \mathbb{N}$.*

Proof. Let $\mathscr{S}$ be a closed class of $\overline{T}$. Fix any two $f, g \in \mathscr{L}(\mathscr{X})$ such that $g\mathbb{I}_{\mathscr{S}} = f\mathbb{I}_{\mathscr{S}}$ and let $\overline{T}_f(\cdot) \coloneqq f + \overline{T}(\cdot)$ and $\overline{T}_g(\cdot) \coloneqq g + \overline{T}(\cdot)$ as before. To prove the statement, we will use an induction argument in $k \in \mathbb{N}$. That the statement holds for $k = 1$ is trivial because, due to U4, we have that $[\overline{T}_f(0)] = f$ and $[\overline{T}_g(0)] = g$. Now suppose that the statement holds for $k = i$ with $i \in \mathbb{N}$. Then, by assumption and by Equation (2), we have that $\tilde{m}_{f,i-1}\mathbb{I}_{\mathscr{S}} = [\overline{T}_f^i(0)]\mathbb{I}_{\mathscr{S}} = [\overline{T}_g^i(0)]\mathbb{I}_{\mathscr{S}} = \tilde{m}_{g,i-1}\mathbb{I}_{\mathscr{S}}$. This allows us to write that, for any $x \in \mathscr{S}$,

$$\tilde{m}_{f,i}(x) = \overline{T}_f\tilde{m}_{f,i-1}(x) = \overline{T}_f(\tilde{m}_{f,i-1}\mathbb{I}_{\mathscr{S}})(x) = \overline{T}_f(\tilde{m}_{g,i-1}\mathbb{I}_{\mathscr{S}})(x)$$
$$= \overline{T}_f\tilde{m}_{g,i-1}(x), \tag{B.1}$$

where the second and last step follow from Lemma 31. Moreover, note that, for any $x \in \mathscr{S}$,

$$\overline{T}_f\tilde{m}_{g,i-1}(x) = (f + \overline{T}\tilde{m}_{g,i-1})(x) = f(x) + \overline{T}\tilde{m}_{g,i-1}(x) = g(x) + \overline{T}\tilde{m}_{g,i-1}(x)$$
$$= \overline{T}_g\tilde{m}_{g,i-1}(x)$$
$$= \tilde{m}_{g,i}(x)$$

where the third step follows from $f\mathbb{I}_{\mathscr{S}} = g\mathbb{I}_{\mathscr{S}}$. Hence, recalling Equation (B.1), we have that $\tilde{m}_{f,i}(x) = \tilde{m}_{g,i}(x)$ for all $x \in \mathscr{S}$, which, due to Equation (2), implies that $[\overline{T}_f^{i+1}(0)]\mathbb{I}_{\mathscr{S}} = \tilde{m}_{f,i}\mathbb{I}_{\mathscr{S}} = \tilde{m}_{g,i}\mathbb{I}_{\mathscr{S}} = [\overline{T}_g^{i+1}(0)]\mathbb{I}_{\mathscr{S}}$ and therefore that the statement holds for $k = i + 1$. $\qquad\square$

Proof of Proposition 10. Consider any closed communication class $\mathscr{S}$ of $\overline{T}$, any $f \in \mathscr{L}(\mathscr{X})$ and any $x \in \mathscr{S}$. It follows from Lemma 36 and Equation (2) that, for any $k \in \mathbb{N}$,

$$\overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \tfrac{1}{k}[\overline{T}_f^k(0)](x) = \tfrac{1}{k}[\overline{T}_{f\mathbb{I}_{\mathscr{S}}}^k(0)](x) = \overline{\mathrm{E}}_{\mathrm{av},k}(f\mathbb{I}_{\mathscr{S}}|x).$$

Moreover, by Lemma 35, we have that $\lim_{k\to+\infty} \tfrac{1}{k}[\overline{T}_f^k(0)](x) = \mu$ for some $\mu \in \mathbb{R}$ and all $x \in \mathscr{S}$, which implies that the upper expectation $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ converges to a constant μ that does not depend on the specific state $x \in \mathscr{S}$. $\quad\square$

Appendix C. Proof of Proposition 11

Lemma 37. *For any $\overline{T}$ with a closed class $\mathscr{S}$, the function $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}$ is non-increasing in $k \in \mathbb{N}$.*

Proof. From Lemma 30, it follows that $\overline{T}\mathbb{I}_{\mathscr{S}^c}(x) = 0$ for all $x \in \mathscr{S}$. Since moreover $0 \le \overline{T}\mathbb{I}_{\mathscr{S}^c} \le 1$ by U4, it follows that $\mathbb{I}_{\mathscr{S}^c} \ge \overline{T}\mathbb{I}_{\mathscr{S}^c}$. Then, using U6′, we deduce that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c} \ge \overline{T}^{k+1}\mathbb{I}_{\mathscr{S}^c}$ for all $k \in \mathbb{N}$. $\qquad\square$

Lemma 38. *For any $\overline{T}$ with an absorbing closed class $\mathscr{S}$, we have that $\lim_{k\to+\infty}\overline{T}^k\mathbb{I}_{\mathscr{S}^c} = 0$.*

Proof. The statement holds if $\mathscr{S}^c = \emptyset$ because then $\mathbb{I}_{\mathscr{S}^c} = 0$, which by U4′ implies that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c} = 0$ for all $k \in \mathbb{N}_0$. So assume that $\mathscr{S}^c$ is non-empty. Since $\mathscr{S}$ is absorbing [A2], there is, for any $x \in \mathscr{S}^c$, an index $k_x \in \mathbb{N}$ such that $\overline{T}^{k_x}\mathbb{I}_{\mathscr{S}^c}(x) < 1$. By Lemma 37 and the fact that $\mathscr{S}$ is closed, we then also have that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}(x) < 1$ for all $x \in \mathscr{S}^c$ and all $k \ge k_x$. Hence, for $k \coloneqq \max_{x \in \mathscr{S}^c} k_x \in \mathbb{N}$, we have that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}(x) < 1$ for all $x \in \mathscr{S}^c$. Now let $\alpha \coloneqq \max_{x \in \mathscr{S}^c} \overline{T}^k\mathbb{I}_{\mathscr{S}^c}(x) < 1$. The set $\mathscr{S}$ is a closed class, so it follows from Lemma 30 that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}(x) = 0$ for all $x \in \mathscr{S}$. Since moreover $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}(x) \le \alpha$ for all $x \in \mathscr{S}^c$, we infer that $\overline{T}^k\mathbb{I}_{\mathscr{S}^c} \le \alpha\mathbb{I}_{\mathscr{S}^c}$. Then, using U6′, U3′ and the non-negativity [U4′] of α, it follows that $\overline{T}^{2k}\mathbb{I}_{\mathscr{S}^c} \le \alpha^2\mathbb{I}_{\mathscr{S}^c}$. Repeating this argument leads us to conclude that $\overline{T}^{\ell k}\mathbb{I}_{\mathscr{S}^c} \le \alpha^\ell\mathbb{I}_{\mathscr{S}^c}$ for all $\ell \in \mathbb{N}$. Since α is a non-negative real such that $\alpha < 1$, this implies that $\lim_{\ell\to+\infty}\overline{T}^{\ell k}\mathbb{I}_{\mathscr{S}^c} \le 0$ and therefore by U4′ that $\lim_{\ell\to+\infty}\overline{T}^{\ell k}\mathbb{I}_{\mathscr{S}^c} = 0$. Then also $\lim_{k\to+\infty}\overline{T}^k\mathbb{I}_{\mathscr{S}^c} = 0$ because $\overline{T}^k\mathbb{I}_{\mathscr{S}^c}$ is non-increasing according to Lemma 37. $\qquad\square$

