The Metric Linear-Time Branching-Time Spectrum on Nondeterministic Probabilistic Processes

Valentina Castiglioni[a,b,1], Michele Loreti[c], Simone Tini[d]

[a] *INRIA Saclay - Ile de France, France*
[b] *Reykjavik University, Iceland*
[c] *University of Camerino, Italy*
[d] *University of Insubria, Italy*

Abstract

Behavioral equivalences were introduced as a simple and elegant proof methodology for establishing whether the behavior of two processes cannot be distinguished by an external observer. The knowledge of observers usually depends on the *observations* that they can make on process behavior. Furthermore, the combination of *nondeterminism and probability* in concurrent systems leads to several interpretations of process behavior. Clearly, different kinds of observations as well as different interpretations lead to different kinds of behavioral relations, such as (bi)simulations, traces and testing. If we restrict our attention to linear properties only, we can identify three main approaches to *trace* and *testing* semantics: the *trace distributions*, the *trace-by-trace* and the *extremal probabilities* approaches. In this paper, we propose novel notions of *behavioral metrics* that are based on the three classic approaches above, and that can be used to measure the disparities in the linear behavior of processes with respect to trace and testing semantics. We study the properties of these metrics, like compositionality (expressed in terms of the *non-expansiveness* property), and we compare their expressive powers. More precisely, we compare them also to (bi)simulation metrics, thus obtaining the first *metric* linear time - branching time *spectrum*.

Keywords: trace metric, testing metric, bisimulation metric, nondeterministic probabilistic processes

1. Introduction

A major task in the development of complex systems is to verify whether an *implementation* of a system meets its *specification*. Typically, in the realm of process calculi, implementation and specification are *processes*, say I and S, formalized with the same language, and the verification task consists in *comparing their behavior*, which can be done at different levels of abstraction, depending on which aspects of the behavior can be ignored or must be captured. If one focuses on linear properties only, processes are usually compared on the basis of the *traces* they can execute, or accordingly to their capacity to pass the same *tests*. This was the main idea behind the study of *trace equivalence* [31] and *testing equivalence* [16].

If we consider also probabilistic aspects of system behavior, reasoning in terms of qualitative equivalences is only partially satisfactory. Any tiny variation in the probability weights will break the equivalence on processes without any further information on the *distance* of their behaviors. Actually, many implementations can only *approximate* the specification; thus, the verification task requires appropriate instruments to measure the *quality* of the approximation. For this reason, we propose to use *hemimetrics* measuring the disparities in process behavior with respect to linear semantics also to *quantify* process verification. We recall that hemimetrics are *asymmetric* distances, and in our setting they will assign a real in $[0, 1]$ to each

[1]This author has been partially supported by the project 'Open Problems in the Equational Logic of Processes' (OPEL) of the Icelandic Research Fund (grant nr. 196050-051).

pair of processes: distance 0 means that the two processes are indistinguishable with respect to the considered semantics; distance 1 means that an observation of the first computation step can distinguish them. Informally, we may see S as a set of *minimum requirements* on system behavior, such as the *lower bounds* on the probabilities to execute given traces or pass given tests. Then, given a *hemimetric* $\mathbf{h}$ expressing trace (resp. testing) semantics, we can set a *tolerance* ε, related to the application context, and transform the verification problem into a *verification up-to-*ε, or ε*-robustness* problem: I is ε*-trace-robust* (resp. ε*-testing-robust*) with respect to S if whenever S can perform a trace (resp. pass a test) with probability p, then I can do the same with probability *at least* $p - \varepsilon$, namely if $\mathbf{h}(S, I) \leq \varepsilon$. Dually, we may see S as an *upper bound* to undesired system behavior, and demand that whenever S can perform a trace (resp. pass a test) with probability p, then I can do the same with probability *at most* $p + \varepsilon$, namely if $\mathbf{h}(I, S) \leq \varepsilon$.

The notion of *behavioral metric* [14, 17, 19, 29, 35, 44] becomes then crucial. In the literature, we can find a wealth of results on the *bisimilarity metric* [17, 19, 44], namely the pseudometric measuring the differences in the behavior of processes accordingly to the bisimulation semantics, but very little has been investigated on linear semantics, especially on quantitative testing semantics.

Our goal. With this paper we aim at bridging this gap. We consider *nondeterministic probabilistic labeled transition systems* (*PTS*) [40], a very general model in which *nondeterminism and probability* coexist, and we discuss the definition of *hemimetrics* and *pseudometrics* suitable to measure the differences in process behavior with respect to trace and testing semantics. We will see that the interplay of probability and nondeterminism leads to some difficulties in defining notions of behavioral distance, as already experienced in the case of equivalences [5]. Informally, such distances are based on the comparison of the probabilities of particular sequences of observations (or *events*) to occur in the two processes. In the PTS such probabilities highly depend on nondeterminism. Consequently, different resolutions of nondeterminism give different probabilities, and thus different distances. Usually, the nondeterministic choices are solved by *schedulers* [28, 39, 45] which, roughly, assign to each process s a set of (fully probabilistic) processes, called *resolutions of* s, representing each a particular way of solving the nondeterministic choices of s and its derivatives. As there is not a univocal method of resolving nondeterminism, specially when combined with probability, schedulers are divided into classes. Therefore, all our distances will be parametric with respect to the considered class of schedulers. For simplicity, we consider *deterministic* and *randomized* schedulers, however an extension to other types of schedulers seems feasible. We will see, for instance, that a distance will be more or less *discriminating* accordingly to whether the choice of the trace, or test, to be analyzed precedes or follows the choices of the scheduler.

For this reason, to give an immediate perception of the differences in the expressive power of the proposed distances, we compare them to (bi)similarity metric semantics, thus obtaining the first *spectrum* of behavioral metrics over processes in the PTS model. Since we consider behavioral metrics for linear and branching semantics, we refer to our spectrum as to the *metric linear time - branching time spectrum.*

The composition of the spectrum. For what concerns the branching time part of our spectrum, we consider bisimulation, ready simulation and simulation semantics [37]. In particular, to mimic the action of randomized schedulers on linear semantics, we also consider *convex* (bi)simulations [41] which are evaluated over *combined transitions* for processes. Interestingly, we will show that the approximation and composition properties already established for (bi)similarity metrics hold also for their convex versions. Our contribution can then be summarized as follows:

1. We formalize the notions of *ready similarity* and *similarity metric.*
2. We introduce the notions of *convex bisimilarity*, *ready similarity* and *similarity metric.*
3. We prove that convex (bi)similarity metrics can be equivalently defined as the limit of a sequence of metrics d_k comparing only the first k computation steps of processes.
4. We prove that all the considered branching metrics are *compositional*, in the sense of *non-expansiveness* [19] with respect to parallel composition, i.e., the quantitative analogue to the congruence property

ensuring that the distance of two composed systems is not greater than the sum of the pairwise distances of their components.

As regards linear semantics, we consider three approaches to *trace semantics*, two known from the literature and a novel one:

(i) The *trace distribution* approach [39], comparing *entire resolutions* created by schedulers by checking if they assign the same probability to the same traces;

(ii) The *trace-by-trace* approach [2], in which firstly we take a trace and then we check if there are resolutions for processes assigning the same probability to it;

(iii) The novel *supremal probabilities* approach, considering for each trace only the suprema of the probabilities assigned to it over all resolutions for the processes.

Similarly, we consider three approaches to *testing semantics*:

(iv) The *may/must* approach [46], in which the extremal probabilities of passing a test are considered;

(v) The *trace-by-trace* approach [5], which is based on a *traced* view of testing and mimics the trace-by-trace approach to trace semantics;

(vi) The novel *supremal probabilities* approach, which can be considered as the adaptation to testing semantics of the supremal probability approach to trace semantics.

For each of these six approaches, and for each class of schedulers, we present a *hemimetric* and a *pseudometric* as the quantitative variant of the related preorder and equivalence in the trace or testing semantics. We stress that in the latter case, to the best of our knowledge, ours is the first attempt in this direction. Our results on linear metrics can then be summarized as follows:

5. We prove that, under each hemimetric/pseudometric, the pairs of processes at distance zero are precisely those related by the corresponding preorder/equivalence.

6. In the case of trace metrics, we prove that the hemimetrics/pseudometrics for trace-by-trace and supremal probabilities semantics are suitable for compositional reasoning, by showing their non-expansiveness.

7. In the case of testing metrics, we prove that all hemimetrics and pseudometrics are non-expansive.

Finally, we proceed to the composition of the metric linear time - branching time spectrum by studying the differences in the expressive powers of all the proposed distances. In particular:

8. We obtain an interesting result in the perspective of an application to process verification: the supremal probabilities semantics defined either on deterministic or randomized schedulers has the same expressive power as the trace-by-trace semantics on randomized schedulers.

9. Under deterministic schedulers, the *supremal* probabilities is the only approach to trace semantics that is comparable with *(bi)simulation* metrics. Nonetheless, the relation with (bi)simulation semantics is regained by the other two approaches when randomized schedulers are considered.

10. The *must testing metric* is comparable with a '*reversed*' *ready similarity metric*: the must distance between s and t is always bounded from above by the ready similarity distance between t and s.

Organization of contents. In Section 2 we review the background. Then, (bi)simulation, trace and testing metrics are discussed, respectively, in Sections 3, 4 and 5. We dedicate Section 6 to the construction of the metric linear time - branching time spectrum. Finally, we discuss related and future work in Section 7.

A preliminary version of this paper appeared as [7]. This special issue version comes with the novel results on convex (bi)simulation metrics and the spectrum in Section 6.

2. Background

In this section we review the preliminary notions on the PTS model that are necessary for our dissertation.

Given an arbitrary set X, a *discrete probability distribution* over X is a mapping $\pi\colon X \to [0,1]$ with $\sum_{x\in X} \pi(x) = 1$. The *support* of π is the set $\mathsf{supp}(\pi) = \{x \in X \mid \pi(x) > 0\}$. By $\Delta(X)$ we denote the set of all *finitely supported* distributions over X, ranged over by $\pi, \pi', \ldots$ Given an element $x \in X$, we let δ_x denote the *Dirac (or point) distribution on* x, defined by $\delta_x(x) = 1$ and $\delta_x(y) = 0$ for all $y \neq x$. For a finite set of indexes I, weights $p_i \in (0,1]$ with $\sum_{i\in I} p_i = 1$ and distributions $\pi_i \in \Delta(X)$ with $i \in I$, the distribution $\sum_{i\in I} p_i\pi_i$ is defined by $(\sum_{i\in I} p_i\pi_i)(x) = \sum_{i\in I} p_i \cdot \pi_i(x)$, for all $x \in X$.

2.1. The PTS model

PTSs [40] extend classical LTSs [34] to model, at the same time, reactive behavior, nondeterminism and probability. In a PTS, the state space is given by a set $\mathcal{S}$ of *processes*, ranged over by $s, t, \ldots$, and the transition steps take processes to probability distributions over processes.

Definition 1 (PTS, [40])**.** A *nondeterministic probabilistic labelled transition system (PTS)* is a triple $(\mathcal{S}, \mathcal{A}, \to)$, where: (i) $\mathcal{S}$ is a countable set of processes, (ii) $\mathcal{A}$ is a countable set of *actions*, and (iii) $\to \subseteq \mathcal{S} \times \mathcal{A} \times \Delta(\mathcal{S})$ is a *transition relation*, where each *transition* $(s, a, \pi) \in \to$ is denoted by $s \xrightarrow{a} \pi$.

The *a-derivatives* of process $s \in \mathcal{S}$ are the distributions $\mathrm{der}(s,a) = \{\pi \mid s \xrightarrow{a} \pi\}$. We write $s \xrightarrow{a}$ if $\mathrm{der}(s,a) \neq \emptyset$ and $s \not\xrightarrow{a}$ otherwise. The *initials* of s are the actions $\mathrm{init}(s) = \{a \in \mathcal{A} \mid s \xrightarrow{a}\}$ that can be performed by s. A PTS is *fully nondeterministic* if every transition has the form $s \xrightarrow{a} \delta_t$, for some $t \in \mathcal{S}$. A PTS is *fully probabilistic* if at most one transition is enabled for each process. Finally, a PTS is *image-finite* [30] if $\mathrm{der}(s,a)$ is finite for each $s \in \mathcal{S}$ and $a \in \mathcal{A}$. We consider only image-finite PTSs.

A *combined transition* [41] is a convex combination of equally labeled transitions, formally defined by $s \xrightarrow{a}_{\mathrm{c}} \pi$ if and only if there are a finite set of indexes I, weights $p_i \in (0,1]$ with $\sum_{i\in I} p_i = 1$ and distributions $\pi_i \in \Delta(\mathcal{S})$ with $i \in I$ such that $s \xrightarrow{a} \pi_i$ for each $i \in I$ and $\pi = \sum_{i\in I} p_i\pi_i$. We let $\mathrm{der}_{\mathrm{ct}}(s,a) = \{\pi \mid s \xrightarrow{a}_{\mathrm{c}} \pi\}$.

Definition 2 (Parallel composition)**.** Let $P_1 = (\mathcal{S}_1, \mathcal{A}, \to_1)$ and $P_2 = (\mathcal{S}_2, \mathcal{A}, \to_2)$ be two PTSs. The (CSP-like [31]) *synchronous parallel composition of* P_1 *and* P_2 is the PTS $P_1 \parallel P_2 = (\mathcal{S}_1 \times \mathcal{S}_2, \mathcal{A}, \to)$, where $(s_1, s_2) \xrightarrow{a} \pi$ if and only if $s_1 \xrightarrow{a}_1 \pi_1$, $s_2 \xrightarrow{a}_2 \pi_2$ and $\pi(s_1', s_2') = \pi_1(s_1') \cdot \pi_2(s_2')$ for all $(s_1', s_2') \in \mathcal{S}_1 \times \mathcal{S}_2$.

With abuse of notation, the notion of parallel composition can be extended to distributions by letting $(\pi_1 \parallel \pi_2)(s) = \pi_1(s_1) \cdot \pi_2(s_2)$, if $s = s_1 \parallel s_2$, and $(\pi_1 \parallel \pi_2)(s) = 0$, if s is not of the form $s = s_1 \parallel s_2$.

2.2. How to express linear semantics

We proceed to recall some notions, mostly from [3–5], necessary to reason on trace and testing semantics. A *computation* is a weighted sequence of process-to-process transitions, denoted by $\twoheadrightarrow$.

Definition 3 (Computation)**.** Let $\twoheadrightarrow \subseteq \mathcal{S} \times \mathcal{A} \times [0,1] \times \mathcal{S}$. A *computation from* s_0 *to* s_n has the form $c := s_0 \overset{a_1,p_1}{\twoheadrightarrow} s_1 \overset{a_2,p_2}{\twoheadrightarrow} \ldots \overset{a_n,p_n}{\twoheadrightarrow} s_n$ where, for all $i = 1, \ldots, n$, there is a transition $s_{i-1} \xrightarrow{a_i} \pi_i$ with $\pi_i(s_i) = p_i$.

Notice that p_i is the *execution probability* of step $s_{i-1} \overset{a_i,p_i}{\twoheadrightarrow} s_i$ conditioned on the selection of the transition $s_{i-1} \xrightarrow{a_i} \pi_i$ at s_{i-1}. We denote by $\Pr(c) = \prod_{i=1}^{n} p_i$ the product of the execution probabilities of the steps in c. A computation c from s is *maximal* if it is not a proper prefix of any other computation from s. We denote by $\mathcal{C}(s)$ (resp. $\mathcal{C}_{\max}(s)$) the set of computations (resp. maximal computations) from s. For any $\mathcal{C} \subseteq \mathcal{C}(s)$, we define $\Pr(\mathcal{C}) = \sum_{c\in\mathcal{C}} \Pr(c)$ whenever none of the computations in $\mathcal{C}$ is a proper prefix of any of the others.

A *trace* is a sequence of actions in $\mathcal{A}$. We denote by $\mathcal{A}^\star$ the set of *finite traces* in $\mathcal{A}$ and we let $\alpha, \beta, \ldots$ range over them. We say that a computation is *compatible* with $\alpha \in \mathcal{A}^\star$ if the sequence of actions labeling the computation steps is equal to α. We denote by $\mathcal{C}(s,\alpha) \subseteq \mathcal{C}(s)$ the set of computations from s that are compatible with α, and by $\mathcal{C}_{\max}(s,\alpha)$ the set $\mathcal{C}_{\max}(s,\alpha) = \mathcal{C}_{\max}(s) \cap \mathcal{C}(s,\alpha)$.

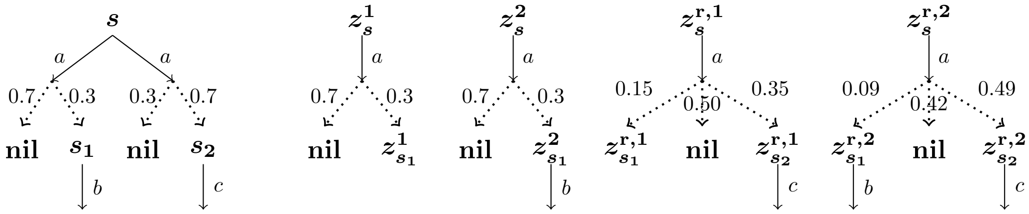

Figure 1: Examples of deterministic and randomized resolutions of process s.

To express linear semantics we need to evaluate and compare the probability of particular sequences of *events* to occur. As in PTSs this probability highly depends also on nondeterminism, *schedulers* [28, 39, 45] (or *adversaries*) resolving it become fundamental. They can be classified into two main classes: *deterministic* and *randomized schedulers* [39]. For each process, a deterministic scheduler selects exactly one transition among the possible ones, or none of them, thus treating all internal nondeterministic choices as distinct. Randomized schedulers allow for a convex combination of the equally labeled transitions. The resolution given by a deterministic scheduler is a fully probabilistic process, whereas from randomized schedulers we get a fully probabilistic process with combined transitions.

Definition 4 (Resolution)**.** Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ and a process $s \in \mathcal{S}$.
We say that a PTS $\mathcal{Z} = (Z, \mathcal{A}, \rightarrow_{\mathcal{Z}})$ is a *deterministic resolution* for s if and only if there is a function $\mathrm{corr}_{\mathcal{Z}} \colon Z \rightarrow \mathcal{S}$ with $s = \mathrm{corr}_{\mathcal{Z}}(z_s)$, for some $z_s \in Z$, and moreover:

(i) If $z \xrightarrow{a}_{\mathcal{Z}} \pi$ then $\mathrm{corr}_{\mathcal{Z}}(z) \xrightarrow{a} \pi'$, with $\pi(z') = \pi'(\mathrm{corr}_{\mathcal{Z}}(z'))$ for all $z' \in Z$.

(ii) If $z \xrightarrow{a_1}_{\mathcal{Z}} \pi_1$ and $z \xrightarrow{a_2}_{\mathcal{Z}} \pi_2$ then $a_1 = a_2$ and $\pi_1 = \pi_2$.

Conversely, $\mathcal{Z}$ is a *randomized resolution* for s if combined transitions are considered, namely (i) is rewritten

(i)' If $z \xrightarrow{a}_{\mathcal{Z}} \pi$ then $\mathrm{corr}_{\mathcal{Z}}(z) \xrightarrow{a}_{\mathrm{c}} \pi'$, with $\pi(z') = \pi'(\mathrm{corr}_{\mathcal{Z}}(z'))$ for all $z' \in Z$.

In both cases, the resolution $\mathcal{Z}$ is *maximal* if it cannot be further extended in accordance with the graph structure of P and the constraints above. For $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$, we denote by $\mathrm{Res}^{\mathrm{x}}(s)$ the set of deterministic/randomized resolutions for s and by $\mathrm{Res}^{\mathrm{x}}_{\max}(s)$ the subset of the maximal resolutions in $\mathrm{Res}^{\mathrm{x}}(s)$.

***Example* 1.** Consider process s in Figure 1. Processes z_s^1 and z_s^2 are two examples of resolutions of s via a deterministic scheduler, whereas the resolutions $z_s^{\mathrm{r},1}$ and $z_s^{\mathrm{r},2}$ are obtained via randomized schedulers. Notice that z_s^1 and z_s^2 are both related to the leftmost a-branch of s: z_s^1 does not select any move for process s_1, giving $\mathrm{init}(z^1_{s_1}) = \emptyset$, whereas z_s^2 selects the b-move of s_1. $z_s^{\mathrm{r},1}$ is obtained by weighting each a-branch of s by 0.5, and not selecting any move for process s_1. Conversely, $z_s^{\mathrm{r},2}$ gives weight 0.3 to the left-most a branch and 0.7 to the right-most one, and enables both the b-move by s_1 and the c-move by s_2. ◄

2.3. Behavioral metrics and their compositional properties

Behavioral equivalences answer the question of whether two processes behave precisely the same way or not with respect to the *observations* that we can make on them. *Behavioral metrics* [14, 17, 19, 42, 44] answer the more general question of measuring the differences in the behavior of processes. Usually, they are defined as 1-bounded *pseudometrics* expressing the behavioral distance on processes, namely they quantify the disparities in the observations that we can make on processes.

A 1-bounded *pseudometric* on a set X is a function $d \colon X \times X \rightarrow [0, 1]$ such that: (i) $d(x, x) = 0$, (ii) $d(x, y) = d(y, x)$, and (iii) $d(x, y) \leq d(x, z) + d(z, y)$, for all $x, y, z \in X$. Then, d is a *hemimetric* if it satisfies (i) and (iii). The *kernel* of a (hemi,pseudo)metric d on X consists in the set of the pairs of elements in X that are at distance 0, namely $\mathit{ker}(d) = \{(x, y) \in X \times X \mid d(x, y) = 0\}$.

As elsewhere in the literature, we will sometimes use the term *metric* in place of pseudometric.

Behavioral metrics are normally parametric with respect to a *discount factor* allowing us to specify how much the distance of future transitions is mitigated [15, 19]. Informally, any difference that can be observed only after a long sequence of computation steps does not have the same impact of the differences that can be witnessed at the beginning of the computation. We will argue that different approaches to the semantics will require different technical formalization of the discount.

We conclude this section by recalling the notion of *non-expansiveness* [19] of a (hemi,pseudo)metric, which is the quantitative analogous to the (pre)congruence property for behavioral equivalences and preorders. Informally, a behavioral distance is non-expansive, and thus *compositional*, if the distance of two composed systems is not greater than the sum of the pairwise distances of their components. Here we consider also the stronger notion of *strict non-expansiveness* [24] that gives tighter bounds on the distance of processes composed in parallel: the distance between $s_1 \parallel t_1$ and $s_2 \parallel t_2$ is bounded by the sum of pairwise distances on the components minus their product. Intuitively, this allows us to avoid double evaluations of distances.

Definition 5 ((Strict) non-expansiveness, [19, 24])**.** A (hemi,pseudo)metric d on $\mathcal{S}$ is *non-expansive* if and only if for all processes $s_1, s_2, t_1, t_2 \in \mathcal{S}$ we have $d(s_1 \parallel s_2, t_1 \parallel t_2) \leq d(s_1, t_1) + d(s_2, t_2)$. Moreover, d is *strictly non-expansive* if $d(s_1 \parallel s_2, t_1 \parallel t_2) \leq d(s_1, t_1) + d(s_2, t_2) - d(s_1, t_1) \cdot d(s_2, t_2)$.

3. (Bi)simulation relations and metrics

In this section we discuss the metrics measuring the disparities in process behavior with respect to *bisimulation semantics*. In Section 3.1 we recall first the equivalences and preorders in the classic approach to probabilistic bisimulation of [37, 40] and their *convex* counterparts, then we consider the corresponding distances. We prove that the quantitative analogous to probabilistic ready similarity and similarity, and the convex (bi)simulation metrics enjoy the strict non-expansiveness property (Theorem 2), which was already established for the quantitative analogous to bisimilarity in [24]. In Section 3.2 we compare the expressive power of these metrics, thus composing the first part of the spectrum (Theorems 3 and 4).

3.1. (Bi)simulation metrics

A *probabilistic bisimulation* is an equivalence relation over $\mathcal{S}$ that equates two processes if they can mimic each other's transitions and evolve to distributions that are related, in turn, by *the same* relation. To formalize this idea, we need to lift relations over processes to relations over distributions. We rely on the notion of *matching*, also referred to in the literature as *coupling* or *weight function*.

Definition 6 (Matching)**.** Assume two sets X and Y. A *matching* for distributions $\pi \in \Delta(X)$, $\pi' \in \Delta(Y)$ is a distribution over the product space $\mathfrak{w} \in \Delta(X \times Y)$ with π and π' as left and right marginal, namely:

$$\text{(i) } \textstyle\sum_{y \in Y} \mathfrak{w}(x,y) = \pi(x), \text{ for all } x \in X, \quad \text{and} \quad \text{(ii) } \textstyle\sum_{x \in X} \mathfrak{w}(x,y) = \pi'(y), \text{ for all } y \in Y.$$

We let $\mathfrak{W}(\pi, \pi')$ denote the set of all matchings for π and π'.

Definition 7 (Relation lifting, [40])**.** Assume two sets X and Y and a relation $\mathcal{R} \subseteq X \times Y$. The *lifting* of $\mathcal{R}$ to a relation $\mathcal{R}^\dagger \subseteq \Delta(X) \times \Delta(Y)$ is defined by letting, for any $\pi \in \Delta(X)$ and $\pi' \in \Delta(Y)$, $\pi \, \mathcal{R}^\dagger \, \pi'$ if and only if there is a matching $\mathfrak{w} \in \mathfrak{W}(\pi, \pi')$ with $x \, \mathcal{R} \, y$ whenever $\mathfrak{w}(x,y) > 0$.

Definition 8 (Probabilistic (bi)simulations, [37, 40])**.** Assume a PTS $(\mathcal{S}, \mathcal{A}, \rightarrow)$. Then:

- A binary relation $\mathcal{R} \subseteq \mathcal{S} \times \mathcal{S}$ is a *probabilistic simulation* if, whenever $s \, \mathcal{R} \, t$: for each transition $s \xrightarrow{a} \pi_s$ there is a transition $t \xrightarrow{a} \pi_t$ such that $\pi_s \, \mathcal{R}^\dagger \pi_t$.
- A probabilistic simulation $\mathcal{R}$ is a *probabilistic ready simulation* if, whenever $s \, \mathcal{R} \, t$, $s \stackrel{a}{\not\rightarrow}$ implies $t \stackrel{a}{\not\rightarrow}$.
- A *probabilistic bisimulation* is a symmetric probabilistic simulation.

The union of all probabilistic simulations (resp.: ready simulations, bisimulations) is the greatest probabilistic simulation (resp.: ready simulation, bisimulation), is denoted by $\sqsubseteq_{\mathrm{s}}^{\mathrm{det}}$ (resp.: $\sqsubseteq_{\mathrm{r}}^{\mathrm{det}}$, $\sim_{\mathrm{b}}^{\mathrm{det}}$), is called *probabilistic similarity* (resp.: *ready similarity*, *bisimilarity*), and is a preorder (resp.: preorder, equivalence). If we consider combined transitions, then we get the *convex similarity* (resp.: *convex ready similarity*, *convex bisimilarity*), denoted by $\sqsubseteq_{\mathrm{s}}^{\mathrm{rand}}$ (resp.: $\sqsubseteq_{\mathrm{r}}^{\mathrm{rand}}$, $\sim_{\mathrm{b}}^{\mathrm{rand}}$).

Bisimulation metrics [17, 19, 44] base on the quantitative analogous to the bisimulation game: two processes can be at some given distance $\varepsilon < 1$ only if they can mimic each other's transitions and evolve to distributions that are, in turn, at a distance $\leq \varepsilon$. To formalize this intuition, we need to lift pseudometrics on processes to pseudometrics on distributions. We rely on the notions of matching and Kantorovich lifting.

Definition 9 (Kantorovich metric, [33])**.** Given a (hemi,pseudo)metric d on a space X, the *Kantorovich lifting* of d is the (hemi,pseudo)metric $\mathbf{K}(d)\colon \Delta(X) \times \Delta(X) \to [0,1]$ defined for all $\pi, \pi' \in \Delta(X)$ by

$$\mathbf{K}(d)(\pi,\pi') = \min_{\mathfrak{w} \in \mathfrak{W}(\pi,\pi')} \sum_{x,y \in X} \mathfrak{w}(x,y) \cdot d(x,y).$$

In the bisimulation game, discounting the difference of future transitions requires a *discount factor* $\lambda \in (0,1]$ such that the distance arising at step n is mitigated by λ^n. Then, bisimulation metrics (resp.: ready simulation metrics, simulation metrics) can be defined as the prefixed points of a suitable functional parametric on λ and defined on the complete lattice $(\mathcal{D}(\mathcal{S}), \preceq)$, with $\mathcal{D}(\mathcal{S})$ the set of the 1-bounded pseudometrics over $\mathcal{S}$ and $d_1 \preceq d_2$ if and only if $d_1(s,t) \leq d_2(s,t)$ for all $s,t \in \mathcal{S}$. For each set $D \subseteq \mathcal{D}(\mathcal{S})$ the supremum and infinimum are defined by $\sup(D)(s,t) = \sup_{d \in D} d(s,t)$ and $\inf(D)(s,t) = \inf_{d \in D} d(s,t)$ for all $s,t \in \mathcal{S}$. Notice that the bottom element of the lattice is the constant function $\mathbf{0}$ with $\mathbf{0}(s,t) = 0$ for all $s,t \in \mathcal{S}$.

Definition 10 ((Bi)simulation metric functional)**.** Assume a discount factor $\lambda \in (0,1]$. The functionals $\mathbf{B}^\lambda, \mathbf{R}^\lambda, \mathbf{S}^\lambda\colon \mathcal{D}(\mathcal{S}) \to \mathcal{D}(\mathcal{S})$ are defined for all functions $d \in \mathcal{D}(\mathcal{S})$ and processes $s,t \in \mathcal{S}$ by

$$\mathbf{B}^\lambda(d)(s,t) = \sup_{a \in \mathcal{A}} \max\Big\{ \sup_{\pi_s \in \mathrm{der}(s,a)} \inf_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t), \sup_{\pi_t \in \mathrm{der}(t,a)} \inf_{\pi_s \in \mathrm{der}(s,a)} \lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t) \Big\}$$

$$\mathbf{R}^\lambda(d)(s,t) = \begin{cases} 1 & \text{if } \mathrm{init}(s) \neq \mathrm{init}(t) \\ \sup_{a \in \mathcal{A}} \sup_{\pi_s \in \mathrm{der}(s,a)} \inf_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t) & \text{otherwise} \end{cases}$$

$$\mathbf{S}^\lambda(d)(s,t) = \sup_{a \in \mathcal{A}} \sup_{\pi_s \in \mathrm{der}(s,a)} \inf_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t)$$

where $\sup \emptyset = 0$ and $\inf \emptyset = 1$.

A pseudometric $d \in \mathcal{D}(\mathcal{S})$ is a *bisimulation metric* if it is a prefixed point of $\mathbf{B}^\lambda$, where $\mathbf{B}^\lambda(d) \preceq d$ ensures that whenever $d(s,t) < 1$, any transition $s \xrightarrow{a} \pi_s$ is mimicked by a transition $t \xrightarrow{a} \pi_t$ with $\lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t) \leq d(s,t)$, and vice versa. Then, the *ready simulation hemimetrics* and the *simulation hemimetrics* are hemimetrics being prefixed points of $\mathbf{R}^\lambda$ and $\mathbf{S}^\lambda$, respectively.

Functional $\mathbf{B}^\lambda$ (resp.: $\mathbf{R}^\lambda$, $\mathbf{S}^\lambda$) will be denoted $\mathbf{B}^{\lambda,\mathrm{det}}$ (resp.: $\mathbf{R}^{\lambda,\mathrm{det}}$, $\mathbf{S}^{\lambda,\mathrm{det}}$) when we intend to stress that we do not consider any combined transition derivable from the PTS, and $\mathbf{B}^{\lambda,\mathrm{rand}}$ (resp.: $\mathbf{R}^{\lambda,\mathrm{rand}}$, $\mathbf{S}^{\lambda,\mathrm{rand}}$) when, conversely, we intend to stress that we consider all the combined transitions derivable from the PTS.

