Università degli Studi dell’Insubria
Dipartimento di Scienza e Alta Tecnologia

Probabilistic Semantics: Metric and Logical Characterizations for Nondeterministic Probabilistic Processes

CANDIDATE:
Valentina Castiglioni

SUPERVISOR:
Prof. Simone Tini

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computer Science and Computational Mathematics

December 2017

Declaration

I, Valentina Castiglioni, declare that this Ph.D thesis entitled "*Probabilistic Semantics: Metric and Logical Characterizations for Nondeterministic Probabilistic Processes*" was carried out by me for the degree of Doctor of Philosophy in Computer Science and Computational Mathematics under the guidance and supervision of Prof. Simone Tini, Department of Science and High Technology, University of Insubria Varese-Como, Italy.

I declare that all the material presented for examination is my own work and has not been written for me, in whole or in part, by any other person.

I also declare that any quotation or paraphrase from the published or unpublished work of another person has been duly acknowledged in the work which I present for examination.

This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma.

Place: Gorla Maggiore

Date: 7 December 2017

(Valentina Castiglioni)

Valentina Castiglioni

Abstract

The constantly growing interest in probabilistic systems pushes to the research of more and more efficient techniques to capture and compare their behavior. In this dissertation we focus on Segala's type systems, namely processes with nondeterminism and probability in the *PTS model*, and we propose novel techniques to study their semantics, in terms of both classic behavioral relations and the more recent behavioral metrics.

For what concerns behavioral relations our main contribution is a *method for decomposing modal formulae* in a probabilistic extension of the Hennessy-Milner logic. This decomposition method will allow us to derive the compositional properties of probabilistic (bi)simulations.

Roughly speaking, in 2004 Bard Bloom, Wan Fokkink and Rob van Glabbeek noticed that the definition of the semantic behavior of processes by means of the *Structural Operational Semantics* (SOS) framework allows for decomposing the satisfaction problem of a formula for a process into the verification of the satisfaction problem of certain formulae for its subprocesses by means of the notion of *ruloid*, namely inference transition rules that are derived from the SOS specification and define the behavior of open processes in terms of the behavior of their variables. Then, they exploited the decomposition of modal formulae to systematically derive expressive (pre)congruence formats for several behavioral equivalences (preorders) from their modal characterizations.

We will extend their approach to the probabilistic setting. In particular, to obtain the decomposition method we will introduce an *SOS-like machinery*, specifying the behavior of *distribution terms* as probability distributions over process terms, that will allow us to decompose the probabilistic modalities proper of the considered probabilistic extension of the Hennessy-Milner logic.

We remark that the one presented in this thesis is actually the first decomposition method proposed for processes in the PTS model.

Then we will focus on behavioral metrics.

First of all we will propose original notions of metrics measuring the disparities in the behavior of processes with respect to (*decorated*) *trace and testing semantics* as expressed by Marco Bernardo, Rocco de Nicola and Michele Loreti: differently from the original approach of Roberto Segala in which traces were *distributions over traces*, they proposed a *trace-by-trace* approach for the definition of the linear semantics of processes in the PTS model which turned out to be compositional and fully backward compatible with the fully-nondeterministic case. By following their approach, we will obtain behavioral metrics capturing these linear semantics and whose kernels will enjoy these desirable properties. In particular, we remark that our metrics for the decorated traces and testing are the first quantitative version of these semantics ever proposed.

To capture the differences in the expressive power of the novel metrics and the ones for probabilistic (bi)simulations we will order them by the relation '*makes processes further than*'. Thus we will obtain the first *spectrum of behavioral metrics* on processes in the PTS model. Interestingly, from this spectrum we will derive an analogous one for the kernels of the considered metrics, ordered by the relation '*make strictly less identification than*'. Our *spectrum of probabilistic relations* is the probabilistic generalization of the linear time - branching time spectrum of van Glabbeek.

Finally we will introduce a novel technique for the *logical characterization* of both behavioral metrics and their kernels, based on the notions of *mimicking formula* and *distance on formulae*. Behavioral relations and modal logics have been successfully employed for the specification and verification of processes. The former ones provide a simple and elegant tool to compare the observable behavior of processes. The latter ones allow for an immediate expression of the desired properties of processes. Since the seminal work of Matthew Hennessy and Robin Milner on their namesake logic, these two approaches are connected by means of *logical characterizations* of behavioral relations: two processes are in relation if and only if they satisfy the same formulae in the considered logic. Starting from this characterization of behavioral relations, in the literature we can find several proposals of characterizations of behavioral metrics following (mostly) the same approach: in general logics equipped with a real-valued semantics are used for the characterization, which is then expressed as $d(s,t) = \sup_{\varphi \in L} |[\varphi](s) - [\varphi](t)|$, where d is the behavioral metric of interest, s and t are two processes, L is the considered logic and $[\varphi](s)$ denotes the value of the formula φ at process s accordingly to the real-valued semantics.

We propose a novel approach that will allow us to obtain the logical characterization of behavioral metrics starting from boolean-valued logics. The idea is the following: 1. Once we have chosen a class L of modal boolean-valued formulae suitable for the considered semantics, for a process s we identify a special formula expressing the relevant properties of s with respect to the considered semantics, called *mimicking formula* of s. This is a formula in L that captures the nondeterministic and probabilistic behavior of the process that is relevant for the considered semantics. 2. Then, we transform the modal logic L into a metric space by introducing a notion of *syntactical distance* on formulae. This is a 1-bounded pseudometric assigning to each pair of formulae a suitable quantitative analogue of their syntactic disparities. 3. We conclude by defining a *logical distance* on processes corresponding to the distance between their mimicking formulae and proving that this logical distance characterizes the considered metric semantics. This kind of characterization will allow us to obtain the first example of a spectrum of behavioral distances on processes obtained directly from modal logics. Moreover, we will show that the kernels of the considered metrics can be characterized by simply comparing the mimicking formulae of processes.

Acknowledgments

Foremost, I would like to thank my advisor, Simone Tini, for his time, patience and guidance. During these years he helped me to grow as a researcher by reading my drafts, discussing ideas, encouraging me and offering valuable advice. Simone has been a generous and brilliant advisor and I couldn't hope to have a better mentor than him.

I would also like to thank Ruggero Lanotte, who always had some good advice for me, and Marco Donatelli, for his humor and friendship.

Many thanks to the *not so* anonymous reviewers, Michele Loreti and Rob van Glabbeek. They have my gratitude for their careful reading of my work and for their valuable comments on it. Their constructive reviews helped me writing a final version of this thesis that I can really be proud of.

Finally, I hope that you would excuse me if for this last paragraph I forget formality and English. . . ma certe cose si possono scrivere solo in italiano.

Vorrei ringraziare tutti i colleghi di Como per aver condiviso con me la folle esperienza del dottorato.
Una menzione speciale, tutta la mia gratitudine, a Mele, Mari, Fabio e Isa per i bellissimi momenti passati insieme, lo scambio di dialetti, le domeniche di MotoGP, il mutuo supporto, e per tutta la pizza!

Infine, non posso non ringraziare la mia famiglia, mia mamma Elia, mio papà Franco e mio fratello Stefano, per il loro supporto incondizionato e per aver creato un'oasi di pace in cui potermi rifugiare di tanto in tanto.

Como, December 2017,
Valentina Castiglioni

Table of Contents

''Ford!'' he said, ''there's an infinite number of monkeys outside who want to talk to us about this script for Hamlet they've worked out.''

Douglas Adams,
The Hitchhiker's Guide to the Galaxy

The story so far:
In the beginning the Universe was created.
This has made a lot of people very angry and
been widely regarded as a bad move.