Lemma 39. *Consider any $\overline{T}$ with an absorbing closed class $\mathscr{S}$. Then, for any $\epsilon > 0$, there is a $k_1 \in \mathbb{N}_0$ such that* $\|\overline{T}^k h - \overline{T}^k(h\mathbb{I}_{\mathscr{S}})\|_\infty \leq \|h\|_\infty \epsilon$ *for all $k \geq k_1$ and all $h \in \mathscr{L}(\mathscr{X})$.*

Proof. Fix any $\epsilon > 0$. Because of Lemma 38, we have that $\lim_{k \to +\infty} \overline{T}^k \mathbb{I}_{\mathscr{S}^c} = 0$. Since $\mathscr{X}$ is finite, this implies that there is an index $k_1 \in \mathbb{N}_0$ such that $0 \leq \overline{T}^k \mathbb{I}_{\mathscr{S}^c} \leq \epsilon$ for all $k \geq k_1$, where we also used the non-negativity [U4'] of $\overline{T}^k \mathbb{I}_{\mathscr{S}^c}$. Hence, multiplying by $\|h\|_\infty$ for any $h \in \mathscr{L}(\mathscr{X})$ and using non-negative homogeneity [U3'], allows us to write that $0 \leq \overline{T}^k(\|h\|_\infty \mathbb{I}_{\mathscr{S}^c}) \leq \|h\|_\infty \epsilon$ for all $k \geq k_1$ and all $h \in \mathscr{L}(\mathscr{X})$. Moreover, note that

$$h\mathbb{I}_{\mathscr{S}} - \|h\|_\infty \mathbb{I}_{\mathscr{S}^c} \leq h \leq h\mathbb{I}_{\mathscr{S}} + \|h\|_\infty \mathbb{I}_{\mathscr{S}^c} \text{ for all } h \in \mathscr{L}(\mathscr{X}).$$

Then, by subsequently applying U7', U6' and U2', we find that

$$\begin{aligned}
\overline{T}^k(h\mathbb{I}_{\mathscr{S}}) - \overline{T}^k(\|h\|_\infty \mathbb{I}_{\mathscr{S}^c}) &\leq \overline{T}^k(h\mathbb{I}_{\mathscr{S}} - \|h\|_\infty \mathbb{I}_{\mathscr{S}^c}) \\
&\leq \overline{T}^k h \leq \overline{T}^k(h\mathbb{I}_{\mathscr{S}}) + \overline{T}^k(\|h\|_\infty \mathbb{I}_{\mathscr{S}^c}),
\end{aligned} \tag{C.1}$$

for all $h \in \mathscr{L}(\mathscr{X})$ and all $k \in \mathbb{N}_0$. Hence, recalling that $0 \leq \overline{T}^k(\|h\|_\infty \mathbb{I}_{\mathscr{S}^c}) \leq \|h\|_\infty \epsilon$ for all $k \geq k_1$ and all $h \in \mathscr{L}(\mathscr{X})$, we indeed find that $\|\overline{T}^k h - \overline{T}^k(h\mathbb{I}_{\mathscr{S}})\|_\infty \leq \|h\|_\infty \epsilon$ for all $k \geq k_1$ and all $h \in \mathscr{L}(\mathscr{X})$. $\square$

For notational convenience, we will henceforth use $\overline{m}_{f,k} := \frac{1}{k} \tilde{m}_{f,k}$ for any $f \in \mathscr{L}(\mathscr{X})$ and any $k \in \mathbb{N}$, to denote the function in $\mathscr{L}(\mathscr{X})$ that takes the value $\overline{m}_{f,k}(x) = \frac{1}{k} \tilde{m}_{f,k}(x) = \overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ in $x \in \mathscr{X}$.

Lemma 40. $\min f \leq \overline{m}_{f,k}(x) = \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) \leq \max f$ *for all $f \in \mathscr{L}(\mathscr{X})$, all $k \in \mathbb{N}$ and all $x \in \mathscr{X}$.*

Proof. Fix any $f \in \mathscr{L}(\mathscr{X})$. Due to the definition of $\overline{m}_{f,k}$, it clearly suffices to prove that $k \min f \leq \tilde{m}_{f,k} \leq k \max f$ for all $k \in \mathbb{N}$. We do this by induction. For $k = 1$, the statement holds trivially because $\tilde{m}_{f,1} = f$ and therefore $\min f \leq \tilde{m}_{f,1} \leq \max f$. Now suppose that the statement holds for $k = i$ with $i \in \mathbb{N}$. Then $i \min f \leq \tilde{m}_{f,i} \leq i \max f$. It then follows from U4 and U6 that

$$i \min f = \overline{T}(i \min f) \leq \overline{T} \tilde{m}_{f,i} \leq \overline{T}(i \max f) = i \max f.$$

By adding f to all the terms, we find that $(i+1) \min f \leq f + \overline{T} \tilde{m}_{f,i} \leq (i+1) \max f$. Since $f + \overline{T} \tilde{m}_{f,i} = \overline{T}_f \tilde{m}_{f,i} = \tilde{m}_{f,i+1}$ by Equation (1), it follows that the statement holds for $k = i + 1$ as well. $\square$

Lemma 41. $\|\overline{T}^k h - \overline{T}_f^k h\|_\infty \leq k\|f\|_\infty$ *for all $f \in \mathscr{L}(\mathscr{X})$, all $h \in \mathscr{L}(\mathscr{X})$ and all $k \in \mathbb{N}_0$.*

Proof. Fix any $f \in \mathscr{L}(\mathscr{X})$. It suffices to show that

$$-k\|f\|_\infty + \overline{T}^k h \leq \overline{T}_f^k h \leq k\|f\|_\infty + \overline{T}^k h \text{ for all } h \in \mathscr{L}(\mathscr{X}) \text{ and all } k \in \mathbb{N}_0. \tag{C.2}$$