Definition 11 ((Bi)simulation (hemi)metric)**.** A pseudometric (resp.: hemimetric, hemimetric) $d \in \mathcal{D}(\mathcal{S})$ is a *bisimulation metric* (resp.: *ready simulation hemimetric*, *simulation hemimetric*) if and only if $\mathbf{B}^{\lambda,\mathrm{det}}(d) \preceq d$ (resp.: $\mathbf{R}^{\lambda,\mathrm{det}}(d) \preceq d$, $\mathbf{S}^{\lambda,\mathrm{det}}(d) \preceq d$). Then, $d \in \mathcal{D}(\mathcal{S})$ is a *convex bisimulation metric* (resp.: *convex ready simulation hemimetric*, *convex simulation hemimetric*) if and only if $\mathbf{B}^{\lambda,\mathrm{rand}}(d) \preceq d$ (resp.: $\mathbf{R}^{\lambda,\mathrm{rand}}(d) \preceq d$, $\mathbf{S}^{\lambda,\mathrm{rand}}(d) \preceq d$).

The monotonicity of $\mathbf{K}$ and of functions sup, inf ensure that all these functionals are monotone. Therefore, accordingly to Tarski's fixed point theorem, they have the least prefixed point.

Definition 12 ((Bi)similarity (hemi)metric)**.** The least prefixed point of $\mathbf{B}^{\lambda,\mathrm{det}}$ (resp.: $\mathbf{R}^{\lambda,\mathrm{det}}$, $\mathbf{S}^{\lambda,\mathrm{det}}$) is denoted by $\mathbf{b}^{\lambda,\mathrm{det}}$ (resp. $\mathbf{r}^{\lambda,\mathrm{det}}$, $\mathbf{s}^{\lambda,\mathrm{det}}$) and called the *bisimilarity metric* (resp.: *ready similarity hemimetric*, *similarity hemimetric*). Analogously, the least prefixed point of $\mathbf{B}^{\lambda,\mathrm{rand}}$ (resp.: $\mathbf{R}^{\lambda,\mathrm{rand}}$, $\mathbf{S}^{\lambda,\mathrm{rand}}$) is denoted by $\mathbf{b}^{\lambda,\mathrm{rand}}$ (resp. $\mathbf{r}^{\lambda,\mathrm{rand}}$, $\mathbf{s}^{\lambda,\mathrm{rand}}$) and called the *convex bisimilarity metric* (resp.: *convex ready similarity hemimetric*, *convex similarity hemimetric*).

Let $\mathbf{F}$ be any of the functionals used in Definitions 11–12. Tarski's theorem ensures also that the least prefixed point of $\mathbf{F}$ coincides with the least fixed point and can be obtained in an iterative fashion, meaning that there exists an ordinal α with $\mathbf{F}^{\alpha}(\mathbf{0}) = \mathbf{F}^{\alpha+1}(\mathbf{0})$. By exploiting the image-finiteness of PTSs, following, e.g., [20, 36], we could easily show that $\mathbf{B}^{\lambda,\mathrm{det}}$, $\mathbf{R}^{\lambda,\mathrm{det}}$, $\mathbf{S}^{\lambda,\mathrm{det}}$ are Scott continuous, thus inferring that their closure ordinal is ω. Unfortunately, this argument does not apply to $\mathbf{B}^{\lambda,\mathrm{rand}}$, $\mathbf{R}^{\lambda,\mathrm{rand}}$, $\mathbf{S}^{\lambda,\mathrm{rand}}$, since combined transitions clearly break image-finiteness. However, by relying on the fact that we consider image-finite PTSs with finitely supported distributions, we can prove the following result of *non-expansiveness* [43] for all six functionals, which is a property ensuring that the closure ordinal is ω [43, Corollary 1].

Proposition 1. *Assume an image finite PTS in which, for each transition* $s \xrightarrow{a} \pi$, π *is a distribution with finite support. Let* $\mathbf{F} \in \{\mathbf{B}^{\lambda,\mathrm{x}}, \mathbf{R}^{\lambda,\mathrm{x}}, \mathbf{S}^{\lambda,\mathrm{x}}\}$ *for* $\lambda \in (0,1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$*. Then, given any* $d_1, d_2 \in \mathcal{D}(\mathcal{S})$ *with* $d_2 \preceq d_1$*, for all* $s,t \in \mathcal{S}$ *we have:*

$$\mathbf{F}(d_1)(s,t) - \mathbf{F}(d_2)(s,t) \leq \sup_{u,v\in\mathcal{S}} (d_1(u,v) - d_2(u,v)).$$

Proof. The proof can be found in Appendix A.1. □

As a consequence, we can associate to each of the six distances introduced in Definition 12 a notion of up-to-k distance, which considers only the discrepancies arising in the first k computation steps.

Definition 13 (Up-to-k (bi)similarity metric)**.** Let $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. We define the *up-to-k bisimilarity metric* $\mathbf{b}_k^{\lambda,\mathrm{x}}$ for $k \in \mathbb{N}$ by $\mathbf{b}_k^{\lambda,\mathrm{x}} = (\mathbf{B}^{\lambda,\mathrm{x}})^k(\mathbf{0})$. Similarly, the *up-to-k ready similarity metric* $\mathbf{r}_k^{\lambda,\mathrm{x}}$ is defined by $\mathbf{r}_k^{\lambda,\mathrm{x}} = (\mathbf{R}^{\lambda,\mathrm{x}})^k(\mathbf{0})$ and the *up-to-k similarity metric* $\mathbf{s}_k^{\lambda,\mathrm{x}}$ is defined by $\mathbf{s}_k^{\lambda,\mathrm{x}} = (\mathbf{S}^{\lambda,\mathrm{x}})^k(\mathbf{0})$.

Let $\mathbf{F} \in \{\mathbf{B}^{\lambda,\mathrm{x}}, \mathbf{R}^{\lambda,\mathrm{x}}, \mathbf{S}^{\lambda,\mathrm{x}}\}$. Since the chain of up-to-k distances $\mathbf{0} \preceq \mathbf{F}(\mathbf{0}) \preceq \mathbf{F}^2(\mathbf{0}) \ldots$ is non-decreasing, and being ω the closure ordinal of $\mathbf{F}$, such a chain converges to the least fixed point.

Proposition 2. *Assume an image finite PTS in which, for each transition* $s \xrightarrow{a} \pi$, π *is a distribution with finite support. Let* $\lambda \in (0,1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ *and* $\mathbf{d} \in \{\mathbf{b}, \mathbf{r}, \mathbf{s}\}$*. Then* $\mathbf{d}^{\lambda,\mathrm{x}} = \lim_{k\to\infty} \mathbf{d}_k^{\lambda,\mathrm{x}}$.

Proof. The proof can be found in Appendix A.2. □

The kernels of the (hemi,pseudo)metrics in Definition 12 are the behavioral relations in Definition 8.

Theorem 1. *Let* $(\mathcal{S}, \mathcal{A}, \to)$ *be a PTS,* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ *and* $\lambda \in (0,1]$*. Then:*

- *The function* $\mathbf{b}^{\lambda,\mathrm{x}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim_{\mathrm{b}}^{\mathrm{x}}$ *as kernel.*
- *The function* $\mathbf{r}^{\lambda,\mathrm{x}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq_{\mathrm{r}}^{\mathrm{x}}$ *as kernel.*
- *The function* $\mathbf{s}^{\lambda,\mathrm{x}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq_{\mathrm{s}}^{\mathrm{x}}$ *as kernel.*

Proof. The first item was proved in [17]. The remaining cases can be proved by analogous arguments. □

Moreover, all these distances are compositional, in the sense of strict non-expansiveness.

Theorem 2. *Let* $(\mathcal{S}, \mathcal{A}, \to)$ *be a PTS,* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ *and* $\lambda \in (0,1]$*. All functions* $\mathbf{b}^{\lambda,\mathrm{x}}$*,* $\mathbf{r}^{\lambda,\mathrm{x}}$ *and* $\mathbf{s}^{\lambda,\mathrm{x}}$ *are strictly non-expansive.*

Proof. The result for $\mathbf{b}^{\lambda,\mathrm{det}}$ is in [24]. The proof of other cases is similar and given in Appendix A.3. □

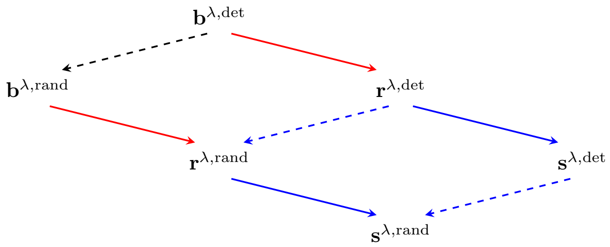

Figure 2: The spectrum of (bi)simulation metrics. An arrow between two distances $d \to d'$ stands for $d > d'$. We use black arrows to compare metrics, blue arrows to compare hemimetrics, dashed arrows to compare the same distance with respect to different classes of schedulers, and red arrows to compare metrics with hemimetrics.

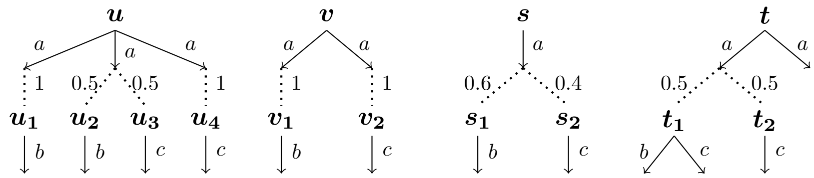

Figure 3: Processes u, v show the strictness of the relation in Theorem 3, and s, t show the strictness of relations in Theorem 4.

3.2. Comparing the distinguishing power of bisimulation semantics

In this section, we compare the distances in Definition 12 obtaining the spectrum in Figure 2. In detail, we order the distances with respect to their distinguishing power: we write $d > d'$ for $d, d' \in \mathcal{D}(\mathcal{S})$ if and only if: (i) $d(s,t) \geq d'(s,t)$ for all processes $s, t \in \mathcal{S}$, and (ii) $d(u,v) > d'(u,v)$ for some processes $u, v \in \mathcal{S}$.

Firstly, we notice that when classic transitions are considered, the three (bi)simulation distances are, in general, more discriminating with respect to their correspondent ones on combined transitions.

Theorem 3. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$ and $\mathbf{d} \in \{\mathbf{b}, \mathbf{r}, \mathbf{s}\}$. Then $\mathbf{d}^{\lambda,\mathrm{rand}} < \mathbf{d}^{\lambda,\mathrm{det}}$.*

Proof. The proof of the non-strict relations $\mathbf{d}^{\lambda,\mathrm{rand}} \leq \mathbf{d}^{\lambda,\mathrm{det}}$ can be found in Appendix A.4. Then the strictness of the relations follows by Example 2. □

***Example* 2.** Consider processes u, v in Figure 3. It is not hard to see that $\mathbf{b}^{\lambda,\mathrm{det}}(u,v) = \mathbf{r}^{\lambda,\mathrm{det}}(u,v) = \mathbf{s}^{\lambda,\mathrm{det}}(u,v) = \lambda \cdot 0.5$. This is due to the central a-branch of u for which $\mathbf{K}(\mathbf{s}^{\lambda,\mathrm{det}})(0.5\delta_{u_2} + 0.5\delta_{u_3}, \delta_{v_i}) = 0.5$ for $i \in \{1,2\}$, since v_i can simulate only one of processes u_2 and u_3. The cases of $\mathbf{r}^{\lambda,\mathrm{det}}$ and $\mathbf{b}^{\lambda,\mathrm{det}}$ are analogous. However, if we allow v to combine its two a-branches, giving weight 0.5 each, we obtain the combined transition $v \xrightarrow{a}_{\mathrm{c}} 0.5\delta_{v_1} + 0.5\delta_{v_2}$, which clearly matches the central a-branch of u with respect to (bi)simulation. Moreover, any combined transition of u can be matched by the combined transitions of v (and vice versa), thus giving $\mathbf{b}^{\lambda,\mathrm{rand}}(u,v) = \mathbf{r}^{\lambda,\mathrm{rand}}(u,v) = \mathbf{s}^{\lambda,\mathrm{rand}}(u,v) = 0$. ◀

Next, we fix the type of transitions that are considered, and we compare the three metric semantics. As one can expect, the distance given by the bisimilarity metric is greater than that given by the ready similarity metric, which is, in turn, greater than that given by the similarity metric.

Theorem 4. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. Then $\mathbf{s}^{\lambda,\mathrm{x}} < \mathbf{r}^{\lambda,\mathrm{x}} < \mathbf{b}^{\lambda,\mathrm{x}}$.*

Proof. The proof of the non-strict relations $\mathbf{r}^{\lambda,\mathrm{x}} \leq \mathbf{b}^{\lambda,\mathrm{x}}$ and $\mathbf{s}^{\lambda,\mathrm{x}} \leq \mathbf{r}^{\lambda,\mathrm{x}}$ is immediate by Definitions 10 and 12. Then the strictness of the relations follows by Example 3. □

***Example* 3.** Consider processes s, t in Figure 3. Firstly, we have $\mathbf{s}^{\lambda,\mathrm{x}}(s,t) = \lambda \cdot 0.1$, which is obtained by comparing the unique a-move for s with the leftmost a-move for t. Clearly, we have $\mathbf{s}^{\lambda,\mathrm{x}}(s_1,t_1) = \mathbf{s}^{\lambda,\mathrm{x}}(s_2,t_1) = \mathbf{s}^{\lambda,\mathrm{x}}(s_2,t_2) = 0$ and $\mathbf{s}^{\lambda,\mathrm{x}}(s_1,t_2) = 1$, thus giving $\mathbf{K}(\mathbf{s}^{\lambda,\mathrm{x}})(0.6\delta_{s_1} + 0.4\delta_{s_2}, 0.5\delta_{t_1} + 0.5\delta_{t_2}) = 0.1$. Secondly, by comparing the same transitions, we get $\mathbf{r}^{\lambda,\mathrm{x}}(s,t) = \lambda \cdot 0.6$. In fact we have $\mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_1) =$

$\mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_1) = \mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_2) = 1$ and $\mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_2) = 0$, thus giving $\mathbf{K}(\mathbf{r}^{\lambda,\mathrm{x}})(0.6\delta_{s_1} + 0.4\delta_{s_2},\ 0.5\delta_{t_1} + 0.5\delta_{t_2}) = 0.6$. Finally, we show that $\mathbf{b}^{\lambda,\mathrm{x}}(s,t) = \lambda$, due to the rightmost a-move for t. In fact, from analogous calculations to those used to evaluate $\mathbf{r}^{\lambda,\mathrm{x}}(s,t)$, we get that the bisimulation distance between the a-move for s and the leftmost a-move for t is $\lambda \cdot 0.6$. However, we have $\mathbf{b}^{\lambda,\mathrm{x}}(s_1,\mathrm{nil}) = \mathbf{b}^{\lambda,\mathrm{x}}(s_2,\mathrm{nil}) = 1$ and thus, when the rightmost a-move for t is considered we get $\mathbf{K}(\mathbf{b}^{\lambda,\mathrm{x}})(0.6\delta_{s_1} + 0.4\delta_{s_2},\ \delta_{\mathrm{nil}}) = 1$. ◀

4. Metrics for traces

In this section we define the metrics for *trace semantics*. We consider three approaches to the combination of nondeterminism and probability: the *trace distribution* (Section 4.1), the *trace-by-trace* (Section 4.2) and the *supremal probabilities* approach (Section 4.3), and we study their compositional properties (Theorems 7 and 9). Then, in Section 4.4 we pursue the composition of our metric spectrum by comparing the distinguishing power of the trace metrics defined for the three approaches (Theorems 10, 11 and 12).

4.1. The trace distribution approach

In the seminal work [39], the observable events characterizing the trace semantics are the so called *trace distributions*, namely the probability measures over traces that are induced by the resolutions of nondeterminism for processes. Hence, in this approach, each resolution for a process, and thus the scheduler, identifies an observable event. Processes $s,t \in \mathcal{S}$ are then *trace distribution equivalent* if, *for any resolution* for s there is a resolution for t inducing *the same trace distribution*, meaning that the execution probability of each trace in the two resolutions is *exactly the same*, and vice versa.

Definition 14 (Trace distribution equivalence [39])**.** Let $(\mathcal{S},\mathcal{A},\to)$ be a PTS and $\mathrm{x} \in \{\mathrm{det},\mathrm{rand}\}$. Processes $s,t \in \mathcal{S}$ are in the *trace distribution preorder*, written $s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}} t$, if:

$$\text{for each resolution } \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \text{ there is a resolution } \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ such that}$$
$$\text{for each trace } \alpha \in \mathcal{A}^\star \colon \Pr(\mathcal{C}(z_s,\alpha)) = \Pr(\mathcal{C}(z_t,\alpha)).$$

Then, s,t are *trace distribution equivalent*, notation $s \sim^{\mathrm{x}}_{\mathrm{Tr,dis}} t$, if and only if $s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}} t$ and $t \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}} s$.

The quantitative analogue to trace distribution equivalence is based on the evaluation of the *differences in the trace distributions* of processes: the distance between processes s,t is *at most* $\varepsilon \geq 0$ if, *for any resolution* for s there is a resolution for t exhibiting *a trace distribution differing at most by* ε. The difference between *trace distributions* is computed as the greatest difference of probabilities for each trace α multiplied by the *discount* related to α, that is $\lambda^{|\alpha|-1}$. We can observe that *the longer* is a trace, *the lower* will be its contribution to the distance between two processes.

Definition 15 (Trace distribution metric)**.** Let $(\mathcal{S},\mathcal{A},\to)$ be a PTS, $\lambda \in (0,1]$ and $\mathrm{x} \in \{\mathrm{det},\mathrm{rand}\}$. The *trace distribution hemimetric* and the *trace distribution metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all processes $s,t \in \mathcal{S}$ by

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|$$
$$\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = \max\left\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(t,s)\right\}.$$

We observe that the expression $\sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|$ used in Definition 15 corresponds to the (weighted) total variation distance between the trace distributions given by the two resolutions $\mathcal{Z}_s$ and $\mathcal{Z}_t$. An equivalent formulation is given, for finite processes, in [11, 42] via the Kantorovich lifting of the discounted discrete metric over traces. The latter is obtained by identifying each maximal resolution of nondeterminism for a process with the probability distribution over (complete) traces that it induces. Then, taken as ground distance the discounted discrete metric over traces, the distance between two such distributions is obtained via the Kantorovich metric.

We now state that trace distribution hemimetrics and metrics are well-defined and that their kernels are the trace distribution preorder and equivalence, respectively.

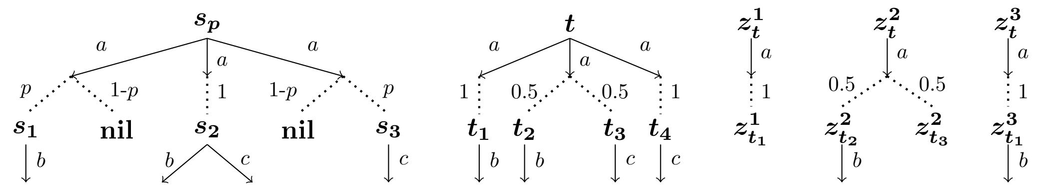

Figure 4: We will evaluate the trace distances between s_p and t with respect to the different approaches, schedulers and parameter $p \in [0, 1]$. In all upcoming examples we will investigate only the traces and the resolutions that are significant for the evaluation of the considered distance.

Theorem 5. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. *Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}}$ *as kernel.*
2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^{\mathrm{x}}_{\mathrm{Tr,dis}}$ *as kernel.*

Proof. The proof can be found in Appendix B.1. □

***Example* 4.** Consider processes s_p and t in Figure 4, with $p \in [0, 1]$. If we focus on deterministic schedulers, the following *trace distributions* can be obtained:

	s_p		t
$(s.1)$	$\{a : 1\}$	$(t.1)$	$\{a : 1\}$
$(s.2)$	$\{a : 1, ab : 1\}$	$(t.2)$	$\{a : 1, ab : 1\}$
$(s.3a)$	$\{a : 1, ab : p\}$	$(t.3a)$	$\{a : 1, ab : 0.5\}$
		$(t.3b)$	$\{a : 1, ab : 0.5, ac : 0.5\}$
$(s.3b)$	$\{a : 1, ac : p\}$	$(t.3c)$	$\{a : 1, ac : 0.5\}$
$(s.4)$	$\{a : 1, ac : 1\}$	$(t.4)$	$\{a : 1, ac : 1\}$

To compute the distance $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s_p, t)$ we have to match each trace distribution π_1 of s_p with a trace distribution π_2 of t that minimizes $\sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1}|\pi_1(\alpha) - \pi_2(\alpha)|$. The latter can be considered as the *distance* between π_1 and π_2. It is easy to see that there is a perfect match between trace distributions $(s.1)$, $(s.2)$ and $(s.4)$ of s_p with the trace distributions $(t.1)$, $(t.2)$ and $(t.4)$ of t, respectively. Let us consider the trace distributions $(s.3a)$ and $(s.3b)$, which are induced by the schedulers selecting the *left-most* and *right-most* a-transitions of s_p, respectively. The distances between these trace distributions of s_p and the ones of t are summarized by in the following table:

	$(t.1)$	$(t.2)$	$(t.3a)$	$(t.3b)$	$(t.3c)$	$(t.4)$
$(s.3a)$	$\lambda \cdot p$	$\lambda \cdot (1-p)$	$\lambda \cdot \lvert p - 0.5 \rvert$	$\lambda \cdot 0.5$	$\lambda \cdot \max\{0.5, p\}$	$\lambda \cdot 1$
$(s.3b)$	$\lambda \cdot p$	$\lambda \cdot 1$	$\lambda \cdot \max\{0.5, p\}$	$\lambda \cdot 0.5$	$\lambda \cdot \lvert p - 0.5 \rvert$	$\lambda \cdot (1-p)$

Then, by simple algebra and by observing that $p \in [0, 1]$, we obtain $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s_p, t) = \lambda \cdot \min\{p, 0.5, 1-p\}$. Similarly, we can easily prove that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t, s_p) = \lambda \cdot 0.5$ and that $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s_p, t) = \lambda \cdot 0.5$.

If we consider randomized schedulers, one can observe that for any $q \in [0, 1]$ the trace distribution $\{a : 1, ab : q, ac : 1-q\}$ can be induced from both s_p and t. This because both s_p and t can perform traces ab and ac with probability 1. This means that a *randomized scheduler* is always able to combine this two resolutions in the appropriate way. Finally, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s_p, t) = \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(t, s_p) = 0$ for any $p \in [0, 1]$. ◀

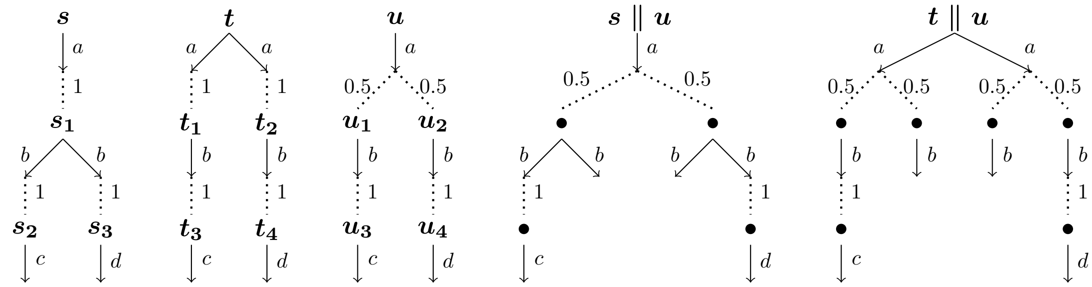

Figure 5: Processes s, t are s.t. $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = 0$. However, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s \parallel u, t \parallel u) = 0.5 \cdot \lambda^2$ and $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s \parallel u, t \parallel u) = 0.25 \cdot \lambda^2$.

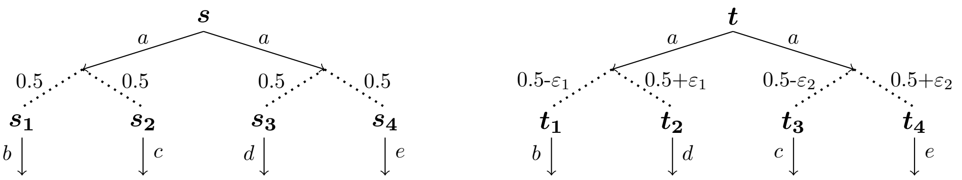

Figure 6: For $\varepsilon_1, \varepsilon_2 \in [0, 0.5]$, we have $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s,t) = \mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) = \lambda \cdot \max(\varepsilon_1, \varepsilon_2)$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) = \lambda \cdot 0.5$ and $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t) = \lambda \cdot \max\{0.25 + \varepsilon_1, 0.25 + \varepsilon_2\}$.

Trace distribution equivalence comes with some desirable properties, such as the full backward compatibility with both the fully nondeterministic and the fully probabilistic cases (cf. [5, Theorem 3.4]). However, it is not a congruence with respect to parallel composition [39], and thus the related metrics cannot be non-expansive. To see this, consider processes s, t in Figure 5. Clearly, we have that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = 0$. However, when we compose each of them in parallel with process u in the same figure, we obtain that $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s \parallel u, t \parallel u) = \lambda^2 \cdot 0.5$ and $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s \parallel u, t \parallel u) = \lambda^2 \cdot 0.25$. This is due to the duplication phenomenon that we witness mixing internal nondeterminism and probability: processes are discriminated by the order of occurrence of the nondeterministic and probabilistic choices.

Moreover, due to the crucial role of the schedulers in the discrimination process, trace distribution distances are sometimes too demanding. Take, for example, processes s, t in Figure 6, with $\varepsilon_1, \varepsilon_2 \in [0, 0.5]$. We have $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) = \lambda \cdot 0.5$ and $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t,s) = \lambda \cdot \max_{i \in \{1,2\}} \max\{0.5 - \varepsilon_i, \varepsilon_i\}$, thus giving $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) = \lambda \cdot 0.5$ for all $\varepsilon_1, \varepsilon_2 \in [0, 0.5]$. However, s and t can perform the same traces with probabilities that differ at most by $\max(\varepsilon_1, \varepsilon_2)$. Specially, for $\varepsilon_1, \varepsilon_2 = 0$, s, t would perform the same traces with exactly the same probability. Hence, it would be reasonable for s and t to be considered equivalent for $\varepsilon_i = 0$, and at a trace distance of $\lambda \cdot \max(\varepsilon_1, \varepsilon_2)$ for $\varepsilon_i \in (0, 0.5]$. This example then suggests to change the notion of observable event: from the trace distributions induced by the schedulers, to the classic notion of *trace*. Following the same approach considered in [2], in the next section a *trace-by-trace* approach is considered.

4.2. The trace-by-trace approach

To overcome some of the issues related to *trace-distribution* approach, in [2] an alternative definition of *trace equivalence* has been proposed that is based on the so called *trace-by-trace* (*tbt*) approach. The idea is to choose first the event that we want to observe, namely *a single trace*, and only as a second step we let the scheduler perform its selection: processes s, t are equivalent with respect to the trace-by-trace approach if *for each trace* α, for each resolution for s there is a resolution for t that assigns to α *exactly the same probability*, and vice versa.

Definition 16 (Tbt-trace equivalence [2, 5])**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. We say that processes $s, t \in \mathcal{S}$ are in the *tbt-trace preorder*, written $s \sqsubseteq^{x}_{\mathrm{Tr,tbt}} t$, if

$$\text{for each trace } \alpha \in \mathcal{A}^\star:$$
$$\text{for each resolution } \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \text{ there is a resolution } \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ with } \Pr(\mathcal{C}(z_s, \alpha)) = \Pr(\mathcal{C}(z_t, \alpha)).$$

Then, $s, t \in \mathcal{S}$ are *tbt-trace equivalent*, notation $s \sim^x_{\mathrm{Tr,tbt}} t$, if and only if $s \sqsubseteq^x_{\mathrm{Tr,tbt}} t$ and $t \sqsubseteq^x_{\mathrm{Tr,tbt}} s$.

In [2] it was proved that tbt-trace equivalences enjoy the congruence property and are full backward compatible with the fully nondeterministic and the fully probabilistic cases.

We introduce now the quantitative analogous to tbt-trace equivalences. Processes s, t are at distance $\varepsilon \geq 0$ if, *for each trace* α, for each resolution for s there is a resolution for t such that the two resolutions assign to α probabilities that *differ at most by* ε, and vice versa. Notably, such difference is multiplied by $\lambda^{|\alpha|-1}$ when a discount $\lambda \in (0, 1]$ is applied. This mitigates the role of traces when their length increases.

Definition 17 (Tbt-trace metric)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $\lambda \in (0, 1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. For each trace $\alpha \in \mathcal{A}^\star$, the function $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ is defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \lambda^{|\alpha|-1} \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))|$$

The *tbt-trace hemimetric* and the *tbt-trace metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \sup_{\alpha \in \mathcal{A}^\star} \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t)$$

$$\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \max\left\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t,s)\right\}.$$

It is not hard to see that for processes in Figure 6 we have $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \lambda \cdot \max(\varepsilon_1, \varepsilon_2)$ (and, in particular, $s \sim^x_{\mathrm{Tr,tbt}} t$ if $\varepsilon_1, \varepsilon_2 = 0$). We show now that tbt-trace hemimetrics and metrics are well-defined and that their kernels are the tbt-trace preorder and equivalence, respectively.

Theorem 6. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$*. Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^x_{\mathrm{Tr,tbt}}$ *as kernel.*
2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^x_{\mathrm{Tr,tbt}}$ *as kernel.*

Proof. The proof can be found in Appendix B.2. □

***Example* 5.** Consider Figure 4. In Example 4 we showed that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s_p, t) = \lambda \cdot \min\{p, |0.5 - p|, 1 - p\}$. In this particular case the two hemimetrics $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s_p, t)$ and $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s_p, t)$ coincide, since each resolution for s_p gives positive probability to at most one of the traces ab and ac, so that quantifying on traces before or after having quantified on resolutions becomes irrelevant.

Let us evaluate now $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(t, s_p)$. To this aim, we focus on trace ab and the resolution $\mathcal{Z}_t$ obtained from the central a-branch of t, for which we have $\Pr(\mathcal{C}(z_t, ab)) = 0.5$. We need the resolution $\mathcal{Z}_{s_p}$ for s_p that minimizes $|0.5 - \Pr(\mathcal{C}(z_{s_p}, ab))|$. Since for any resolution $\mathcal{Z}_{s_p}$ for s_p we have $\Pr(\mathcal{C}(z_{s_p}, ab)) \in \{0, p, 1\}$, we infer that the resolution $\mathcal{Z}_{s_p}$ we are looking for satisfies $\Pr(\mathcal{C}(z_{s_p}, ab)) = p$. By considering also the other resolutions for ab and, then, the other traces, we can check that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(t, s_p) = \lambda \cdot |0.5 - p|$. In Example 4 we showed that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t, s_p) = \lambda \cdot 0.5$ for all $p \in [0, 1]$. Hence, we get $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t, s_p) = \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(t, s_p)$ for $p \in \{0, 1\}$, and $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t, s_p) > \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(t, s_p)$ for $p \in (0, 1)$. This disparity is due to the fact that the trace distributions approach forced us to match the resolution for t assigning positive probability to both ab and ac, whereas in the trace-by-trace approach one never considers two traces at the same time.

The same argument used in Example 4 allows us to conclude that with randomized schedulers we have $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s_p, t) = \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(t, s_p) = 0$. ◄

We conclude this section by stating that tbt-trace distances are strictly non-expansive. As a corollary, we re-obtain the (pre)congruence properties for their kernels (proved in [5]).

Theorem 7. *All distances* $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$*,* $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$*,* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$*,* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$ *are strictly non-expansive.*

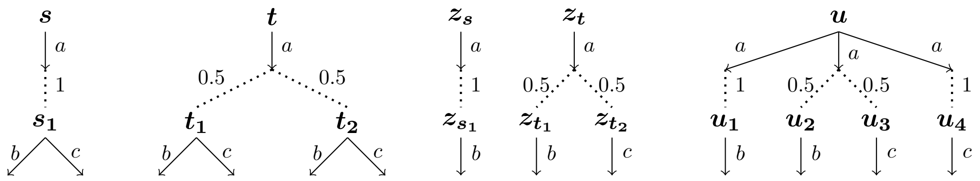

Figure 7: Processes s and t are distinguished by $\sim^{\det}_{\mathrm{Tr,tbt}}$, but related by $\sim^{\det}_{\mathrm{Tr,sup}}$. We remark that t and u are related by all the relations in the three approaches to trace semantics.