Douglas Adams,
The Restaurant at the End of the Universe

Chapter 1

Introduction

Nowadays we live surrounded by computing systems: education, healthcare, business, security, telecommunications, military systems, social interactions, nearly every aspect of our lives is (or will be soon) closely related to several devices and concurrent systems. Their purpose is to communicate with the external environment and with other systems and devices to quickly and properly react to our requests. The growth in the number of connected systems and of their tasks is inevitably proportional to a dramatic increase in the size and complexity of these systems. Consequently, to study and analyze their behavior and verify their correctness we need to *abstract* away from unimportant details related to their computation. This means that we need formal notation and models allowing us to understand the behavior of such systems and thus to develop algorithms for their analysis, simulation, comparison and verification. For this reason we study *formal methods*, namely mathematical theories and techniques aimed at equipping models and specification languages with a formal syntax and semantics, thus allowing for the specification and verification of the considered systems also with respect to several application fields.

Classic formal methods are closely related to the functional behavior of systems. Roughly speaking, we observe system behavior in order to check whether the proposed model fulfills its computational tasks. However, also due to the increasing complexity of systems, functional behavior alone is not sufficient for a proper and efficient modeling. All these systems are characterized by a variety of uncertainties which are, moreover, of different natures, as empirical uncertainties due to incomplete knowledge on system design and probabilistic uncertainties due to random physical events. Clearly, for a correct system analysis we need to take all these approximations and uncertainties into account for modeling and verification. As an extremely simple example, consider the sender-receiver system constituted by an Xbox controller and the console itself. More precisely, in this system we have a user, the player, giving inputs and requests to one device, the controller, which has to correctly react to the user's inputs and forward them to another device, the console, which has to output the correct answer to the user's requests. So, by a traditional analysis we will obtain that, for instance, whenever we move the left stick to the right then Master Chief[1] will move to the right as well. However, knowing whether Chief will react to our command after 0.1 seconds or after 1 second would make the difference between him finding cover and live to save Earth or being hit by a plasma blast and let the Covenant conquer the World.

In a more formal and general fashion, we should say that when interacting with any computing system the user would prefer knowing which is the probability that an error may occur in the transmission of an input, or how many times a message has to be retransmitted in order to be delivered correctly, rather than having solely a functional analysis stating that the system may reach a deadlock state. Therefore, researchers started to look for techniques for specifying and verifying also *extra-functional* aspects of systems behavior like the quantitative aspects of it.

As a result, a wealth of probabilistic, stochastic, real-time and hybrid models have been proposed (see among others [1,27,28,105,121,124,145,161]) along with behavioral relations [10,29,51,56,70,109,123] and logics [2,57,59,66,68,129,134,140] describing and comparing their behavior.
Probabilistic systems can be thought of as to discrete time - discrete space Markov Chains. They have been proved useful to model and study the reliability of systems, their fairness and to introduce randomization into distributed algorithms, by means of which we can obtain efficient solutions to problems otherwise unsolvable in a deterministic setting [11,145] and that can be also used to introduce additional security measures against information leakage [46]. Moreover, probabilistic models have been also applied to study security and privacy issues [5,9,45,47,76] as well as to evaluate the performance of systems [6,86,87].
Stochastic models are mainly obtained as the abstraction of Markov processes, characterized by a continuous time - continuous state space. Thus, they have been successfully employed to model physical [71], biological [49,52] and chemical [39,40] phenomena along with performance modeling of computer systems [36,105] and robotics [113].
Real-time systems are a particular type of stochastic systems in which correctness depends not only on the logical result of the computation but also on the time at which the results are produced. These features make them suitable for industrial applications [124].
Hybrid systems combine probabilistic and stochastic models to reduce the state space explosion problem that affects the later ones. Informally, they consist in identifying the part of the system that can be approximated by a continuous model while proposing a discrete abstraction for the remaining part. For this reason they have been successfully applied to study system performance [35,36] as well as to model biological systems [88].

In this thesis we will focus on probabilistic models.

1.1 Research context and contribution

With this thesis we aim to give our contribution to the studies on the semantics of probabilistic reactive communicating concurrent systems, henceforth *probabilistic processes*.

[1] All rights reserved (and many thanks) to Bungie and Microsoft Studios for Chief and the Halo Universe.

To fulfill our purpose we will need to investigate the following four building stones of concurrency theory and formal methods: 1. Semantic models, expressing the behavior of processes; 2. Behavioral relations and metrics, allowing for comparing the behavior of distinct processes; 3. Logical characterizations, allowing for checking whether the process behaves as requested; 4. Structural Operational Semantics, allowing for the definition of the semantics of processes and to guarantee compositional reasoning.

A model for probabilistic processes

In the last three decades, researchers proposed several different semantic models for probabilistic processes. For instance, just to name a few, we recall the *reactive*, *generative* and *stratified* models (see [161] for a survey on the three of them) constructed as probabilistic extensions of the classic labeled transition systems (LTS) [111], along with discrete time Markov chains (MCs) [100, 150], Markov decision processes (MDPs) [107] and so on. In this thesis we will consider the general semantic model of *nondeterministic probabilistic labeled transition systems* (PTSs) [145], which extend classic LTSs and discrete time MCs to allow us to model the behavior of those processes in which nondeterminism and probability coexist. In detail, a transition step takes a process to a probability distribution over processes and for each transition label, modeling a specific kind of action step, a process may nondeterministically choose among several transitions with that label.

The phenomenon of *nondeterminism* was introduced in systems modeling to express the fact that the process reached by another one by performing a nonempty sequence of transition steps is not necessarily determined. On one hand, this means that the behavior of a process will depend also on its interactions with the external environment, either it be another process or a user requesting a particular task to it, and in this case we talk about *external nondeterminism*. This phenomenon is usually interpreted as a choice among executable transitions steps with distinct labels for a process. On the other hand, it means that there may be some internal computations of the process that are neither observable nor controllable by the external environment. This is called *internal nondeterminism* and it is usually related to the presence of distinct equally labeled transition steps executable by the same process. Processes in the PTS model are characterized by both internal and external nondeterminism, whereas the majority of the other semantic models proposed in the literature present only external nondeterminism, as for instance happens in the cases of the reactive model and discrete time MCs.

Probability has been introduced in formal methods basically in terms of a probabilistic choice on possible process behaviors. In the case of PTSs this is modeled by considering (labeled) transition steps taking a process to a probability distribution over processes. There has been a long-standing discussion on whether probability could completely substitute nondeterminism. For instance, in *generative* models [161] probability *controls* nondeterminism in the sense that process-to-process transition steps are considered and each process has a certain probability to execute one of the possible transition steps for it. More in general, the idea would be to treat nondeterminism as a probability distribution in which the weights can change at every repetition of the experiment [16]. Our opinion, which is in agreement to the current interpretation, is that nondeterminism and probability are two distinct concepts obeying their own laws and thus we need to model both of them and in separate ways. For instance, nondeterminism is fundamental in formal methods in which the main issue is the functionality of systems, like establishing deadlock-freedom or modeling the independent behavior of processes in interleaving parallel composition, and although the presence of probability can certainly lead to a more accurate study of these features it would not justify the elimination of nondeterminism. A substitution of nondeterminism with probability would in fact subsume additional knowledge on the system behavior and also impose some constraints on it. Moreover the abstraction necessary to define process semantics introduces additional uncertainty about the behavior of the system and thus it is necessary to have an abstract semantics combining together probabilistic and nondeterministic steps.

For all these reasons we decided to consider the semantic model of PTSs.

Behavioral relations and metrics

As we have previously outlined, two main objectives of concurrency theory are the *specification* and *verification* of processes: to specify a process means to define its desired behavior, whereas verifying it means to prove that the actual behavior of the process is equal to the one specified for it. For this reason we need to establish a criterion to determine whether the behavior of two or more processes is the same, or more generally to compare the behavior of distinct processes.