To do so, we will use an induction argument in k. The inequalities above clearly hold for all $h \in \mathscr{L}(\mathscr{X})$ and $k = 0$. Now suppose that they hold for all $h \in \mathscr{L}(\mathscr{X})$ and all $k \in \{0, \cdots, i\}$, with $i \in \mathbb{N}_0$. Then we have that

$$\overline{T}_f^{i+1} h \leq \overline{T}_f(i\|f\|_\infty + \overline{T}^i h) = i\|f\|_\infty + \overline{T}_f(\overline{T}^i h) \leq (i+1)\|f\|_\infty + \overline{T}^{(i+1)} h,$$

for all $h \in \mathscr{L}(\mathscr{X})$, where the first step follows from the induction hypothesis for $k = i$ and the monotonicity [T2] of $\overline{T}_f$, the second from the constant additivity [T1] of $\overline{T}_f$, and the third from the definition of $\overline{T}_f$. In an analogous way, we find that

$$\begin{aligned}
\overline{T}_f^{i+1} h &\geq \overline{T}_f(-i\|f\|_\infty + \overline{T}^i h) = -i\|f\|_\infty + \overline{T}_f(\overline{T}^i h) \\
&\geq -(i+1)\|f\|_\infty + \overline{T}^{(i+1)} h,
\end{aligned}$$

26

for all $h \in \mathcal{L}(\mathcal{X})$, where the first step follows once more from the induction hypothesis for $k = i$ and the monotonicity [T2] of $\overline{T}_f$, the second from the constant additivity [T1] of $\overline{T}_f$, and the third from the definition of $\overline{T}_f$. Both inequalities together establish that the statement holds for $k = i + 1$, thereby concluding the induction step. $\qquad\square$

Lemma 42. $\lim_{k \to +\infty} \|\overline{T}^\ell \overline{m}_{f,k} - \overline{m}_{f,k+\ell}\|_\infty = 0$ *for all* $f \in \mathcal{L}(\mathcal{X})$ *and all* $\ell \in \mathbb{N}_0$.

Proof. Fix any $f \in \mathcal{L}(\mathcal{X})$, any $\ell \in \mathbb{N}_0$ and any $\epsilon > 0$. Let $k \in \mathbb{N}$ be such that $k \geq \ell \|f\|_\infty / \epsilon$ and let $h \coloneqq \overline{T}_f^k(0) = k\overline{m}_{f,k}$. Then

$$\epsilon \geq \tfrac{\ell}{k} \|f\|_\infty \geq \tfrac{1}{k} \big\| \overline{T}^\ell h - \overline{T}_f^\ell h \big\|_\infty = \big\| \overline{T}^\ell \tfrac{1}{k} h - \tfrac{1}{k} \overline{T}_f^\ell h \big\|_\infty \\ = \big\| \overline{T}^\ell \overline{m}_{f,k} - \tfrac{1}{k} \overline{T}_f^\ell h \big\|_\infty, \tag{C.3}$$

where the second step follows from Lemma 41 and the third follows from the non-negative homogeneity [U3′] of $\overline{T}^\ell$. Moreover, we also have that

$$\big\| \tfrac{1}{k} \overline{T}_f^\ell h - \overline{m}_{f,k+\ell} \big\|_\infty = \big\| \tfrac{k+\ell}{k} \overline{m}_{f,k+\ell} - \overline{m}_{f,k+\ell} \big\|_\infty = \tfrac{\ell}{k} \|\overline{m}_{f,k+\ell}\|_\infty \\ \leq \tfrac{\ell}{k} \|f\|_\infty \leq \epsilon,$$

where the second to last step follows from Lemma 40. Combining this with Equation (C.3) and using the triangle inequality, we get that $\|\overline{T}^\ell \overline{m}_{f,k} - \overline{m}_{f,k+\ell}\|_\infty \leq 2\epsilon$. Since this holds for any $\epsilon > 0$ and all $k \geq \ell \|f\|_\infty / \epsilon$, we indeed have that $\lim_{k \to +\infty} \|\overline{T}^\ell \overline{m}_{f,k} - \overline{m}_{f,k+\ell}\|_\infty = 0$. $\qquad\square$

Proposition 43. *Consider any* $\overline{T}$ *with a closed class* $\mathcal{S}$ *that is absorbing. Then, for any* $f \in \mathcal{L}(\mathcal{X})$ *and any* $y \in \mathcal{X}$, *we have that*

$$\min_{x \in \mathcal{S}} \liminf_{k \to +\infty} \tfrac{1}{k} [\overline{T}_f^k(0)](x) \leq \liminf_{k \to +\infty} \tfrac{1}{k} [\overline{T}_f^k(0)](y) \leq \limsup_{k \to +\infty} \tfrac{1}{k} [\overline{T}_f^k(0)](y) \leq \max_{x \in \mathcal{S}} \limsup_{k \to +\infty} \tfrac{1}{k} [\overline{T}_f^k(0)](x).$$

Proof. First note that the proposition is trivially satisfied if $f = 0$ because Lemma 40 and Equation (2) then imply that $\tfrac{1}{k} [\overline{T}_f^k(0)](x) = \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = 0$ for all $x \in \mathcal{X}$ and $k \in \mathbb{N}$. So suppose that $f \neq 0$, fix any $\epsilon > 0$ and let $\epsilon_1 \coloneqq (1/\|f\|_\infty)\epsilon$. Choose ℓ_1 such that Lemma 39 holds with ϵ_1; that is, in such a way that

$$\|\overline{T}^\ell(h\mathbb{I}_\mathcal{S}) - \overline{T}^\ell h\|_\infty \leq \epsilon_1 \|h\|_\infty \text{ for all } \ell \geq \ell_1 \text{ and all } h \in \mathcal{L}(\mathcal{X}). \tag{C.4}$$

Then, in particular, for any $\ell \geq \ell_1$, we have that $\big\| \overline{T}^\ell(\overline{m}_{f,k}\mathbb{I}_\mathcal{S}) - \overline{T}^\ell \overline{m}_{f,k} \big\|_\infty \leq \epsilon_1 \|\overline{m}_{f,k}\|_\infty \leq \epsilon_1 \|f\|_\infty = \epsilon$ for all $k \in \mathbb{N}$, where the second inequality follows from Lemma 40. Fix any such $\ell \geq \ell_1$. Furthermore, recall Lemma 42, which guarantees that there is some $k_1 \in \mathbb{N}$ such that $\|\overline{T}^\ell \overline{m}_{f,k} - \overline{m}_{f,k+\ell}\|_\infty \leq \epsilon$ for all $k \geq k_1$. Combining this with the inequality above, we get that $\|\overline{T}^\ell(\overline{m}_{f,k}\mathbb{I}_\mathcal{S}) - \overline{m}_{f,k+\ell}\|_\infty \leq 2\epsilon$ for all $k \geq k_1$ or, equivalently, that

$$\overline{T}^\ell(\overline{m}_{f,k}\mathbb{I}_\mathcal{S}) - 2\epsilon \leq \overline{m}_{f,k+\ell} \leq \overline{T}^\ell(\overline{m}_{f,k}\mathbb{I}_\mathcal{S}) + 2\epsilon \text{ for all } k \geq k_1. \tag{C.5}$$