Proof. The proof can be found in Appendix B.3. □

The trace-by-trace approach improves on trace distribution approach since it supports equivalences and metrics that are compositional. Moreover, by focusing on traces instead of resolutions, the trace-by-trace approach puts processes in Figure 6 in the expected relations. However, we argue here that trace-by-trace approach on deterministic schedulers still gives some questionable results. Take, for example, processes s, t in Figure 7. We believe that these processes should be equivalent in any semantics approach, since, after performing the action a, they reach two distributions that should be identified, as they assign total probability 1 to states with an identical behavior. But, if we consider the trace ab, the resolution $\mathcal{Z}_t \in \mathrm{Res}^{\det}(t)$ in Figure 7 is such that $\Pr(\mathcal{C}(z_t, ab)) = 0.5$, whereas the unique resolution for s assigning positive probability to ab is $\mathcal{Z}_s$ in Figure 7, for which $\Pr(\mathcal{C}(z_s, ab)) = 1$. Hence no resolution in $\mathrm{Res}^{\det}(s)$ matches $\mathcal{Z}_t$ on trace ab, thus giving $\mathbf{m}^{\lambda,\det}_{\mathrm{Tr,tbt}}(s.t) = \lambda \cdot 0.5$ and, consequently, $s \not\sim^{\det}_{\mathrm{Tr,tbt}} t$. This motivates to look for an alternative approach that allows us to equate processes in Figure 7 and, at the same time, preserves all the desirable properties of the tbt-trace semantics.

4.3. The supremal probabilities approach

The solution proposed in this section takes inspiration from the *extremal probabilities* approach proposed in [3], which bases on the comparison, for each trace α, of both suprema and infima execution probabilities, with respect to resolutions, of α: two processes are equated if they assign the same extremal probabilities to all traces. However, reasoning on infima may cause some arguable results. In particular, it is unclear whether such infima should be evaluated over the whole class of resolutions or over a restricted class, as for instance the resolutions in which the considered trace is actually executed. Besides, desirable properties like the backward compatibility and compositionality are not guaranteed. For all these reasons, we find it more reasonable to define a notion of trace equivalence, and a related metric, based on the comparison of supremal probabilities only.

Notice that, if we focus on *verification*, the comparison of supremal probabilities becomes natural. To exemplify, we let the classical non-probabilistic case guide us. To verify whether a process t satisfies the specification S, we check that whenever S can execute a particular trace, then so does t. Actually, only *positive information* is considered: if there is a resolution for S in which a given trace is executed, then this information is used to verify the equivalence. Still, resolutions in S in which such a trace is not enabled are not considered. The same principle should hold for PTSs: a process should perform all the traces enabled in S and it should do it with *at least* the same probability, in the perspective that the quantitative behavior expressed in the specification expresses the minimal requirements on process behavior.

Focusing on supremal probabilities means relaxing the tbt-trace approach by simply requiring for equivalent processes s, t that *for each trace* α and resolution $\mathcal{Z}_s$ there is a resolution for t assigning to α *at least* the same probability given by $\mathcal{Z}_s$, and vice versa.

Definition 18 (Sup-trace equivalence)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS and $\mathrm{x} \in \{\det, \mathrm{rand}\}$. We say that processes $s, t \in \mathcal{S}$ are in the *sup-trace preorder*, written $s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}} t$, if

$$\text{for each trace } \alpha \in \mathcal{A}^\star : \quad \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \leq \sup_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \Pr(\mathcal{C}(z_t, \alpha)).$$

Then, $s, t \in \mathcal{S}$ are *sup-trace equivalent*, notation $s \sim^{\mathrm{x}}_{\mathrm{Tr,sup}} t$, if and only if $s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}} t$ and $t \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}} s$.

We stress that all good properties of trace-by-trace approach, as the backward compatibility with the fully nondeterministic and fully probabilistic cases and the strict non-expansiveness of the metric with respect to parallel composition, are preserved by the supremal probabilities approach (Proposition 3 and Theorem 9 below). Let $\sim^{\mathbf{N}}_{\mathrm{Tr}}$ denote the trace equivalence on fully nondeterministic systems [6] and $\sim^{\mathbf{P}}_{\mathrm{Tr}}$ denote the one on fully-probabilistic systems [32].

Proposition 3. *Assume a PTS* $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ *and processes* $s, t \in \mathcal{S}$. *Then:*

1. *If* P *is fully-nondeterministic, then* $s \sim^{\mathrm{det}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathrm{rand}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{N}}_{\mathrm{Tr}} t$.
2. *If* P *is fully-probabilistic, then* $s \sim^{\mathrm{det}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathrm{rand}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{P}}_{\mathrm{Tr}} t$.

Proof. The proof can be found in Appendix B.4. □

The idea behind the quantitative analogue of sup-trace equivalence is that two processes are at distance $\varepsilon \geq 0$ if, *for each trace* α, the difference in *supremal* execution probabilities with respect to the resolutions of nondeterminism for the two processes multiplied by $\lambda^{|\alpha|-1}$ is at most ε.

Definition 19 (Sup-trace metric)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $\lambda \in (0, 1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. For each trace $\alpha \in \mathcal{A}^\star$, the function $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ is defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) = \max\Big\{0, \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \Pr(\mathcal{C}(z_t, \alpha))\Big)\Big\}.$$

The *sup-trace hemimetric* and the *sup-trace metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ defined for all processes $s, t \in \mathcal{S}$ by

$$\begin{aligned}
\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) &= \sup_{\alpha \in \mathcal{A}^\star} \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) \\
\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) &= \max\Big\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(t,s)\Big\}.
\end{aligned}$$

We can show that sup-trace hemimetrics and metrics are well-defined and that their kernels are the sup-trace preorders and equivalences, respectively.

Theorem 8. *Assume a PTS* $(\mathcal{S}, \mathcal{A}, \rightarrow)$, $\lambda \in (0, 1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. *Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$, *with* $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}}$ *as kernel.*
2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$, *with* $\sim^{\mathrm{x}}_{\mathrm{Tr,sup}}$ *as kernel.*

Proof. The proof can be found in Appendix B.5. □

We conclude this section by showing that sup-trace distances are strictly non-expansive. As a corollary, we infer the (pre)congruence property of their kernels.

Theorem 9. *All distances* $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$ *are strictly non-expansive.*

Proof. The proof can be found in Appendix B.6. □

Remark 1. We can show that the upper bounds to the distance of composed processes provided in Theorems 7 and 9 are tight, namely for each distance d considered in these theorems, there are processes s_1, s_2, t_1, t_2 with $d((s_1, s_2), (t_1, t_2)) = d(s_1, t_1) + d(s_2, t_2) - d(s_1, t_1) \cdot d(s_2, t_2)$. Indeed, for z_s, z_t in Figure 7, with $\lambda = 1$, we have $d(z_s, z_t) = 0.5$ and $d((z_s, z_s), (z_t, z_t)) = 0.75 = 0.5 + 0.5 - 0.5 \cdot 0.5$.

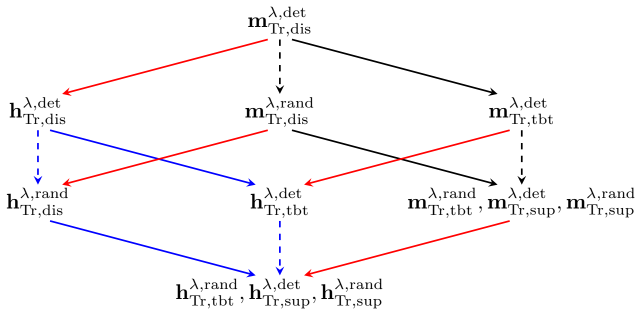

Figure 8: The spectrum of trace distances.

4.4. Comparing the distinguishing power of trace metrics

So far, we have discussed the properties of trace-based behavioral distances under different approaches. Our aim is now to compare these distances with respect to their distinguishing power. By applying the same ordering relation used in Section 3.2, we will obtain the spectrum in Figure 8.

For trace distributions and trace-by-trace semantics, the distances evaluated on deterministic schedulers are more discriminating than their randomized analogues.

Theorem 10. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$*,* $\mathrm{y} \in \{\mathrm{dis}, \mathrm{tbt}\}$ *and* $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$*. Then* $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,y}} < \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Tr,y}}$ *.*

Proof. In Appendix B.7 we show that $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,y}} \leq \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Tr,y}}$. Moreover, we can consider the processes s, t in Figure 7 with $s \sim^{\mathrm{rand}}_{\mathrm{Tr,dis}} t$, $s \sim^{\mathrm{rand}}_{\mathrm{Tr,tbt}} t$, $s \not\sim^{\mathrm{det}}_{\mathrm{Tr,dis}} t$, $s \not\sim^{\mathrm{det}}_{\mathrm{Tr,tbt}} t$ that witness the strictness of the relations. □

As a corollary of Theorem 10, by using the relations between distances and equivalences in Theorems 5 and 6, we re-obtain the relations $\sim^{\mathrm{det}}_{\mathrm{Tr,dis}} \subset \sim^{\mathrm{rand}}_{\mathrm{Tr,dis}}$ and $\sim^{\mathrm{det}}_{\mathrm{Tr,tbt}} \subset \sim^{\mathrm{rand}}_{\mathrm{Tr,tbt}}$ proved in [5]. Moreover, also the analogous results for preorders follow.

As one can expect, the metrics on trace distributions are more discriminating than their corresponding ones in the trace-by-trace approach.

Theorem 11. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$*,* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ *and* $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$*. Then* $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$*.*

Proof. The proof can be found in Appendix B.8 where we show that $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \leq \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$. The strictness of the relations follows by considering the processes in Figure 6. □

As a corollary, by using the kernel relations given in Theorems 5 and 6, we re-obtain the relation $\sim^{\mathrm{x}}_{\mathrm{Tr,dis}} \subset \sim^{x}_{\mathrm{Tr,tbt}}$ proved in [5] and we get $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}} \subset \sqsubseteq^{x}_{\mathrm{Tr,tbt}}$.

Theorem 10 states that the distances on deterministic schedulers are more discriminating than those on randomized ones, and Theorem 11 states that the distances on trace distributions are more discriminating than those in the trace-by-trace approach. A natural question is how $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}$ and $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$ are related.

***Example* 6.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}$ *and* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$*.*

Consider processes s, t in Figure 6. We have $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s, t) = \lambda \cdot \max\{0.25 + \varepsilon_1, 0.25 + \varepsilon_2\}$ and $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s, t) = \lambda \cdot \max(\varepsilon_1, \varepsilon_2)$. Conversely, for s, t in Figure 7 we have $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s, t) = 0$ and $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s, t) = \lambda \cdot 0.5$. ◀

We focus now on supremal probabilities approach, that comes with a particularly interesting result: the sup-trace metric on deterministic schedulers coincides with tbt-trace metrics on randomized schedulers. Moreover, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$ coincides also with its randomized version.

Theorem 12. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$ *and* $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$*. Then* $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} = \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}} = \mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$.

Proof. The proof can be found in Appendix B.9. □

This result is fundamental in the perspective of the application of our trace metrics to process verification: by comparing solely the suprema execution probabilities of the linear properties of interest we get the same expressive power as a pairwise comparison of the probabilities in all possible randomized resolutions of nondeterminism.

Clearly, Theorem 12 together with the kernel relations from Theorems 8 and 6 imply that the relations for the supremal probabilities semantics coincide with those for the tbt-trace semantics with respect to randomized schedulers, ie. $\sqsubseteq^{\mathrm{det}}_{\mathrm{Tr,sup}}=\sqsubseteq^{\mathrm{rand}}_{\mathrm{Tr,sup}}=\sqsubseteq^{\mathrm{rand}}_{\mathrm{Tr,tbt}}$ and $\sim^{\mathrm{det}}_{\mathrm{Tr,sup}}=\sim^{\mathrm{rand}}_{\mathrm{Tr,sup}}=\sim^{\mathrm{rand}}_{\mathrm{Tr,tbt}}$.

5. Metrics for testing

Testing semantics [16] compares processes according to their capacity to *pass* a test. The latter is a PTS equipped with a distinguished state indicating the *success* of the test.

Definition 20 (NPT)**.** A *nondeterministic probabilistic test transition system* (NPT) is a finite PTS $(\mathbf{O}, \mathcal{A}, \rightarrow)$ where $\mathbf{O}$ is a set of processes, called *tests*, containing a distinguished *success process* $\surd$ with no outgoing transitions. We say that a computation from $o \in \mathbf{O}$ is *successful* if and only if its last state is $\surd$.

Given a process s and a test o, we can consider the *interaction system* among the two. This models the response of the process to the application of the test, so that s *passes* the test o if there is a computation in the interaction system that reaches $\surd$. Informally, the interaction system is the result of the parallel composition of the process with the test.

Definition 21 (Interaction system)**.** The *interaction system* of a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ and an NPT $O = (\mathbf{O}, \mathcal{A}, \rightarrow_{\mathbf{O}})$ is the PTS $P \parallel O = (\mathcal{S} \times \mathbf{O}, \mathcal{A}, \rightarrow')$ where: (i) $(s, o) \in \mathcal{S} \times \mathbf{O}$ is called a *configuration* and is *successful* if and only if $o = \surd$; (ii) a computation from $(s, o) \in \mathcal{S} \times \mathbf{O}$ is *successful* if and only if its last configuration is successful; (iii) $\rightarrow'$ is obtained from $\rightarrow$ and $\rightarrow_{\mathbf{O}}$ as described in Definition 2.

For a configuration (s, o) and a resolution $\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{x}}(s, o)$, with $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$, we let $\mathbf{SC}(z_{s,o})$ be the set of *successful computations* from $z_{s,o}$. Then, for a trace $\alpha \in \mathcal{A}^\star$, $\mathbf{SC}(z_{s,o}, \alpha)$ is the set of α-compatible successful computations from $z_{s,o}$.

Probabilistic testing semantics should compare processes with respect to their probability to pass a test. In this Section we consider three approaches to it: (i) the *may/must* (Section 5.1), (ii) the *trace-by-trace* (Section 5.2), and (iii) the *supremal probabilities* (Section 5.3). For each approach, we present (hemi,pseudo)metrics that measure the differences in the behavior of processes when they interact with tests. We study the non-expansiveness of these distances (Theorems 14, 16 and 18), and, in Section 5.4 we compare their discriminating powers (Theorems 19 and 20). To the best of our knowledge, ours is the first attempt in this direction.

5.1. The may/must approach

In the original work on nondeterministic systems [16], testing equivalence was defined via the *may* and *must* preorders. The former expresses the ability of processes to pass a test. The latter expresses the impossibility to fail a test. When also probability is considered, these two preorders are defined, respectively, in terms of *suprema* and *infima* success probabilities [46].

Definition 22 (May/must testing equivalence, [46])**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \rightarrow_O)$ an NPT and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. We say that processes $s, t \in \mathcal{S}$ are in the *may testing preorder*, written $s \sqsubseteq^{x}_{\mathrm{Te,may}} t$, if for each test $o \in \mathbf{O}$

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o})) \le \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o})).$$

Then, $s, t \in \mathcal{S}$ are *may testing equivalent*, written $s \sim^{x}_{\mathrm{Te,may}} t$, if and only if $s \sqsubseteq^{x}_{\mathrm{Te,may}} t$ and $t \sqsubseteq^{x}_{\mathrm{Te,may}} s$.

The notions of *must testing preorder*, $\sqsubseteq^{x}_{\mathrm{Te,must}}$, and *must testing equivalence*, $\sim^{x}_{\mathrm{Te,must}}$, are obtained by replacing the suprema in $\sqsubseteq^{x}_{\mathrm{Te,may}}$ with infima.

Finally, we say that $s, t \in \mathcal{S}$ are in the *may/must testing preorder*, written $s \sqsubseteq^{x}_{\mathrm{Te,mM}} t$, if $s \sqsubseteq^{x}_{\mathrm{Te,may}} t$ and $s \sqsubseteq^{x}_{\mathrm{Te,must}} t$. They are *may/must testing equivalent*, written $s \sim^{x}_{\mathrm{Te,mM}} t$, if and only if $s \sqsubseteq^{x}_{\mathrm{Te,mM}} t$ and $t \sqsubseteq^{x}_{\mathrm{Te,mM}} s$.

The quantitative analogue to may/must testing equivalence bases on the evaluation of the differences in the extremal success probabilities. The may (resp. must) distance between $s, t \in \mathcal{S}$ is *at most* $\varepsilon \geq 0$ if the difference in the suprema (resp. infima) success probabilities with respect to all resolutions of nondeterminism for s and t is *at most* ε. For a correct evaluation of the discounted distances, also with respect to the other metrics discussed in the paper, we need to consider success probabilities in a step-by-step fashion. Given any maximal resolution $\mathcal{Z}_{s,o}$ for (s, o), we let $\Pr^n(\mathbf{SC}(z_{s,o}))$ be the probability of $z_{s,o}$ to reach a successful configuration in exactly n computation steps. Notice that $\Pr(\mathbf{SC}(z_{s,o})) = \sum_{n=1}^{\infty} \Pr^n(\mathbf{SC}(z_{s,o}))$. Then, accordingly to the general discounting policy, for each $n \geq 1$, we will apply a discount of λ^{n-1} to $\Pr^n(\mathbf{SC}(\cdot))$.

Definition 23. [May/must testing metric] Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \to_O)$ an NPT, $\lambda \in (0, 1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. For each test $o \in \mathbf{O}$, the function $\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ is defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t) = \max\Big\{0, \Big(\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr^n(\mathbf{SC}(z_{s,o})) - \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr^n(\mathbf{SC}(z_{t,o})) \Big) \Big\}.$$

Function $\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,must}} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ is obtained by replacing the suprema in $\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}$ with infima. Given $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}\}$, the y *testing hemimetric* and the y *testing metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ defined for all processes $s, t \in \mathcal{S}$ by

$$\begin{aligned} \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}(s,t) &= \sup_{o \in \mathbf{O}} \mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,y}}(s,t) \\ \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}(s,t) &= \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}(t,s)\}. \end{aligned}$$

The *may/must testing hemimetric* and the *may/must testing metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ defined for all processes $s, t \in \mathcal{S}$ by

$$\begin{aligned} \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}(s,t) &= \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t)\} \\ \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}(s,t) &= \max\{\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t), \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t)\}. \end{aligned}$$

We now state that may/must testing hemimetrics and metrics are well-defined and that their kernels are the may/must testing preorder and equivalence, respectively.

Theorem 13. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0, 1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}, \mathrm{mM}\}$:*

1. *The function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{\mathrm{x}}_{\mathrm{Te,y}}$ as kernel.*
2. *The function $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{\mathrm{x}}_{\mathrm{Te,y}}$ as kernel.*

Proof. The proof can be found in Appendix C.1. □

***Example* 7.** Consider processes t, u in Figure 7 and their interactions with test o_1 in Figure 9. Let $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. If we compare the infima success probabilities, we get $\inf_{\mathcal{Z}_{t,o_1} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o_1)} \Pr(\mathbf{SC}(z_{t,o_1})) = 1$ since (t, o_1) has only one maximal resolution corresponding to (t, o_1) itself and that with probability 1 reaches $\surd$. Still, $\inf_{\mathcal{Z}_{u,o_1} \in \mathrm{Res}^{\mathrm{x}}_{\max}(u,o_1)} \Pr(\mathbf{SC}(z_{u,o_1})) = 0$, given by the maximal resolution corresponding to $(u, o_1) \xrightarrow{a}\!\!\!\!\rightarrow$ nil. Hence, $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, u) = |\lambda \cdot 1 - \lambda \cdot 0| = \lambda$. Conversely, to evaluate the may testing distance between t, u, consider their interactions with the test o_2 in the same figure. Due to the duplication phenomenon induced by o_2 on u, we get $\sup_{\mathcal{Z}_{t,o_2} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o_2)} \Pr(\mathbf{SC}(z_{t,o_2})) = 1$ and $\sup_{\mathcal{Z}_{u,o_2} \in \mathrm{Res}^{\mathrm{x}}_{\max}(u,o_2)} \Pr(\mathbf{SC}(z_{u,o_2})) = 0.5$, from which we obtain $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s, t) = 0.5 \cdot \lambda$. ◀

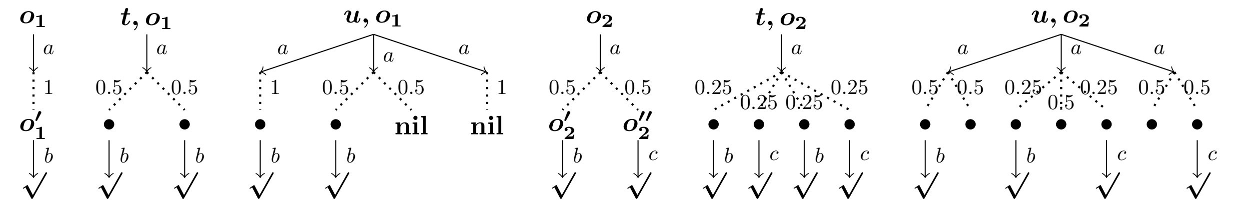

Figure 9: We use the tests o_1, o_2 to evaluate the distance between processes s, t, u in Figure 7 with respect to testing semantics. • represents a generic configuration in the interaction system. In all upcoming examples we will consider only the tests and traces that are significant for the evaluations of the testing metrics.

We can finally observe that the may/must testing (hemi)metrics are non-expansive. As a corollary, we re-obtain the (pre)congruence properties of their kernels.

Theorem 14. *Let* $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}, \mathrm{mM}\}$. *All distances* $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Te,y}}$, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,y}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,y}}$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,y}}$ *are non-expansive.*

Proof. The proof can be found in Appendix C.2. □

5.2. The trace-by-trace approach

In [5] it was proved that the may/must testing is fully backward compatible with the restricted class of processes only if the same restriction is applied to the class of tests, namely if we consider respectively fully nondeterministic and fully probabilistic tests only. This is due to the duplication ability of nondeterministic probabilistic tests. However, by applying the trace-by-trace approach to testing semantics, we regain the full backward compatibility with respect to all tests (cf. [5, Thm. 5.4]).

Definition 24 (Tbt-testing equivalence, [5])**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \rightarrow_O)$ an NPT and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. We say that processes $s, t \in \mathcal{S}$ are in *tbt-testing preorder*, written $s \sqsubseteq^{x}_{\mathrm{Te,tbt}} t$, if

$$\text{for each test } o \in \mathbf{O} \text{ and trace } \alpha \in \mathcal{A}^\star:$$
$$\text{for each } \mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o) \text{ there is } \mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o) \text{ with } \Pr(\mathbf{SC}(z_{s,o},\alpha)) = \Pr(\mathbf{SC}(z_{t,o},\alpha)).$$

Then, $s, t \in \mathcal{S}$ are *tbt-testing equivalent*, notation $s \sim^{x}_{\mathrm{Te,tbt}} t$, if and only if $s \sqsubseteq^{x}_{\mathrm{Te,tbt}} t$ and $t \sqsubseteq^{x}_{\mathrm{Te,tbt}} s$.

The definition of the *tbt-testing metric* naturally follows from Definition 17.

Definition 25 (Tbt-testing metric)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \rightarrow_O)$ an NPT, $\lambda \in (0,1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. For each test $o \in \mathbf{O}$ and trace $\alpha \in \mathcal{A}^\star$, the function $\mathbf{h}^{o,\alpha,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}} : \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{o,\alpha,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = \lambda^{|\alpha|-1} \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} |\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))|$$

The *tbt-testing hemimetric* and the *tbt-testing metric* are the functions $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}} : \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = \sup_{o \in \mathbf{O}} \sup_{\alpha \in \mathcal{A}^\star} \mathbf{h}^{o,\alpha,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t)$$
$$\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = \max\left\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(t,s)\right\}.$$

We now state that tbt-testing hemimetrics and metrics are well-defined and that their kernels are the tbt-testing preorder and equivalence, respectively.

Theorem 15. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0,1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. *Then:*

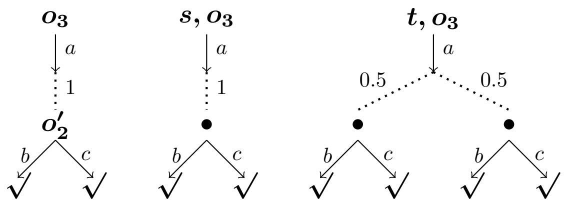

Figure 10: Processes s, t are such that $\mathbf{m}^{\lambda,\text{det}}_{\text{Te,tbt}}(s,t) = 0.5 \cdot \lambda$ and $\mathbf{m}^{\lambda,\text{rand}}_{\text{Te,tbt}}(s,t) = 0$.

1. *The function* $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,tbt}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^{x}_{\text{Te,tbt}}$ *as kernel.*

2. *The function* $\mathbf{m}^{\lambda,\text{x}}_{\text{Te,tbt}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^{x}_{\text{Te,tbt}}$ *as kernel.*

Proof. The proof can be found in Appendix C.3. □

***Example* 8.** Consider processes s, t in Figure 7 and their interactions with test o_3 in Figure 10. The same reasoning detailed in the last paragraph of Section 4.2, gives $\mathbf{m}^{\lambda,\text{det}}_{\text{Te,tbt}}(s,t) = \lambda \cdot 0.5$ and $\mathbf{m}^{\lambda,\text{rand}}_{\text{Te,tbt}}(s,t) = 0$. ◀

When the *tbt*-approach is used to define testing metrics, we get a refinement of the non-expansiveness property to strict non-expansiveness. As a corollary, we re-obtain the (pre)congruence properties of their kernels (proved in [5]).

Theorem 16. *All distances* $\mathbf{h}^{\lambda,\text{det}}_{\text{Te,tbt}}$, $\mathbf{h}^{\lambda,\text{rand}}_{\text{Te,tbt}}$, $\mathbf{m}^{\lambda,\text{det}}_{\text{Te,tbt}}$, $\mathbf{m}^{\lambda,\text{rand}}_{\text{Te,tbt}}$ *are strictly non-expansive.*

Proof. The proof can be found in Appendix C.4. □

5.3. The supremal probabilities approach

If we focus on verification, we can use the testing semantics to verify whether a process will behave as intended by its specification in all possible environments, as modeled by the interaction with tests. Informally, we could see each test as a set of requests of the environment to the system: the ones ending in the success state are those that must be answered. The interaction of the specification with the test then tells us whether the system is able to provide those answers. Thus, an implementation has to guarantee *at least* all the answers provided by the specification. For this reason we decided to introduce also a *supremal probabilities* variant of testing semantics: for each test and for each trace we compare the suprema with respect to all resolutions of nondeterminism of the probabilities of processes to reach success by performing the considered trace.

Definition 26 (Sup-testing equivalence)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \rightarrow_O)$ an NPT and $\text{x} \in \{\text{det}, \text{rand}\}$. We say that processes $s, t \in \mathcal{S}$ are in the *sup-testing preorder*, written $s \sqsubseteq^{\text{x}}_{\text{Te,sup}} t$, if for each test $o \in \mathbf{O}$ and trace $\alpha \in \mathcal{A}^\star$

$$\sup_{\mathcal{Z}_{s,o} \in \text{Res}^{\text{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o}, \alpha)) \leq \sup_{\mathcal{Z}_{t,o} \in \text{Res}^{\text{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o}, \alpha)).$$

Then, $s, t \in \mathcal{S}$ are *sup-testing equivalent*, notation $s \sim^{\text{x}}_{\text{Te,sup}} t$, if and only if $s \sqsubseteq^{\text{x}}_{\text{Te,sup}} t$ and $t \sqsubseteq^{\text{x}}_{\text{Te,sup}} s$.

We obtain the sup-*testing metric* as a direct adaptation to tests of Definition 19.

Definition 27 (Sup-testing metric)**.** Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $(\mathbf{O}, \mathcal{A}, \rightarrow_O)$ an NPT, $\lambda \in (0, 1]$ and $\text{x} \in \{\text{det}, \text{rand}\}$. For each test $o \in \mathbf{O}$ and trace $\alpha \in \mathcal{A}^\star$, the function $\mathbf{h}^{o,\alpha,\lambda,\text{x}}_{\text{Te,sup}} : \mathcal{S} \times \mathcal{S} \to [0, 1]$ is defined for all processes $s, t \in \mathcal{S}$ by

$$\mathbf{h}^{o,\alpha,\lambda,\text{x}}_{\text{Te,sup}}(s,t) = \max\Big\{0, \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_{s,o} \in \text{Res}^{\text{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o}, \alpha)) - \sup_{\mathcal{Z}_{t,o} \in \text{Res}^{\text{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o}, \alpha))\Big)\Big\}.$$

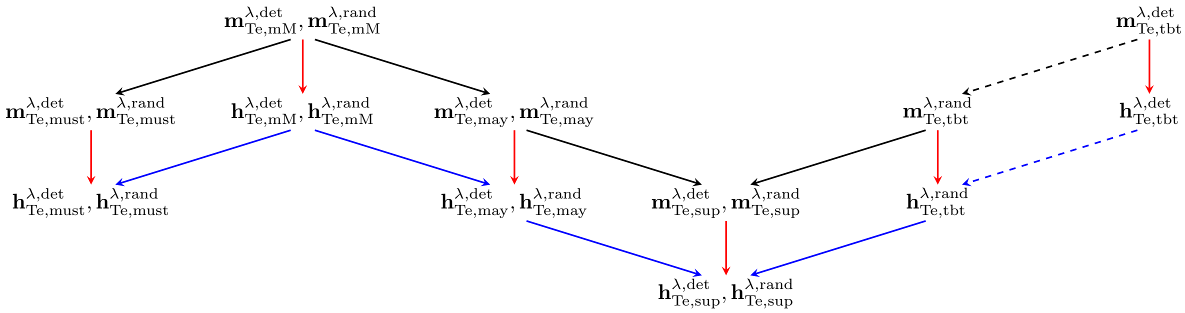

Figure 11: The spectrum of testing metrics

The *sup-testing hemimetric* and the *sup-testing metric* are the functions $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}, \mathbf{m}^{\lambda,\text{x}}_{\text{Te,sup}} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all processes $s,t \in \mathcal{S}$ by

$$\mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}(s,t) = \sup_{o \in \mathbf{O}} \sup_{\alpha \in \mathcal{A}^\star} \mathbf{h}^{o,\alpha,\lambda,\text{x}}_{\text{Te,sup}}(s,t)$$
$$\mathbf{m}^{\lambda,\text{x}}_{\text{Te,sup}}(s,t) = \max\{\mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}(s,t), \mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}(t,s)\}.$$

We now state that sup-testing hemimetrics and metrics are well-defined and that their kernels are the sup-testing preorder and equivalence, respectively.

Theorem 17. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$ and $\text{x} \in \{\text{det}, \text{rand}\}$. Then:*

1. *The function $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{\text{x}}_{\text{Te,sup}}$ as kernel.*
2. *The function $\mathbf{m}^{\lambda,\text{x}}_{\text{Te,sup}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{\text{x}}_{\text{Te,sup}}$ as kernel.*

Proof. The proof can be found in Appendix C.5. □

Finally, we can show that both $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,sup}}$ and $\mathbf{m}^{\lambda,\text{x}}_{\text{Te,sup}}$ are strictly non-expansive. As a corollary, we obtain the (pre)congruence properties of their kernels.

Theorem 18. *All distances $\mathbf{h}^{\lambda,\text{det}}_{\text{Te,sup}}$, $\mathbf{h}^{\lambda,\text{rand}}_{\text{Te,sup}}$, $\mathbf{m}^{\lambda,\text{det}}_{\text{Te,sup}}$, $\mathbf{m}^{\lambda,\text{rand}}_{\text{Te,sup}}$ are strictly non-expansive.*

Proof. The proof can be found in Appendix C.6. □

Remark 2. For all distances d considered in Theorems 14, 16, 18 and processes z_s, z_t in Figure 7, with $\lambda = 1$, we have $d(z_s, z_t) = 0.5$ and $d(z_s \parallel z_s, z_t \parallel z_t) = 0.75 = 0.5 + 0.5 - 0.5 \cdot 0.5$. Hence, the upper bounds to the distance between composed processes provided in Theorems 16 and 18 are tight. We leave as a future work the analogous result for distances considered in Theorem 14.

5.4. Comparing the distinguishing power of testing metrics

In this section we compare the testing distances with respect to their distinguishing power, thus obtaining the spectrum in Figure 11.

In the may/must and in the supremal probabilities approach, the distance evaluated on deterministic and randomized schedulers coincide. Notice that in the supremal probabilities approach, we already observed this fact with trace semantics (cf. Theorem 12). As regards the tbt-testing semantics, the distances evaluated on deterministic schedulers are more discriminating than their randomized analogues, analogously to what happens in the case of trace semantics (cf. Theorem 10).