Behavioral relations have been proposed with this exact purpose: comparing the semantics of processes. They consist of *behavioral equivalences* and *behavioral preorders*, where a preorder is a relation that is reflexive and transitive, while an equivalence is a preorder which is also symmetric. Equivalences are usually used to establish whether two processes are indistinguishable for behavior, whereas preorders are mostly used to establish process refinements with respect to behavior. However, it is clear that to compare the semantics of processes we need first to observe them and our observational power may depend on our particular interests as well as on the environment in which they are operating. Hence, in the literature we can find several notions of behavioral relations based on the *observations* that an external observer can make on the process, as (bi)simulations [131], (decorated) traces [106] and testing [62].

The same wealth of notions can be found, even enriched, in the probabilistic setting where the choice on how probabilities have to be taken into account and compared play a fundamental rôle in the definition of equivalences and preorders. We refer the interested reader to [101] for a survey on the different notions of *probabilistic bisimulation* proposed in the literature and to [29, 30] for a spectrum of probabilistic relations containing *probabilistic (bi)simulations*, *probabilistic (decorated) traces equivalences* and several notions of *probabilistic testing equivalences*.

A common feature to behavioral relations in the non-probabilistic and probabilistic settings is that they relate processes that behave *exactly the same*. However, the values of the probability weights assigned in the PTS usually derive from statistical samplings or measures on physical systems and thus they are inevitably subject to errors and approximations. Consequently, one can be interested to know whether the behavior of two processes is similar *up-to* some tolerance or, more simply, *how far* the behavior of two processes is apart. This led to the introduction of the so called *behavioral metrics* [14, 53, 59, 61, 64, 72, 73, 96, 114, 148], which are 1-bounded distances measuring the disparities in the quantitative properties of processes with respect to a chosen semantics and have been showed to provide a robust semantics for probabilistic processes [72, 91, 94, 157]. In particular, *bisimulation metrics* are the quantitative analogue to probabilistic bisimulation, namely they are 1-bounded pseudometrics quantifying the disparities with respect to bisimulation of processes and whose kernel is probabilistic bisimulation.

Our contribution

The majority of the contributions on metric semantics that we can find in the literature are focused on the quantitative versions of the probabilistic bisimulation from [123] and of the probabilistic trace equivalence of [144]. We will propose original notions of metrics measuring the disparities in the behavior of processes with respect to (*decorated*) *trace* and *testing semantics* as expressed by [29]: differently from the original approach of [144] in which traces were *distributions over traces*, [29] proposed a *trace-by-trace* approach for the definition of the linear semantics of processes in the PTS model which, differently from the approach in [144], turned out to be compositional and fully backward compatible with the fully-nondeterministic case. By following their approach, we will obtain behavioral metrics capturing these linear semantics and whose kernels will enjoy the aforementioned desirable properties. In particular, we remark that our metrics for decorated trace and testing metric are the first quantitative version of these semantics ever proposed.

To capture the differences in the expressive power of the novel metrics and the ones for probabilistic (bi)simulations we will order them by the relation '*makes processes further than*'. Thus we will obtain the first *spectrum of behavioral metrics* on processes in the PTS model. Interestingly, from this spectrum we will derive an analogous one for the kernels of the considered metrics, ordered by the relation '*make strictly less identification than*'. Our *spectrum of probabilistic relations* is the probabilistic generalization of the linear time - branching time spectrum of [159].

The contribution provided by this thesis with regard to behavioral metrics can be then summarized as follows:

- ⋆ We provide original notions of behavioral metrics capturing the (decorated) traces semantics and the testing semantics.
- ⋆ We compare the expressive power of the proposed metrics and the ones for probabilistic (bi)similarities in the first spectrum of behavioral distances on processes in the PTS model.
- ⋆ We prove that the kernels of the proposed metrics satisfy some important properties, like compositionality and the full backward compatibility with the fully-nondeterministic and fully-probabilistic cases.
- ⋆ We order the obtained kernels in a spectrum on probabilistic relations including (bi)similarities, (decorated) traces and testing semantics.

Logical characterizations

Behavioral relations and modal logics have been successfully employed for the specification and verification of processes. The former ones provide a simple and elegant tool to compare the observable behavior of processes. The latter ones allow us to express the properties of processes and thus their specification.

Since the work in [102] on the Hennessy-Milner logic (HML), these two approaches are connected by means of *logical characterizations* of behavioral equivalences consisting in proving both the fact that the logic is as expressive as the equivalence and the fact that the equivalence preserves the logical properties of processes. More precisely, a logic is said to be *adequate* for an equivalence if two processes are equivalent if and only if they satisfy the same formulae in the logic, and a logic is said to be *expressive* for an equivalence if for each process s we can identify a *characteristic formula* ϕ_s [97] such that the processes equivalent to s are exactly those satisfying ϕ_s. In the literature we can find several examples of logics that are adequate for probabilistic relations and that consider different semantic models: the model of reactive probabilistic transition systems in the seminal work [123], probabilistic automata in [104, 137], PTSs in [32, 66, 68], labeled Markov processes (LMP) in [56, 70] and continuous-time continuous-space LMP in [129]. Expressive characterizations in the context of PTSs can be found in [68], while [143] considers probabilistic automata. Notice that the majority of the classes of formulae used for these characterizations are obtained as extensions of HML or of the μ-calculus [112] with probabilistic choice modalities or quantitative versions of the diamond modality, which allow for expressing the quantitative properties of processes.

When behavioral metrics instead of relations are considered, logical characterizations proposed in the literature are usually obtained by considering a suitable class of real-valued formulae and then expressing the considered behavioral metric d in terms of the *total variation distance* on the value of formulae at processes, that is

$$d(s,t) = \sup_{\phi \in L} | [\phi](s) - [\phi](t) |$$

where L is the considered logic and $[\phi](s)$ denotes the value of the formula ϕ at processes s accordingly to the real-valued semantics of L. Examples of this kind of logical characterizations can be found in [75, 157] on PTSs, in [59] on Metric Transition Systems (MTS) and in [61] on deterministic game structures. Along with this general approach to the logical characterization of metrics, [14] proposed an alternative technique: they consider the boolean-valued logic LTL and only at a later time they assign a real-valued semantics to it, based on the quantitative properties of processes.

Our contribution

Our contribution to this topic consists in a novel approach to the logical characterization of behavioral metrics and their kernels on the PTS model that will be obtained by means of minimal boolean logics. For what concerns the behavioral metrics our characterization technique can be outlined as follows: 1. Once we have chosen a class L of modal boolean-valued formulae suitable for the considered semantics, for a process s we identify a special formula called *mimicking formula* of s. This is a formula in L that captures the (nondeterministic and probabilistic) behavior of the process that is relevant for the considered semantics. **2.** Then, we transform the modal logic L into a metric space by introducing a notion of *syntactical distance* on formulae. This is a 1-bounded distance assigning to each pair of formulae a suitable quantitative analogue of their syntactic disparities. **3.** We conclude by defining a *logical distance* on processes corresponding to the distance between their mimicking formulae and proving that this logical distance characterizes the considered metric semantics. We will apply this technique to all the behavioral distances proposed in our dissertation and thus it will allow us to obtain the first example of a spectrum of behavioral distances on processes obtained directly from modal logics. Moreover, we will show that the kernels of the considered metrics can be characterized by simply comparing the mimicking formulae of processes with respect to a proper notion of structural equivalence of formulae.

Summarizing, our approach to the logical characterization of behavioral metrics and relations comes with four important features:

- ⋆ We use the same (proper) boolean-valued logic to characterize both the chosen behavioral metric and its kernel. Thus if we have a model checking tool built on that particular boolean-logic, then we should be able to exploit it to obtain a model checking tool for behavioral metrics. Notice that, up-to our knowledge, no model checking tool has been developed so far for real-valued formulae.

- ⋆ The distance between two processes can be obtained by simply looking at their mimicking formulae, without analyzing any other formula in the logic. This should favor the development of new algorithms for model checking as well as for a direct computation of the behavioral metric.