Now, as a consequence of the definitions of the limit inferior and the limit superior operators and the fact that $\overline{m}_{f,k}$ is bounded due to Lemma 40, there is, for any $x \in \mathcal{S}$, an index $k_x \in \mathbb{N}$ such that

$$\liminf_{k' \to +\infty} \overline{m}_{f,k'}(x) - \epsilon \leq \overline{m}_{f,k}(x) \leq \limsup_{k' \to +\infty} \overline{m}_{f,k'}(x) + \epsilon \text{ for all } k \geq k_x.$$

Then, if we let $a \coloneqq \min_{x \in \mathcal{S}} \liminf_{k' \to +\infty} \overline{m}_{f,k'}(x)$ and $b \coloneqq \max_{x \in \mathcal{S}} \limsup_{k' \to +\infty} \overline{m}_{f,k'}(x)$, we certainly have that

$a - \epsilon \leq \overline{m}_{f,k}(x) \leq b + \epsilon$ for any $x \in \mathscr{S}$ and all $k \geq k_x$. Hence,

$$a\mathbb{I}_{\mathscr{S}} - \epsilon \leq \overline{m}_{f,k}\mathbb{I}_{\mathscr{S}} \leq b\mathbb{I}_{\mathscr{S}} + \epsilon \text{ for all } k \geq k_2 := \max_{x \in \mathscr{S}} k_x \in \mathbb{N},$$

which by U6$'$ and U5$'$ implies that

$$\overline{T}^\ell(a\mathbb{I}_{\mathscr{S}}) - \epsilon \leq \overline{T}^\ell(\overline{m}_{f,k}\mathbb{I}_{\mathscr{S}}) \leq \overline{T}^\ell(b\mathbb{I}_{\mathscr{S}}) + \epsilon \text{ for all } k \geq k_2.$$

Together with Equation (C.5), this implies that

$$\overline{T}^\ell(a\mathbb{I}_{\mathscr{S}}) - 3\epsilon \leq \overline{m}_{f,k+\ell} \leq \overline{T}^\ell(b\mathbb{I}_{\mathscr{S}}) + 3\epsilon \text{ for all } k \geq \max\{k_1, k_2\}. \tag{C.6}$$

Now, recall Equation (C.4) and observe that therefore

$$\left\| \overline{T}^\ell(a\mathbb{I}_{\mathscr{S}}) - a \right\|_\infty = \left\| \overline{T}^\ell(a\mathbb{I}_{\mathscr{S}}) - \overline{T}^\ell a \right\|_\infty \leq \epsilon_1 \|a\|_\infty,$$

where the first step follows from U4$'$. Moreover, we have that $\|a\|_\infty = |a| \leq \|f\|_\infty$ due to Lemma 40, so we conclude that $\|\overline{T}^\ell(a\mathbb{I}_{\mathscr{S}}) - a\|_\infty \leq \epsilon_1 \|f\|_\infty = \epsilon$. In a completely analogous way, we can deduce that $\|\overline{T}^\ell(b\mathbb{I}_{\mathscr{S}}) - b\|_\infty \leq \epsilon$. So, in particular, we have that $a - \epsilon \leq \overline{T}^\ell(a\mathbb{I}_{\mathscr{S}})$ and that $\overline{T}^\ell(b\mathbb{I}_{\mathscr{S}}) \leq b + \epsilon$. Applying these inequalities to Equation (C.6), we obtain that $a - 4\epsilon \leq \overline{m}_{f,k+\ell} \leq b + 4\epsilon$ for all $k \geq \max\{k_1, k_2\}$. Hence,

$$a - 4\epsilon \leq \liminf_{k \to +\infty} \overline{m}_{f,k} \leq \limsup_{k \to +\infty} \overline{m}_{f,k} \leq b + 4\epsilon$$

Since this holds for any $\epsilon > 0$, we conclude that $a \leq \liminf_{k \to +\infty} \overline{m}_{f,k} \leq \limsup_{k \to +\infty} \overline{m}_{f,k} \leq b$, which by the definition of a and b, and the fact that, due to Equation (2), $\overline{m}_{f,k} = \frac{1}{k}[\overline{T}_f^k(0)]$ for all $k \in \mathbb{N}_0$, implies the desired statement. $\qquad\square$

Proof of Proposition 11. Assume that $\overline{T}$ satisfies (TCA) and let $\mathscr{R}$ be the corresponding top class. Fix any $f \in \mathscr{L}(\mathscr{X})$. According to Lemma 25, $\mathscr{R}$ is a closed communication class, which by Proposition 10 implies that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \lim_{k \to +\infty} \overline{m}_{f,k}(x) = \mu$ for some $\mu \in \mathbb{R}$ and all $x \in \mathscr{R}$. Since $\mathscr{R}$ is closed and absorbing [A2], we can then apply Proposition 43—taking into account that $\frac{1}{k}[\overline{T}_f^k(0)] = \overline{m}_{f,k}$ for all $k \in \mathbb{N}$ due to Equation (2)—to find that, for all $y \in \mathscr{X}$, $\mu \leq \liminf_{k \to +\infty} \overline{m}_{f,k}(y) \leq \limsup_{k \to +\infty} \overline{m}_{f,k}(y) \leq \mu$. Hence, we have that $\lim_{k \to +\infty} \overline{m}_{f,k} = \mu$ and conclude that $\overline{T}$ is weakly ergodic. $\qquad\square$

Appendix D. Proof of Propositions 12 and 13

In the following proofs, as well as in the proofs of Appendix E, we will often implicitly use the fact that any given top class $\mathscr{R}$ is necessarily closed; see Lemma 25.

Proof of Proposition 12. Suppose that $\overline{T}$ does not have a top class. Then due to Corollary 26, there are (at least) two closed communication classes $\mathscr{S}_1$ and $\mathscr{S}_2$. Consider any two $c_1, c_2 \in \mathbb{R}$ such that $c_1 \neq c_2$, and let $f := c_1\mathbb{I}_{\mathscr{S}_1} + c_2\mathbb{I}_{\mathscr{S}_2}$. Since $f\mathbb{I}_{\mathscr{S}_1} = c_1\mathbb{I}_{\mathscr{S}_1}$, Lemma 36 implies that $\frac{1}{k}[\overline{T}_f^k(0)]\mathbb{I}_{\mathscr{S}_1} = \frac{1}{k}[\overline{T}_{c_1}^k(0)]\mathbb{I}_{\mathscr{S}_1}$ for all $k \in \mathbb{N}$ or, by Equation (2), that $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \overline{\mathrm{E}}_{\mathrm{av},k}(c_1|x)$ for all $x \in \mathscr{S}_1$ and all $k \in \mathbb{N}$. By Lemma 40, we know that for any $x \in \mathscr{S}_1$ and any $k \in \mathbb{N}$, $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \overline{\mathrm{E}}_{\mathrm{av},k}(c_1|x) = c_1$. Hence, $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = c_1$ for all $x \in \mathscr{S}_1$. In a completely analogous way, we can deduce that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = c_2$ for all $x \in \mathscr{S}_2$. By assumption, $c_1 \neq c_2$, so we can conclude that the upper expectation $\overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ with $f = c_1\mathbb{I}_{\mathscr{S}_1} + c_2\mathbb{I}_{\mathscr{S}_2}$, does not converge to a constant that is equal for all $x \in \mathscr{X}$. Hence, the upper transition operator $\overline{T}$ is not weakly ergodic. $\qquad\square$