Theorem 19. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$, $\text{y} \in \{\text{may}, \text{must}, \text{mM}\}$ and $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$:*

1. $\mathbf{d}^{\lambda,\text{rand}}_{\text{Te,y}} = \mathbf{d}^{\lambda,\text{det}}_{\text{Te,y}}$.

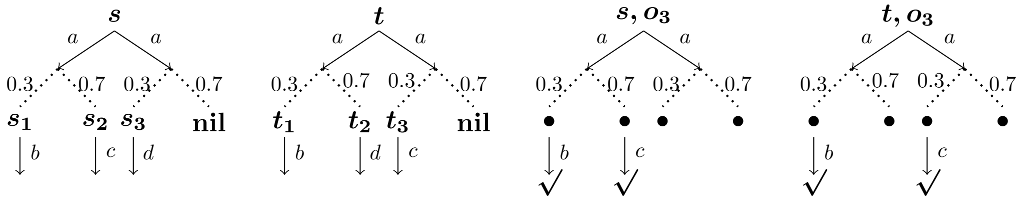

Figure 12: Processes s, t and their interaction systems with the test o_3 in Figure 10.

2. $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$.

3. $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Te,sup}} = \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Te,sup}}$.

Proof. The proof of the relations in Theorem 19.1 and 19.3 and of the non strict relation $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} \leq \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$ in Theorem 19.2 can be found in Appendix C.7. Then, the strict version of the relation in Theorem 19.2 follows from Example 8. □

From Theorem 19, by using the kernel relations in Theorems 13 and 15, we regain relations $\sim^{\mathrm{rand}}_{\mathrm{Te,may}} = \sim^{\mathrm{det}}_{\mathrm{Te,may}}$, $\sim^{\mathrm{rand}}_{\mathrm{Te,must}} = \sim^{\mathrm{det}}_{\mathrm{Te,must}}$, $\sim^{\mathrm{rand}}_{\mathrm{Te,mM}} = \sim^{\mathrm{det}}_{\mathrm{Te,mM}}$, $\sim^{\mathrm{det}}_{\mathrm{Te,tbt}} \subset \sim^{\mathrm{rand}}_{\mathrm{Te,tbt}}$, and their analogues on preorders, proved in [5]. From Theorem 17 we get $\sqsubseteq^{\mathrm{rand}}_{\mathrm{Te,sup}} = \sqsubseteq^{\mathrm{det}}_{\mathrm{Te,sup}}$ and $\sim^{\mathrm{rand}}_{\mathrm{Te,sup}} = \sim^{\mathrm{det}}_{\mathrm{Te,sup}}$.

The metrics on the may/must approach are more discriminating that their corresponding ones in the supremal probabilities approach. As already observed in the trace semantics (cf. Theorem 11), the metrics on trace-by-trace approach are more discriminating that their corresponding ones in the supremal probabilities approach.

Theorem 20. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ *and* $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$*:*

1. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$ *and* $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$.

2. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$.

3. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Proof. Given any relation $d' < d$ in Theorem 20, the proof of the non-strict relation $d' \leq d$ is presented in Appendix C.8. Then, the strict relation $d' < d$ follows from: (i) Example 9 for Theorem 20.1; (ii) Example 12 for Theorem 20.2; (iii) Example 13 for Theorem 20.3. □

The following Examples prove the strictness of the inequalities in Theorem 20 and the (possible) non comparability of the testing (hemi)metrics. For simplicity, we consider only the cases of the metrics.

***Example* 9.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$ *with* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$.

This example provides the strictness of the relation in Theorem 20.1 by showing that the may metric is not comparable to the must metric.

In Example 7 we showed that for t, u in Figure 7 from their interaction with the test o_1 in Figure 9 we obtain $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, u) = \lambda$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(t, u) = 0.5 \cdot \lambda$.

Consider now s, t and their interactions in Figure 12 with the test o_3 from Figure 10. Clearly, we have $\sup_{\mathcal{Z}_{s,o_3} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o_3)} \Pr(\mathbf{SC}(z_{s,o_3})) = 1$ and $\sup_{\mathcal{Z}_{t,o_3} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o_3)} \Pr(\mathbf{SC}(z_{t,o_3})) = 0.3$ and thus $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s, t) = 0.7 \cdot \lambda$. Conversely, if we consider infima success probabilities, we have $\inf_{\mathcal{Z}_{s,o_3} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o_3)} \Pr(\mathbf{SC}(z_{s,o_3})) = 0$ and $\inf_{\mathcal{Z}_{t,o_3} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o_3)} \Pr(\mathbf{SC}(z_{t,o_3})) = 0.3$. Thus, $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s, t) = 0.3 \cdot \lambda$. ◀

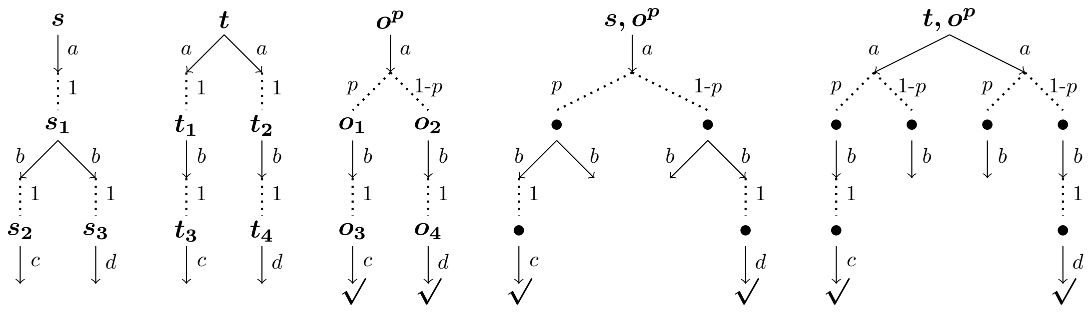

Figure 13: Processes s, t are such that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = 0.5 \cdot \lambda$, as witnessed by the test $o^{1/2}$.

***Example* 10.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$ *with* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$ *and* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

We start with $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$. Form Example 7 we know that for t, u in Figure 7 it holds $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t,u) = \lambda$. Since both t and u have maximal resolutions giving probability 1 to either ab or ac, we get $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}(t,u) = 0$. Consider now s, t in Figure 12. In Example 9 we showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = 0.3 \cdot \lambda$. From the interaction systems in Figure 12, by considering the suprema success probabilities of trace ac, we obtain $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} = 0.4 \cdot \lambda$.

Next we deal with the tbt-testing metrics. Consider s, t in Figure 13 and the family of tests $O = \{o^p \mid p \in (0,1)\}$, each duplicating the actions b in the interaction with s and t. For each $o^p \in O$, $\inf_{\mathcal{Z}_{s,o^p} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o^p)} \Pr(\mathbf{SC}(z_{s,o^p})) = 0$ and $\inf_{\mathcal{Z}_{t,o^p} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o^p)} \Pr(\mathbf{SC}(z_{t,o^p})) = \min\{p, 1-p\}$, thus giving $\mathbf{h}^{o^p,\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t,s) = \lambda^2 \cdot \min\{p, 1-p\}$. One can then easily check that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t,s) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = \lambda^2 \cdot \sup_{p \in (0,1)} \min\{p, 1-p\} = 0.5 \cdot \lambda^2$. Conversely, as the tbt-testing metric compares the success probabilities related to the execution of a single trace per time, we get $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0$. In the case of randomized schedulers, all the resolutions for t, o^p combining the two a-moves can be matched by s, o^p by combining the b-moves and vice versa. Consider now s, t in Figure 12. Even under randomized schedulers, the tbt-testing distance on them is given by the difference in the success probability of the trace ac (or equivalently ad) and thus $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0.4 \cdot \lambda$. However, we have already showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = 0.3 \cdot \lambda$. ◀

***Example* 11.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$ *with* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Consider processes s, t in Figure 13. In Example 10 we showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0$. However, the same reasoning giving $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = 0.5 \cdot \lambda^2$, can be applied on suprema success probabilities, thus giving $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t) = 0.5 \cdot \lambda^2$. Consider now t, u in Figure 7 and their interactions with test o_1 in Figure 9. As we consider maximal resolutions only, for both classes of schedulers, the success probability of trace ab evaluates to 1 on t, o_1, whereas on u, o_1 it evaluates to 0, due to the maximal resolution corresponding to the rightmost a-branch. Hence $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(t,u) = \lambda$, whereas we already showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(t,u) = 0.5 \cdot \lambda$. ◀

***Example* 12.** *Strictness of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$.

Consider s, t in Figure 12. In Example 9 we have shown that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t) = 0.7 \cdot \lambda$. However, since the supremal probability approach to testing proceeds in a trace-by-trace fashion, the sup-testing distance is given by the difference in the success probability of the trace ac (or ad) and thus $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}(s,t) = 0.4 \cdot \lambda$. ◀

***Example* 13.** *Strictness of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Consider now t, u in Figure 7 and their interactions with test o_1 in Figure 9. In Example 11 we have shown that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(t,u) = \lambda$. However, one can easily check that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}(t,u) = 0$.

We stress that the strictness of the relation in Theorem 20.3 is due to the restriction to maximal resolutions, necessary to reason on testing semantics. ◀

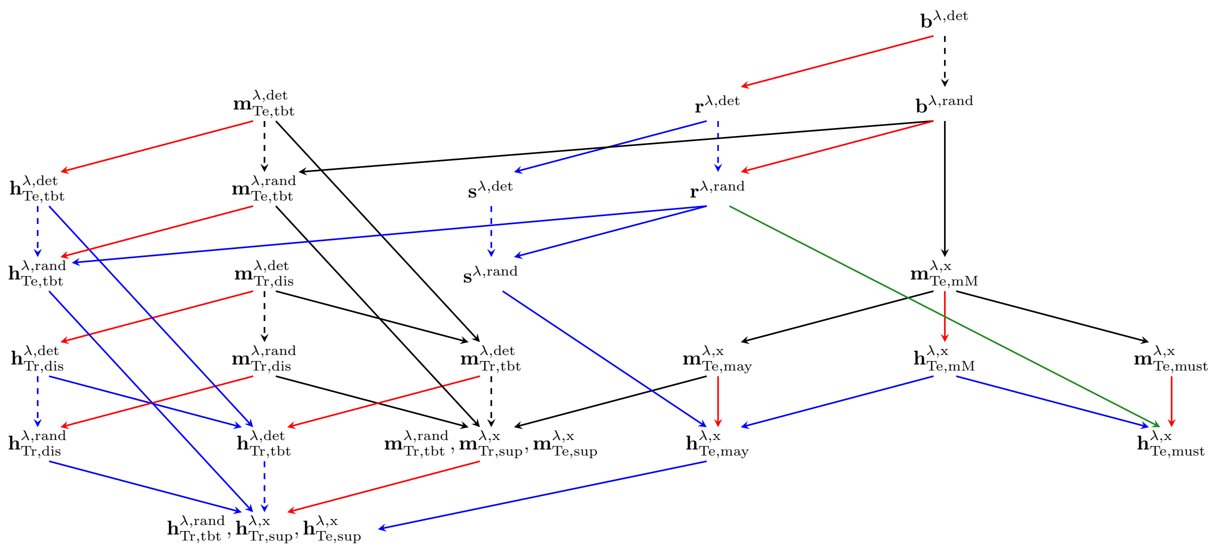

Figure 14: The metric linear time - branching time spectrum. The green arrow $\mathbf{r}^{\lambda,\text{rand}} \rightarrow \mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}$ specifies that the relation $\mathbf{r}^{\lambda,\text{rand}} > \mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}$ holds by reversing the order of processes.

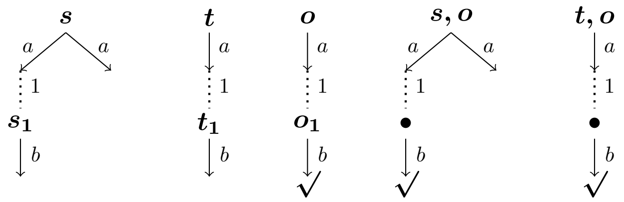

Figure 15: Processes s, t are such that $\mathbf{r}^{\lambda,\text{x}}(s,t) = \mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(t,s) = \lambda$ and $\mathbf{r}^{\lambda,\text{x}}(t,s) = \mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(s,t) = 0$.

6. The metric linear time - branching time spectrum

In this section we compare the distinguishing power of all the metrics discussed so far and combine the spectra obtained in Sections 3–5 into the first *metric linear time - branching time spectrum* presented in Figure 14. The connections among the three spectra are stated in Theorem 21.

An interesting result regards the relation between the ready similarity and the must testing metric semantics. We have that for all processes $s, t \in \mathcal{S}$ it holds that $\mathbf{r}^{\lambda,\text{x}}(s,t) \geq \mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(t,s)$, whereas $\mathbf{r}^{\lambda,\text{x}}(s,t)$ and $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(s,t)$ are, in general, not comparable. As shown in the following Example 14 this process *inversion* is due to the fact that for $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(t,s) > 0$ the differences in infima success probabilities of t and s allow us to detect the possible presence of actions that are performed only by t, which means that t cannot ready simulate s, namely $\mathbf{r}^{\lambda,\text{x}}(s,t) > 0$.

***Example* 14.** Consider processes s, t in Figure 15. Clearly, $\mathbf{r}^{\lambda,\text{x}}(s,t) = \lambda$, due to the presence of the transition $s \xrightarrow{a} \delta_{\text{nil}}$. Conversely, as t corresponds to the leftmost a-branch of s, we have $\mathbf{r}^{\lambda,\text{x}}(t,s) = 0$.

Consider now the interaction of s, t with the test o in the same figure, that tests for the ability of performing the trace ab. We have that $\inf_{\mathcal{Z}_{s,o} \in \text{Res}^{\text{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o})) = 0$, due to the maximal computation $(s,o) \xrightarrow{a} \delta_{(\text{nil},o_1)}$, whereas $\inf_{\mathcal{Z}_{t,o} \in \text{Res}^{\text{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o})) = 1$, as (t,o) has only one maximal resolution which corresponds to a successful computation. Hence, we infer $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(s,t) = 0$ and $\mathbf{h}^{\lambda,\text{x}}_{\text{Te,must}}(t,s) = \lambda$. ◀

Theorem 21. *Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $\lambda \in (0,1]$, $\text{x} \in \{\text{det}, \text{rand}\}$ and $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$:*

1. $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}} < \mathbf{s}^{\lambda,\mathrm{rand}}$.

2. *For all* $s, t \in \mathcal{S}$ *it holds* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, s) \le \mathbf{r}^{\lambda,\mathrm{rand}}(s, t)$*, and there are* $u, v \in \mathcal{S}$ *such that* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(u, v) < \mathbf{r}^{\lambda,\mathrm{rand}}(v, u)$.

3. $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.

4. $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{r}^{\lambda,\mathrm{rand}}$ *and* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.

5. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

6. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} = \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$.

Proof. Given any relation $d' < d$ in Theorem 21, the proof of the non-strict relation $d' \le d$ is presented in Appendix D.1. Then, the strict relation $d' < d$ follows from: (i) Example 15 for Theorem 21.1; (ii) Example 17 for Theorem 21.2; (iii) Example 18 for Theorem 21.3; (iv) Example 20 for Theorem 21.4; (v) Example 23 for Theorem 21.5. □

The following examples prove the strictness of the inequalities in Theorem 21 and the non comparability of the distances shown in Figure 14. For simplicity, when possible, we consider only the cases of the metrics.

***Example* 15.** *Strictness of* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}} < \mathbf{s}^{\lambda,\mathrm{rand}}$.

Consider processes s, t in Figure 13. From Examples 10 and 11 we can infer that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s, t) = 0.5 \cdot \lambda^2$. However, it is easy to check that $\mathbf{s}^{\lambda,\mathrm{x}}(s, t) = \lambda$, due to the nondeterministic choice of s_1. ◀

***Example* 16.** *Non comparability of* $\mathbf{s}^{\lambda,\mathrm{x}}$ *with* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$ *and* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Consider processes s, t in Figure 13. In previous Example 15, we obtained that $\mathbf{s}^{\lambda,\mathrm{x}}(s, t) = \lambda$. However, it is easy to check that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s, t) = \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s, t) = 0$.

Consider now processes s, t in Figure 15. In Example 14 we showed that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, s) = \lambda$ and by applying the same reasoning we obtain $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s, t) = \lambda$ (notice the inversion of the two processes). However, since process nil is simulated by any process, we have that $\mathbf{s}^{\lambda,\mathrm{x}}(s, t) = \mathbf{s}^{\lambda,\mathrm{x}}(t, s) = 0$. ◀

***Example* 17.** *Existence of* $s, t \in \mathcal{S}$ *such that* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, s) < \mathbf{r}^{\lambda,\mathrm{rand}}(s, t)$.

Consider processes s, t in Figure 13. In Example 10 we showed that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t, s) = 0.5 \cdot \lambda^2$. However, we clearly have that $\mathbf{r}^{\lambda,\mathrm{x}}(s, t) = \lambda$. ◀

***Example* 18.** *Strictness of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.

The same reasoning used in previous Examples 15 and 17 allows us to conclude that for processes s, t in Figure 13 we have $\mathbf{b}^{\lambda,\mathrm{x}}(s, t) = \lambda$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}(s, t) = 0.5 \cdot \lambda^2$. ◀

The following example stresses the non comparability of bisimilarity metric with respect to trace distribution and (testing) trace-by-trace metrics under deterministic schedulers. Conversely, as a consequence of Theorem 21, the novel supremal probabilities approach to trace and testing metrics fits in with bisimilarity as expected.

***Example* 19.** *Non comparability of* $\mathbf{b}^{\lambda,\mathrm{det}}$ *with* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$ *and* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$.

As already discussed at the end of Section 4.2 and in Example 8 processes s, t in Figure 7 are such that $\mathbf{b}^{\lambda,\mathrm{det}}(s, t) = 0$, whereas $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s, t) = \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s, t) = \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}(s, t) = 0.5 \cdot \lambda$. Conversely, by considering processes s, t in Figure 15 we have that $\mathbf{b}^{\lambda,\mathrm{det}}(s, t) = \lambda$, whereas clearly $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s, t) = \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s, t) = 0$. Similarly, processes s, t in Figure 13 are such that $\mathbf{b}^{\lambda,\mathrm{det}}(s, t) = \lambda$, whereas in Example 10 we derived $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}(s, t) = 0$. ◀

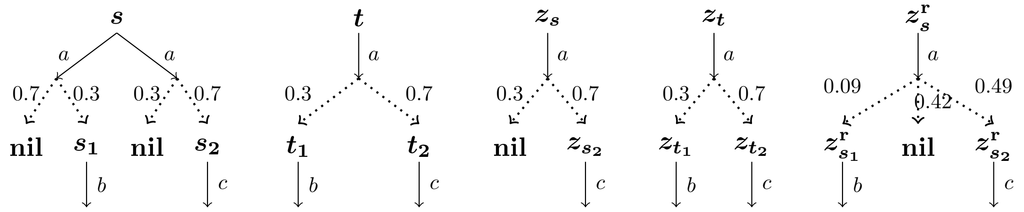

Figure 16: Processes s, t are such that $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) = 0.3 \cdot \lambda$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t) = 0.21 \cdot \lambda$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0$.

***Example* 20.** *Strictness of* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.

Consider processes s, t in Figure 13. Clearly we have that $\mathbf{b}^{\lambda,\mathrm{rand}}(s,t) = \lambda$, whereas in Example 10 we derived $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t) = 0$. ◀

***Example* 21.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$ *with* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}$ *and* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$.

For processes s, t in Figure 13, we showed, in Example 11, that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t) = 0.5 \cdot \lambda^2$. However, as both processes have a single resolution each allowing them to execute either trace *abc* or *abd*, we can infer that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = 0$. Notice that this also shows the strictness of the relation $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}} < \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$. Consider now s, t in Figure 7. As discussed in Section 4.2 we have that $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}} = \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s,t) = 0.5 \cdot \lambda$. However, one can easily check that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t) = 0$. ◀

***Example* 22.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$ *with* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$, $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ *and* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$.

Consider processes t, u in Figure 7. Clearly, $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(t,u) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t,u) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(t,u) = 0$. However, in Example 7 we showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t,u) = \lambda$. Consider now s, t in Figure 12. From Example 9 we have $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t) = 0.3 \cdot \lambda$, but clearly $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) = 0.4 \cdot \lambda$. ◀

***Example* 23.** *Strictness of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} < \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Consider processes t, u in Figure 7 and their interactions with the test o_1 in Figure 9. Clearly, we have $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t,u) = 0$, whereas $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(t,u) = \lambda$, as discussed in Example 11. ◀

***Example* 24.** *Non comparability of* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$ *and* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$.

Consider processes s, t in Figure 16. We have that $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) = 0.3 \cdot \lambda$ in that s will use the resolution $\mathcal{Z}_s$, in the same Figure, corresponding to its rightmost a-branch to match the resolution $\mathcal{Z}_t$ for t corresponding to the process itself, and reported in the same Figure. Interestingly, we can show that this distance is lowered when randomized schedulers are considered. In fact if we match the resolution $\mathcal{Z}_t$ with the randomized resolution $\mathcal{Z}^{\mathrm{r}}_s$ in Figure 16, obtained by applying weights 0.3 and 0.7 to the resolutions for s corresponding, respectively, to its leftmost and rightmost a-branches, we obtain that $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t) = 0.21 \cdot \lambda$. However, it is easy to see that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t) = 0$, since success probabilities are evaluated in the trace-by-trace fashion.

Consider now processes t, u in Figure 7 and their interactions with the test o_1 in Figure 9. We already noticed in Figure 7 that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(t,u) = 0$. Still, in Example 11 we showed that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(t,u) = \lambda$. ◀

7. Related and future work

Trace metrics have been thoroughly studied on quantitative systems, as testified by the spectrum of distances, defined as the generalization of a chosen trace distance, in [22] and the one on Metric Transition Systems (MTSs) in [14]. In [22], the spectrum is obtained by applying to LTSs the theory of quantitative Ehrenfeucht-Fraïssé games [21, 23]. However, our results on PTSs cannot be obtained from the ones in [22] since the considered metric semantics are quite different and the *well-behavedness* property assumed for the metrics in [22] does not hold for distances on PTSs. Notably, in [14] the trace distance is based on a *propositional distance* defined over *valuations* of atomic propositions that characterize the MTS. Although such valuation could play the role of the probability distributions in the PTS, it is unclear whether we could combine the ground distance on atomic propositions and the propositional distance, to obtain trace distances comparable to ours. In [1, 13] trace metrics on Markov Chains (MCs) are defined as total variation distances on the cones generated by traces. As in MCs probability depends only on the current state and not on nondeterminism, our quantification over resolutions would be trivial on MCs, giving a total variation distance as well.

Although ours is the first proposal of a metric expressing testing semantics, testing equivalences for probabilistic processes have been studied also in [2, 3, 18]. In detail, [18] proposed notions of probabilistic *may/must testing* for a Kleisli lifting of the PTS model, i.e., the transition relation is lifted to a relation $(\to)^\dagger \subseteq (\Delta(\mathcal{S}) \times \mathcal{A} \times \Delta(\mathcal{S}))$ taking distributions over processes to distributions over processes. Interestingly, they prove that the so obtained may testing preorder coincides with *forward similarity* [40], namely the simulation preorder obtained on those lifted transitions, and that the must testing preorder coincides with the *forward failure similarity*, obtained as the lifting of the failure simulation preorder. Again, the disparity in the two models prevents us from thoroughly comparing the proposed testing relations. Our intuition is that the metrics defined for *forward* (bi)simulation semantics would result into less discriminating metrics with respect to those for (bi)simulation semantics and that, due to the duplication phenomenon introduced by probabilistic tests, the equality between non-zero forward and testing metrics could no longer be guaranteed. We leave as future work a thorough investigation on this issue.

As future work, we aim at extending the metric spectrum to *decorated* trace semantics and to metrics on different semantic models, and to study their logical characterizations and compositional properties, along the same line proposed in [8–10, 12, 25–27] for bisimulation semantics. Further, in light of the relation among the supremal and randomized trace-by-trace metrics, we aim to provide efficient algorithms for the evaluation of the proposed metrics and to develop a tool for quantitative process verification: we will use the distance between a process and its specification to quantify how much that process satisfies a given property.

Acknowledgements

We thank the anonymous reviewers for their valuable comments that helped us to improve the paper.

References

[1] Bacci, G., Bacci, G., Larsen, K. G., Mardare, R., 2015. Converging from branching to linear metrics on Markov chains. In: Proc. ICTAC 2015. pp. 349–367.

[2] Bernardo, M., De Nicola, R., Loreti, M., 2012. Revisiting trace and testing equivalences for nondeterministic and probabilistic processes. In: Proc. FoSSaCS 2012. pp. 195–209.

[3] Bernardo, M., De Nicola, R., Loreti, M., 2013. The spectrum of strong behavioral equivalences for nondeterministic and probabilistic processes. In: Proc. QAPL 2013. pp. 81–96.

[4] Bernardo, M., De Nicola, R., Loreti, M., 2014. Relating strong behavioral equivalences for processes with nondeterminism and probabilities. Theor. Comput. Sci. 546, 63–92.

[5] Bernardo, M., De Nicola, R., Loreti, M., 2014. Revisiting trace and testing equivalences for nondeterministic and probabilistic processes. Logical Methods in Computer Science 10 (1).

[6] Brookes, S. D., Hoare, C. A. R., Roscoe, A. W., 1984. A theory of communicating sequential processes. J. ACM 31 (3), 560–599.

[7] Castiglioni, V., 2018. Trace and testing metrics on nondeterministic probabilistic processes. In: Proc. EXPRESS/SOS 2018. Vol. 276 of EPTCS. pp. 19–36.

[8] Castiglioni, V., Gebler, D., Tini, S., 2016. Logical characterization of bisimulation metrics. In: Proc. QAPL'16. Vol. 227 of EPTCS. pp. 44–62.

[9] Castiglioni, V., Gebler, D., Tini, S., 2016. Modal decomposition on nondeterministic probabilistic processes. In: Proc. CONCUR 2016. Vol. 49 of LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, pp. 36:1–36:15.

[10] Castiglioni, V., Gebler, D., Tini, S., 2018. SOS-based modal decomposition on nondeterministic probabilistic processes. Logical Methods in Computer Science 14 (2).

[11] Castiglioni, V., Tini, S., 2017. Logical characterization of trace metrics. In: Proc. QAPL@ETAPS 2017. Vol. 250 of EPTCS. pp. 39–74.

[12] Castiglioni, V., Tini, S., 2019. Logical characterization of branching metrics for nondeterministic probabilistic transition systems. Inf. and Comp.

[13] Daca, P., Henzinger, T. A., Křetínský, J., Petrov, T., 2016. Linear distances between Markov chains. In: Proc. CONCUR 2016. pp. 20:1–20:15.
[14] de Alfaro, L., Faella, M., Stoelinga, M., 2009. Linear and branching system metrics. IEEE Trans. Software Eng. 35 (2), 258–273.
[15] de Alfaro, L., Henzinger, T. A., Majumdar, R., 2003. Discounting the Future in Systems Theory. In: Proc. ICALP'03. ICALP '03. Springer, pp. 1022–1037.
[16] De Nicola, R., Hennessy, M., 1984. Testing equivalences for processes. Theor. Comput. Sci. 34, 83–133.
[17] Deng, Y., Chothia, T., Palamidessi, C., Pang, J., 2006. Metrics for action-labelled quantitative transition systems. Electr. Notes Theor. Comput. Sci. 153 (2), 79–96.
[18] Deng, Y., van Glabbeek, R. J., Hennessy, M., Morgan, C., 2008. Characterising testing preorders for finite probabilistic processes. Logical Methods in Computer Science 4 (4).
[19] Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P., 2004. Metrics for labelled Markov processes. Theor. Comput. Sci. 318 (3), 323–354.
[20] Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P., 2002. The metric analogue of weak bisimulation for probabilistic processes. In: Proc. LICS 2002. pp. 413–422.
[21] Ehrenfeucht, A., 1961. An application of games to the completeness problem for formalized theories. Fundamenta Mathematicae 49, 129–141.
[22] Fahrenberg, U., Legay, A., 2014. The quantitative linear-time-branching-time spectrum. Theor. Comput. Sci. 538, 54–69.
[23] Fraïssé, R., 1954. Sur quelques classifications des systèms de relations. Publications Scientifiques de l'Université d'Alger, Série A 1, 35–182.
[24] Gebler, D., Larsen, K. G., Tini, S., 2016. Compositional bisimulation metric reasoning with probabilistic process calculi. Logical Methods in Computer Science 12 (4).
[25] Gebler, D., Tini, S., 2013. Compositionality of approximate bisimulation for probabilistic systems. In: Proc. EXPRESS/SOS 2013. Vol. 120 of EPTCS. pp. 32–46.
[26] Gebler, D., Tini, S., 2014. Fixed-point characterization of compositionality properties of probabilistic processes combinators. In: Proc. EXPRESS/SOS 2014. Vol. 160 of EPTCS. pp. 63–78.
[27] Gebler, D., Tini, S., 2018. SOS specifications for uniformly continuous operators. Journal of Computer and System Sciences 92, 113–151.
[28] Georgievska, S., Andova, S., 2012. Probabilistic may/must testing: retaining probabilities by restricted schedulers. Formal Asp. Comput. 24 (4-6), 727–748.
[29] Giacalone, A., Jou, C.-C., Smolka, S. A., 1990. Algebraic reasoning for probabilistic concurrent systems. In: Proc. IFIP Work, Conf. on Programming, Concepts and Methods. pp. 443–458.
[30] Hermanns, H., Parma, A., Segala, R., Wachter, B., Zhang, L., 2011. Probabilistic logical characterization. Inf. Comput. 209 (2), 154–172.
[31] Hoare, A., 1985. Communicating Sequential Processes. Prentice-Hall.
[32] Jou, C., Smolka, S. A., 1990. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In: Proc. CONCUR '90. Vol. 458 of Lecture Notes in Computer Science. pp. 367–383.
[33] Kantorovich, L. V., 1942. On the transfer of masses. Doklady Akademii Nauk 37 (2), 227–229.
[34] Keller, R. M., 1976. Formal verification of parallel programs. Commun. ACM 19 (7), 371–384.
[35] Kwiatkowska, M. Z., Norman, G., 1996. Probabilistic metric semantics for a simple language with recursion. In: Proc. MFCS'96. pp. 419–430.
[36] Lanotte, R., Tini, S., 2018. Weak bisimulation metrics in models with nondeterminism and continuous state spaces. In: Proc. ICTAC 2018. Vol. 11187 of LNCS. Springer, pp. 292–312.
[37] Larsen, K. G., Skou, A., 1991. Bisimulation through probabilistic testing. Inf. Comput. 94 (1), 1–28.
[38] Panangaden, P., 2009. Labelled Markov Processes. Imperial College Press.
[39] Segala, R., 1995. A compositional trace-based semantics for probabilistic automata. In: Proc. CONCUR '95. pp. 234–248.
[40] Segala, R., 1995. Modeling and verification of randomized distributed real-time systems. Ph.D. thesis, MIT.
[41] Segala, R., Lynch, N. A., 1995. Probabilistic simulations for probabilistic processes. Nord. J. Comput. 2 (2), 250–273.
[42] Song, L., Deng, Y., Cai, X., 2007. Towards automatic measurement of probabilistic processes. In: Proc. QSIC 2007. pp. 50–59.
[43] van Breugel, F., 2012. On behavioural pseudometrics and closure ordinals. Inf. Process. Lett. 112 (19), 715–718.
[44] van Breugel, F., Worrell, J., 2005. A behavioural pseudometric for probabilistic transition systems. Theor. Comput. Sci. 331 (1), 115–142.
[45] Wolovick, N., Johr, S., 2006. A characterization of meaningful schedulers for continuous-time markov decision processes. In: Proc. FORMATS 2006. pp. 352–367.
[46] Yi, W., Larsen, K. G., 1992. Testing probabilistic and nondeterministic processes. In: Proc. PSTV'92 IFIP Transactions C-8. pp. 47–61.