- ⋆ To establish whether two processes are equivalent, or related by a preorder, we simply need to investigate the relation between their mimicking formulae, without analyzing any other formula in the logic.

- ⋆ Our approach can be easily generalized to the relations in the probabilistic weak linear time branching time spectrum and to the metrics expressing them.

Structural Operational Semantics

Structural Operational Semantics (SOS) [138] is nowadays considered the standard framework to define the operational semantics of processes. Briefly, processes are represented as terms over a proper algebra, giving the abstract syntax of the considered language, and their transition steps are derived from a *transition system specification* (TSS) [138], namely a set of inference rules of the form

$$\frac{\text{premises}}{\text{conclusion}}$$

whose intuitive meaning is that whenever the premises are satisfied, then the transition step constituting the conclusion can be deduced. More precisely, the set of transitions that can be deduced, or *proved*, from the TSS constitutes the LTS *generated* by the TSS [160].

Equipping processes with a semantics, however, is not the only application of the SOS framework. One of the main concerns in the development of a meta-theory of process languages is to guarantee their compositionality, that is to prove the compatibility of the language operators with the behavioral relation chosen for the application context. In algebraic terms, this compatibility is known as the *congruence* (resp. *precongruence*) *property* of the considered behavioral equivalence (resp. preorder) $\mathcal{R}$, which consists in verifying whether

$$\text{whenever } t_i \,\mathcal{R}\, t_i' \text{ for all } i = 1, \dots, \mathfrak{n} \text{ then } f(t_1, \dots, t_\mathfrak{n}) \,\mathcal{R}\, f(t_1', \dots, t_\mathfrak{n}') \text{ for any operator } f.$$

Thus, the importance of the congruence property is in that it guarantees that the substitution of a subcomponent of the system with an equivalent one does not affect the behavior of the system. The SOS framework plays a crucial rôle in supporting the compositional reasoning and verification: a *rule* (or *specification*) *format* (see [3] for a survey), is a set of syntactical constraints over inference rules ensuring the desired semantic properties of the transition system derived from them. Thus, one can prove useful results, as the (pre)congruence property, for a whole class of languages at the same time. Many formats have been developed to guarantee that a behavioral equivalence (resp. preorder) is a congruence (resp. precongruence) for all language operators defined by inference rules satisfying the considered format. For instance, the *De Simone* format [63] ensures that trace equivalence is a congruence, the *GSOS* format [34] works for bisimilarity and in [33] the *ntyft/ntxt* format [98] is reduced to the *ready trace* format and its variants to guarantee that decorated trace preorders are precongruences. In the probabilistic setting, considering only congruence formats proposed on TSSs generating PTSs, we can find a few generalizations of the most common formats as the *PGSOS* format [26] and the *ntμfθ/ntμxθ* format [55], for both of which probabilistic bisimilarity has been proven to be a congruence, and in [126] a probabilistic version of the RBB safe format from [81] for (rooted) branching bisimilarity is proposed.

Disregarding for a while whether probability is considered or not, the main question that needs to be answered is "*How can we derive compositional results for a relation from a rule format?*" One possible answer, the one that will be pursued in this thesis, is to exploit the logical characterization of the considered relation. In [119, 125] a *compositional proof system* for HML was provided. The authors observed that since we want systems implementations to be correct with respect to their specifications, it would be much easier to reason in a *implementation by contexts* fashion: instead of extracting an implementation for the complete system from the specification, it would be preferable to implement first the behavior of subcomponents, which is extremely simple if compared to the one of the whole system and so its verification. Thus, to obtain the correctness of the whole system we need to establish what properties each subcomponent should satisfy in order to guarantee that the system in which they are combined (or more generally their *context*) will satisfy some given property (by the specification). Since the specification of a system can be expressed in terms of modal formulae, the above statement can be reduced to establish whether given a formula ϕ and a context C there are formulae $\phi_1 \dots \phi_n$ such that

$$\text{whenever } x_i \models \phi_i \text{ for each } i = 1, \dots, n \text{ then } C[x_1, \dots, x_n] \models \phi. \tag{1.1}$$

The analogy with the congruence property should be clear and it is in fact the reason why the technique used to obtain this result has been referred to as a compositional proof system. Roughly speaking, to obtain it in [119, 125] the authors exploited an SOS machinery used to specify contexts [118]: by means of *action transducers* they reformulate a given TSS into a TSS in De Simone format from which the formulae ϕ_i required in (1.1) are derived by means of *property transformers*.

Inspired by these works, [33] introduced *modal decomposition methods*. The underlying idea is the same: reducing the satisfaction problem of a formula for a process to verifying whether its subprocesses satisfy certain formulae obtained from its decomposition. This is obtained by the notions of *ruloids* [34] (an enhanced version of the action transducers of [119, 125]), namely derived inference rules deducing the behavior of process terms directly from the behavior of the variables occurring in them, and of *decomposition mappings* (the property transformers of [119, 125]) associating to each pair term, formula (t, ϕ) the set of formulae that the variables in t have to satisfy to guarantee that t satisfies ϕ. But the contribution of [33] and the subsequent works [80, 82–85] is not only related to the definition of the decomposition methods but also to their application. In fact they show that by combining the logical characterization of a relation, the decomposition of such a logic and a rule format for the relation it is possible to systematically derive a (pre)congruence format for that relation directly from its modal characterization. Briefly, it is enough to guarantee that the construction of the class of ruloids from the considered TSS preserves the syntactical constraints of the format. At the same time the modal decomposition has to preserve the logical characterization, that is formulae in the characterizing class L have to be decomposed into formulae (equivalent to formulae) in L. Then, from the compositional result (1.1) related to the modal decomposition, we can immediately derive the congruence property.

Our contribution

We propose the first extension of the research line of [33, 80, 83, 84] to behavioral relations defined on the PTS model, that is we provide an SOS-driven decomposition method allowing us to derive (pre)congruence formats for probabilistic equivalences and preoders directly from their logical characterizations. As an example we will analyze in detail the case of strong probabilistic (bi)similarities, and we will provide a schema to generalize our results to the weak case. We will consider (subclasses of) a probabilistic extension of HML with a probabilistic choice modality which provides an adequate characterization of probabilistic (bi)similarities [66]. Our decomposition on classic HML operators will be standard, but to deal with the probabilistic choice modality and its decomposition we will introduce an SOS-like machinery, called *distribution specification*, by which we syntactically represent probability distributions as distributions over terms. Thus, we will provide a class of ruloids built over PGSOS rules and a class of Σ-distribution ruloids built on this new distribution specification and we will exploit both of them to define the decomposition of formulae. The congruence results, stating that probabilistic bisimilarity and ready similarity are (pre)congruences for all operators defined in a specification in PGSOS format and that probabilistic similarity is a precongruence for all operators defined in a specification in *positive* PGSOS format, are then derived by combining the modal decomposition with the logical characterization.

Our contribution can be summarized as follows:

- ⋆ We present new logical characterizations of probabilistic ready similarity and similarity obtained by means of two sublogics of the probabilistic extension of HML from [66] characterizing probabilistic bisimilarity.
- ⋆ We define an SOS-like machinery for the specification of the probabilistic behavior of processes, which can support the decomposition of any modal logic for PTSs.
- ⋆ We develop a method for decomposing formulae equipped with a probabilistic choice modality.
- ⋆ We derive (pre)congruence formats for probabilistic bisimilarity, ready similarity and similarity by exploiting our decomposition method and their logical characterizations.
- ⋆ We sketch how the proposed decomposition method can be generalized to derive congruence formats for probabilistic weak semantics.

1.2 Organization of the thesis

We conclude this introductive Chapter by briefly describing the contents of the upcoming Chapters, in which we will develop the contributions presented in the previous Section.