28

Lemma 44. *Consider any $\overline{T}$ that has a top class $\mathscr{R}$ but that does not satisfy (TCA). Then there is a non-empty subset $A \subseteq \mathscr{R}^c$ such that $\mathbb{I}_A \leq \overline{T}\mathbb{I}_A$.*

Proof. If $\overline{T}$ has a top class $\mathscr{R}$ that is not absorbing, then there is at least one $x \in \mathscr{R}^c$ such that $\overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) = 1$ for all $k \in \mathbb{N}$. Let $A \subseteq \mathscr{R}^c$ be the set of all such states $x \in \mathscr{R}^c$. If $A = \mathscr{R}^c$ then, since $\overline{T}\mathbb{I}_{\mathscr{R}^c}(x) = 1$ for all $x \in A = \mathscr{R}^c$ and since $\overline{T}\mathbb{I}_{\mathscr{R}^c}(x) = 0$ for all $x \in \mathscr{R}$ by Lemma 30, we have that $\overline{T}\mathbb{I}_{\mathscr{R}^c} = \mathbb{I}_{\mathscr{R}^c}$. Hence, in that case, the statement holds. In the remainder of the proof, we can therefore assume that $A \subset \mathscr{R}^c$, implying that $\mathscr{R}^c \setminus A$ is non-empty.

Observe that by the definition of A there is for any $x \in \mathscr{R}^c \setminus A$, an index $k_x \in \mathbb{N}$ such that $\overline{T}^{k_x}\mathbb{I}_{\mathscr{R}^c}(x) \neq 1$, and therefore, due to U4′, also that $\overline{T}^{k_x}\mathbb{I}_{\mathscr{R}^c}(x) < 1$. By Lemma 37, it follows that then also $\overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) < 1$ for all $x \in \mathscr{R}^c \setminus A$ and all $k \geq k_x$. Hence, for $k := \max_{x \in \mathscr{R}^c \setminus A} k_x \in \mathbb{N}$, we have that $\overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) < 1$ for all $x \in \mathscr{R}^c \setminus A$. Let $\alpha := \max_{x \in \mathscr{R}^c \setminus A} \overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) < 1$. Since $\overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) = 0$ for all $x \in \mathscr{R}$ due to Lemma 30 and $0 \leq \alpha$ due to U4′, we infer that $\overline{T}^k\mathbb{I}_{\mathscr{R}^c}(x) \leq \alpha$ for all $x \in \mathscr{R} \cup (\mathscr{R}^c \setminus A) = A^c$ or, equivalently, that $\mathbb{I}_{A^c}\overline{T}^k\mathbb{I}_{\mathscr{R}^c} \leq \alpha\mathbb{I}_{A^c}$. Hence,

$$\overline{T}^k\mathbb{I}_{\mathscr{R}^c} = \mathbb{I}_A\overline{T}^k\mathbb{I}_{\mathscr{R}^c} + \mathbb{I}_{A^c}\overline{T}^k\mathbb{I}_{\mathscr{R}^c} \leq \mathbb{I}_A\overline{T}^k\mathbb{I}_{\mathscr{R}^c} + \alpha\mathbb{I}_{A^c} = \mathbb{I}_A + \alpha\mathbb{I}_{A^c},$$

using the definition of A for the last equality. It follows that

$$\overline{T}^{k+1}\mathbb{I}_{\mathscr{R}^c} = \overline{T}\,\overline{T}^k\mathbb{I}_{\mathscr{R}^c} \leq \overline{T}\big(\mathbb{I}_A + \alpha\mathbb{I}_{A^c}\big) = \overline{T}\big(\alpha + (1-\alpha)\mathbb{I}_A\big) = \alpha + (1-\alpha)\overline{T}\mathbb{I}_A,$$

where we used monotonicity [U6] for the inequality and U5 and U3 for the last equality. Multiplying with $\mathbb{I}_A$ yields $\mathbb{I}_A\overline{T}^{k+1}\mathbb{I}_{\mathscr{R}^c} \leq \alpha\mathbb{I}_A + (1-\alpha)\mathbb{I}_A\overline{T}\mathbb{I}_A$ and therefore, since the definition of A implies that $\mathbb{I}_A = \mathbb{I}_A\overline{T}^{k+1}\mathbb{I}_{\mathscr{R}^c}$, we find that $\mathbb{I}_A \leq \alpha\mathbb{I}_A + (1-\alpha)\mathbb{I}_A\overline{T}\mathbb{I}_A$, or equivalently, that $(1-\alpha)\mathbb{I}_A \leq (1-\alpha)\mathbb{I}_A\overline{T}\mathbb{I}_A$. Since $1-\alpha > 0$, it follows that $\mathbb{I}_A \leq \mathbb{I}_A\overline{T}\mathbb{I}_A$, which implies that $\mathbb{I}_A \leq \overline{T}\mathbb{I}_A$ because $\overline{T}\mathbb{I}_A$ is non-negative [U4]. □

Proof of Proposition 13. Suppose that $\overline{T}$ has a top class $\mathscr{R}$, but that $\mathscr{R}$ is not absorbing. Then Lemma 44 guarantees that there is a non-empty subset $A \subseteq \mathscr{R}^c$ such that $\mathbb{I}_A \leq \overline{T}\mathbb{I}_A$. Now consider any $x \in A$ and any $y \in \mathscr{R}$. We will show that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|x) = 1$ and that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|y) = 0$, implying that $\overline{T}$ cannot be weakly ergodic.

To prove that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|x) = 1$, we first show by induction that $\tilde{m}_{\mathbb{I}_A,k} \geq k\mathbb{I}_A$ for all $k \in \mathbb{N}$. By definition, we have that $\tilde{m}_{\mathbb{I}_A,1} = \mathbb{I}_A$, which establishes our induction base. To prove the induction step, assume that the inequality holds for $k = i$, with $i \in \mathbb{N}$, so $\tilde{m}_{\mathbb{I}_A,i} \geq i\mathbb{I}_A$. Then according to the recursive expression (1),

$$\tilde{m}_{\mathbb{I}_A,i+1} = \mathbb{I}_A + \overline{T}\tilde{m}_{\mathbb{I}_A,i} \geq \mathbb{I}_A + \overline{T}(i\mathbb{I}_A) \geq (i+1)\mathbb{I}_A,$$

where the second step follows from the induction hypothesis and the monotonicity [U6] of $\overline{T}$, and the last from U3 together with the fact that $\mathbb{I}_A \leq \overline{T}\mathbb{I}_A$. This implies that the inequality holds for $k = i+1$ as well, hence finalising our induction argument. Taking into account Equation (2), we conclude that $\overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|x) = \frac{1}{k}\tilde{m}_{\mathbb{I}_A,k}(x) \geq \mathbb{I}_A(x)$ for all $k \in \mathbb{N}$. Due to Lemma 40 and since $x \in A$, this implies that $\overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|x) = 1$ for all $k \in \mathbb{N}$. Hence, $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|x) = 1$.