Appendix A. Proofs of Section 3

Appendix A.1. Proof of Proposition 1

Proposition 1. *Assume an image finite PTS in which, for each transition $s \xrightarrow{a} \pi$, π is a distribution with finite support. Let $\mathbf{F} \in \{\mathbf{B}^{\lambda,\mathrm{x}}, \mathbf{R}^{\lambda,\mathrm{x}}, \mathbf{S}^{\lambda,\mathrm{x}}\}$ for $\lambda \in (0,1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. Then, given any $d_1, d_2 \in \mathcal{D}(\mathcal{S})$ with $d_2 \preceq d_1$, for all $s,t \in \mathcal{S}$ we have:*

$$\mathbf{F}(d_1)(s,t) - \mathbf{F}(d_2)(s,t) \leq \sup_{u,v \in \mathcal{S}} (d_1(u,v) - d_2(u,v)).$$

Proof of Proposition 1. We expand only the case of $\mathbf{S}^{\lambda,\mathrm{rand}}$. The cases of the other functionals can be obtained similarly. Hence, the proof obligation instantiates as

$$\mathbf{S}^{\lambda,\mathrm{rand}}(d_1)(s,t) - \mathbf{S}^{\lambda,\mathrm{rand}}(d_2)(s,t) \leq \sup_{u,v \in \mathcal{S}} (d_1(u,v) - d_2(u,v)). \tag{A.1}$$

First of all, notice that since we are considering an image finite PTS in which all distributions have finite support, we are guaranteed that also all the distributions that are target of combined transitions will have finite support. This is fundamental to guarantee that the minimum over the matchings for such distributions in the evaluations of the Kantorovich distance is always achieved. Then we have

$$\begin{aligned} & \mathbf{S}^{\lambda,\mathrm{rand}}(d_1)(s,t) - \mathbf{S}^{\lambda,\mathrm{rand}}(d_2)(s,t) \\ = & \sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\pi_s, \pi_t) - \sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_2)(\pi_s, \pi_t) \end{aligned} \tag{A.2}$$

By definition of supremum, given $\varepsilon_1 > 0$ there is a combined transition $s \xrightarrow{a}_{\mathrm{c}} \tilde{\pi}_s$ such that $\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\pi_s, \pi_t) < \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\tilde{\pi}_s, \pi_t) + \varepsilon_1$. Therefore

$$(\mathrm{A.2}) < \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\tilde{\pi}_s, \pi_t) - \sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_2)(\pi_s, \pi_t) + \varepsilon_1 \tag{A.3}$$

By substituting the supremum over combined transitions for s with respect to the lifting of d_2 with the arbitrary combined transition $s \xrightarrow{a}_{\mathrm{c}} \tilde{\pi}_s$ from the previous step, we get

$$(\mathrm{A.3}) \leq \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\tilde{\pi}_s, \pi_t) - \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_2)(\tilde{\pi}_s, \pi_t) + \varepsilon_1 \tag{A.4}$$

By definition of infimum, given $\varepsilon_2 > 0$ there is a combined transition $t \xrightarrow{a}_{\mathrm{c}} \tilde{\pi}_t$ such that $\inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_2)(\tilde{\pi}_s, \pi_t) > \lambda \cdot \mathbf{K}(d_2)(\tilde{\pi}_s, \tilde{\pi}_t) - \varepsilon_2$. Therefore

$$(\mathrm{A.4}) < \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(d_1)(\tilde{\pi}_s, \pi_t) - \lambda \cdot \mathbf{K}(d_2)(\tilde{\pi}_s, \tilde{\pi}_t) + \varepsilon_1 + \varepsilon_2 \tag{A.5}$$

By substituting the infimum over combined transitions for t with respect to the lifting of d_1 with the arbitrary combined transition $t \xrightarrow{a}_{\mathrm{c}} \tilde{\pi}_t$ from the previous step, and letting $\varepsilon = \varepsilon_1 + \varepsilon_2$, we get

$$\begin{aligned} (\mathrm{A.5}) & \leq \lambda \cdot \mathbf{K}(d_1)(\tilde{\pi}_s, \tilde{\pi}_t) - \lambda \cdot \mathbf{K}(d_2)(\tilde{\pi}_s, \tilde{\pi}_t) + \varepsilon \\ & = \lambda \cdot \min_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s, \tilde{\pi}_t)} \sum_{s',t' \in \mathcal{S}} \mathfrak{w}(s',t') \cdot d_1(s',t') - \lambda \cdot \min_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s, \tilde{\pi}_t)} \sum_{s',t' \in \mathcal{S}} \mathfrak{w}(s',t') \cdot d_2(s',t') + \varepsilon \end{aligned} \tag{A.6}$$

By choosing $\tilde{\mathfrak{w}} = \arg\min_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s, \tilde{\pi}_t)} \sum_{s',t' \in \mathcal{S}} \mathfrak{w}(s',t') \cdot d_2(s',t')$, we get

$$(\mathrm{A.6}) = \lambda \cdot \left(\min_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s, \tilde{\pi}_t)} \sum_{s',t' \in \mathcal{S}} \mathfrak{w}(s',t') \cdot d_1(s',t') - \sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') \cdot d_2(s',t') \right) + \varepsilon \tag{A.7}$$

By substituting the minimum over the matchings for $\tilde{\pi}_s, \tilde{\pi}_t$ with respect to d_1 with the matching $\tilde{\mathfrak{w}}$ form the previous step, we get

$$
\begin{aligned}
(A.7) &\leq \lambda \cdot \left(\sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') \cdot d_1(s',t') - \sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') \cdot d_2(s',t') \right) + \varepsilon \\
&= \lambda \cdot \sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') \cdot (d_1(s',t') - d_2(s',t')) + \varepsilon \\
&\leq \lambda \cdot \sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') \cdot \sup_{s,t \in \mathcal{S}} (d_1(s,t) - d_2(s,t)) + \varepsilon
\end{aligned} \tag{A.8}
$$

By noticing that $\lambda \leq 1$ and $\sum_{s',t' \in \mathcal{S}} \tilde{\mathfrak{w}}(s',t') = 1$, we get

$$
(A.8) \leq \sup_{u,v \in \mathcal{S}} (d_1(u,v) - d_2(u,v)) + \varepsilon.
$$

Since the inequality $\mathbf{S}^{\lambda,\mathrm{rand}}(d_1)(s,t) - \mathbf{S}^{\lambda,\mathrm{rand}}(d_2)(s,t) < \sup_{u,v \in \mathcal{S}}(d_1(u,v) - d_2(u,v)) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that Equation (A.1) holds. □

Appendix A.2. Proof of Proposition 2

Proposition 2. *Assume an image finite PTS in which, for each transition $s \xrightarrow{a} \pi$, π is a distribution with finite support. Let $\lambda \in (0,1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\mathbf{d} \in \{\mathbf{b}, \mathbf{r}, \mathbf{s}\}$. Then $\mathbf{d}^{\lambda,\mathrm{x}} = \lim_{k \to \infty} \mathbf{d}_k^{\lambda,\mathrm{x}}$.*

Proof of Proposition 2. We expand only the case of convex bisimulation metrics. The proofs for ready simulation and simulation metrics are analogous.

Since $\mathbf{B}^{\lambda,\mathrm{rand}}$ is monotone and its closure ordinal is ω, due to Proposition 1 and [43, Corollary 1], we can immediately infer that $\lim_{k\to\infty} \mathbf{b}_k^{\lambda,\mathrm{rand}} = \mathbf{b}_\omega^{\lambda,\mathrm{rand}}$ and that $\mathbf{b}_\omega^{\lambda,\mathrm{rand}}$ is a fixed point for $\mathbf{B}^{\lambda,\mathrm{rand}}$. Moreover, by an easy induction over $k \in \mathbb{N}$, we can prove that $\mathbf{b}^{\lambda,\mathrm{rand}} \geq \mathbf{b}_k^{\lambda,\mathrm{rand}}$ for all $k \in \mathbb{N}$. In particular, $\mathbf{b}^{\lambda,\mathrm{rand}} \geq \mathbf{b}_\omega^{\lambda,\mathrm{rand}}$. Therefore, by uniqueness of the least fixed point, we can conclude that $\mathbf{b}_\omega^{\lambda,\mathrm{rand}} = \mathbf{b}^{\lambda,\mathrm{rand}}$. □

Appendix A.3. Proof of Theorem 2

Theorem 2. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\lambda \in (0,1]$. All functions $\mathbf{b}^{\lambda,\mathrm{x}}$, $\mathbf{r}^{\lambda,\mathrm{x}}$ and $\mathbf{s}^{\lambda,\mathrm{x}}$ are strictly non-expansive.*

Proof of Theorem 2. We prove the thesis for $\mathbf{r}^{\lambda,\mathrm{x}}$. The proof for $\mathbf{s}^{\lambda,\mathrm{x}}$ is analogous. First we need to introduce the notion of precongruence closure for hemimetric $\mathbf{r}^{\lambda,\mathrm{x}}$ with respect to operator $\|$ as the quantitative analogue of the well-known concept of precongruence closure of a process preorder. We define the precongruence closure of $\mathbf{r}^{\lambda,\mathrm{x}}$ for operator $\|$ as an hemimetric $d \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined by

$$
d(s,t) = \begin{cases} \min\left\{ d(s_1,t_1) + d(s_2,t_2) - d(s_1,t_1) \cdot d(s_2,t_2), \mathbf{r}^{\lambda,\mathrm{x}}(s,t) \right\} & \text{if } \left[\begin{array}{l} s = s_1 \parallel s_2 \wedge t = t_1 \parallel t_2 \wedge \\ \mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_1), \mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_2) < 1 \end{array} \right. \\ \mathbf{r}^{\lambda,\mathrm{x}}(s,t) & \text{otherwise} \end{cases}
$$

We notice that, by construction, d satisfies the properties $d \preceq \mathbf{r}^{\lambda,\mathrm{x}}$ and $d(s_1 \parallel s_2, t_1 \parallel t_2) \leq d(s_1,s_2) + d(t_1,t_2) - d(s_1,s_2) \cdot d(t_1,t_2)$. Hence, it remains to show that $\mathbf{r}^{\lambda,\mathrm{x}} \preceq d$, thus giving $\mathbf{r}^{\lambda,\mathrm{x}} = d$. Since $\mathbf{r}^{\lambda,\mathrm{x}}$ is the least prefixed point of $\mathbf{R}^\lambda$, to show $\mathbf{r}^{\lambda,\mathrm{x}} \preceq d$ it is enough to prove that d is a prefixed point of $\mathbf{R}^\lambda$.

To prove that $\mathbf{R}^\lambda(d) \preceq d$ we need to show that d satisfies the conditions of the ready bisimulation metrics, namely for all $s, t \in \mathcal{S}$ with $d(s,t) < 1$ we need to show the following two properties:

$$
\forall\, s \xrightarrow{a} \pi_s \; \exists\, t \xrightarrow{a} \pi_t \text{ with } \lambda \cdot \mathbf{K}(d)(\pi_s, \pi_t) \leq d(s,t) \tag{A.9}
$$

$$\text{if } s \stackrel{a}{\not\longrightarrow} \text{ then } t \stackrel{a}{\not\longrightarrow} . \tag{A.10}$$

We prove Equations (A.9) and (A.10) by induction over the overall number k of the occurrences of operator $\|$ in s and t.

Consider the base case $k = 0$. By definition of d, we have that $d(s,t) = \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$. Let us start with Equation A.9. Since $d(s,t) < 1$ and $d(s,t) = \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$, we infer $\mathbf{r}^{\lambda,\mathrm{x}}(s,t) < 1$, therefore we are sure that each transition $s \xrightarrow{a} \pi_s$ is mimicked by some transition $t \xrightarrow{a} \pi_t$ for some distribution $\pi_t \in \Delta(\mathcal{S})$ such that $\lambda \cdot \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{x}})(\pi_s, \pi_t) \leq \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$. Since $\mathbf{K}$ is monotone, from $d \preceq \mathbf{r}^{\lambda,\mathrm{x}}$ we infer $\mathbf{K}(d) \preceq \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{x}})$. Therefore we conclude

$$\lambda \cdot \mathbf{K}(d)(\pi_s,\pi_t) \leq \lambda \cdot \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{x}})(\pi_s,\pi_t) \leq \mathbf{r}^{\lambda,\mathrm{x}}(s,t) = d(s,t)$$

which confirms that Equation (A.9) holds for s and t. Equation (A.10) follows by observing that if $s \stackrel{a}{\not\longrightarrow}$ and $\mathbf{r}^{\lambda,\mathrm{x}}(s,t) < 1$ then we have $t \stackrel{a}{\not\longrightarrow}$.

Consider the inductive step $k > 0$. If s is not of the form $s = s_1 \| s_2$, or t is not of the form $t = t_1 \| t_2$, then by definition of d we have $d(s,t) = \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$ and Equations (A.9) and (A.10) follow precisely as in the base case $k = 0$. Otherwise, if both $s = s_1 \| s_2$ and $t = t_1 \| t_2$, then we distinguish two subcases:

- $d(s,t) = \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$, with $\mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_1) = 1$ or $\mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_2) = 1$ or $\mathbf{r}^{\lambda,\mathrm{x}}(s,t) < d(s_1,t_2) + d(s_2,t_2) - d(s_1,t_2) \cdot d(s_2,t_2)$.
- $d(s,t) = d(s_1,t_1) + d(s_2,t_2) - d(s_1,t_1) \cdot d(s_2,t_2)$, with $\mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_1) < 1$ and $\mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_2) < 1$ and $\mathbf{r}^{\lambda,\mathrm{x}}(s,t) \geq d(s_1,t_1) + d(s_2,t_2) - d(s_1,t_1) \cdot d(s_2,t_2)$.

In subcase $d(s,t) = \mathbf{r}^{\lambda,\mathrm{x}}(s,t)$, Equations (A.9) and (A.10) follow precisely as in the base case $k = 0$.

Consider the subcase $d(s,t) = d(s_1,t_1) + d(s_2,t_2) - d(s_1,t_1) \cdot d(s_2,t_2)$. Let us start with Equation (A.9). We note that a transition $s \xrightarrow{a} \pi$ derives from transitions $s_1 \xrightarrow{a} \pi_s^1$ and $s_2 \xrightarrow{a} \pi_s^2$ with $\pi_s = \pi_s^1 \| \pi_s^2$. By $\mathbf{r}^{\lambda,\mathrm{x}}(s_1,t_1) < 1$, $\mathbf{r}^{\lambda,\mathrm{x}}(s_2,t_2) < 1$ and $d \preceq \mathbf{r}^{\lambda,\mathrm{x}}$, we get $d(s_1,t_1) < 1$ and $d(s_2,t_2) < 1$. By the inductive hypothesis we get that there are also transitions $t_1 \xrightarrow{a} \pi_t^1$ and $t_2 \xrightarrow{a} \pi_t^2$ with $\lambda \cdot \mathbf{K}(d)(\pi_s^1,\pi_t^1) \leq d(s_1,t_1)$ and $\lambda \cdot \mathbf{K}(d)(\pi_s^2,\pi_t^2) \leq d(s_2,t_2)$. Clearly, we can infer that there is also the transition $t_1 \| t_2 \xrightarrow{a} \pi_t^1 \| \pi_t^2$. For $i \in \{1,2\}$, let $\mathfrak{w}_i \in \mathfrak{W}(\pi_s^i,\pi_t^i)$ be the optimal matching for $\mathbf{K}(d)(\pi_s^i,\pi_t^i)$. Notice that, for $i \in \{1,2\}$, since by definition we have $(\pi_s^i \| \pi_t^i)(s_i' \| t_i') = \pi_s^i(s_i') \cdot \pi_t^i(t_i')$, then $\mathfrak{w}_1 \cdot \mathfrak{w}_2$ is trivially a matching for $\pi_s^1 \| \pi_s^2$ and $\pi_t^1 \| \pi_t^2$. Then we have

$$\begin{aligned}&\lambda \cdot \mathbf{K}(d)(\pi_s^1 \| \pi_s^2, \pi_t^1 \| \pi_t^2)\\ &= \lambda \cdot \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s^1\cdot\pi_s^2,\pi_t^1\cdot\pi_t^2)} \sum_{s_i',t_i' \in \mathcal{S}, i\in\{1,2\}} \mathfrak{w}(s_1' \| s_2', t_1' \| t_2') \cdot d(s_1' \| s_2', t_1' \| t_2')\end{aligned} \tag{A.11}$$

By construction of d, we get

$$\text{(A.11)} \leq \lambda \cdot \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s^1\cdot\pi_s^2,\pi_t^1\cdot\pi_t^2)} \sum_{s_i',t_i' \in \mathcal{S}, i\in\{1,2\}} \mathfrak{w}(s_1' \| s_2', t_1' \| t_2') \cdot (d(s_1',t_1') + d(s_2',t_2') - d(s_1',t_1') \cdot d(s_2',t_2')) \tag{A.12}$$

By choosing an arbitrary matching, namely $\mathfrak{w}_1 \cdot \mathfrak{w}_2$, in place of the optimal matching for $\pi_s^1 \cdot \pi_t^1$ and $\pi_s^2 \cdot \pi_t^2$, we get

$$\text{(A.12)} \leq \lambda \cdot \sum_{s_i',t_i' \in \mathcal{S}, i\in\{1,2\}} \mathfrak{w}_1(s_1',t_1') \cdot \mathfrak{w}_2(s_2',t_2') \cdot (d(s_1',t_1') + d(s_2',t_2') - d(s_1',t_1') \cdot d(s_2',t_2')) \tag{A.13}$$

By the independence of $\mathfrak{w}_1$ and $\mathfrak{w}_2$ and the fact that $\sum_{s_i',t_i'} \mathfrak{w}_i(s_i',t_i') = 1$ for $i \in \{1,2\}$, we get

$$\text{(A.13)} = \lambda \cdot \Big(\sum_{s_1' \in \mathsf{supp}(\pi_s^1), t_1' \in \mathsf{supp}(\pi_t^1)} \mathfrak{w}_1(s_1',t_1') \cdot d(s_1',t_1') + \sum_{s_2' \in \mathsf{supp}(\pi_s^2), t_2' \in \mathsf{supp}(\pi_t^2)} \mathfrak{w}_2(s_2',t_2') \cdot d(s_2',t_2') -$$

$$\Big(\sum_{s_1'\in\mathsf{supp}(\pi_s^1),t_1'\in\mathsf{supp}(\pi_t^1)} \mathfrak{w}_1(s_1',t_1')\cdot d(s_1',t_1')\Big)\cdot\Big(\sum_{s_2'\in\mathsf{supp}(\pi_s^2),t_2'\in\mathsf{supp}(\pi_t^2)} \mathfrak{w}_2(s_2',t_2')\cdot d(s_2',t_2')\Big)\Big) \tag{A.14}$$

By the choice of $\mathfrak{w}_1$ and $\mathfrak{w}_2$, we get

$$\begin{aligned}(A.14) &= \lambda\cdot\big(\mathbf{K}(d)(\pi_s^1,\pi_t^1)+\mathbf{K}(d)(\pi_s^2,\pi_t^2)-\mathbf{K}(d)(\pi_s^1,\pi_t^1)\cdot\mathbf{K}(d)(\pi_s^2,\pi_t^2)\big)\\ &\le \lambda\cdot\mathbf{K}(d)(\pi_s^1,\pi_t^1)+\lambda\cdot\mathbf{K}(d)(\pi_s^2,\pi_t^2)-\big(\lambda\cdot\mathbf{K}(d)(\pi_s^1,\pi_t^1)\big)\cdot\big(\lambda\cdot\mathbf{K}(d)(\pi_s^2,\pi_t^2)\big)\end{aligned} \tag{A.15}$$

By the inductive hypothesis and the fact that both $\lambda\cdot\mathbf{K}(d)$ and d are bounded by 1, we get

$$\begin{aligned}(A.15) &\le d(s_1,t_1)+d(s_2,t_2)-d(s_1,t_1)\cdot d(s_2,t_2)\\ &= d(s,t).\end{aligned}$$

Thus, Equation (A.9) is satisfied for d in this case. Consider now Equation (A.10). If $s \stackrel{a}{\not\to}$ then $s_1 \stackrel{a}{\not\to}$ or $s_2 \stackrel{a}{\not\to}$. By the inductive hypothesis, we infer that $t_1 \stackrel{a}{\not\to}$ or $t_2 \stackrel{a}{\not\to}$, from which we can immediately derive that $t \stackrel{a}{\not\to}$, thus confirming that also Equation (A.10) holds. □

Appendix A.4. Proof of Theorem 3

In order to prove Theorem 3 we need to recall that the Kantorovich metric is subadditive with respect to convex combinations of distributions.

Proposition 4 ([38])**.** *Let (X,d) be any metric space and I a finite set of indexes for which $\pi_i,\pi_i'\in\Delta(X)$ and $p_i\in(0,1]$ with $\sum_{i\in I}p_i=1$. Then, $\mathbf{K}(d)(\sum_{i\in I}p_i\pi_i,\sum_{i\in I}p_i\pi_i')\le\sum_{i\in I}p_i\,\mathbf{K}(d)(\pi_i,\pi_i')$.*

Theorem 3. *Let $(\mathcal{S},\mathcal{A},\to)$ be a PTS, $\lambda\in(0,1]$ and $\mathbf{d}\in\{\mathbf{b},\mathbf{r},\mathbf{s}\}$. Then $\mathbf{d}^{\lambda,\mathrm{rand}}<\mathbf{d}^{\lambda,\mathrm{det}}$.*

Proof of Theorem 3. We expand only the case of the similarity metrics. The proofs for ready similarity and bisimilarity metrics can be obtained by similar arguments.

We proceed by induction on $k\in\mathbb{N}$ to show that

$$\mathbf{s}_k^{\lambda,\mathrm{det}}\ge\mathbf{s}_k^{\lambda,\mathrm{rand}} \text{ for all } k\in\mathbb{N} \tag{A.16}$$

then the thesis $\mathbf{s}^{\lambda,\mathrm{det}}\ge\mathbf{s}^{\lambda,\mathrm{rand}}$ will follow by Equation (A.16), Proposition 2 and the monotonicity of the limit.

Consider the base case $k=0$. Given arbitrary processes $s,t\in\mathcal{S}$, we have $\mathbf{s}_0^{\lambda,\mathrm{det}}=\mathbf{s}_0^{\lambda,\mathrm{rand}}=\mathbf{0}$, and thus Equation (A.16) holds in this case.

Consider now the inductive step $k>0$. We have

$$\begin{aligned}\mathbf{s}_k^{\lambda,\mathrm{det}}(s,t) &= \sup_{a\in\mathcal{A}}\ \sup_{s\xrightarrow{a}\pi_s}\ \inf_{t\xrightarrow{a}\pi_t}\ \lambda\cdot\mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{det}})(\pi_s,\pi_t)\\ &\ge \sup_{a\in\mathcal{A}}\ \sup_{s\xrightarrow{a}\pi_s}\ \inf_{t\xrightarrow{a}\pi_t}\ \lambda\cdot\mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_s,\pi_t)\\ &\ge \sup_{a\in\mathcal{A}}\ \sup_{s\xrightarrow{a}\pi_s}\ \inf_{t\xrightarrow{a}_{\mathrm{c}}\pi_t}\ \lambda\cdot\mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_s,\pi_t)\\ &= \sup_{a\in\mathcal{A}}\ \sup_{s\xrightarrow{a}_{\mathrm{c}}\pi_s}\ \inf_{t\xrightarrow{a}_{\mathrm{c}}\pi_t}\ \lambda\cdot\mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_s,\pi_t)\\ &= \mathbf{s}_k^{\lambda,\mathrm{rand}}(s,t)\end{aligned}$$

where:

- the second step follows by the inductive hypothesis and the monotonicity of $\mathbf{K}$;
- the third step follows by the fact that by evaluating the infimum over a wider class of transitions we can obtain a better matching of the transitions by s;
- by letting $f(s,a,\pi_s) = \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_s, \pi_t)$, the fourth step derives from
  - $\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} f(s,a,\pi_s) \geq \sup_{s \xrightarrow{a} \pi'_s} f(s,a,\pi'_s)$. This derives directly from the fact that we are evaluating the supremum over a wider class of transitions.
  - $\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} f(s,a,\pi_s) \geq \sup_{s \xrightarrow{a} \pi'_s} f(s,a,\pi'_s)$. This follows since each a-combined transition from s consists in a convex combination of the distributions reached by a-transitions from s. More formally, for each $s \xrightarrow{a}_{\mathrm{c}} \pi_s$ there is a set of indexes I s.t. $\pi_s = \sum_{i \in I} p_i \pi_s^i$ for weights $p_i \in (0,1]$ with $\sum_{i \in I} p_i = 1$ and distributions $\pi_s^i \in \mathrm{der}(s,a)$. Thus, given any $\varepsilon > 0$, by definition of supremum and Proposition 4, we have

$$
\begin{aligned}
\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} f(s,a,\pi_s) &< f(s,a,\pi_\varepsilon) + \varepsilon \\
&= \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_\varepsilon, \pi_t) + \varepsilon \\
&= \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\sum_{i \in I} p_i \pi_\varepsilon^i, \pi_t) + \varepsilon \\
&\leq \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \sum_{i \in I} p_i \cdot \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_\varepsilon^i, \pi_t) + \varepsilon \\
&\leq \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \max_{i \in I} \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_\varepsilon^i, \pi_t) + \varepsilon \\
&\leq \max_{i \in I} \inf_{t \xrightarrow{a}_{\mathrm{c}} \pi_t} \lambda \cdot \mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\pi_\varepsilon^i, \pi_t) + \varepsilon \\
&\leq \sup_{s \xrightarrow{a} \pi'_s} f(s,a,\pi'_s) + \varepsilon.
\end{aligned}
$$

    Since $\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} f(s,a,\pi_s) < \sup_{s \xrightarrow{a} \pi'_s} f(s,a,\pi'_s) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that $\sup_{s \xrightarrow{a}_{\mathrm{c}} \pi_s} f(s,a,\pi_s) \geq \sup_{s \xrightarrow{a} \pi'_s} f(s,a,\pi'_s)$.

Hence, we can conclude that Equation (A.16) holds also in this case. □

Appendix B. Proofs of Section 4

Appendix B.1. Proof of Theorem 5

Theorem 5. *Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $\lambda \in (0,1]$ and $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. Then:*

1. *The function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}}$ as kernel.*

2. *The function $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{\mathrm{x}}_{\mathrm{Tr,dis}}$ as kernel.*

Proof of Theorem 5. We expand only the case of hemimetrics and their kernels. The proof for metrics and their kernels simply follows by symmetrization.

We start with proving that the function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ defined by

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}} = \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))|$$

is an hemimetric. We reason as follows:

- the function
$$\sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))|$$
is a pseudometric over resolutions, since the function $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by $f(x,y) = |x-y|$ is a pseudometric over reals, a pseudometrics multiplied by a constant is a pseudometrics, and the supremum of a set of pseudometrics is a pseudometrics;

- for any pseudometric d on resolutions, the function over processes assigning to s and t the value
$$\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} d(z_s, z_t)$$
is the asymmetric version of the Hausdorff distance, and thus an hemimetric.

Then, the 1-boundedness property of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ follows by $\lambda \in (0,1]$ and

$$\begin{aligned} & \Pr(\mathcal{C}(z_s, \alpha)) \le 1 \text{ for all } \alpha \in \mathcal{A}^\star, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \\ \Rightarrow\ & |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))| \le 1 \text{ for all } \alpha \in \mathcal{A}^\star, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s), \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \\ \Rightarrow\ & \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))| \le 1. \end{aligned}$$

Finally, we prove that the kernel of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ is $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}}$ by

$$\begin{aligned} & s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,dis}} t \\ \Longleftrightarrow\ & \forall\, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)\ \ \exists\, \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ s.t. } \forall\, \alpha \in \mathcal{A}^\star\ \Pr(\mathcal{C}(z_s, \alpha)) = \Pr(\mathcal{C}(z_t, \alpha)) \\ \Longleftrightarrow\ & \forall\, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)\ \ \exists\, \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ s.t. } \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))| = 0 \\ \Longleftrightarrow\ & \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s, \alpha)) - \Pr(\mathcal{C}(z_t, \alpha))| = 0 \\ \Longleftrightarrow\ & \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) = 0. \end{aligned}$$

□

Appendix B.2. Proof of Theorem 6

Theorem 6. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0, 1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$*. Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^{x}_{\mathrm{Tr,tbt}}$ *as kernel.*
2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^{x}_{\mathrm{Tr,tbt}}$ *as kernel.*

Proof of Theorem 6. We show that for $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$, the function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{x}_{\mathrm{Tr,tbt}}$ as kernel. From this result, directly from the definition of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ and of $\sim^{x}_{\mathrm{Tr,tbt}}$, follows that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{x}_{\mathrm{Tr,tbt}}$ as kernel.

The proof that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ is a 1-bounded hemimetric in analogous to the proof for $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ (see Appendix B.1) being a 1-bounded hemimetric.

For the kernel, we have

$$
\begin{aligned}
& s \sqsubseteq^{x}_{\mathrm{Tr,tbt}} t \\
\iff & \forall\, \alpha \in \mathcal{A}^{\star} \;\; \forall\, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \;\; \exists\, \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ s.t. } \Pr(\mathcal{C}(z_s,\alpha)) = \Pr(\mathcal{C}(z_t,\alpha)) \\
\iff & \forall\, \alpha \in \mathcal{A}^{\star} \;\; \forall\, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \;\; \exists\, \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ s.t. } \lambda^{|\alpha|-1}|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| = 0 \\
\iff & \forall\, \alpha \in \mathcal{A}^{\star} \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \lambda^{|\alpha|-1}|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| = 0 \\
\iff & \sup_{\alpha \in \mathcal{A}^{\star}} \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \lambda^{|\alpha|-1}|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| = 0 \\
\iff & \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = 0.
\end{aligned}
$$

□

Appendix B.3. Proof of Theorem 7

Theorem 7. *All distances* $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$ *are strictly non-expansive.*

Proof of Theorem 7. We start with $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$. Assume $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1,t_1) = \varepsilon_1$ and $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2,t_2) = \varepsilon_2$. Then, for each trace $\alpha \in \mathcal{A}^{\star}$, we have $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1,t_1) \le \varepsilon_1$ and $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2,t_2) \le \varepsilon_2$, thus implying:

- $\sup_{\mathcal{Z}_{s_1} \in \mathrm{Res}^{\mathrm{x}}(s_1)} \inf_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} |\Pr(\mathcal{C}(z_{s_1},\alpha)) - \Pr(\mathcal{C}(z_{t_1},\alpha))| \le \frac{\varepsilon_1}{\lambda^{|\alpha|-1}}$,
- $\sup_{\mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)} \inf_{\mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2)} |\Pr(\mathcal{C}(z_{s_2},\alpha)) - \Pr(\mathcal{C}(z_{t_2},\alpha))| \le \frac{\varepsilon_2}{\lambda^{|\alpha|-1}}$.

Therefore, by definition of infimum, given any $\delta_1 > 0$ and $\delta_2 > 0$, we can infer that:

1. $\forall\, \mathcal{Z}_{s_1} \in \mathrm{Res}^{\mathrm{x}}(s_1)\, \exists\, \mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1) :\; |\Pr(\mathcal{C}(z_{s_1},\alpha)) - \Pr(\mathcal{C}(z_{t_1},\alpha))| < \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} + \delta_1$,
2. $\forall\, \mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)\, \exists\, \mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2) :\; |\Pr(\mathcal{C}(z_{s_2},\alpha)) - \Pr(\mathcal{C}(z_{t_2},\alpha))| < \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} + \delta_2$.

For simplicity, let us denote process $s_1 \parallel s_2$ by s and process $t_1 \parallel t_2$ by t. We have to show that

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) \le \varepsilon_1 + \varepsilon_2 - \varepsilon_1 \cdot \varepsilon_2$$

and thus that for each trace $\alpha \in \mathcal{A}^{\star}$ we have $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) \le \frac{\varepsilon_1+\varepsilon_2-\varepsilon_1\cdot\varepsilon_2}{\lambda^{|\alpha|-1}}$, namely

$$\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| \le \frac{\varepsilon_1 + \varepsilon_2 - \varepsilon_1 \cdot \varepsilon_2}{\lambda^{|\alpha|-1}}.$$

This is equivalent to show that for each $\delta > 0$ and for each resolution $\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)$ there is some resolution $\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)$ satisfying

$$|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| < \frac{\varepsilon_1 + \varepsilon_2 - \varepsilon_1 \cdot \varepsilon_2}{\lambda^{|\alpha|-1}} + \delta. \tag{B.1}$$

Assume any resolution $\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)$. Then, by definition of parallel composition, there are resolutions $\mathcal{Z}_{s_1} \in \mathrm{Res}^{\mathrm{x}}(s_1)$ and $\mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)$ such that $\mathcal{Z}_s = \mathcal{Z}_{s_1} \parallel \mathcal{Z}_{s_2}$ and $\Pr(\mathcal{C}(z_s,\alpha)) = \Pr(\mathcal{C}(z_{s_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{s_2},\alpha))$. Consider then the two resolutions $\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)$ and $\mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2)$ satisfying items 1 and 2 above with respect to $\mathcal{Z}_{s_1}$ and $\mathcal{Z}_{s_2}$, respectively, with $\delta_1 = \delta_2 = \frac{\delta}{2}$. By definition of parallel composition, $\mathcal{Z}_t = \mathcal{Z}_{t_1} \parallel \mathcal{Z}_{t_2}$ is a resolution for t. On the basis of the relations $\Pr(\mathcal{C}(z_{s_1},\alpha))\, \mathit{op_1}\, \Pr(\mathcal{C}(z_{t_1},\alpha))$ and $\Pr(\mathcal{C}(z_{s_2},\alpha))\, \mathit{op_2}\, \Pr(\mathcal{C}(z_{t_2},\alpha))$ with $\mathit{op_1}, \mathit{op_2} \in \{\geq, \leq\}$, we distinguish four cases and we show that in all cases Equation B.1 holds.