Chapter 2 We dedicate this Chapter to the review of the basic notions and notation that we will use throughout the thesis along with some simple preliminary results.

Chapter 3 In this Chapter we propose an SOS-based method for decomposing modal formulae capturing the probabilistic (bi)simulation semantics of processes with nondeterminism and probability. In detail, we will consider (subclasses of) a probabilistic extension of the Hennessy-Milner logic developed in [66] for the PTS model, and therefore equipped with modalities allowing for the specification of the quantitative properties of processes. In essence, this means that some formulae are evaluated on probability distributions over processes. In order to decompose this kind of formulae, we introduce an SOS-like machinery, called *distribution specification*, in which we syntactically represent open distribution terms as probability distributions over open process terms. By our decomposition, we obtain (pre)congruence formats for probabilistic bisimilarity, ready similarity and similarity.

A preliminary version of this chapter appeared as [42].

Chapter 4 In this Chapter we propose original notions of behavioral (hemi)metrics capturing the ready simulation, simulation, (decorated) trace and testing semantics for processes with nondeterminism and probability. We study the relations among them and also the bisimilarity metric and we compare all these metrics with respect to their distinguishing power thus obtaining the first *spectrum* of behavioral metrics on processes in the PTS model partially ordered by the relation 'makes processes farther than'. Further, we derive an analogous *spectrum* on the probabilistic equivalences and preorders constituting the kernels of these metrics, which can be in turn partially ordered by the relation 'makes strictly less identifications than'.

Chapter 5 In this Chapter we propose a *logical characterization of branching metrics* (bisimulation, ready simulation and simulation metrics) on image finite processes, obtained by the probabilistic variant of the Hennessy-Milner logic considered in Chapter 3 enriched with variables, whose semantics is defined following the equational μ-calculus approach [120]. Our characterization is based on the novel notions of *mimicking formula* and *distance on formulae*. The former are a weak version of characteristic formulae and allow us to characterize also probabilistic (ready) similarity and bisimilarity. The latter are 1-bounded pseudometrics and hemimetrics on formulae measuring their syntactical disparities. The characterization is then obtained by showing that the chosen behavioral distance between two processes corresponds to the distance between their mimicking formulae, called *logical distance*, for the considered semantics.

A preliminary version of this chapter appeared as [41].

Chapter 6 In this Chapter we exploit the method introduced in Chapter 5 to obtain a *logical characterization of linear metrics* (decorated traces and testing metrics) on image finite processes. Clearly, by means of mimicking formulae we are also able to characterize the kernels of these metrics. Moreover, we conclude the Chapter by providing the first example of a *spectrum* on behavioral metrics on processes in the PTS model obtained directly from modal logics. More precisely, this spectrum will be the *logical* analogous of the spectrum proposed in Chapter 4: we consider the *logical distances* on processes introduced in Chapter 5 and in the first part of Chapter 6 and we order them with respect to the relation 'makes processes farther than'.

Chapter 7 We conclude our dissertation with this Chapter, in which we briefly summarize the results we have obtained and we discuss their potential future developments along with new research objectives.

It has long been an axiom of mine that the little things are infinitely the most important.

Sir Arthur Conan Doyle,
The Memoirs of Sherlock Holmes

Chapter 2

Background

In this chapter we introduce the basic notions and notation that we will use throughout the thesis as well as some preliminary results.

Briefly, in Section 2.1 we recall Mathematical notions including fixed points theory, probability spaces and metric spaces with a special focus on the Kantorovich and Hausdorff metrics. Then, in the three following Sections, we introduce the main characters of this dissertation. In Section 2.2 we define the *nondeterministic probabilistic transition systems* (the PTS model) [145] along with some references to the *Structural Operational Semantics* (SOS) [139] theory required to reason about them. In Section 2.3 we discuss behavioral equivalences and we introduce the *bisimilarity metric* [64, 72, 157]. Finally, in Section 2.4 we recall a few basic notions on *logical characterizations* of behavioral equivalences along with the modal logic $\mathcal{L}$, which has been defined in [66] to characterize probabilistic bisimilarity.

2.1 Mathematical background

Complete lattices and fixed points

Let X be any set. We denote by $\mathcal{P}(X)$ the power set of X. Let $\preceq \subseteq X \times X$ be a binary relation such that the pair $(X, \preceq)$ is a partially ordered set. For any $X' \subseteq X$ an *upper bound* is an element $x \in X$ such that $x' \preceq x$ for all $x' \in X'$ and $\tilde{x} \in X$ is the *supremum* of X' if $\tilde{x}$ is the least upper bound of X', that is $\tilde{x}$ is an upper bound of X' and $\tilde{x} \preceq x$ for all upper bounds x of X'. Conversely, a *lower bound* of X' is an element $x \in X$ such that $x \preceq x'$ for all $x' \in X'$ and we call *infimum* the greatest lower bound of X'. Finally, we say that $C \subseteq X$ is a *chain* in X if for all $x, y \in C$ either $x \preceq y$ or $y \preceq x$.

Definition 2.1. A partially ordered set $(X, \preceq)$ is said to be a *complete lattice* if each subset of X admits a supremum and an infimum in X.

Let $(X, \preceq)$ be a complete lattice. We denote the suprema of the subsets of X by the *join* operator $\bigsqcup \colon \mathcal{P}(X) \to X$ and we denote their infima by the *meet* operator $\bigsqcap \colon \mathcal{P}(X) \to X$. Moreover, we denote by $\bot$ the *bottom element* of X, namely $\bot = \bigsqcap X$. Dually, we denote by $\top$ the *top element* of X, namely $\top = \bigsqcup X$.

Definition 2.2. Let $(X, \preceq_X)$ and $(Y, \preceq_Y)$ be two partially ordered sets and consider a function $f \colon X \to Y$. We say that f is *monotone* if it preserves the order over X, namely whenever $x_1 \preceq_X x_2$ then $f(x_1) \preceq_Y f(x_2)$.

A function $f \colon X \to Y$ is called an *endofunction* whenever $X = Y$. For an endofunction f on $(X, \preceq)$ an element $x \in X$ is (i) a *pre-fixed point* of f if $f(x) \preceq x$; (ii) a *post-fixed point* of f if $x \preceq f(x)$; (iii) a *fixed point* of f if $x = f(x)$.

Knaster-Tarski fixed point theorem ([151]). *Let $(X, \preceq)$ be a complete lattice and let f be a monotone endofunction on X. Then the set of fixed points of f is a complete lattice $(F, \preceq)$ and the least and greatest fixed points are such that:*

$$\bigsqcap F = \bigsqcap \{x \in X \mid f(x) \preceq x\} \qquad \text{and} \qquad \bigsqcup F = \bigsqcup \{x \in X \mid x \preceq f(x)\}.$$

We will use inductive definitions of relations or functions, also known as *Kleene chains*. To guarantee their convergence to their fixed points we exploit the *Kleene fixed point theorem* and a notion of *continuity* for functions on complete lattices.

Definition 2.3 (Kleene chain)**.** Let $(X, \preceq)$ be a complete lattice and let f be a monotone endofunction on X. The *ascending Kleene chain* of f is the chain

$$\bot \preceq f(\bot) \preceq f\big(f(\bot)\big) \preceq \ldots \preceq f^n(\bot) \preceq \ldots$$

obtained by iterating f on the bottom element $\bot$ of X. Analogously, the *descending Kleene chain* of f is the chain obtained by iterating f on the top element of the lattice.

Definition 2.4 (Scott-continuity, [130])**.** Given two complete lattices X and Y, a function $f \colon X \to Y$ is *Scott-continuous* if it preserves the suprema of all nonempty chains in X, that is if for every chain C in X it holds that $\bigsqcup f(C) = f(\bigsqcup C)$.