It remains to prove that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|y) = 0$. Because $A \subseteq \mathscr{R}^c$, we have that $\mathbb{I}_A\mathbb{I}_{\mathscr{R}} = 0 = 0\mathbb{I}_{\mathscr{R}}$, and since $\mathscr{R}$ is a closed communication class, Lemma 36 implies that $\frac{1}{k}[\overline{T}^k_{\mathbb{I}_A}(0)]\mathbb{I}_{\mathscr{R}} = \frac{1}{k}[\overline{T}^k_0(0)]\mathbb{I}_{\mathscr{R}}$ for all $k \in \mathbb{N}$. Hence, by Equation (2), we have that $\overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|y) = \overline{\mathrm{E}}_{\mathrm{av},k}(0|y)$ for all $k \in \mathbb{N}$, and therefore, by Lemma 40, that $\overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|y) = 0$ for all $k \in \mathbb{N}$. As a consequence, $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(\mathbb{I}_A|y) = 0$. □

Appendix E. Proof of the results in Section 8

For any upper transition operator $\overline{T}$, any transition matrix T, any $f \in \mathscr{L}(\mathscr{X})$ and any non-empty $\mathscr{S} \subseteq \mathscr{X}$, we define the maps $T_{f,\mathscr{S}}$, $\overline{T}_{\mathscr{S}}$ and $T_{\mathscr{S}}$ on $\mathscr{L}(\mathscr{S})$ similarly to how we previously defined the map $\overline{T}_{f,\mathscr{S}}$; so $T_{f,\mathscr{S}}h :=$

$(T_f h^\uparrow)|_{\mathscr{S}}$, $\overline{T}_{\mathscr{S}} h := (\overline{T} h^\uparrow)|_{\mathscr{S}}$ and $T_{\mathscr{S}} h := (T h^\uparrow)|_{\mathscr{S}}$ for all $h \in \mathscr{L}(\mathscr{S})$. In the following statements, we implicitly assume that $\mathscr{T}$ is a separately specified set of transition matrices, that $\overline{T}$ is the associated upper transition operator and that f is any function in $\mathscr{L}(\mathscr{X})$.

Lemma 45. *For any closed class $\mathscr{S}$ in $\mathscr{G}(\overline{T})$, we have that $\{T_{\mathscr{S}} : T \in \mathscr{T}\}$ is a separately specified set of $|\mathscr{S}| \times |\mathscr{S}|$ transition matrices and that $\overline{T}_{\mathscr{S}}$ is the corresponding upper transition operator.*

Proof. It follows trivially from its definition that, for any $T \in \mathscr{T}$, the map $T_{\mathscr{S}}$ is represented by a $|\mathscr{S}| \times |\mathscr{S}|$ matrix where the (i,j)-th element is equal to the (i,j)-th element in T for all $i, j \in \{1, \cdots, |\mathscr{S}|\}$. So, to ensure that $T_{\mathscr{S}}$ is row-stochastic—and can therefore be called a transition matrix—we only have to check whether the elements in each row sum to 1. To see that this is indeed the case, note that, due to Lemma 17 and the fact that $\mathscr{S}$ is closed in $\mathscr{G}(\overline{T})$, the class $\mathscr{S}$ is closed in $\mathscr{G}(T)$. Then Lemma 30 implies that $T\mathbb{I}_{\mathscr{S}^c}(x) = 0$ for all $x \in \mathscr{S}$ and therefore, since T is a transition matrix (which is linear), that $T\mathbb{I}_{\mathscr{S}}(x) = \sum_{y \in \mathscr{S}} T(x,y) = 1$ for all $x \in \mathscr{S}$. Hence, recalling our previous considerations, we also have that $\sum_{y \in \mathscr{S}} T_{\mathscr{S}}(x,y) = 1$ for all $x \in \mathscr{S}$, which allows us to call $T_{\mathscr{S}}$ a transition matrix. Furthermore, that $\{T_{\mathscr{S}} : T \in \mathscr{T}\}$ is separately specified follows immediately from the fact that $\mathscr{T}$ is separately specified and that, as explained above, for any $T \in \mathscr{T}$, the matrix $T_{\mathscr{S}}$ is simply a restriction of T to the first $|\mathscr{S}|$ dimensions. Finally, to see that $\overline{T}_{\mathscr{S}}$ is the upper expectation operator corresponding to the set $\mathscr{T}' := \{T_{\mathscr{S}} : T \in \mathscr{T}\}$—meaning that $[\overline{T}_{\mathscr{S}} h](x) = \sup_{T' \in \mathscr{T}'} T' h(x)$ for all $h \in \mathscr{L}(\mathscr{S})$ and all $x \in \mathscr{S}$—it suffices to observe that, for any $h \in \mathscr{L}(\mathscr{S})$ and any $x \in \mathscr{S}$,

$$[\overline{T}_{\mathscr{S}} h](x) = [\overline{T} h^\uparrow](x) = \sup_{T \in \mathscr{T}} T h^\uparrow(x) = \sup_{T \in \mathscr{T}} T_{\mathscr{S}} h(x) = \sup_{T' \in \mathscr{T}'} T' h(x). \qquad \square$$

Lemma 46. *Consider any closed communication class $\mathscr{S}$ in $\mathscr{G}(\overline{T})$ and let μ be the unique eigenvalue of $\overline{T}_{f,\mathscr{S}}$. Then, for any $\epsilon > 0$, there is a $T \in \mathscr{T}$, such that*

$$\mu - \epsilon \le \liminf_{k \to +\infty} \tfrac{1}{k} [T_f^k(0)]|_{\mathscr{S}} \le \mu.$$

Proof. Due to Lemma 45 and the fact that $\mathscr{S}$ is closed in $\mathscr{G}(\overline{T})$, we have that $\mathscr{T}' := \{T_{\mathscr{S}} : T \in \mathscr{T}\}$ is a separately specified set of transition matrices and that $\overline{T}' := \overline{T}_{\mathscr{S}}$ is the corresponding upper transition operator. Furthermore, if we let $g := f|_{\mathscr{S}}$, then we can write that $\overline{T}_{f,\mathscr{S}} h = f|_{\mathscr{S}} + (\overline{T} h^\uparrow)|_{\mathscr{S}} = g + \overline{T}' h = \overline{T}'_g h$ for all $h \in \mathscr{L}(\mathscr{S})$. Now, because of Lemma 34 and the fact that $\mathscr{S}$ is a closed communication class in $\mathscr{G}(\overline{T})$, the map $\overline{T}_{f,\mathscr{S}}$ has a unique eigenvalue μ, which implies that μ is also the unique eigenvalue of the map $\overline{T}'_g$. Recalling that $\overline{T}'$ is the upper transition operator corresponding to the separately specified set $\mathscr{T}'$ of transition matrices, we obtain from Lemma 15 that, for any $\epsilon > 0$, there is a $T' \in \mathscr{T}'$ such that

$$\mu - \epsilon \le \liminf_{k \to +\infty} \tfrac{1}{k} [(T'_g)^k(0)] \le \mu. \tag{E.1}$$