- $\mathit{op_1}, \mathit{op_2} = \geq$. Hence, we have $\Pr(\mathcal{C}(z_{t_1},\alpha)) > \Pr(\mathcal{C}(z_{s_1},\alpha)) - \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} - \delta_1$ and $\Pr(\mathcal{C}(z_{t_2},\alpha)) > \Pr(\mathcal{C}(z_{s_2},\alpha)) - \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} - \delta_2$. We derive Equation B.1 by

$$\begin{aligned}&|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|\\ &= \Pr(\mathcal{C}(z_{s_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{s_2},\alpha)) - \Pr(\mathcal{C}(z_{t_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{t_2},\alpha))\end{aligned} \tag{B.2}$$

By $\Pr(\mathcal{C}(z_{t_1},\alpha)), \Pr(\mathcal{C}(z_{t_2},\alpha)) \leq 1$, we get

$$\begin{aligned}\text{(B.2)} &< \Pr(\mathcal{C}(z_{s_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{s_2},\alpha)) - \left(\Pr(\mathcal{C}(z_{s_1},\alpha)) - \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} - \frac{\delta}{2}\right) \cdot \left(\Pr(\mathcal{C}(z_{s_2},\alpha)) - \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} - \frac{\delta}{2}\right)\\ &< \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} + \frac{\delta}{2} + \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} + \frac{\delta}{2} - \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} \cdot \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} - \frac{\delta}{2} \cdot \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} - \frac{\delta}{2} \cdot \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} - \frac{\delta}{2} \cdot \frac{\delta}{2}\\ &< \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} + \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} + \delta - \frac{\varepsilon_1\varepsilon_2}{\lambda^{2(|\alpha|-1)}}\end{aligned} \tag{B.3}$$

By $\lambda \leq 1$, we get

$$\begin{aligned}\text{(B.3)} &< \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} + \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} + \delta - \frac{\varepsilon_1\varepsilon_2}{\lambda^{|\alpha|-1}}\\ &= \frac{\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2}{\lambda^{|\alpha|-1}} + \delta.\end{aligned}$$

- $\mathit{op1}, \mathit{op2} = \leq$. This case is analogous to the previous one.

- $\mathit{op1} = \geq$ and $\mathit{op2} = \leq$. In this case, we have $\Pr(\mathcal{C}(z_{t_1},\alpha)) > \Pr(\mathcal{C}(z_{s_1},\alpha)) - \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} - \delta_1$ and $\Pr(\mathcal{C}(z_{s_2},\alpha)) > \Pr(\mathcal{C}(z_{t_2},\alpha)) - \frac{\varepsilon_2}{\lambda^{|\alpha|-1}} - \delta_2$. We can distinguish two cases:

  - $\Pr(\mathcal{C}(z_s,\alpha)) \geq \Pr(\mathcal{C}(z_t,\alpha))$. Then we have

$$\begin{aligned}&|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|\\ &= \Pr(\mathcal{C}(z_{s_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{s_2},\alpha)) - \Pr(\mathcal{C}(z_{t_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{t_2},\alpha))\\ &\leq \Pr(\mathcal{C}(z_{s_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{t_2},\alpha)) - \Pr(\mathcal{C}(z_{t_1},\alpha)) \cdot \Pr(\mathcal{C}(z_{t_2},\alpha))\\ &= \Pr(\mathcal{C}(z_{t_2},\alpha)) \cdot (\Pr(\mathcal{C}(z_{s_1},\alpha)) - \Pr(\mathcal{C}(z_{s_{2,\alpha}})))\\ &\leq \Pr(\mathcal{C}(z_{s_1},\alpha)) - \Pr(\mathcal{C}(z_{s_{2,\alpha}}))\\ &< \frac{\varepsilon_1}{\lambda^{|\alpha|-1}} + \frac{\delta}{2}\\ &< \frac{\varepsilon_1 + \varepsilon_2 - \varepsilon_1\varepsilon_2}{\lambda^{|\alpha|-1}} + \delta.\end{aligned}$$

  - $\Pr(\mathcal{C}(z_s,\alpha)) \leq \Pr(\mathcal{C}(z_t,\alpha))$. This case follows analogously to the previous one.

  In both cases we obtained that Equation (B.1) holds.

- $\mathit{op1} = \leq$ and $\mathit{op2} = \geq$. This case is analogous to the previous one.

Summarizing, in all four cases Equation (B.1) holds.

Consider now $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$, with $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. The strict non-expansiveness of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ follows by

$$\begin{aligned} & \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}((s_1 \parallel s_2, t_1 \parallel t_2)) \\ = \; & \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1 \parallel s_2, t_1 \parallel t_2), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1 \parallel t_2, s_1 \parallel s_2)\} \end{aligned} \tag{B.4}$$

By the strict non-expansiveness of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ proved above, we get

$$\begin{aligned} \text{(B.4)} \leq \; & \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2), \\ & \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) + \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(t_2, s_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_2, s_2)\} \end{aligned} \tag{B.5}$$

By $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \leq \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ and the 1-boundedness of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$, we get

$$\begin{aligned} \text{(B.5)} \leq \; & \max\{\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2) - \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) \cdot \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}(s_2, t_2), \\ & \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_2, s_2) - \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_2, s_2)\} \end{aligned} \tag{B.6}$$

By the same reasons as above, we get

$$\begin{aligned} \text{(B.6)} \leq \; & \max\{\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) + \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2) - \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) \cdot \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2), \\ & \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) + \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_2, s_2) - \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_1, s_1) \cdot \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t_2, s_2)\} \end{aligned} \tag{B.7}$$

By the symmetry of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$, we get

$$\text{(B.7)} = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) + \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2) - \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_1, t_1) \cdot \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s_2, t_2).$$

□

Appendix B.4. Proof of Proposition 3

Proposition 3. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ and processes $s, t \in \mathcal{S}$. Then:*

1. *If P is fully-nondeterministic, then $s \sim^{\mathrm{det}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathrm{rand}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{N}}_{\mathrm{Tr}} t$.*
2. *If P is fully-probabilistic, then $s \sim^{\mathrm{det}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathrm{rand}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{P}}_{\mathrm{Tr}} t$.*

Proof of Proposition 3.

1. In the fully nondeterministic setting the execution probability of each trace is either 0 or 1. Therefore, for each trace $\alpha \in \mathcal{A}^\star$ and for each process s, we have that $\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) = 1$ if and only if s performs α; $\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) = 0$ otherwise. Since $s \sim^{\mathbf{N}}_{\mathrm{Tr}} t$ if and only if s and t perform the same traces, it is then immediate to conclude that $s \sim^{\mathrm{x}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{N}}_{\mathrm{Tr}} t$.
2. In the fully probabilistic setting each process has a single maximal resolution, which is the process itself, thus implying that the equality on the supremal execution probability of each trace $\alpha \in \mathcal{A}^\star$ becomes an equality on the execution probability of α on such maximal resolutions. Since $s \sim^{\mathbf{P}}_{\mathrm{Tr}} t$ if and only if s and t perform the same traces with the same probability, we can immediately conclude that $s \sim^{\mathrm{x}}_{\mathrm{Tr,sup}} t \Leftrightarrow s \sim^{\mathbf{P}}_{\mathrm{Tr}} t$.

□

Appendix B.5. Proof of Theorem 8

Theorem 8. *Assume a PTS* $(\mathcal{S}, \mathcal{A}, \rightarrow)$, $\lambda \in (0, 1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. *Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}}$ *as kernel.*

2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^{\mathrm{x}}_{\mathrm{Tr,sup}}$ *as kernel.*

Proof of Theorem 8.

We show that for $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$, the function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}}$ as kernel. Directly from definition of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ and of $\sim^{\mathrm{x}}_{\mathrm{Tr,sup}}$, we also have that $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{\mathrm{x}}_{\mathrm{Tr,sup}}$ as kernel.

To prove that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ is a 1-bounded hemimetric it is enough to show that for each trace $\alpha \in \mathcal{A}^\star$, the function $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ is a 1-bounded hemimetric, that is we need to show that

1. $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s, s) = 0$ for each $s \in \mathcal{S}$.
2. $\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1, s_2) \leq \mathbf{h}^{\alpha,\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}(s_1, s_3) + \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_3, s_2)$ for each $s_1, s_2, s_3 \in \mathcal{S}$.

The first item is immediate by Def. 19. Let us prove the triangular inequality. We can distinguish two cases.

- $\sup_{\mathcal{Z}_1 \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_1, \alpha)) \leq \sup_{\mathcal{Z}_2 \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_2, \alpha))$. Hence we have
$$\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1, s_2) = 0 \leq \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1, s_3) + \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_3, s_2).$$
- $\sup_{\mathcal{Z}_1 \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_1, \alpha)) > \sup_{\mathcal{Z}_2 \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_2, \alpha))$. Hence we have
$$\begin{aligned}
&\mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1, s_2)\\
&= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1 \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_1, \alpha)) - \sup_{\mathcal{Z}_2 \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_2, \alpha))\Big)\\
&= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1 \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_1, \alpha)) - \sup_{\mathcal{Z}_2 \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_2, \alpha)) \pm \sup_{\mathcal{Z}_3 \in \mathrm{Res}^{\mathrm{x}}(s_3)} \Pr(\mathcal{C}(z_3, \alpha))\Big)\\
&= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1 \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_1, \alpha)) - \sup_{\mathcal{Z}_3 \in \mathrm{Res}^{\mathrm{x}}(s_3)} \Pr(\mathcal{C}(z_3, \alpha))\Big) +\\
&\quad \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_3 \in \mathrm{Res}^{\mathrm{x}}(s_3)} \Pr(\mathcal{C}(z_3, \alpha)) - \sup_{\mathcal{Z}_2 \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_2, \alpha))\Big)\\
&\leq \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1, s_3) + \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_3, s_2).
\end{aligned}$$

The 1-boundedness property follows by $\lambda \in (0, 1]$ and

$$\begin{aligned}
&\Pr(\mathcal{C}(z_s, \alpha)) \leq 1 \text{ for all } \alpha \in \mathcal{A}^\star\\
\Rightarrow &\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \leq 1 \text{ for all } \alpha \in \mathcal{A}^\star, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)\\
\Rightarrow &\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \Pr(\mathcal{C}(z_t, \alpha)) \leq 1 \text{ for all } \alpha \in \mathcal{A}^\star, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s), \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t).
\end{aligned}$$

For the kernel, we have

$$\begin{aligned}
s \sqsubseteq^{\mathrm{x}}_{\mathrm{Tr,sup}} t &\iff \forall\, \alpha \in \mathcal{A}^\star \;\; \forall\, \mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s) \;\; \exists\, \mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t) \text{ s.t. } \Pr(\mathcal{C}(z_s, \alpha)) \leq \Pr(\mathcal{C}(z_t, \alpha))\\
&\iff \forall\, \alpha \in \mathcal{A}^\star \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \leq \sup_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \Pr(\mathcal{C}(z_t, \alpha))\\
&\iff \forall\, \alpha \in \mathcal{A}^\star\; \mathbf{h}^{\alpha\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s, t) = 0
\end{aligned}$$

$$\iff \sup_{\alpha \in \mathcal{A}^\star} \mathbf{h}^{\alpha\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) = 0$$
$$\iff \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s,t) = 0.$$

□

Appendix B.6. Proof of Theorem 9

Theorem 9. *All distances $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$ are strictly non-expansive.*

Proof of Theorem 9. We start with $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$, with $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$. Our aim is to prove that for all processes $s_1, s_2, t_1, t_2 \in \mathcal{S}$ it holds that

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2)$$

Notice that if $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) = 0$ then there is nothing to prove. Hence assume that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) > 0$. If suffices to prove that for all $\varepsilon > 0$ we have

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) < \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon \tag{B.8}$$

Fixed any $\varepsilon > 0$, in order to prove Equation B.8, we recall that by definition of supremum there is a trace α_ε with $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) < \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) + \varepsilon$. Let $\varepsilon > 0$. We have

$$\begin{aligned}
&\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) \\
&< \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1 \parallel s_2, t_1 \parallel t_2) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \left(\sup_{\mathcal{Z}_{s_1 \parallel s_2} \in \mathrm{Res}^{\mathrm{x}}(s_1 \parallel s_2)} \Pr(\mathcal{C}(z_{s_1\parallel s_2}, \alpha_\varepsilon)) - \sup_{\mathcal{Z}_{t_1 \parallel t_2} \in \mathrm{Res}^{\mathrm{x}}(t_1 \parallel t_2)} \Pr(\mathcal{C}(z_{t_1\parallel t_2}, \alpha_\varepsilon)) \right) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \left(\sup_{\mathcal{Z}_{s_1} \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1}, \alpha_\varepsilon)) \cdot \sup_{\mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2}, \alpha_\varepsilon)) + \right. \\
&\qquad \left. - \sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{det}}(st_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \sup_{\mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2)} \Pr(\mathcal{C}(z_{t_2}, \alpha_\varepsilon)) \right) + \varepsilon.
\end{aligned} \tag{B.9}$$

with the last equality follows by the definition of resolution. We can distinguish three cases:

1. $\sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \geq \sup_{\mathcal{Z}_{s_1} \in \mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1}, \alpha_\varepsilon))$. In this case we have:

$$\begin{aligned}
&(\text{B.9}) \\
&\leq \lambda^{|\alpha_\varepsilon|-1} \left(\sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \sup_{\mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2}, \alpha_\varepsilon)) + \right. \\
&\qquad \left. - \sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \sup_{\mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2)} \Pr(\mathcal{C}(z_{t_2}, \alpha_\varepsilon)) \right) + \varepsilon \\
&= \sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \lambda^{|\alpha_\varepsilon|-1} \left(\sup_{\mathcal{Z}_{s_2} \in \mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2}, \alpha_\varepsilon)) - \sup_{\mathcal{Z}_{t_2} \in \mathrm{Res}^{\mathrm{x}}(t_2)} \Pr(\mathcal{C}(z_{t_2}, \alpha_\varepsilon)) \right) + \varepsilon \\
&\leq \sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon \\
&\leq \sup_{\mathcal{Z}_{t_1} \in \mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1}, \alpha_\varepsilon)) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon \\
&\leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon
\end{aligned}$$

$\leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon.$

thus giving Equation (B.8), with the first step derived by inequality $\sup_{\mathcal{Z}_{t_1}\in\mathrm{Res}^{\mathrm{x}}(t_1)} \Pr(\mathcal{C}(z_{t_1},\alpha_\varepsilon)) \geq \sup_{\mathcal{Z}_{s_1}\in\mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1},\alpha_\varepsilon))$, the third step by the definition of $\mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$, the fourth step by the definition of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$, the fifth step by $\Pr(\mathcal{C}(z_{t_1},\alpha_\varepsilon)) \leq 1$ and the last step by the 1-boundedness of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$.

2. $\sup_{\mathcal{Z}_{t_2}\in\mathrm{Res}^{\mathrm{x}}(t_2)} \Pr(\mathcal{C}(z_{t_2},\alpha_\varepsilon)) \geq \sup_{\mathcal{Z}_{s_2}\in\mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2},\alpha_\varepsilon))$. This case is analogous to the previous one and also gives Equation (B.8).
3. $\sup_{\mathcal{Z}_{s_i}\in\mathrm{Res}^{\mathrm{x}}(s_i)} \Pr(\mathcal{C}(z_{s_i},\alpha_\varepsilon)) > \sup_{\mathcal{Z}_{t_i}\in\mathrm{Res}^{\mathrm{x}}(t_i)} \Pr(\mathcal{C}(z_{t_i},\alpha_\varepsilon))$, for $i = 1,2$. In this case, since we have $\mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_i,t_i) = \lambda^{|\alpha_\varepsilon|-1}(\sup_{\mathcal{Z}_{s_i}\in\mathrm{Res}^{\mathrm{x}}(s_i)} \Pr(\mathcal{C}(z_{s_i},\alpha_\varepsilon)) - \sup_{\mathcal{Z}_{t_i}\in\mathrm{Res}^{\mathrm{x}}(t_i)} \Pr(\mathcal{C}(z_{t_i},\alpha_\varepsilon)))$ and $\mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_i,t_i) \leq \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}(s_i,t_i)$, we infer

$$\sup_{\mathcal{Z}_{t_i}\in\mathrm{Res}^{\mathrm{x}}(t_i)} \Pr(\mathcal{C}(z_{t_i},\alpha_\varepsilon)) \geq \sup_{\mathcal{Z}_{s_i}\in\mathrm{Res}^{\mathrm{det}}(s_i)} \Pr(\mathcal{C}(z_{s_i},\alpha_\varepsilon)) - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_i,t_i)}{\lambda^{|\alpha_\varepsilon|-1}}$$

which allows us to derive

(B.9)

$$\begin{aligned}
&\leq \varepsilon + \lambda^{|\alpha_\varepsilon|-1}\Bigg(\sup_{\mathcal{Z}_{s_1}\in\mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1},\alpha_\varepsilon)) \cdot \sup_{\mathcal{Z}_{s_2}\in\mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2},\alpha_\varepsilon)) + \\
&\quad - \left(\sup_{\mathcal{Z}_{s_1}\in\mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1},\alpha_\varepsilon)) - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1)}{\lambda^{|\alpha_\varepsilon|-1}}\right) \cdot \left(\sup_{\mathcal{Z}_{s_2}\in\mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2},\alpha_\varepsilon)) - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2)}{\lambda^{|\alpha_\varepsilon|-1}}\right)\Bigg) \\
&= \varepsilon + \lambda^{|\alpha_\varepsilon|-1}\Bigg(\sup_{\mathcal{Z}_{s_1}\in\mathrm{Res}^{\mathrm{x}}(s_1)} \Pr(\mathcal{C}(z_{s_1},\alpha_\varepsilon)) \cdot \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2)}{\lambda^{|\alpha_\varepsilon|-1}} + \\
&\quad + \sup_{\mathcal{Z}_{s_2}\in\mathrm{Res}^{\mathrm{x}}(s_2)} \Pr(\mathcal{C}(z_{s_2},\alpha_\varepsilon)) \cdot \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1)}{\lambda^{|\alpha_\varepsilon|-1}} - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1)}{\lambda^{|\alpha_\varepsilon|-1}} \cdot \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2)}{\lambda^{|\alpha_\varepsilon|-1}}\Bigg) \\
&\leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}(s_2,t_2) + \varepsilon
\end{aligned}$$

thus giving Equation (B.8) also in this case.

We conclude by observing that the strict non-expansiveness of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ can be proved by exploiting the strict non-expansiveness of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$ exactly as the strict non-expansiveness of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ was proved by exploiting the strict non-expansiveness of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}$ (see Theorem 7). □

Appendix B.7. Proof of Theorem 10

Theorem 10. *Let $(\mathcal{S},\mathcal{A},\to)$ be a PTS, $\lambda \in (0,1]$, $\mathrm{y} \in \{\mathrm{dis},\mathrm{tbt}\}$ and $\mathbf{d} \in \{\mathbf{h},\mathbf{m}\}$. Then $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,y}} < \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Tr,y}}$.*

Proof of Theorem 10. We show that in the trace distribution approach $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}} \leq \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}$ and $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}} \leq \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}$. The fact that in the trace-by-trace approach we have $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} \leq \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$ and $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} \leq \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,tbt}}$ follows analogously.

We start by proving that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}} \leq \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}$. Given arbitrary processes $s,t \in \mathcal{S}$, we have

$$\begin{aligned}
\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t) &= \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \inf_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1}|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| \\
&\geq \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \inf_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1}|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|
\end{aligned}$$

$$= \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{rand}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|$$
$$= \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t)$$

where:

- the second step follows by $\mathrm{Res}^{\mathrm{det}}(t) \subseteq \mathrm{Res}^{\mathrm{rand}}(t)$ and the fact that by evaluating the infimum over a wider class of resolutions we can obtain a better approximation of the resolutions in $\mathrm{Res}^{\mathrm{det}}(s)$
- by letting $f(z_s) = \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|$, the third step immediately derives from
  - $\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{rand}}(s)} f(z_s) \geq \sup_{\mathcal{Z}'_s \in \mathrm{Res}^{\mathrm{det}}(s)} f(z'_s)$. This derives directly from $\mathrm{Res}^{\mathrm{det}}(s) \subseteq \mathrm{Res}^{\mathrm{rand}}(s)$ and the properties of suprema.
  - $\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{rand}}(c)} f(z_s) \leq \sup_{\mathcal{Z}'_s \in \mathrm{Res}^{\mathrm{det}}(s)} f(z'_s)$. This follows since the randomization consists in a convex combination of the distributions reached by equally labeled transitions. More formally, for each $\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{rand}}(s)$ there is a set of indexes I s.t. for each $\alpha \in \mathcal{A}^\star$ $\Pr(\mathcal{C}(z_s,\alpha)) = \sum_{i\in I} p_i \Pr(\mathcal{C}(z^i_s,\alpha))$ for weights $p_i \in (0,1]$ with $\sum_{i\in I} p_i = 1$ and deterministic resolutions $\mathcal{Z}^i_s \in \mathrm{Res}^{\mathrm{det}}(s)$. Thus, given any $\varepsilon > 0$ and by definition of supremum

$$\begin{aligned}
\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{rand}}(s)} f(z_s) &< f(z^\varepsilon_s) + \varepsilon \\
&= \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} |\Pr(\mathcal{C}(z^\varepsilon_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon \\
&= \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} |\sum_{i \in I^\varepsilon} p_i \Pr(\mathcal{C}(z^i_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon \\
&\leq \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \sum_{i \in I^\varepsilon} p_i |\Pr(\mathcal{C}(z^i_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon \\
&\leq \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \max_{i \in I^\varepsilon} |\Pr(\mathcal{C}(z^i_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon \\
&\leq \max_{i \in I^\varepsilon} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{rand}}(t)} \sup_{\alpha \in \mathcal{A}^\star} |\Pr(\mathcal{C}(z^i_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon \\
&\leq \sup_{\mathcal{Z}'_s \in \mathrm{Res}^{\mathrm{det}}(s)} f(z'_s) + \varepsilon.
\end{aligned}$$

Then, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t) = \max\{\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(s,t), \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,dis}}(t,s)\} \leq \max\{\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t), \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(t,s)\} = \mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Tr,dis}}(s,t)$.

Finally, the processes s,t in Figure 7 with $s \sim^{\mathrm{rand}}_{\mathrm{Tr,dis}} t$, $s \sim^{\mathrm{rand}}_{\mathrm{Tr,tbt}} t$, $s \not\sim^{\mathrm{det}}_{\mathrm{Tr,dis}} t$, $s \not\sim^{\mathrm{det}}_{\mathrm{Tr,tbt}} t$ witness the strictness of all four relations. □

Appendix B.8. Proof of Theorem 11

Theorem 11. *Let $(\mathcal{S},\mathcal{A},\rightarrow)$ be a PTS, $\lambda \in (0,1]$, $\mathrm{x} \in \{\mathrm{det},\mathrm{rand}\}$ and $\mathbf{d} \in \{\mathbf{h},\mathbf{m}\}$. Then $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$.*

Proof of Theorem 11. Let x in $\{\mathrm{det},\mathrm{rand}\}$. We start with hemimetrics and show that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$. Given any processes $s,t \in \mathcal{S}$, consider the trace-by-trace hemimetric $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \sup_{\alpha\in\mathcal{A}^*} \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t)$. We recall that by definition of supremum, for each $\varepsilon > 0$ there is a trace $\alpha_\varepsilon \in \mathcal{A}^\star$ such that $\sup_{\alpha\in\mathcal{A}^\star} \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) < \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) + \varepsilon$. Hence, given any $\varepsilon > 0$, we have

$$\begin{aligned}
\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) &= \sup_{\alpha\in\mathcal{A}^*} \mathbf{h}^{\alpha,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) \\
&< \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) + \varepsilon \\
&= \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \lambda^{|\alpha_\varepsilon|-1} |\Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))| + \varepsilon
\end{aligned}$$

$$\leq \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{x}}(s)} \inf_{\mathcal{Z}_t \in \mathrm{Res}^{\mathrm{x}}(t)} \sup_{\alpha \in \mathcal{A}^\star} \lambda^{|\alpha|-1} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))| + \varepsilon$$
$$= \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t) + \varepsilon.$$

Since the inequality $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) < \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t)+\varepsilon$ holds for all $\varepsilon > 0$ we can conclude that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}$ as required. For the metrics, we get $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t) = \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}}(t,s)\} \leq \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(t,s)\} = \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Tr,dis}}(s,t)$.

Finally, we note that the processes in Figure 6 witness the strictness of all four relations. □

Appendix B.9. Proof of Theorem 12

Theorem 12. *Let $(\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS, $\lambda \in (0,1]$ and $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$. Then $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} = \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}} = \mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$.*

Proof of Theorem 12. We expand only the proof of the first item, namely for the case of hemimetrics. The results on pseudometrics can be obtained as a direct consequence.

We start by showing that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}} = \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$. To this aim it is enough to prove that for each trace $\alpha \in \mathcal{A}^\star$ we have

$$\sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) = \sup_{\mathcal{Z}'_s \in \mathrm{Res}^{\mathrm{rand}}(s)} \Pr(\mathcal{C}(z'_s,\alpha)). \tag{B.10}$$

First of all we notice that $\mathrm{Res}^{\mathrm{rand}}(s) = \mathrm{Res}^{\mathrm{det}}(s) \cup (\mathrm{Res}^{\mathrm{rand}}(s) \setminus \mathrm{Res}^{\mathrm{det}}(s))$, that is the set of randomized resolutions for a process is given by the disjoint union of the set of the deterministic resolutions for that process with the set of resolutions which are not deterministic. Thus,

$$\sup_{\mathcal{Z}'_s \in \mathrm{Res}^{\mathrm{rand}}(s)} \Pr(\mathcal{C}(z'_s,\alpha)) = \max\Big\{ \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)), \sup_{\mathcal{Z}''_s \in \mathrm{Res}^{\mathrm{rand}}(s)\setminus \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z''_s,\alpha)) \Big\}.$$

As a consequence, to prove Equation (B.10) it is enough to prove that

$$\sup_{\mathcal{Z}''_s \in \mathrm{Res}^{\mathrm{rand}}(s)\setminus \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z''_s,\alpha)) \leq \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)). \tag{B.11}$$

By definition of supremum, for each $\varepsilon > 0$ there is a $\mathcal{Z}_\varepsilon \in \mathrm{Res}^{\mathrm{rand}}(s) \setminus \mathrm{Res}^{\mathrm{det}}(s)$ such that

$$\sup_{\mathcal{Z}''_s \in \mathrm{Res}^{\mathrm{rand}}(s)\setminus \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z''_s,\alpha)) < \Pr(\mathcal{C}(z_\varepsilon,\alpha)) + \varepsilon.$$

Then, given any $\varepsilon > 0$, we have

$$\begin{aligned}
\sup_{\mathcal{Z}''_s \in \mathrm{Res}^{\mathrm{rand}}(s)\setminus \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z''_s,\alpha)) &< \Pr(\mathcal{C}(z_\varepsilon,\alpha)) + \varepsilon \\
&= \varepsilon + \sum_{i \in I_\varepsilon} p_i \Pr(\mathcal{C}(z_i,\alpha)) \\
&\leq \varepsilon + \sum_{i \in I_\varepsilon} p_i \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) \\
&= \sup_{\mathcal{Z}_s \in \mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) + \varepsilon
\end{aligned}$$

where the z_i in the second step are the deterministic schedulers combined in the randomization by $\mathcal{Z}_\varepsilon$. Since the inequality holds for all $\varepsilon > 0$, this concludes the proof of Equation (B.11).

Next we prove that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} \geq \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$. The property $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) \geq \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t)$ is immediate if $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) = 0$. Assume that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) > 0$. Given any $0 < \varepsilon < \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t)$, by the definition of supremum and by $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) = \sup_{\alpha\in\mathcal{A}^*} \mathbf{h}^{\alpha,\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t)$ we infer that there exists a trace α_ϵ with $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) < \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) + \varepsilon$. This allows us to derive

$$\begin{aligned}
\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) &< \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) + \varepsilon \\
&= \max\left\{0, \lambda^{|\alpha_\varepsilon|-1}\left(\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{rand}}(c)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\right)\right\} + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1}\left(\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{rand}}(c)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\right) + \varepsilon \\
&= \inf_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \lambda^{|\alpha_\varepsilon|-1}\left(\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{rand}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\right) + \varepsilon \\
&\le \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{rand}}(s)} \inf_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \lambda^{|\alpha_\varepsilon|-1} \left|\Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\right| + \varepsilon \\
&\le \sup_{\alpha\in\mathcal{A}^\star} \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{rand}}(s)} \inf_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \lambda^{|\alpha_\varepsilon|-1} \left|\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))\right| + \varepsilon \\
&= \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) + \varepsilon
\end{aligned}$$

with the third step by $\mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}(s,t) > 0$. Then, since the inequality holds for all $\varepsilon > 0$ we can infer that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} \ge \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,sup}}$ as required.

Finally, we prove that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}} \ge \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$. Notice that since in $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}$ we consider the difference in the execution probabilities of the processes on one trace per time, the action of randomized schedulers for a process s can be subsumed by saying that they can assign to each trace α any probability that can be expressed as $p \cdot \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha))$ with $p \in (0,1]$. Thus, for a given $\varepsilon > 0$, we have

$$\begin{aligned}
&\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) \\
&< \mathbf{h}^{\alpha_\varepsilon,\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}}(s,t) + \varepsilon \\
&= \sup_{\mathcal{Z}'_s\in\mathrm{Res}^{\mathrm{rand}}(s)} \inf_{\mathcal{Z}'_t\in\mathrm{Res}^{\mathrm{rand}}(t)} \lambda^{|\alpha_\varepsilon|-1} |\Pr(\mathcal{C}(z'_s,\alpha_\varepsilon)) - \Pr(\mathcal{C}(z'_t,\alpha_\varepsilon))| + \varepsilon \\
&= \sup_{p\in(0,1]} \inf_{q\in(0,1]} \lambda^{|\alpha_\varepsilon|-1} |p \cdot \sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - q \cdot \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))| + \varepsilon.
\end{aligned} \tag{B.12}$$

We can distinguish two cases:

- $\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) \le \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))$. In this case, for each $p \in (0,1]$, $q = p \cdot \frac{\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon))}{\sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))}$ gives (B.12) $= 0$ and thus $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Tr,tbt}} \le \mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Tr,sup}}$ immediately follows.
- $\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) > \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))$. Then the sup-inf distance in (B.12) is obtained by choosing $p = q = 1$ as p has to maximize the difference, whereas q has to minimize it. Thus we get

$$\begin{aligned}
\text{(B.12)} &= \lambda^{|\alpha_\varepsilon|-1} |\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))| + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1}\left(\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\right) + \varepsilon \\
&\le \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1}\left(\sup_{\mathcal{Z}_s\in\mathrm{Res}^{\mathrm{det}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}^{\mathrm{det}}(t)} \Pr(\mathcal{C}(z_t,\alpha))\right) + \varepsilon
\end{aligned}$$

$$= \mathbf{h}^{\lambda,\text{det}}_{\text{Tr},\text{sup}}(s,t) + \varepsilon$$

and since the inequality holds for all $\varepsilon > 0$ we can infer that $\mathbf{h}^{\lambda,\text{rand}}_{\text{Tr},\text{tbt}} \leq \mathbf{h}^{\lambda,\text{det}}_{\text{Tr},\text{sup}}$ as required.