Kleene fixed point theorem. *Let $(X, \preceq)$ be a complete lattice, and let $f \colon X \to X$ be a Scott-continuous endofunction. Then the least fixed point of f is the supremum of the ascending Kleene chain of f.*

Remark 2.1. Scott-co-continuity is referred to the preservation of all infima, namely a function f is *Scott-co-continuous* if $\bigsqcap f(C) = f(\bigsqcap C)$ for every chain C in X. Therefore, the dual statement of the Kleene fixed point theorem can be easily obtained for Scott-co-continuous functions over Kleene descending chains.

Metric spaces

For a set X, a non-negative function $d \colon X \times X \to \mathbb{R}^+$ is said to be a *metric* on X whenever it satisfies the following axioms:

1. *Identity of the indiscernibles*: $d(x, y) = 0$ if and only if $x = y$, for all $x, y \in X$.

2. *Symmetry*: $d(x,y) = d(y,x)$, for all $x,y \in X$.

3. *Triangular inequality*: $d(x,y) \le d(x,z) + d(z,y)$, for all $x,y,z \in X$.

By relaxing these axioms we can obtain several notions of *generalized* metric. In particular, in this thesis we will consider the notions of *pseudometric* and *hemimetric*. We say that $d\colon X \times X \to \mathbb{R}^+$ is a *pseudometric* on X if the identity of the indiscernibles is substituted by

1'. $d(x,x) = 0$ for all $x \in X$.

If the symmetry axiom is dropped, d is said to be a *quasimetric*. Finally, we say that d is a *hemimetric* (or equivalently a *pseudoquasimetric*) on X if it is a non-symmetric pseudometric.

Given a metric (resp: pseudometric, hemimetric) d on X, the pair (X,d) is called *metric space* (resp: *pseudometric space*, *hemimetric space*). Moreover, the *kernel* of a metric (resp: pseudometric, hemimetric) d on X is defined as the set of pairs of elements in X which are at distance 0, namely $ker(d) = \{(x,y) \in X \times X \mid d(x,y) = 0\}$.

Finally, to stress the fact that a metric (resp: pseudometric, hemimetric) on a set is bounded from above, we introduce the l-boundedness property: for $l \in \mathbb{R}^+$, we say that a metric (resp: pseudometric, hemimetric) d on X is *l-bounded* if and only if $d(x,y) \le l$ for all $x,y \in X$. The we say that the metric space (X,d) is *bounded* if d is l-bounded for some finite positive real l.

Probability spaces

Let X be a countable set. Probability distributions over X are mappings $\pi\colon X \to [0,1]$ with $\sum_{x \in X} \pi(x) = 1$ that assign to each $x \in X$ its probability $\pi(x)$. Each element $x \in X$ is called *event*, the set X is the *space of events* and the pair (X,π) is a *probability space*.

For a probability distribution π over X we denote by $\mathsf{supp}(\pi)$ the support of π, namely $\mathsf{supp}(\pi) = \{x \in X \mid \pi(x) > 0\}$. By $\Delta(X)$ we denote the set of all *finitely supported* probability distributions over X. We let $\pi, \pi', \dots$ range over $\Delta(X)$. We remark that in this thesis we will consider only probability distributions with *finite* support.

For $x \in X$ we denote by δ_x the *Dirac distribution* defined by $\delta_x(x) = 1$ and $\delta_x(y) = 0$ for $x \neq y$. The convex combination $\sum_{i \in I} p_i \pi_i$ of a family $\{\pi_i\}_{i \in I}$ of probability distributions $\pi_i \in \Delta(X)$ with $p_i \in (0,1]$ and $\sum_{i \in I} p_i = 1$ is defined by $(\sum_{i \in I} p_i \pi_i)(x) = \sum_{i \in I}(p_i \pi_i(x))$ for all $x \in X$. Finally, we say that two distributions $\pi_1, \pi_2 \in \Delta(X)$ are equal, notation $\pi_1 = \pi_2$ if for all $x \in X$ we have $\pi_1(x) = \pi_2(x)$.

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

- for each $x \in X$, $\sum_{y \in Y} \mathfrak{w}(x,y) = \pi(x)$, and
- for each $y \in Y$, $\sum_{x \in X} \mathfrak{w}(x,y) = \pi'(y)$.

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

The Kantorovich and Hausdorff metrics

Assume a metric space (X,d). We are interested in defining a distance on the probability distributions over X. Intuitively, such a distance can be obtained by a proper *lifting* of d, which is usually referred to as the *ground distance*. In the literature we can find several proposals for metrics on probability distributions (see [142] for a survey). However, to study the distances on probabilistic systems, researchers have focused on the *Kantorovich* (or Wasserstein) metric [110]. The interest in this particular metric stems from optimal transport theory [163] and its close relation to linear programming. Roughly speaking, accordingly to the Monge-Kantorovich formulation of the optimal transport problem, the *optimal transport cost* for the shipment of goods from producers to consumers, whose respective spatial distributions are modeled by probability distributions, is given by the function

$$C(\pi,\pi') = \inf_{\mathfrak{w}\in\mathfrak{W}(\pi,\pi')} \sum_{x,y\in X} \mathfrak{w}(x,y)\cdot c(x,y)$$

where $c(x,y)$ is the *cost* for transporting one unit of mass from x to y. Indeed, in this setting a matching $\mathfrak{w} \in \mathfrak{W}(\pi,\pi')$ may be understood as a transportation schedule describing the shipment of probability mass from π to π'. Whenever the function $c(x,y)$ is defined in terms of the ground distance d, the optimal transport cost function C becomes a distance over probability distributions on X: the *Kantorovich metric*. For simplicity, we define the *Kantorovich lifting functional* on a generic metric d. However, nothing would change in considering pseudometrics or hemimetrics.

Definition 2.5 (Kantorovich metric, [110])**.** Assume a metric space (X,d). The *Kantorovich lifting* of d is the metric $\mathbf{K}(d)\colon \Delta(X)\times\Delta(X) \to [0,1]$ defined for all $\pi,\pi'\in\Delta(X)$ by

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

For any metric d, we call $\mathbf{K}(d)$ the *Kantorovich metric*.

We remark that accordingly to the original definition, we should have defined the Kantorovich metric as the *infimum* over the matchings for π and π', for any $\pi,\pi'\in\Delta(X)$ and 1-bounded metric d. However, the assumption of having only probability distributions with a finite support guarantees that this infimum is always achieved, since there can be only finitely many matchings between the two distributions, and therefore it is indeed a minimum. As a consequence, the continuity of the lifting functional $\mathbf{K}$ is guaranteed [154].

The Kantorovich metric satisfies some desirable properties that are particularly feasible for its use in computer science. For instance, it can be evaluated in polynomial time [12, 136, 158] and moreover it preserves the properties of the ground distance, as shown in the following Proposition. We remark that the validity of this result would be trivial on *Polish spaces*, namely complete separable metric spaces. However, the (pseudo,hemi)metric spaces that we will consider in this thesis are not guaranteed to be Polish.

Proposition 2.1. *Assume a metric (resp. pseudometric, hemimetric) space (X,d). Then also $(\Delta(X),\mathbf{K}(d))$ is a metric (resp. pseudometric, hemimetric) space. Moreover, if d is l-bounded then also $\mathbf{K}(d)$ is l-bounded.*

Proof. We present only the proof for the general case of d being a metric. The cases of pseudometrics and hemimetrics are subcases of it and they can be easily derived. We need to show that $\mathbf{K}(d)$ satisfies the three axioms of metrics.