Furthermore, let $T \in \mathscr{T}$ be such that $T_{\mathscr{S}} = T'$; there is at least one such a T because $T' \in \mathscr{T}' = \{T_{\mathscr{S}} : T \in \mathscr{T}\}$. Then, similarly as before, we have that $T_{f,\mathscr{S}} h = f|_{\mathscr{S}} + (T h^\uparrow)|_{\mathscr{S}} = g + T' h = T'_g h$ for all $h \in \mathscr{L}(\mathscr{S})$. In particular, we infer that $[T_{f,\mathscr{S}}^k(0)] = [(T'_g)^k(0)]$ for all $k \in \mathbb{N}$. Moreover, since $\mathscr{S}$ is closed in $\mathscr{G}(\overline{T})$, the class $\mathscr{S}$ is also closed in $\mathscr{G}(T)$ due to Lemma 17, so Lemma 32 implies that $[T_{f,\mathscr{S}}^k(0)] = [T_f^k(0)]|_{\mathscr{S}}$ for all $k \in \mathbb{N}$. As a result, we obtain that $[(T'_g)^k(0)] = [T_f^k(0)]|_{\mathscr{S}}$ for all $k \in \mathbb{N}$, which, by plugging this back into Equation (E.1), implies the desired statement. $\qquad \square$

Proof of Proposition 19. Fix any $f \in \mathscr{L}(\mathscr{X})$, any $\epsilon > 0$ and suppose that $\overline{T}$ satisfies (TCA) with top class $\mathscr{R}$. Due

to Lemma 46 and the fact that $\mathscr{R}$ is a closed communication class in $\mathscr{G}(\overline{T})$, there is a $T \in \mathscr{T}$ such that

$$\mu - \epsilon \leq \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)](x) \text{ for all } x \in \mathscr{R}, \tag{E.2}$$

where μ is the unique eigenvalue of $\overline{T}_{f,\mathscr{R}}$. Furthermore, note that $\mathscr{R}$ is also an absorbing closed class in $\mathscr{G}(T)$ because of Lemma 18, so Proposition 43 implies that

$$\min_{x \in \mathscr{R}} \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)](x) \leq \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)](y) \text{ for all } y \in \mathscr{X}.$$

Combining this with Equation (E.2), we infer that $\mu - \epsilon \leq \liminf_{k \to +\infty} \tfrac{1}{k}[T_f^k(0)](x)$ for all $x \in \mathscr{X}$, and therefore certainly that $\mu - \epsilon \leq \liminf_{k \to +\infty} \sup_{T \in \mathscr{T}} \tfrac{1}{k}[T_f^k(0)](x)$ for all $x \in \mathscr{X}$. This holds for any $\epsilon > 0$, so we conclude that

$$\mu \leq \liminf_{k \to +\infty} \sup_{T \in \mathscr{T}} \tfrac{1}{k}[T_f^k(0)](x) = \liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \text{ for all } x \in \mathscr{X},$$

where the last equality follows from Equation (4). Now recall that μ is the unique eigenvalue of $\overline{T}_{f,\mathscr{R}}$ and therefore, that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \mu$ for all $x \in \mathscr{X}$. Indeed, this can easily be deduced from our earlier results and the fact that $\mathscr{R}$ is an absorbing top class in $\mathscr{G}(\overline{T})$; Lemma 35 and Equation (2) imply that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \mu$ for all $x \in \mathscr{R}$, which then also implies that $\lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|y) = \mu$ for all $y \in \mathscr{R}^c$ because $\overline{T}$ is weakly ergodic due to Proposition 11. Since moreover, for all $x \in \mathscr{X}$ and all $k \in \mathbb{N}$, $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \leq \overline{\mathrm{E}}_{\mathrm{av},k}(f|x)$ because $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}} \subseteq \mathscr{P}_{\mathscr{T}}^{\mathrm{ci}} \subseteq \mathscr{P}_{\mathscr{T}}^{\mathrm{ei}}$, we obtain that

$$\mu \leq \liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \leq \limsup_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) \leq \limsup_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \lim_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}(f|x) = \mu \text{ for all } x \in \mathscr{X}.$$

This implies that $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is weakly ergodic and that $\overline{\mathrm{E}}_{\mathrm{av},\infty}^{\mathrm{ri}}(f) = \overline{\mathrm{E}}_{\mathrm{av},\infty}(f)$. $\qquad\square$

Proof of Proposition 20. Suppose that $\mathscr{G}(\overline{T})$ does not have a top class. Due to Corollary 26, there are (at least) two, disjoint, closed communication classes $\mathscr{S}_1$ and $\mathscr{S}_2$. Consider any two $c_1, c_2 \in \mathbb{R}$ such that $c_1 \neq c_2$, and let $f := c_1 \mathbb{I}_{\mathscr{S}_1} + c_2 \mathbb{I}_{\mathscr{S}_2}$. Furthermore, fix any $T \in \mathscr{T}$ and observe that, because of Lemma 17, the classes $\mathscr{S}_1$ and $\mathscr{S}_2$ are also closed (and obviously disjoint) in $\mathscr{G}(T)$. Since $f \mathbb{I}_{\mathscr{S}_1} = c_1 \mathbb{I}_{\mathscr{S}_1}$, Lemma 36 implies that $\tfrac{1}{k}[T_f^k(0)]\mathbb{I}_{\mathscr{S}_1} = \tfrac{1}{k}[T_{c_1}^k(0)]\mathbb{I}_{\mathscr{S}_1}$ for all $k \in \mathbb{N}$. Moreover, it follows trivially from the definition of the map T_{c_1} and the fact that T is a transition matrix (or, using U4), that $T_{c_1}^k(0) = k c_1$ for all $k \in \mathbb{N}$, so we have that $\tfrac{1}{k}[T_{c_1}^k(0)]\mathbb{I}_{\mathscr{S}_1} = c_1 \mathbb{I}_{\mathscr{S}_1}$ for all $k \in \mathbb{N}$. Taking our earlier considerations into account, we obtain that $\tfrac{1}{k}[T_f^k(0)]\mathbb{I}_{\mathscr{S}_1} = c_1 \mathbb{I}_{\mathscr{S}_1}$ for all $k \in \mathbb{N}$. In a completely analogous way, we can deduce that $\tfrac{1}{k}[T_f^k(0)]\mathbb{I}_{\mathscr{S}_2} = c_2 \mathbb{I}_{\mathscr{S}_2}$ for all $k \in \mathbb{N}$. Hence, for any $x \in \mathscr{S}_1$ and any $y \in \mathscr{S}_2$, we have that $\tfrac{1}{k}[T_f^k(0)](x) = c_1$ and $\tfrac{1}{k}[T_f^k(0)](y) = c_2$ for all $k \in \mathbb{N}$. This holds for any $T \in \mathscr{T}$, so we infer that $\sup_{T \in \mathscr{T}} \tfrac{1}{k}[T_f^k(0)](x) = c_1$ and $\sup_{T \in \mathscr{T}} \tfrac{1}{k}[T_f^k(0)](y) = c_2$ for all $k \in \mathbb{N}$. By Equation (4), this implies that $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x) = c_1$ and $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|y) = c_2$ for all $k \in \mathbb{N}$, and since $c_1 \neq c_2$ by assumption, we conclude that $\overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(f|x')$ with $f = c_1 \mathbb{I}_{\mathscr{S}_1} + c_2 \mathbb{I}_{\mathscr{S}_2}$ does not converge to a constant that is equal for all $x' \in \mathscr{X}$. $\qquad\square$