□

Appendix C. Proofs of Section 5

Appendix C.1. Proof of Theorem 13

Theorem 13. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}, \mathrm{mM}\}$:*

1. *The function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{\mathrm{x}}_{\mathrm{Te,y}}$ as kernel.*
2. *The function $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,y}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{\mathrm{x}}_{\mathrm{Te,y}}$ as kernel.*

Proof of Theorem 13. The proof is analogous to that of Theorem 8. □

Appendix C.2. Proof of Theorem 14

Theorem 14. *Let $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}, \mathrm{mM}\}$. All distances $\mathbf{h}^{\omega,\mathrm{det}}_{\mathrm{Te,y}}$, $\mathbf{h}^{\omega,\mathrm{rand}}_{\mathrm{Te,y}}$, $\mathbf{m}^{\omega,\mathrm{det}}_{\mathrm{Te,y}}$, $\mathbf{m}^{\omega,\mathrm{rand}}_{\mathrm{Te,y}}$ are non-expansive.*

Proof of Theorem 14. Assume $x \in \{\mathrm{det}, \mathrm{rand}\}$. We expand only the case of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$, since the case for $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$ can be obtained analogously, the cases of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}, \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}$ follow as direct consequences of the result on the respective hemimetrics and the cases of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$ follow from the previous ones.

First of all, we notice that since we are considering fully synchronous parallel compositions, for any $s, t \in \mathcal{S}$ and $o \in \mathbf{O}$, we have

$$\begin{aligned}\sup_{\mathcal{Z}_{s\|t,o}\in\mathrm{Res}^{\mathrm{x}}(s\|t,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s\|t,o})) &= \sup_{\mathcal{Z}_{s,t\|o}\in\mathrm{Res}^{\mathrm{x}}(s,t\|o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s,t\|o}))\\ &= \sup_{\mathcal{Z}_{t,s\|o}\in\mathrm{Res}^{\mathrm{x}}(t,s\|o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{t,s\|o})). \end{aligned} \tag{C.1}$$

We can proceed now to prove that for any $s_1, s_2, t_1, t_2 \in \mathcal{S}$ and $o \in \mathbf{O}$

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1, t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_2, t_2). \tag{C.2}$$

We recall that by definition of supremum, given $\varepsilon > 0$ there is a test $o_\varepsilon \in \mathbf{O}$ such that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) < \mathbf{h}^{o_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) + \varepsilon$. Hence, to prove Equation (C.2) it is enough to prove that, for all $\varepsilon > 0$ it holds

$$\mathbf{h}^{o_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1, t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_2, t_2). \tag{C.3}$$

For simplicity of notation, let $o_\varepsilon = o$. Clearly if $\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) = 0$, then there is nothing to prove. Hence, assume that $\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2) > 0$. We have

$$\begin{aligned} &\mathbf{h}^{o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1 \parallel s_2, t_1 \parallel t_2)\\ =& \sup_{\mathcal{Z}_{s_1\|s_2,o}\in\mathrm{Res}^{\mathrm{x}}(s_1\|s_2,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s_1\|s_2,o})) - \sup_{\mathcal{Z}_{t_1\|t_2,o}\in\mathrm{Res}^{\mathrm{x}}(t_1\|t_2,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{t_1\|t_2,o}))\\ =& \left(\sup_{\mathcal{Z}_{s_1\|s_2,o}\in\mathrm{Res}^{\mathrm{x}}(s_1\|s_2,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s_1\|s_2,o})) - \sup_{\mathcal{Z}_{s_2\|t_1,o}\in\mathrm{Res}^{\mathrm{x}}(s_2\|t_1,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s_2\|t_1,o}))\right) +\\ &\left(\sup_{\mathcal{Z}_{s_2\|t_1,o}\in\mathrm{Res}^{\mathrm{x}}(s_2\|t_1,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s_2\|t_1,o})) - \sup_{\mathcal{Z}_{t_1\|t_2,o}\in\mathrm{Res}^{\mathrm{x}}(t_1\|t_2,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{t_1\|t_2,o}))\right) \end{aligned} \tag{C.4}$$

By Equation (C.1), we get

$$\text{(C.4)} = \left(\sup_{\mathcal{Z}_{s_1,s_2\|o}\in\mathrm{Res}^{\mathrm{x}}(s_1,s_2\|o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s_1,s_2\|o})) - \sup_{\mathcal{Z}_{t_1,s_2\|o}\in\mathrm{Res}^{\mathrm{x}}(t_1,s_2\|o)} \sum_{n=1}^{\infty} \lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{t_1,s_2\|o}))\right) +$$

$$\left(\sup_{\mathcal{Z}_{s_2,t_1\|o}\in\mathrm{Res}^{\mathrm{x}}(s_2,t_1\|o)}\sum_{n=1}^{\infty}\lambda^{n-1}\Pr^n(\mathbf{SC}(z_{s_2,t_1\|o})) - \sup_{\mathcal{Z}_{t_2,t_1\|o}\in\mathrm{Res}^{\mathrm{x}}(t_2,t_1\|o)}\sum_{n=1}^{\infty}\lambda^{n-1}\Pr^n(\mathbf{SC}(z_{t_2,t_1\|o}))\right)$$
$$= \mathbf{h}^{s_2\|o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1,t_1) + \mathbf{h}^{t_1\|o,\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_2,t_2)$$
$$\leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s_2,t_2).$$

□

Appendix C.3. Proof of Theorem 15

Theorem 15. *Let $(\mathcal{S},\mathcal{A},\rightarrow)$ be a PTS, $\lambda\in(0,1]$ and $\mathrm{x}\in\{\mathrm{det},\mathrm{rand}\}$. Then:*

1. *The function $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$ is a 1-bounded hemimetric on $\mathcal{S}$, with $\sqsubseteq^{x}_{\mathrm{Te,tbt}}$ as kernel.*
2. *The function $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$ is a 1-bounded pseudometric on $\mathcal{S}$, with $\sim^{x}_{\mathrm{Te,tbt}}$ as kernel.*

Proof of Theorem 15. The proof is analogous to that of Theorem 6. □

Appendix C.4. Proof of Theorem 16

Theorem 16. *All distances $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$, $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}$, $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$, $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}$ are strictly non-expansive.*

Proof of Theorem 16. We expand only the case of the hemimetrics $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$, with $\mathrm{x}\in\{\mathrm{det},\mathrm{rand}\}$. The cases of pseudometrics $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$ are an immediate consequence of the same property for their asymmetric versions.

We have to prove that for all $s_1,s_2,t_1,t_2\in\mathcal{S}$ we have:

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2, t_1\parallel t_2) \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)\cdot\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2).$$

Let $s_1,s_2,t_1,t_2\in\mathcal{S}$. Indeed $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2) = \sup_{o\in\mathbf{O}}\sup_{\alpha\in\mathcal{A}^\star}\mathbf{h}^{o,\alpha,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2)$. Hence, by definition of supremum, for each $\varepsilon>0$ there are a test $o_\varepsilon\in\mathbf{O}$ and a trace $\alpha_\varepsilon\in\mathcal{A}^\star$ such that

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2) < \mathbf{h}^{o_\varepsilon,\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2)+\varepsilon.$$

We will then show that:

$$\mathbf{h}^{o_\varepsilon,\alpha_\varepsilon,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2) \leq \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)\cdot\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2). \tag{C.5}$$

The thesis will then follow by the fact that the inequality $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2)\leq\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)\cdot\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)+\varepsilon$ holds for any $\varepsilon>0$.

For simplicity we denote o_ε by o and α_ε by α. We have that:

$$\begin{aligned}&\mathbf{h}^{o,\alpha,\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1\parallel s_2,t_1\parallel t_2)\\ &= \lambda^{|\alpha|-1}\sup_{\mathcal{Z}_{s_1\|s_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_1\|s_2,o)}\ \inf_{\mathcal{Z}_{t_1\|t_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t_1\|t_2,o)}|\Pr(\mathbf{SC}(z_{s_1\|s_2,o},\alpha))-\Pr(\mathbf{SC}(z_{t_1\|t_2,o},\alpha))|.\end{aligned} \tag{C.6}$$

We can observe that for any $\mathcal{Z}_{s_1\|s_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_1\parallel s_2,o)$ there exist $\mathcal{Z}_{s_1,o_\alpha}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_1,o_\alpha)$ and $\mathcal{Z}_{s_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_2,o)$ such that $\Pr(\mathbf{SC}(z_{s_1\|s_2,o},\alpha))=\Pr(\mathbf{SC}(z_{s_1,o_\alpha},\alpha))\cdot\Pr(\mathbf{SC}(z_{s_2,o},\alpha))$, where o_α is the *deterministic* test reaching the successful state just after the trace α. Considering t_1 and t_2 in place of, respectively, s_1 and s_2, we obtain an analogous result. Hence:

$$\begin{aligned}\text{(C.6)} = \lambda^{|\alpha|-1}&\sup_{\mathcal{Z}_{s_1,o_\alpha}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_1,o_\alpha)}\ \sup_{\mathcal{Z}_{s_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s_2,o)}\ \inf_{\mathcal{Z}_{t_1,o_\alpha}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t_1,o_\alpha)}\ \inf_{\mathcal{Z}_{t_2,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t_2,o)}\\ &|\Pr(\mathbf{SC}(z_{s_1,o_\alpha},\alpha))\cdot\Pr(\mathbf{SC}(z_{s_2,o},\alpha))-\Pr(\mathbf{SC}(z_{t_1,o_\alpha},\alpha))\cdot\Pr(\mathbf{SC}(z_{t_2,o},\alpha))|\end{aligned}$$

$$< \lambda^{|\alpha|-1} \inf_{\mathcal{Z}_{t_1,o_\alpha} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t_1,o_\alpha)} \inf_{\mathcal{Z}_{t_2,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t_2,o)}$$

$$|\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \cdot \Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) - \Pr(\mathbf{SC}(z_{t_1,o_\alpha},\alpha)) \cdot \Pr(\mathbf{SC}(z_{t_2,o},\alpha))| + \varepsilon'$$

$$\leq \lambda^{|\alpha|-1} \cdot |\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \cdot \Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) - \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha)) \cdot \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha))| + \varepsilon' \quad \text{(C.7)}$$

where

- the second step follows by definition of supremum with respect to a chosen ε';
- the third step follows by noticing that by definition of infimum, for each $\delta > 0$
  - for the chosen resolution $\mathcal{Z}^{\varepsilon'}_{s_1,o_\alpha} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s_1,o_\alpha)$ there is a resolution $\mathcal{Z}^{\delta}_{t_1,o_\alpha} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t_1,o_\alpha)$ such that $|\Pr(\mathcal{C}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) - \Pr(\mathcal{C}(z^{\delta}_{t_1,o_\alpha},\alpha))| < \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} + \delta$,
  - for the chosen resolution $\mathcal{Z}^{\varepsilon'}_{s_2,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s_2,o)$ there is a resolution $\mathcal{Z}^{\delta}_{t_2,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(t_2,o)$ such that $|\Pr(\mathcal{C}(z^{\varepsilon'}_{s_2,o},\alpha)) - \Pr(\mathcal{C}(z^{\delta}_{t_2,o},\alpha))| < \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} + \delta$.

We can then distinguish four cases:

- $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \geq \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha))$ and $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) \geq \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha))$. Hence:

$$\begin{aligned}
\text{(C.7)} \leq & \lambda^{|\alpha|-1} \cdot \Big(\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \cdot \Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) - \\
& \Big(\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} - \delta\Big) \cdot \Big(\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} - \delta\Big)\Big) + \varepsilon' \\
= & \lambda^{|\alpha|-1} \cdot \Bigg(\Big(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} + \delta\Big) \cdot \Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) + \Big(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} + \delta\Big) \cdot \Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) - \\
& \Big(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} + \delta\Big) \cdot \Big(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} + \delta\Big)\Bigg) + \varepsilon' \\
\leq & \lambda^{|\alpha|-1} \cdot \Bigg(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} + \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} - \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} \cdot \frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2)}{\lambda^{|\alpha|-1}} + 2 \cdot \delta\Bigg) + \varepsilon' \\
\leq & \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) + \varepsilon' + 2 \cdot \delta
\end{aligned}$$

  and since the inequality holds for any $\varepsilon', \delta > 0$, we can conclude that Equation C.5 holds in this case.

- The case of $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \leq \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha))$ and $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) \leq \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha))$ is analogous to the previous one.

- $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \geq \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha))$ and $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) \leq \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha))$. Hence:

$$\begin{aligned}
\text{(C.7)} \leq & \lambda^{|\alpha|-1} \cdot \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha)) \cdot \Big(\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) - \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha))\Big) + \varepsilon' \\
\leq & \lambda^{|\alpha|-1} \cdot \Bigg(\frac{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1)}{\lambda^{|\alpha|-1}} + \delta\Bigg) + \varepsilon' \\
\leq & \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) + \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) - \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_1,t_1) \cdot \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s_2,t_2) + \varepsilon' + \delta
\end{aligned}$$

  and since the inequality holds for any $\varepsilon', \delta > 0$, we can conclude that Equation C.5 holds also in this case.

- The case of $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_1,o_\alpha},\alpha)) \leq \Pr(\mathbf{SC}(z^{\delta}_{t_1,o_\alpha},\alpha))$ and $\Pr(\mathbf{SC}(z^{\varepsilon'}_{s_2,o},\alpha)) \geq \Pr(\mathbf{SC}(z^{\delta}_{t_2,o},\alpha))$ is analogous to the previous one.

□

Appendix C.5. Proof of Theorem 17

Theorem 17. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0,1]$ *and* $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$*. Then:*

1. *The function* $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$ *is a 1-bounded hemimetric on* $\mathcal{S}$*, with* $\sqsubseteq^{\mathrm{x}}_{\mathrm{Te,sup}}$ *as kernel.*

2. *The function* $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$ *is a 1-bounded pseudometric on* $\mathcal{S}$*, with* $\sim^{\mathrm{x}}_{\mathrm{Te,sup}}$ *as kernel.*

Proof of Theorem 17. The proof is analogous to that of Theorem 8. □

Appendix C.6. Proof of Theorem 18

Theorem 18. *All distances* $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Te,sup}}$*,* $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,sup}}$*,* $\mathbf{m}^{\lambda,\mathrm{det}}_{\mathrm{Te,sup}}$*,* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,sup}}$ *are strictly non-expansive.*

Proof of Theorem 18. Similarly to the other theorems above we prove only the case for $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$ with $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$, since the other is a direct consequence of the definition of $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}$.

This case can be proved by following the same reasoning as for the proof of Theorem 16 and by observing that for each $s_1, s_2 \in \mathcal{S}$, $o \in \mathbf{O}$ and $\alpha \in \mathcal{A}^\star$:

$$
\begin{aligned}
& \sup_{\mathcal{Z}_{s_1\|s_2,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s_1\|s_2,o)} \Pr(\mathbf{SC}(z_{s_1\|s_2,o}, \alpha)) \\
= & \sup_{\mathcal{Z}_{s_1,o_\alpha} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s_1,o_\alpha)} \Pr(\mathbf{SC}(z_{s_1,o_\alpha}, \alpha)) \cdot \sup_{\mathcal{Z}_{s_2,o} \in \mathrm{Res}^{\mathrm{x}}_{\max}(s_2,o)} \Pr(\mathbf{SC}(z_{s_2,o}, \alpha))
\end{aligned}
$$

□

Appendix C.7. Proof of Theorem 19

Theorem 19. *Let* $(\mathcal{S}, \mathcal{A}, \rightarrow)$ *be a PTS,* $\lambda \in (0,1]$*,* $\omega \colon \mathbf{O} \rightarrow (0,1]$*,* $\mathrm{y} \in \{\mathrm{may}, \mathrm{must}, \mathrm{mM}\}$ *and* $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$*:*

1. $\mathbf{d}^{\omega,\mathrm{rand}}_{\mathrm{Te,y}} = \mathbf{d}^{\omega,\mathrm{det}}_{\mathrm{Te,y}}$.

2. $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Te,tbt}}$.

3. $\mathbf{d}^{\lambda,\mathrm{rand}}_{\mathrm{Te,sup}} = \mathbf{d}^{\lambda,\mathrm{det}}_{\mathrm{Te,sup}}$.

Proof of Theorem 19.1. We expand only the case of must hemimetrics. The proofs for the other cases follow by applying an analogous reasoning.

To prove that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Te,must}} = \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,must}}$, it is enough to prove that for any process $s \in \mathcal{S}$ and for each test $o \in \mathbf{O}$ it holds that $\inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o})) = \inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o}))$. Let $s \in \mathcal{S}$ be an arbitrary process and $o \in \mathbf{O}$ an arbitrary test.

First of all we notice that, by the properties of infima, $\mathrm{Res}^{\mathrm{det}}_{\max}(s,o) \subseteq \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)$ implies

$$
\inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr\nolimits^n(\mathbf{SC}(z_{s,o})) \geq \inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr\nolimits^n(\mathbf{SC}(z_{s,o})).
$$

Next we prove that also the opposite inequality holds, namely that

$$
\inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr\nolimits^n(\mathbf{SC}(z_{s,o})) \leq \inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr\nolimits^n(\mathbf{SC}(z_{s,o})).
$$

We have

$$
\inf_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1} \cdot \Pr\nolimits^n(\mathbf{SC}(z_{s,o}))
$$

$$= \inf_{\substack{\sum_{i\in I} p_i z^i_{s,o} \\ z^i_{s,o}\in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)}} \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(\sum_{i\in I} p_i z^i_{s,o}))$$

$$= \inf_{\substack{\sum_{i\in I} p_i z^i_{s,o} \\ z^i_{s,o}\in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)}} \sum_{i\in I} p_i \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z^i_{s,o}))$$

$$\geq \inf_{p_i\in(0,1],\sum_{i\in I}p_i=1} \sum_{i\in I} p_i \inf_{\mathcal{Z}_{s,o}\in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{s,o}))$$

$$= \inf_{\mathcal{Z}_{s,o}\in \mathrm{Res}^{\mathrm{det}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{s,o})).$$

□

Proof of Theorem 19.2. The proof is similar to that of Theorem 10 in Appendix B.7. □

Proof of Theorem 19.3. The proof is similar to that of Theorem 19.1. □

Appendix C.8. Proof of Theorem 20

Theorem 20. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0, 1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$:*

1. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$ *and* $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$.
2. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$.
3. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.

Proof of Theorem 20.1. The thesis follows directly from definition of $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$ and $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}$:

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}(s,t) = \max\{\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t), \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t)\}$$
$$\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}}(s,t) = \max\{\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t), \mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(s,t)\}.$$

□

Proof of Theorem 20.2. We show that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}$. The proof for the case of the metrics is similar. Given a test $o \in \mathbf{O}$, we can consider the test $o \downarrow \alpha$ where all the states that in o that are not reachable via α are made unsuccessful. We have that:

$$\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}}(s,t)$$
$$= \sup_{o\in\mathbf{O}} \max\Big\{0, \Big(\sup_{\mathcal{Z}_{s,o}\in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{s,o}) - \sup_{\mathcal{Z}_{t,o}\in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \sum_{n=1}^{\infty} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{t,o})\Big)\Big\}$$
$$> \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \max\Big\{0, \Big(\sup_{\mathcal{Z}_{s,o\downarrow\alpha}\in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o\downarrow\alpha)} \lambda^{|\alpha|-1}\Pr^{|\alpha|}(\mathbf{SC}(z_{s,o\downarrow\alpha}) - \sup_{\mathcal{Z}_{t,o\downarrow\alpha}\in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o\downarrow\alpha)} \lambda^{|\alpha|-1}\Pr^{|\alpha|}(\mathbf{SC}(z_{t,o\downarrow\alpha})\Big)\Big\}$$
$$= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \max\Big\{0, \Big(\sup_{\mathcal{Z}_{s,o\downarrow\alpha}\in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o\downarrow\alpha)} \lambda^{|\alpha|-1}\Pr(\mathbf{SC}(z_{s,o\downarrow\alpha},\alpha) - \sup_{\mathcal{Z}_{t,o\downarrow\alpha}\in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o\downarrow\alpha)} \lambda^{|\alpha|-1}\Pr(\mathbf{SC}(z_{t,o\downarrow\alpha},\alpha)\Big)\Big\}$$
$$= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \max\Big\{0, \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_{s,o}\in \mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o},\alpha) - \sup_{\mathcal{Z}_{t,o}\in \mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o},\alpha)\Big)\Big\}$$
$$= \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}(s,t).$$

□

Proof of Theorem 20.3. We show that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} < \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$. The proof for the case of the metrics is similar.

$$\begin{aligned}
&\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}}(s,t)\\
&= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \max\Big\{0, \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o},\alpha)) - \sup_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o},\alpha))\Big)\Big\}\\
&< \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \max\Big\{0, \lambda^{|\alpha|-1}\Big(\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))\Big)\Big\}\\
&< \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{x}}_{\max}(t,o)} \lambda^{|\alpha|-1}\Big|\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))\Big|\\
&= \mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}(s,t).
\end{aligned}$$

□

Appendix D. Proofs of Section 6

Appendix D.1. Proof of Theorem 21

Theorem 21. *Let $(\mathcal{S}, \mathcal{A}, \to)$ be a PTS, $\lambda \in (0,1]$, $\mathrm{x} \in \{\mathrm{det}, \mathrm{rand}\}$ and $\mathbf{d} \in \{\mathbf{h}, \mathbf{m}\}$:*

1. $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,may}} < \mathbf{s}^{\lambda,\mathrm{rand}}$.
2. *For all $s, t \in \mathcal{S}$ it holds $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(t,s) \leq \mathbf{r}^{\lambda,\mathrm{rand}}(s,t)$, and there are $u, v \in \mathcal{S}$ such that $\mathbf{h}^{\lambda,\mathrm{x}}_{\mathrm{Te,must}}(u,v) < \mathbf{r}^{\lambda,\mathrm{rand}}(v,u)$.*
3. $\mathbf{m}^{\lambda,\mathrm{x}}_{\mathrm{Te,mM}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.
4. $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{r}^{\lambda,\mathrm{rand}}$ *and* $\mathbf{m}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}} < \mathbf{b}^{\lambda,\mathrm{rand}}$.
5. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,tbt}} < \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,tbt}}$.
6. $\mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Te,sup}} = \mathbf{d}^{\lambda,\mathrm{x}}_{\mathrm{Tr,sup}}$.

Proof of Theorem 21.1. Since in Theorem 19 we proved that $\mathbf{h}^{\lambda,\mathrm{det}}_{\mathrm{Te,may}} = \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,may}}$, to prove the thesis it is enough to show that $\mathbf{s}^{\lambda,\mathrm{rand}} \geq \mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,may}}$.

With abuse of notation, given $k \in \mathbb{N}$, we write

$$\begin{aligned}&\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t)\\ &= \sup_{o \in \mathbf{O}} \max\left\{0, \Big(\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \sum_{n=1}^{k} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s,o})) - \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t,o)} \sum_{n=1}^{k} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t,o}))\Big)\right\}\end{aligned}$$

so that $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t)$ takes into account only the differences of s, t that can be tested in the first k steps. Notice that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) = \lim_{k\to\infty} \mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t)$. Therefore, to prove the thesis, we prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ \mathbf{s}^{\lambda,\mathrm{rand}}_k(s,t) \geq \mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t). \tag{D.1}$$

The thesis will then follow by Proposition 2 and the monotonicity of the limit.

We proceed by induction over $k \in \mathbb{N}$.

Consider the base case $k = 1$. It is easy to check that

$$\mathbf{s}^{\lambda,\mathrm{rand}}_1(s,t) = \mathbf{h}^{1,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) = \begin{cases} 1 & \text{if } \mathrm{init}(s) \not\subseteq \mathrm{init}(t) \\ 0 & \text{otherwise} \end{cases}$$

and thus Equation (D.1) directly follows.

Consider now the inductive step $k > 1$. If $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) = 0$, then there is nothing to prove. Moreover, notice that $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) = 1$ iff s and t are distinguished by a test of depth 1, namely iff $\mathrm{init}(s) \not\subseteq \mathrm{init}(t)$ and thus iff $\mathbf{s}^{\lambda,\mathrm{rand}}_k(s,t) = 1$. Hence assume that $0 < \mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) < 1$. We have

$$\begin{aligned}&\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t)\\ &= \sup_{o \in \mathbf{O}} \left(\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \sum_{n=1}^{k} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s,o})) - \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t,o)} \sum_{n=1}^{k} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t,o}))\right)\end{aligned} \tag{D.2}$$

By definition of supremum, given $\varepsilon > 0$ there is a test $o_\varepsilon \in \mathbf{O}$ such that $\sup_{o\in\mathbf{O}} \mathbf{h}^{k,o,\lambda,\text{rand}}_{\text{Te,may}}(s,t) < \mathbf{h}^{k,o_\varepsilon,\lambda,\text{rand}}_{\text{Te,may}}(s,t) + \varepsilon$. Therefore

$$(\text{D.2}) < \sup_{\mathcal{Z}_{s,o_\varepsilon}\in\text{Res}^{\text{rand}}_{\max}(s,o_\varepsilon)} \sum_{n=1}^{k} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{s,o_\varepsilon})) - \sup_{\mathcal{Z}_{t,o_\varepsilon}\in\text{Res}^{\text{rand}}_{\max}(t,o_\varepsilon)} \sum_{n=1}^{k} \lambda^{n-1}\cdot \Pr^n(\mathbf{SC}(z_{t,o_\varepsilon})) + \varepsilon \quad (\text{D.3})$$

By rewriting the first transition step of each resolutions explicitly in terms of the probability distribution that is reached, we get

$$\begin{aligned}(\text{D.3}) = & \sup_{a\in\mathcal{A}} \sup_{\pi_{s,o_\varepsilon}\in\text{der}_{\text{ct}}(s,o_\varepsilon,a)} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\pi_{s,o_\varepsilon}(s',o')\cdot \sup_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr^n(\mathbf{SC}(z_{s',o'}))\right) + \right.\\ & \left. + \sum_{s'\in\mathcal{S}} \lambda\cdot\pi_{s,o_\varepsilon}(s',\surd)\right\} + \\ & - \sup_{a\in\mathcal{A}} \sup_{\pi_{t,o_\varepsilon}\in\text{der}_{\text{ct}}(t,o_\varepsilon,a)} \left\{ \sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\pi_{t,o_\varepsilon}(t',o')\cdot \sup_{\mathcal{Z}_{t',o'}\in\text{Res}^{\text{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr^n(\mathbf{SC}(z_{t',o'}))\right) + \right.\\ & \left. + \sum_{t'\in\mathcal{S}} \lambda\cdot\pi_{t,o_\varepsilon}(t',\surd)\right\} + \varepsilon \qquad (\text{D.4})\end{aligned}$$

By definition of supremum, given any $\varepsilon_1 > 0$ we let $\tilde{\pi}_s$ and $\tilde{\pi}_{o_\varepsilon}$ be such that

$$\begin{aligned}& \sup_{\pi_{s,o_\varepsilon}\in\text{der}_{\text{ct}}(s,o_\varepsilon,a)} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\pi_{s,o_\varepsilon}(s',o') \sup_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\Pr^n(\mathbf{SC}(z_{s',o'}))\right) + \sum_{s'\in\mathcal{S}}\lambda\pi_{s,o_\varepsilon}(s',\surd)\right\} \\ & < \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\tilde{\pi}_s(s')\tilde{\pi}_{o_\varepsilon}(o') \sup_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\Pr^n(\mathbf{SC}(z_{s',o'}))\right) + \sum_{s'\in\mathcal{S}}\lambda\tilde{\pi}_s(s')\tilde{\pi}_{o_\varepsilon}(\surd) + \varepsilon_1\end{aligned}$$

Then we let $\varepsilon' = \varepsilon + \varepsilon_1$, and we get

$$\begin{aligned}(\text{D.4}) < & \sup_{a\in\mathcal{A}} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\tilde{\pi}_s(s')\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \sup_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr^n(\mathbf{SC}(z_{s',o'}))\right) + \right.\\ & \left. + \sum_{s'\in\mathcal{S}} \lambda\cdot\tilde{\pi}_s(s')\cdot\tilde{\pi}_{o_\varepsilon}(\surd)\right\} + \\ & - \sup_{a\in\mathcal{A}} \sup_{\pi_{t,o_\varepsilon}\in\text{der}_{\text{ct}}(t,o_\varepsilon,a)} \left\{ \sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\pi_{t,o_\varepsilon}(t',o')\cdot \sup_{\mathcal{Z}_{t',o'}\in\text{Res}^{\text{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr^n(\mathbf{SC}(z_{t',o'}))\right) + \right.\\ & \left. + \sum_{t'\in\mathcal{S}} \lambda\cdot\pi_{t,o_\varepsilon}(t',\surd)\right\} + \varepsilon' \qquad (\text{D.5})\end{aligned}$$

We choose an arbitrary distribution $\tilde{\pi}_{t,o_\varepsilon}$ to substitute the supremum. To choose such a distribution we exploit the distribution $\tilde{\pi}_{o_\varepsilon}$ selected at the previous step and then we exploit the definition of infimum which guarantees that for each $\varepsilon_2 > 0$ there is a distribution $\tilde{\pi}_t$ such that $\inf_{\pi_t\in\text{der}_{\text{ct}}(t,a)} \mathbf{K}(\mathbf{s}^{\lambda,\text{rand}}_{k-1})(\tilde{\pi}_s,\pi_t) > \mathbf{K}(\mathbf{s}^{\lambda,\text{rand}}_{k-1})(\tilde{\pi}_s,\tilde{\pi}_t) - \varepsilon_2$. Then we let $\varepsilon'' = \varepsilon' - \varepsilon_2$, and we get

$$(\text{D.5}) < \sup_{a\in\mathcal{A}} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\tilde{\pi}_s(s')\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \sup_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr^n(\mathbf{SC}(z_{s',o'}))\right) + \right.$$

$$+ \sum_{s' \in \mathcal{S}} \lambda \cdot \tilde{\pi}_s(s') \cdot \tilde{\pi}_{o_\varepsilon}(\surd) \Bigg\} +$$

$$- \sup_{a \in \mathcal{A}} \Bigg\{ \sum_{t' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \Bigg(\lambda \cdot \tilde{\pi}_t(t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \sup_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t',o'})) \Bigg) +$$

$$+ \sum_{t' \in \mathcal{S}} \lambda \cdot \tilde{\pi}_t(t') \cdot \tilde{\pi}_{o_\varepsilon}(\surd) \Bigg\} + \varepsilon'' \tag{D.6}$$

By noticing that since $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s,t) < 1$, we are guaranteed that $\sum_{s' \in \mathcal{S}} \lambda \tilde{\pi}_s(s') \tilde{\pi}_{o_\varepsilon}(\surd) = \lambda \tilde{\pi}_{o_\varepsilon}(\surd) = \sum_{t' \in \mathcal{S}} \lambda \tilde{\pi}_t(t') \tilde{\pi}_{o_\varepsilon}(\surd)$, we get

$$(\mathrm{D.6}) = \sup_{a \in \mathcal{A}} \Bigg\{ \sum_{s' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \Bigg(\lambda \cdot \tilde{\pi}_s(s') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \sup_{\mathcal{Z}_{s',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s',o'})) \Bigg) \Bigg\} +$$

$$- \sup_{a \in \mathcal{A}} \Bigg\{ \sum_{t' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \Bigg(\lambda \cdot \tilde{\pi}_t(t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \sup_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t',o'})) \Bigg) \Bigg\} + \varepsilon'' \tag{D.7}$$

By choosing $\mathfrak{w} = \mathrm{argmin}_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s, \tilde{\pi}_t)} \mathbf{K}(\mathbf{s}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t)$, we get

$$\begin{aligned}
(\mathrm{D.7}) = & \sup_{a \in \mathcal{A}} \Bigg\{ \sum_{s' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \Bigg(\lambda \cdot \big(\sum_{t' \in \mathcal{S}} \mathfrak{w}(s',t') \big) \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \sup_{\mathcal{Z}_{s',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s',o'})) \Bigg) \Bigg\} + \\
& - \sup_{a \in \mathcal{A}} \Bigg\{ \sum_{t' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \Bigg(\lambda \cdot \big(\sum_{s' \in \mathcal{S}} \mathfrak{w}(s',t') \big) \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \sup_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t',o'})) \Bigg) \Bigg\} + \varepsilon'' \\
= & \sup_{a \in \mathcal{A}} \sum_{s',t' \in \mathcal{S}, o' \neq \surd \in \mathbf{O}} \lambda \cdot \mathfrak{w}(s',t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \Bigg(\sup_{\mathcal{Z}_{s',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s',o'})) + \\
& \qquad - \sup_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t',o'})) \Bigg) + \varepsilon'' \\
\leq & \sup_{a \in \mathcal{A}} \sum_{s',t' \in \mathcal{S}} \lambda \cdot \mathfrak{w}(s',t') \cdot \sup_{o \in \mathbf{O}} \cdot \Bigg(\sup_{\mathcal{Z}_{s',o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s',o)} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{s',o})) + \\
& \qquad - \sup_{\mathcal{Z}_{t',o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o)} \sum_{n=1}^{k-1} \lambda^{n-1} \cdot \mathrm{Pr}^n(\mathbf{SC}(z_{t',o})) \Bigg) + \varepsilon'' \\
= & \sup_{a \in \mathcal{A}} \sum_{s',t' \in \mathcal{S}} \lambda \cdot \mathfrak{w}(s',t') \cdot \mathbf{h}^{k-1,\lambda,\mathrm{rand}}_{\mathrm{Te,may}}(s',t') + \varepsilon''
\end{aligned} \tag{D.8}$$

By induction over $k-1$ we get

$$(\mathrm{D.8}) \leq \sup_{a \in \mathcal{A}} \sum_{s',t' \in \mathcal{S}} \lambda \cdot \mathfrak{w}(s',t') \cdot \mathbf{s}^{\lambda,\mathrm{rand}}_{k-1}(s',t') + \varepsilon'' \tag{D.9}$$

By the choice of $\mathfrak{w}$ we get

$$(\mathrm{D.9}) = \sup_{a \in \mathcal{A}} \lambda \cdot \mathbf{K}(\mathbf{s}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t) + \varepsilon'' \tag{D.10}$$

By the choice of $\tilde{\pi}_t$ and $\tilde{\pi}_s$ we get

$$\begin{aligned}(D.10) &\leq \sup_{a\in\mathcal{A}} \sup_{\pi_s\in\mathrm{der}_{\mathrm{ct}}(s,a)} \inf_{\pi_t\in\mathrm{der}_{\mathrm{ct}}(t,a)} \lambda\cdot\mathbf{K}(\mathbf{s}_{k-1}^{\lambda,\mathrm{rand}})(\tilde{\pi}_s,\tilde{\pi}_t)+\varepsilon'' \\ &= \mathbf{s}_k^{\lambda,\mathrm{rand}}(s,t)+\varepsilon''.\end{aligned}$$

Since $\mathbf{h}_{\mathrm{Te,may}}^{k,\lambda,\mathrm{rand}}(s,t) < \mathbf{s}_k^{\lambda,\mathrm{rand}}(s,t)+\varepsilon$ holds for all $\varepsilon>0$, we can conclude that Equation (D.1) holds. □

Proof of Theorem 21.2. Since in Theorem 19 we proved that $\mathbf{h}_{\mathrm{Te,must}}^{\lambda,\mathrm{det}} = \mathbf{h}_{\mathrm{Te,must}}^{\lambda,\mathrm{rand}}$, to prove the thesis it is enough to show that $\mathbf{r}^{\lambda,\mathrm{rand}}(s,t) \geq \mathbf{h}_{\mathrm{Te,must}}^{\lambda,\mathrm{rand}}(t,s)$.