1. *Identity of indiscernibles.* Firstly we show that for all $\pi \in \Delta(X)$ we have $\mathbf{K}(d)(\pi,\pi) = 0$. Notice that the function $\bar{\mathfrak{w}}$ defined for all $x, y \in X$ by

$$\bar{\mathfrak{w}}(x,y) = \begin{cases} \pi(x) & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

is a matching for the pair π,π. Then we have

$$\begin{aligned} \mathbf{K}(d)(\pi,\pi) &= \min_{\mathfrak{w} \in \mathfrak{W}(\pi,\pi)} \sum_{x,y \in X} \mathfrak{w}(x,y) d(x,y) \\ &\leq \sum_{x,y \in X} \bar{\mathfrak{w}}(x,y) d(x,y) \\ &= \sum_{x \in X} \pi(x) d(x,x) \\ &= 0 \end{aligned}$$

where the last step follows by $d(x,x) = 0$ for all $x \in X$ as d is a metric by hypothesis.

Assume that $\mathbf{K}(d)(\pi_1,\pi_2) = 0$. We need to show that this implies that $\pi_1 = \pi_2$. Let $\tilde{\mathfrak{w}}$ be an optimal matching for π_1,π_2 with respect to d, namely

$$\tilde{\mathfrak{w}} = \arg \min_{\mathfrak{w} \in \mathfrak{W}(\pi_1,\pi_2)} \sum_{x,y \in X} \mathfrak{w}(x,y) d(x,y).$$

Then we have

$$\begin{aligned} \mathbf{K}(d)(\pi_1,\pi_2) &= \sum_{x,y \in X} \tilde{\mathfrak{w}}(x,y) d(x,y) \\ &= 0 \quad \text{iff} \\ &\qquad \tilde{\mathfrak{w}}(x,y) > 0 \text{ iff } d(x,y) = 0 \\ &\qquad\qquad\qquad\quad \text{iff } x = y \end{aligned}$$

where the last condition follows by the fact that d is a metric by hypothesis. Therefore, as π_1 us the left marginal of $\tilde{\mathfrak{w}}$ and π_2 is its right marginal, we obtain that for all $x \in X$

$$\pi_1(x) = \sum_{y \in X} \tilde{\mathfrak{w}}(x,y) = \tilde{\mathfrak{w}}(x,x) = \sum_{y \in X} \tilde{\mathfrak{w}}(y,x) = \pi_2(x).$$

2. *Symmetry.* First of all we observe that given two arbitrary mappings $\mathfrak{w}, \mathfrak{w}' \colon X \times X \to [0,1]$ with $\mathfrak{w}(x,y) = \mathfrak{w}'(y,x)$ for all $x, y \in X$, it holds that

$$\mathfrak{w} \in \mathfrak{W}(\pi_1,\pi_2) \text{ iff } \mathfrak{w}' \in \mathfrak{W}(\pi_2,\pi_1). \tag{2.1}$$

Then we have

$$\begin{aligned}
\mathbf{K}(d)(\pi_1,\pi_2) &= \min_{\mathfrak{w}\in\mathfrak{W}(\pi_1,\pi_2)} \sum_{x,y\in X} \mathfrak{w}(x,y)d(x,y) \\
&= \min_{\mathfrak{w}\in\mathfrak{W}(\pi_1,\pi_2)} \sum_{x,y\in X} \mathfrak{w}(x,y)d(y,x) && (d \text{ is metric}) \\
&= \min_{\mathfrak{w}'\in\mathfrak{W}(\pi_2,\pi_1)} \sum_{x,y\in X} \mathfrak{w}'(y,x)d(y,x) && (\text{by Equation (2.1)}) \\
&= \mathbf{K}(d)(\pi_2,\pi_1).
\end{aligned}$$

3. *Triangular inequality.* Let us prove that for all distributions $\pi_1,\pi_2,\pi_3 \in \Delta(X)$

$$\mathbf{K}(d)(\pi_1,\pi_2) \le \mathbf{K}(d)(\pi_1,\pi_3) + \mathbf{K}(d)(\pi_3,\pi_2). \tag{2.2}$$

For simplicity, let I, J, H be three finite sets of indexes s.t. $\mathsf{supp}(\pi_1) = \{x_i \mid i \in I\}$, $\mathsf{supp}(\pi_2) = \{x_j \mid j \in J\}$ and $\mathsf{supp}(\pi_3) = \{x_h \mid h \in H\}$. Let $\mathfrak{w}_{1,3} \in \mathfrak{W}(\pi_1,\pi_3)$ be an optimal matching for π_1,π_3, namely

$$\mathbf{K}(d)(\pi_1,\pi_3) = \sum_{i\in I, h\in H} \mathfrak{w}_{1,3}(x_i,x_h)d(x_i,x_h)$$

and let $\mathfrak{w}_{2,3} \in \mathfrak{W}(\pi_2,\pi_3)$ be an optimal matching for π_2,π_3, that is

$$\mathbf{K}(d)(\pi_3,\pi_2) = \sum_{j\in J, h\in H} \mathfrak{w}_{2,3}(x_h,x_j)d(x_h,x_j).$$

Consider now the function $f\colon I \times J \times H \to [0,1]$ defined by

$$f(i,j,h) = \mathfrak{w}_{1,3}(x_i,x_h) \cdot \mathfrak{w}_{2,3}(x_h,x_j) \cdot \frac{1}{\pi_3(x_h)}.$$

Then, we have

$$\begin{aligned}
\sum_{j\in J} f(i,j,h) &= \sum_{j\in J} \mathfrak{w}_{1,3}(x_i,x_h) \cdot \mathfrak{w}_{2,3}(x_h,x_j) \cdot \frac{1}{\pi_3(x_h)} \\
&= \mathfrak{w}_{1,3}(x_i,x_h) \cdot \frac{1}{\pi_3(x_h)} \sum_{j\in J} \mathfrak{w}_{2,3}(x_h,x_j) \\
&= \mathfrak{w}_{1,3}(x_i,x_h) \cdot \frac{1}{\pi_3(x_h)} \cdot \pi_3(x_h) && (\text{by } \mathfrak{w}_{2,3} \in \mathfrak{W}(\pi_2\,\pi_3)) \\
&= \mathfrak{w}_{1,3}(x_i,x_h)
\end{aligned}$$

namely the projection of f over the first and third components coincides with the optimal matching for π_1,π_3. Similarly, we obtain that

$$\sum_{i\in I} f(i,j,h) = \mathfrak{w}_{2,3}(x_h,x_j)$$

namely the projection of f over the second and third components coincides with the optimal matching for π_2, π_3. Moreover, it holds that

$$\begin{aligned}
\sum_{j\in J, h\in H} f(i,j,h) &= \sum_{j\in J, h\in H} \mathfrak{w}_{1,3}(x_i,x_h)\cdot \mathfrak{w}_{2,3}(x_h,x_j)\cdot \frac{1}{\pi_3(x_h)} \\
&= \sum_{h\in H} \mathfrak{w}_{1,3}(x_i,x_h)\cdot \frac{1}{\pi_3(x_h)} \sum_{j\in J} \mathfrak{w}_{2,3}(x_h,x_j) \\
&= \sum_{h\in H} \mathfrak{w}_{1,3}(x_i,x_h)\cdot \frac{1}{\pi_3(x_h)} \cdot \pi_3(x_h) \\
&= \sum_{h\in H} \mathfrak{w}_{1,3}(x_i,x_h) \\
&= \pi_1(x_i) && (\text{by } \mathfrak{w}_{1,3} \in \mathfrak{W}(\pi_1,\pi_3))
\end{aligned}$$

and by similar calculations we obtain

$$\sum_{i\in I, h\in H} f(i,j,h) = \pi_2(x_j)$$

that is $f(i,j,h)$ is a matching in $\mathfrak{W}(\pi_1,\pi_2)$. Therefore,

$$\begin{aligned}
&\mathbf{K}(d)(\pi_1,\pi_2) \\
&= \min_{\mathfrak{w}\in\mathfrak{W}(\pi_1,\pi_2)} \sum_{i\in I, j\in J} \mathfrak{w}(x_i,x_j) d(x_i,x_j) \\
&\le \sum_{i\in I, j\in J, h\in H} f(i,j,h) d(x_i,x_j) \\
&\le \sum_{i\in I, j\in J, h\in H} f(i,j,h)\big(d(x_i,x_h) + d(x_h,x_j)\big) && (d \text{ is metric}) \\
&= \sum_{i\in I, j\in J, h\in H} f(i,j,h) d(x_i,x_h) + \sum_{i\in I, j\in J, h\in H} f(i,j,h) d(x_h,x_j) \\
&= \sum_{i\in I, h\in H} \Big(\sum_{j\in J} f(i,j,h)\Big)\cdot d(x_i,x_h) + \sum_{j\in J, h\in H} \Big(\sum_{i\in I} f(i,j,h)\Big)\cdot d(x_h,x_j) \\
&= \sum_{i\in I, h\in H} \mathfrak{w}_{1,3}(x_i,x_h) d(x_i,x_h) + \sum_{j\in J, h\in H} \mathfrak{w}_{2,3}(x_h,x_j) d(x_h,x_j) \\
&= \mathbf{K}(d)(\pi_1,\pi_3) + \mathbf{K}(d)(\pi_3,\pi_2)
\end{aligned}$$

which completes the proof of Eq. (2.2).