Proof of Proposition 21. Suppose that $\mathscr{G}(\overline{T})$ has a top class $\mathscr{R}$, but that $\overline{T}$ does not satisfy (TCA). We show that $\mathscr{P}_{\mathscr{T}}^{\mathrm{ri}}$ is not weakly ergodic.

Since $\overline{T}$ does not satisfy (TCA), there is, according to Lemma 44, a non-empty set $A \subseteq \mathscr{R}^c$ such that $\mathbb{I}_A \leq \overline{T}\mathbb{I}_A$. Since $\mathscr{T}$ is separately specified, S1 ensures that, for any $\epsilon > 0$, there is a $T \in \mathscr{T}$ such that

$$\mathbb{I}_A - \epsilon \leq \overline{T}\mathbb{I}_A - \epsilon \leq T\mathbb{I}_A. \tag{E.3}$$

Then it is easy to see, using an induction argument, that $[T_{\mathbb{I}_A}^k(0)] \geq k\mathbb{I}_A - \tfrac{k(k-1)}{2}\epsilon$ for all $k \in \mathbb{N}$. Indeed, the inequality

31

trivially holds for $k = 1$ because $[T_{\mathbb{I}_A}(0)] = \mathbb{I}_A$ due to U4. To prove the induction step, suppose that $[T_{\mathbb{I}_A}^k(0)] \geq k\mathbb{I}_A - \frac{k(k-1)}{2}\epsilon$ for some $k \in \mathbb{N}$. Then, using the linearity, monotonicity and constant additivity of T and the fact that $\frac{k(k-1)}{2} = \sum_{\ell=1}^{k-1} \ell$, we find that

$$[T_{\mathbb{I}_A}^{k+1}(0)] = \mathbb{I}_A + T[T_{\mathbb{I}_A}^k(0)] \geq \mathbb{I}_A + T[k\mathbb{I}_A - \tfrac{k(k-1)}{2}\epsilon] = \mathbb{I}_A + kT\mathbb{I}_A - \tfrac{k(k-1)}{2}\epsilon \geq \mathbb{I}_A + k(\mathbb{I}_A - \epsilon) - \tfrac{k(k-1)}{2}\epsilon$$

$$= (k+1)\mathbb{I}_A - k\epsilon - \sum_{\ell=1}^{k-1} \ell\epsilon = (k+1)\mathbb{I}_A - \tfrac{k(k+1)}{2}\epsilon,$$

which establishes the induction step and therefore, the fact that $[T_{\mathbb{I}_A}^k(0)] \geq k\mathbb{I}_A - \frac{k(k-1)}{2}\epsilon$ for all $k \in \mathbb{N}$. Hence, for any $x \in A$, we have that $\sup_{T' \in \mathcal{T}} \frac{1}{k}[(T'_{\mathbb{I}_A})^k(0)](x) \geq \frac{1}{k}[T_{\mathbb{I}_A}^k(0)](x) \geq 1 - \frac{(k-1)}{2}\epsilon$ for all $k \in \mathbb{N}$. Since this is true for any $\epsilon > 0$, we infer that $\sup_{T \in \mathcal{T}} \frac{1}{k}[T_{\mathbb{I}_A}^k(0)](x) \geq 1$ for all $k \in \mathbb{N}$ and therefore, by Equation (4), that

$$\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(\mathbb{I}_A | x) \geq 1 \text{ for all } x \in A.$$

On the other hand, we have that $\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(\mathbb{I}_A | y) \leq 0$ for all $y \in \mathcal{R}$. Indeed, because $A \subseteq \mathcal{R}^c$, we have that $\mathbb{I}_A \mathbb{I}_{\mathcal{R}} = 0 = 0\mathbb{I}_{\mathcal{R}}$, and since $\mathcal{R}$ is a closed communication class in $\mathcal{G}(\overline{T})$, Lemma 36 implies that $\frac{1}{k}[\overline{T}_{\mathbb{I}_A}^k(0)]\mathbb{I}_{\mathcal{R}} = \frac{1}{k}[\overline{T}_0^k(0)]\mathbb{I}_{\mathcal{R}}$ for all $k \in \mathbb{N}$. Since $\frac{1}{k}[\overline{T}_0^k(0)] = 0$ due to U4, we therefore have that $\frac{1}{k}[\overline{T}_{\mathbb{I}_A}^k(0)](y) = 0$ for all $y \in \mathcal{R}$ and all $k \in \mathbb{N}$. Moreover, for any $T \in \mathcal{T}$, note that $Th \leq \overline{T}h$ and therefore $T_{\mathbb{I}_A} h \leq \overline{T}_{\mathbb{I}_A} h$ for all $h \in \mathcal{L}(\mathcal{X})$, which implies by the monotonicity [T2] of $T_{\mathbb{I}_A}$ that $[T_{\mathbb{I}_A}^k(0)] \leq [\overline{T}_{\mathbb{I}_A}^k(0)]$ for all $k \in \mathbb{N}$. Together with our previous considerations, we find that $\frac{1}{k}[T_{\mathbb{I}_A}^k(0)](y) \leq \frac{1}{k}[\overline{T}_{\mathbb{I}_A}^k(0)](y) = 0$ for any $T \in \mathcal{T}$, all $y \in \mathcal{R}$ and all $k \in \mathbb{N}$. As a result,

$$\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(\mathbb{I}_A | y) \overset{(4)}{=} \liminf_{k \to +\infty} \sup_{T \in \mathcal{T}} \frac{1}{k}[T_{\mathbb{I}_A}^k(0)](y) \leq 0 \text{ for all } y \in \mathcal{R}.$$

So, we have that $\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(\mathbb{I}_A | x) \geq 1$ and $\liminf_{k \to +\infty} \overline{\mathrm{E}}_{\mathrm{av},k}^{\mathrm{ri}}(\mathbb{I}_A | y) \leq 0$ for all $x \in A$ and all $y \in \mathcal{R}$, implying that $\mathscr{P}_{\mathcal{T}}^{\mathrm{ri}}$ cannot be weakly ergodic. □