With abuse of notation, given $k\in\mathbb{N}$, we write

$$\begin{aligned}&\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s) \\ &= \sup_{o\in\mathbf{O}} \max\left\{0, \left(\inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(t,o)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t,o})) - \inf_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(s,o)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s,o}))\right)\right\}\end{aligned}$$

so that $\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s)$ takes into account only the differences of t,s that can be tested in the first k steps. Notice that $\mathbf{h}_{\mathrm{Te,must}}^{\lambda,\mathrm{rand}}(t,s) = \lim_{k\to\infty}\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s)$. Therefore, to prove the thesis, we prove the stronger property that

$$\text{for each } k\in\mathbb{N},\ \mathbf{r}_k^{\lambda,\mathrm{rand}}(s,t) \geq \mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s). \tag{D.11}$$

The thesis will then follow by Proposition 2 and the monotonicity of the limit.

We proceed by induction over $k\in\mathbb{N}$.

Consider the base case $k=1$. It is easy to check that

$$\mathbf{r}_1^{\lambda,\mathrm{rand}}(s,t) = \mathbf{h}_{\mathrm{Te,must}}^{1,\lambda,\mathrm{rand}}(t,s) = \begin{cases}1 & \text{if } \mathrm{init}(s)\neq\mathrm{init}(t)\\ 0 & \text{otherwise}\end{cases}$$

and thus Equation (D.11) directly follows.

Consider now the inductive step $k>1$. If $\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s)=0$, then there is nothing to prove. Moreover, notice that $\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s)=1$ iff t and s are distinguished by a test of depth 1, namely iff $\mathrm{init}(s)\neq\mathrm{init}(t)$ and thus iff $\mathbf{r}_k^{\lambda,\mathrm{rand}}(s,t)=1$. Hence assume that $0<\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s)<1$. We have

$$\begin{aligned}&\mathbf{h}_{\mathrm{Te,must}}^{k,\lambda,\mathrm{rand}}(t,s) \\ &= \sup_{o\in\mathbf{O}} \left(\inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(t,o)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t,o})) - \inf_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(s,o)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s,o}))\right)\end{aligned} \tag{D.12}$$

By definition of supremum, for each $\varepsilon>0$ there is a test $o_\varepsilon\in\mathbf{O}$, such that $\sup_{o\in\mathbf{O}}\mathbf{h}_{\mathrm{Te,must}}^{k,o,\lambda,\mathrm{rand}}(s,t) < \mathbf{h}_{\mathrm{Te,must}}^{k,o_\varepsilon,\lambda,\mathrm{rand}}(s,t)+\varepsilon$. Therefore

$$(D.12) < \inf_{\mathcal{Z}_{t,o_\varepsilon}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(t,o_\varepsilon)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t,o_\varepsilon})) - \inf_{\mathcal{Z}_{s,o_\varepsilon}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(s,o_\varepsilon)} \sum_{n=1}^{k}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s,o_\varepsilon}))+\varepsilon \tag{D.13}$$

By rewriting the first transition step of each resolutions explicitly in terms of the probability distribution that is reached, we get

$$(D.13) = \sup_{a\in\mathcal{A}} \inf_{\pi_{t,o_\varepsilon}\in\mathrm{der}_{\mathrm{ct}}(t,o_\varepsilon,a)} \left\{ \sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\pi_{t,o_\varepsilon}(t',o')\cdot \inf_{\mathcal{Z}_{t',o'}\in\mathrm{Res}_{\max}^{\mathrm{rand}}(t',o')} \sum_{n=1}^{k-1}\lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t',o'}))\right) + \right.$$

$$+\sum_{t'\in\mathcal{S}}\lambda\cdot\pi_{t,o_\varepsilon}(t',\surd)\Bigg\}+$$

$$-\sup_{a\in\mathcal{A}}\inf_{\pi_{s,o_\varepsilon}\in\mathrm{der}_{\mathrm{ct}}(s,o_\varepsilon,a)}\Bigg\{\sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\cdot\pi_{s,o_\varepsilon}(s',o')\cdot\inf_{\mathcal{Z}_{s',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\cdot\mathrm{Pr}^n(\mathbf{SC}(z_{s',o'}))\Bigg)+$$

$$+\sum_{s'\in\mathcal{S}}\lambda\cdot\pi_{s,o_\varepsilon}(s',\surd)\Bigg\}+\varepsilon \tag{D.14}$$

By definition of infimum, given any $\varepsilon_1>0$ we let $\tilde\pi_s$ and $\tilde\pi_{o_\varepsilon}$ be such that

$$\inf_{\pi_{s,o_\varepsilon}\in\mathrm{der}_{\mathrm{ct}}(s,o_\varepsilon,a)}\Bigg\{\sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\pi_{s,o_\varepsilon}(s',o')\sup_{\mathcal{Z}_{s',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s',o'}))\Bigg)+\sum_{s'\in\mathcal{S}}\lambda\pi_{s,o_\varepsilon}(s',\surd)\Bigg\}$$
$$>\sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\tilde\pi_s(s')\tilde\pi_{o_\varepsilon}(o')\sup_{\mathcal{Z}_{s',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\mathrm{Pr}^n(\mathbf{SC}(z_{s',o'}))\Bigg)+\sum_{s'\in\mathcal{S}}\lambda\tilde\pi_s(s')\tilde\pi_{o_\varepsilon}(\surd)-\varepsilon_1$$

Then we let $\varepsilon'=\varepsilon-\varepsilon_1$, we get

$$(\mathrm{D.14})<\sup_{a\in\mathcal{A}}\inf_{\pi_{t,o_\varepsilon}\in\mathrm{der}_{\mathrm{ct}}(t,o_\varepsilon,a)}\Bigg\{\sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\cdot\pi_{t,o_\varepsilon}(t',o')\cdot\inf_{\mathcal{Z}_{t',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\cdot\mathrm{Pr}^n(\mathbf{SC}(z_{t',o'}))\Bigg)+$$

$$+\sum_{t'\in\mathcal{S}}\lambda\cdot\pi_{t,o_\varepsilon}(t',\surd)\Bigg\}$$

$$-\sup_{a\in\mathcal{A}}\Bigg\{\sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\cdot\tilde\pi_s(s')\cdot\tilde\pi_{o_\varepsilon}(o')\cdot\inf_{\mathcal{Z}_{s',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\cdot\mathrm{Pr}^n(\mathbf{SC}(z_{s',o'}))\Bigg)+$$

$$+\sum_{s'\in\mathcal{S}}\lambda\cdot\tilde\pi_s(s')\cdot\tilde\pi_{o_\varepsilon}(\surd)\Bigg\}+\varepsilon' \tag{D.15}$$

We choose an arbitrary distribution $\tilde\pi_{t,o_\varepsilon}$ to substitute the infimum. To choose such a distribution we exploit the distribution $\tilde\pi_{o_\varepsilon}$ selected at the previous step and then we exploit the definition of infimum which guarantees that for each $\varepsilon_2>0$ there is a distribution $\tilde\pi_t$ such that $\inf_{\pi_t\in\mathrm{der}_{\mathrm{ct}}(t,a)}\mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde\pi_s,\pi_t)>\mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde\pi_s,\tilde\pi_t)-\varepsilon_2$. Then we let $\varepsilon''=\varepsilon'-\varepsilon_2$, and we get

$$(\mathrm{D.15})<\sup_{a\in\mathcal{A}}\Bigg\{\sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\cdot\tilde\pi_t(t')\cdot\tilde\pi_{o_\varepsilon}(o')\cdot\inf_{\mathcal{Z}_{t',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\cdot\mathrm{Pr}^n(\mathbf{SC}(z_{t',o'}))\Bigg)+$$

$$+\sum_{t'\in\mathcal{S}}\lambda\cdot\tilde\pi_t(t')\cdot\tilde\pi_{o_\varepsilon}(\surd)\Bigg\}+$$

$$-\sup_{a\in\mathcal{A}}\Bigg\{\sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}}\Bigg(\lambda\cdot\tilde\pi_s(s')\cdot\tilde\pi_{o_\varepsilon}(o')\cdot\inf_{\mathcal{Z}_{s',o'}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s',o')}\sum_{n=1}^{k-1}\lambda^{n-1}\cdot\mathrm{Pr}^n(\mathbf{SC}(z_{s',o'}))\Bigg)+$$

$$+\sum_{s'\in\mathcal{S}}\lambda\cdot\tilde\pi_s(s')\cdot\tilde\pi_{o_\varepsilon}(\surd)\Bigg\}+\varepsilon'' \tag{D.16}$$

By noticing that since $\mathbf{h}^{k,\lambda,\text{rand}}_{\text{Te,must}}(s,t) < 1$, we are guaranteed that $\sum_{s'\in\mathcal{S}} \lambda\tilde{\pi}_s(s')\tilde{\pi}_{o_\varepsilon}(\surd) = \lambda\tilde{\pi}_{o_\varepsilon}(\surd) = \sum_{t'\in\mathcal{S}} \lambda\tilde{\pi}_t(t')\tilde{\pi}_{o_\varepsilon}(\surd)$, we get

$$(\text{D.16}) = \sup_{a\in\mathcal{A}} \left\{ \sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\tilde{\pi}_t(t')\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \inf_{\mathcal{Z}_{t',o'}\in\text{Res}^{\text{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t',o'}))\right)\right\} +$$
$$- \sup_{a\in\mathcal{A}} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\tilde{\pi}_s(s')\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \inf_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s',o'}))\right)\right\} + \varepsilon'' \quad (\text{D.17})$$

By choosing $\mathfrak{w} = \text{argmin}_{\mathfrak{w}\in\mathfrak{W}(\tilde{\pi}_s,\tilde{\pi}_t)} \mathbf{K}(\mathbf{r}^{\lambda,\text{rand}}_{k-1})(\tilde{\pi}_s,\tilde{\pi}_t)$, we get

$$(\text{D.17}) = \sup_{a\in\mathcal{A}} \left\{ \sum_{t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\big(\sum_{s'\in\mathcal{S}}\mathfrak{w}(s',t')\big)\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \inf_{\mathcal{Z}_{t',o'}\in\text{Res}^{\text{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t',o'}))\right)\right\} +$$
$$- \sup_{a\in\mathcal{A}} \left\{ \sum_{s'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \left(\lambda\cdot\big(\sum_{t'\in\mathcal{S}}\mathfrak{w}(s',t')\big)\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot \inf_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s',o'}))\right)\right\} + \varepsilon''$$
$$= \sup_{a\in\mathcal{A}} \sum_{s',t'\in\mathcal{S},o'\neq\surd\in\mathbf{O}} \lambda\cdot\mathfrak{w}(s',t')\cdot\tilde{\pi}_{o_\varepsilon}(o')\cdot\Bigg(\inf_{\mathcal{Z}_{t',o'}\in\text{Res}^{\text{rand}}_{\max}(t',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t',o'})) +$$
$$- \inf_{\mathcal{Z}_{s',o'}\in\text{Res}^{\text{rand}}_{\max}(s',o')} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s',o'}))\Bigg) + \varepsilon''$$
$$\leq \sup_{a\in\mathcal{A}} \sum_{s',t'\in\mathcal{S}} \lambda\cdot\mathfrak{w}(s',t')\cdot\sup_{o\in\mathbf{O}}\cdot\Bigg(\inf_{\mathcal{Z}_{t',o}\in\text{Res}^{\text{rand}}_{\max}(t',o)} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{t',o})) +$$
$$- \inf_{\mathcal{Z}_{s',o}\in\text{Res}^{\text{rand}}_{\max}(s',o)} \sum_{n=1}^{k-1} \lambda^{n-1}\cdot\Pr{}^n(\mathbf{SC}(z_{s',o}))\Bigg) + \varepsilon''$$
$$= \sup_{a\in\mathcal{A}} \sum_{s',t'\in\mathcal{S}} \lambda\cdot\mathfrak{w}(s',t')\cdot\mathbf{h}^{k-1,\lambda,\text{rand}}_{\text{Te,must}}(t',s') + \varepsilon'' \quad (\text{D.18})$$

By induction over $k-1$ we get

$$(\text{D.18}) \leq \sup_{a\in\mathcal{A}} \sum_{s',t'\in\mathcal{S}} \lambda\cdot\mathfrak{w}(s',t')\cdot\mathbf{r}^{\lambda,\text{rand}}_{k-1}(s',t') + \varepsilon'' \quad (\text{D.19})$$

By the choice of $\mathfrak{w}$ we get

$$(\text{D.19}) = \sup_{a\in\mathcal{A}} \lambda\cdot\mathbf{K}(\mathbf{r}^{\lambda,\text{rand}}_{k-1})(\tilde{\pi}_s,\tilde{\pi}_t) + \varepsilon'' \quad (\text{D.20})$$

By the choice of $\tilde{\pi}_t$ and $\tilde{\pi}_s$, we get

$$(\text{D.20}) \leq \sup_{a\in\mathcal{A}} \sup_{\pi_s\in\text{der}_{\text{ct}}(s,a)} \inf_{\pi_t\in\text{der}_{\text{ct}}(t,a)} \lambda\cdot\mathbf{K}(\mathbf{r}^{\lambda,\text{rand}}_{k-1})(\tilde{\pi}_s,\tilde{\pi}_t) + \varepsilon''$$
$$= \mathbf{r}^{\lambda,\text{rand}}_{k}(s,t) + \varepsilon''.$$

Since $\mathbf{h}^{k,\lambda,\text{rand}}_{\text{Te,must}}(t,s) < \mathbf{r}^{\lambda,\text{rand}}_{k}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that Equation (D.11) holds. □

Proof of Theorem 21.3. The relation $\mathbf{b}^{\lambda,\text{rand}} \geq \mathbf{m}^{\lambda,\text{x}}_{\text{Te,mM}}$ is an immediate consequence of Theorem 4, Theorem 21.1, Theorem 21.2 and Definition 23. □

Proof of Theorem 21.4. We expand only the case of hemimetrics. The case of metrics will then follow by symmetrization.

With abuse of notation, given $k \in \mathbb{N}$, we write

$$\begin{aligned}&\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t)\\ &= \sup_{o\in\mathbf{O}} \sup_{\substack{\alpha\in\mathcal{A}^\star\\ |\alpha|\le k}} \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t,o)} \lambda^{|\alpha|-1}|\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))|\end{aligned}$$

so that $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t)$ takes into account only the differences of s,t that can be tested in the first k steps. Notice that $\mathbf{h}^{\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t) = \lim_{k\to\infty} \mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t)$. Therefore, to prove the thesis, we prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ \mathbf{r}^{\lambda,\mathrm{rand}}_{k}(s,t) \ge \mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t). \tag{D.21}$$

The thesis will then follow by Proposition 2 and the monotonicity of the limit.

We proceed by induction over $k \in \mathbb{N}$.

Consider the base case $k = 1$. It is easy to check that

$$\mathbf{r}^{\lambda,\mathrm{rand}}_{1}(s,t) = \mathbf{h}^{1,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t) = \begin{cases}1 & \text{if } \mathrm{init}(s) \ne \mathrm{init}(t)\\ 0 & \text{otherwise}\end{cases}$$

and thus Equation (D.21) directly follows.

Consider now the inductive step $k > 1$. If $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t) = 0$, then there is nothing to prove. Hence assume that $\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(s,t) > 0$. We have

$$\begin{aligned}&\mathbf{h}^{k,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(t,s)\\ &= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star,|\alpha|\le k} \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t,o)} \lambda^{|\alpha|-1}\cdot|\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))|\end{aligned} \tag{D.22}$$

By definition of supremum, for each $\varepsilon > 0$ there are $o_\varepsilon \in \mathbf{O}$, $\alpha_\varepsilon \in \mathcal{A}^\star$ and $\mathcal{Z}^{\varepsilon}_{s,o_\varepsilon} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s,o_\varepsilon)$ such that

$$\begin{aligned}&\sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star,|\alpha|\le k} \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t,o)} \lambda^{|\alpha|-1}\cdot|\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))|\\ &< \inf_{\mathcal{Z}_{t,o_\varepsilon}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t,o_\varepsilon)} \lambda^{|\alpha_\varepsilon|-1}\cdot|\Pr(\mathbf{SC}(z^{\varepsilon}_{s,o_\varepsilon},\alpha_\varepsilon)) - \Pr(\mathbf{SC}(z_{t,o_\varepsilon},\alpha_\varepsilon))| + \varepsilon.\end{aligned}$$

We can assume wlog. that $\alpha_\varepsilon = a\alpha'$ for some $\alpha' \in \mathcal{A}^\star$ s.t. $|\alpha'| \le k-1$. We remark that $z^{\varepsilon}_{s,o_\varepsilon}$ is of the form $\tilde{z}_s \parallel \tilde{z}_{o_\varepsilon}$ for some $\tilde{\mathcal{Z}}_s \in \mathrm{Res}^{\mathrm{rand}}(s)$, $\tilde{\mathcal{Z}}_{o_\varepsilon} \in \mathrm{Res}^{\mathrm{rand}}(o_\varepsilon)$. We denote by $\tilde{\pi}_s \in \mathrm{der}_{\mathrm{ct}}(s,a)$ the distribution reached by $\tilde{z}_s$ via the execution of a, and by $\tilde{\pi}_{o_\varepsilon}$ the analogous for z_{o_ε}. Therefore

$$(\mathrm{D.22}) < \inf_{\mathcal{Z}_{t,o_\varepsilon}\in\mathrm{Res}^{\mathrm{rand}}_{\max}(t,o_\varepsilon)} \lambda^{|\alpha_\varepsilon|-1}\cdot|\Pr(\mathbf{SC}(z^{\varepsilon}_{s,o_\varepsilon},\alpha_\varepsilon)) - \Pr(\mathbf{SC}(z_{t,o_\varepsilon},\alpha_\varepsilon))| + \varepsilon \tag{D.23}$$

We choose an arbitrary resolution $\mathcal{Z}^{\varepsilon}_{t,o_\varepsilon} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t, o_\varepsilon)$ to substitute the infimum. We construct $\mathcal{Z}^{\varepsilon}_{t,o_\varepsilon}$ as follows. We exploit the distribution $\tilde{\pi}_s$ from the previous step and then we exploit the definition of infimum which guarantees that for each $\varepsilon_1 > 0$ there is a distribution $\tilde{\pi}_t$ such that $\inf_{\pi_t \in \mathrm{der}_{\mathrm{ct}}(t,a)} \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \pi_t) > \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t) - \varepsilon_1$. Then we let $\mathfrak{w} = \mathrm{argmin}_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi_s}, \tilde{\pi}_t)} \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t)$. Choose $\varepsilon_2 > 0$. For each $s' \in \mathsf{supp}(\tilde{\pi}_s), t' \in \mathsf{supp}(\tilde{\pi}_t), o' \in \mathsf{supp}(\pi_{o_\varepsilon})$, we let $\mathcal{Z}^{\varepsilon_2}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t', o')$ be such that

$$\begin{aligned}&\inf_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \lambda^{|\alpha'|-1} \cdot |\Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) - \Pr(\mathbf{SC}(z_{t',o'}, \alpha'))| \\ &> |\Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) - \Pr(\mathbf{SC}(z^{\varepsilon_2}_{t',o'}, \alpha'))| - \varepsilon_2.\end{aligned}$$

Then we let $z^{\varepsilon}_{t,o_\varepsilon} \xrightarrow{a} \pi$ with $\pi = \sum_{t' \in \mathcal{S}, o' \in \mathbf{O}} \tilde{\pi}_t(t') \tilde{\pi}_{o_\varepsilon}(o') \delta_{z^{\varepsilon_2}_{t',o'}}$. Therefore

$$\begin{aligned}\text{(D.23)} &\leq \lambda^{|\alpha_\varepsilon|-1} \cdot |\Pr(\mathbf{SC}(z^{\varepsilon}_{s,o_\varepsilon}, \alpha_\varepsilon)) - \Pr(\mathbf{SC}(z^{\varepsilon}_{t,o_\varepsilon}, \alpha_\varepsilon))| + \varepsilon \\ &= \lambda^{|\alpha_\varepsilon|-1} \cdot \left| \sum_{s' \in \mathcal{S}, o' \in \mathbf{O}} \tilde{\pi}_s(s') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) - \sum_{t' \in \mathcal{S}, o' \in \mathbf{O}} \tilde{\pi}_t(t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \Pr(\mathbf{SC}(z^{\varepsilon_2}_{t',o'}, \alpha')) \right| \end{aligned} \quad \text{(D.24)}$$

By the choice of $\mathfrak{w}$, we get

$$\begin{aligned}\text{(D.24)} &= \lambda^{|\alpha_\varepsilon|-1} \cdot \Bigg| \sum_{s' \in \mathcal{S}, o' \in \mathbf{O}} (\sum_{t' \in \mathcal{S}} \mathfrak{w}(s', t')) \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) + \\ &\quad - \sum_{t' \in \mathcal{S}, o' \in \mathbf{O}} (\sum_{s' \in \mathcal{S}} \mathfrak{w}(s', t')) \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \Pr(\mathbf{SC}(z^{\varepsilon_2}_{t',o'}, \alpha')) \Bigg| + \varepsilon \\ &\leq \lambda \cdot \sum_{s', t' \in \mathcal{S}, o' \in \mathbf{O}} \mathfrak{w}(s', t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \cdot \lambda^{|\alpha'|-1} \cdot |\Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) - \Pr(\mathbf{SC}(z^{\varepsilon_2}_{t',o'}, \alpha'))| + \varepsilon \end{aligned} \quad \text{(D.25)}$$

By the choice of each $z^{\varepsilon_2}_{t',o'}$, by letting $\varepsilon' = \varepsilon + \varepsilon_2$, we get

$$\begin{aligned}\text{(D.25)} &< \lambda \cdot \sum_{s', t' \in \mathcal{S}, o' \in \mathbf{O}} \mathfrak{w}(s', t') \cdot \tilde{\pi}_{o_\varepsilon}(o') \inf_{\mathcal{Z}_{t',o'} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o')} \lambda^{|\alpha'|-1} \cdot |\Pr(\mathbf{SC}(z^{\varepsilon}_{s',o'}, \alpha')) - \Pr(\mathbf{SC}(z_{t',o'}, \alpha'))| + \varepsilon' \\ &\leq \lambda \cdot \sum_{s', t' \in \mathcal{S}, o' \in \mathbf{O}} \mathfrak{w}(s', t') \cdot \sup_{o \in \mathbf{O}} \sup_{\alpha \in \mathcal{A}^\star, |\alpha| \leq k-1} \\ &\qquad \sup_{\mathcal{Z}_{s',o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(s',o)} \inf_{\mathcal{Z}_{t',o} \in \mathrm{Res}^{\mathrm{rand}}_{\max}(t',o)} \lambda^{|\alpha|-1} \cdot |\Pr(\mathbf{SC}(z_{s',o}, \alpha)) - \Pr(\mathbf{SC}(z_{t',o}, \alpha))| + \varepsilon' \\ &= \sum_{s', t' \in \mathcal{S}} \lambda \cdot \mathfrak{w}(s', t') \cdot \mathbf{h}^{k-1,\lambda,\mathrm{rand}}_{\mathrm{Te,tbt}}(t', s') + \varepsilon' \end{aligned} \quad \text{(D.26)}$$

By induction over $k - 1$, we get

$$\text{(D.26)} \leq \sum_{s', t' \in \mathcal{S}} \lambda \cdot \mathfrak{w}(s', t') \cdot \mathbf{r}^{\lambda,\mathrm{rand}}_{k-1}(s', t') + \varepsilon' \quad \text{(D.27)}$$

By the choice of $\mathfrak{w}$, we get

$$\text{(D.27)} = \lambda \cdot \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t) + \varepsilon' \quad \text{(D.28)}$$

By the choice of $\tilde{\pi}_t$ and $\tilde{\pi}_s$, by letting $\varepsilon'' = \varepsilon' + \varepsilon_1$, we get

$$\text{(D.28)} < \sup_{a \in \mathcal{A}} \sup_{\pi_s \in \mathrm{der}_{\mathrm{ct}}(s,a)} \inf_{\pi_t \in \mathrm{der}_{\mathrm{ct}}(t,a)} \lambda \cdot \mathbf{K}(\mathbf{r}^{\lambda,\mathrm{rand}}_{k-1})(\tilde{\pi}_s, \tilde{\pi}_t) + \varepsilon''$$

$$= \mathbf{r}_k^{\lambda,\text{rand}}(s,t) + \varepsilon''.$$

Since $\mathbf{h}_{\text{Te,tbt}}^{k,\lambda,\text{rand}}(s,t) < \mathbf{r}_k^{\lambda,\text{rand}}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that Equation (D.21) holds. □

Proof of Theorem 21.5. The thesis follows by noticing that each trace $\alpha \in \mathcal{A}^\star$ corresponds to a particular test, o_α, that has only one α-compatible maximal computation. For completeness, in the case of hemimetrics, we have

$$\begin{aligned}
&\mathbf{h}_{\text{Tr,tbt}}^{\lambda,\text{x}}(s,t)\\
&= \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \sup_{\mathcal{Z}_s\in\text{Res}^{\text{x}}(s)} \inf_{\mathcal{Z}_t\in\text{Res}^{\text{x}}(t)} |\Pr(\mathcal{C}(z_s,\alpha)) - \Pr(\mathcal{C}(z_t,\alpha))|\\
&= \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \sup_{\mathcal{Z}_{s,o_\alpha}\in\text{Res}^{\text{x}}_{\max}(s,o_\alpha)} \inf_{\mathcal{Z}_{t,o_\alpha}\in\text{Res}^{\text{x}}(t,o_\alpha)} |\Pr(\mathbf{SC}(z_{s,o_\alpha},\alpha)) - \Pr(\mathbf{SC}(z_{t,o_\alpha},\alpha))|\\
&\leq \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \sup_{\mathcal{Z}_{s,o}\in\text{Res}^{\text{x}}_{\max}(s,o)} \inf_{\mathcal{Z}_{t,o}\in\text{Res}^{\text{x}}(t,o)} |\Pr(\mathbf{SC}(z_{s,o},\alpha)) - \Pr(\mathbf{SC}(z_{t,o},\alpha))|\\
&= \mathbf{h}_{\text{Te,tbt}}^{\lambda,\text{x}}(s,t).
\end{aligned}$$

□

Proof of Theorem 21.6. We expand only the case of hemimetrics. The case of metrics can be obtain by symmetrization.

Firstly we notice that $\mathbf{h}_{\text{Tr,sup}}^{\lambda,\text{x}}(s,t) \leq \mathbf{h}_{\text{Te,sup}}^{\lambda,\text{x}}(s,t)$ immediately follows by noticing that each trace $\alpha \in \mathcal{A}^\star$ corresponds to a particular test.

Hence, to prove the thesis, it is enough to show that $\mathbf{h}_{\text{Te,sup}}^{\lambda,\text{x}}(s,t) \leq \mathbf{h}_{\text{Tr,sup}}^{\lambda,\text{x}}(s,t)$. Clearly, if $\mathbf{h}_{\text{Te,sup}}^{\lambda,\text{x}}(s,t) = 0$, then there is nothing to prove. So, assume $\mathbf{h}_{\text{Te,sup}}^{\lambda,\text{x}}(s,t) > 0$. We have

$$\begin{aligned}
&\mathbf{h}_{\text{Te,sup}}^{\lambda,\text{x}}(s,t)\\
&= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \left(\sup_{\mathcal{Z}_{s,o}\in\text{Res}^{\text{x}}_{\max}(s,o)} \Pr(\mathbf{SC}(z_{s,o},\alpha)) - \sup_{\mathcal{Z}_{t,o}\in\text{Res}^{\text{x}}_{\max}(t,o)} \Pr(\mathbf{SC}(z_{t,o},\alpha)) \right)\\
&= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \Bigg(\sup_{\mathcal{Z}_s\in\text{Res}^{\text{x}}(s)} \sup_{\mathcal{Z}_o\in\text{Res}^{\text{x}}_{\max}(o)} \Pr(\mathcal{C}(z_s,\alpha))\Pr(\mathbf{SC}(z_o,\alpha)) +\\
&\qquad\qquad - \sup_{\mathcal{Z}_t\in\text{Res}^{\text{x}}(t)} \sup_{\mathcal{Z}_o\in\text{Res}^{\text{x}}_{\max}(o)} \Pr(\mathcal{C}(z_t,\alpha))\Pr(\mathbf{SC}(z_o,\alpha)) \Bigg)\\
&= \sup_{o\in\mathbf{O}} \sup_{\alpha\in\mathcal{A}^\star} \sup_{\mathcal{Z}_o\in\text{Res}^{\text{x}}_{\max}(o)} \Pr(\mathbf{SC}(z_o,\alpha))\lambda^{|\alpha|-1} \left(\sup_{\mathcal{Z}_s\in\text{Res}^{\text{x}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t\in\text{Res}^{\text{x}}(t)} \Pr(\mathcal{C}(z_t,\alpha)) \right)\\
&= \sup_{\alpha\in\mathcal{A}^\star} \lambda^{|\alpha|-1} \left(\sup_{\mathcal{Z}_s\in\text{Res}^{\text{x}}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t\in\text{Res}^{\text{x}}(t)} \Pr(\mathcal{C}(z_t,\alpha)) \right)\\
&= \mathbf{h}_{\text{Tr,sup}}^{\lambda,\text{x}}(s,t).
\end{aligned}$$

□