To conclude we need to show that whenever d is l-bounded then also $\mathbf{K}(d)$ is l-bounded. We have

$$\begin{aligned}
\mathbf{K}(d)(\pi_1,\pi_2) &= \min_{\mathfrak{w}\in\mathfrak{W}(\pi_1,\pi_2)} \sum_{x,y\in X} \mathfrak{w}(x,y) d(x,y) \\
&\le \min_{\mathfrak{w}\in\mathfrak{W}(\pi_1,\pi_2)} \sum_{x,y\in X} \mathfrak{w}(x,y)\cdot l && (d \text{ is } l\text{-bounded}) \\
&= l && (\textstyle\sum_{x,y\in X} \mathfrak{w}(x,y) = 1 \text{ for all } \mathfrak{w})
\end{aligned}$$

■

So far, we have seen that given a metric space (X,d) through the Kantorovich functional we can lift the ground distance d to a distance on $\Delta(X)$. However, we will be also interested in evaluating distances between *sets* of probability distributions. For this reason we introduce the *Hausdorff metric*, which allows us to lift the ground distance d to a distance on $\mathcal{P}(X)$.

First of all, we recall the notion of *distance function*, namely the distance between a point x and a set Y, given by the distance between x and the element of Y which is closest to it with respect to the ground distance.

Definition 2.6 (Distance function)**.** Consider a metric space (X,d). Let $Y \subseteq X$. Given any $x \in X$ we denote by $d(x,Y)$ the *distance between the point x and the set Y* defined by

$$d(x,Y) = \inf_{y \in Y} d(x,y).$$

The Hausdorff metric between sets X, Y is then obtained as the maximum between the supremum over X of the distance functions between any point of X and the set Y, and the symmetric version obtained by switching the rôles of X and Y. Intuitively, the Hausdorff metric models the longest distance a player can be forced to travel by an adversary who chooses a point in one of the two sets, from where they must travel to the other set.

Definition 2.7 (Hausdorff metric)**.** Let $d\colon X \times X \to \mathbb{R}^+$ be a metric. The *Hausdorff lifting* of d is the metric $\mathbf{H}(d)\colon \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}^+$ defined for all $X_1, X_2 \subseteq X$ by

$$\mathbf{H}(d)(X_1,X_2) = \max\left\{ \sup_{x_1 \in X_1} \inf_{x_2 \in X_2} d(x_1,x_2),\ \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} d(x_2,x_1) \right\}$$

For any metric d, we call $\mathbf{H}(d)$ the *Hausdorff metric*.

We remark that by convention we assume $\inf_{\emptyset} = \sup_{x,y \in X} d(x,y)$ and $\sup_{\emptyset} = 0$.

The Hausdorff lifting preserves the properties of the ground distance d. Notice that this proposition would be trivial on compact metric spaces. However, the (pseudo)metric spaces that we will consider in this thesis are not guaranteed to be compact.

Proposition 2.2. *Assume a bounded metric space (X,d). Then, given $\mathcal{C}(X) \subseteq \mathcal{P}(X)$ set of subsets of X which are closed with respect to the topology induced by d, we have that $(\mathcal{C}(X), \mathbf{H}(d))$ is a metric space. Moreover if (X,d) is a pseudometric space, then $(\mathcal{P}(X), \mathbf{H}(d))$ is a pseudometric space. Finally, if d is l-bounded then also $\mathbf{H}(d)$ is l-bounded.*

Proof. We will show only the general case for d being a metric. Although the cases of pseudometrics and hemimetrics can be easily derived from it, a few remarks are needed. Firstly, when considering pseudometrics (resp. hemimetrics) we can drop the requirement on sets to be closed, as this property in necessary only to guarantee the satisfaction of the identity of the indiscernibles axiom. Moreover, even if d is a hemimetric, $\mathbf{H}(d)$ will be a pseudometric as the symmetry of $\mathbf{H}(d)$ is imposed by the definition of the Hausdorff functional and does not depend on the symmetry properties of d.

1. *Identity of the indiscernibles.* We recall that closed subsets of a metric space are closed with respect to converging sequences, which means that they contain all their limit

points. Thus, as suprema e infima are limit points, we have

$$\begin{aligned}
&\mathbf{H}(d)(X_1, X_2)\\
&= \max\left\{ \sup_{x_1 \in X_1} \inf_{x_2 \in X_2} d(x_1, x_2),\ \sup_{x_2 \in X_2} \inf_{x_1 \in X_1} d(x_2, x_1) \right\}\\
&= \max\left\{ \max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2),\ \max_{x_2 \in X_2} \min_{x_1 \in X_1} d(x_2, x_1) \right\}\\
&= 0 \text{ iff } \max_{x_1 \in X_1} \min_{x_2 \in X_2} d(x_1, x_2) = 0 \text{ and } \max_{x_2 \in X_2} \min_{x_1 \in X_1} d(x_2, x_1) = 0\\
&\quad \text{iff } \forall\, x_1 \in X_1\, \exists x_2 \in X_2 \text{ st. } d(x_1, x_2) = 0 \text{ and } \forall\, x_2 \in X_2\, \exists x_1 \in X_1 \text{ st. } d(x_2, x_1) = 0\\
&\quad \text{iff } \forall\, x_1 \in X_1\, \exists x_2 \in X_2 \text{ st. } x_1 = x_2 \text{ and } \forall\, x_2 \in X_2\, \exists x_1 \in X_1 \text{ st. } x_2 = x_1\\
&\quad \text{iff } X_1 = X_2
\end{aligned}$$

where the second last step follows by d being a metric.

2. *Symmetry.* As previously outlined, the symmetry of $\mathbf{H}(d)$ is immediate from the definition of the Hausdorff functional (Definition 2.7) and does not actually depend on the properties of d.

3. *Triangular inequality.* We present the proof for the general case in which the subsets are not necessarily closed, namely we prove that for all sets $X_1, X_2, X_3 \in \mathcal{P}(X)$

$$\mathbf{H}(d)(X_1, X_2) \leq \mathbf{H}(d)(X_1, X_3) + \mathbf{H}(d)(X_3, X_2). \tag{2.3}$$

For simplicity, we let $X_1 = \{x_i \mid i \in I\}$, $X_2 = \{x_j \mid j \in J\}$ and $X_3 = \{x_h \mid h \in H\}$. Firstly, we notice that clearly

$$\sup_{i \in I} \inf_{h \in H} d(x_i, x_h) \leq \mathbf{H}(d)(X_1, X_3) \tag{2.4}$$

$$\sup_{h \in H} \inf_{j \in J} d(x_h, x_j) \leq \mathbf{H}(d)(X_3, X_2). \tag{2.5}$$

As a first step, we aim to show that

$$\sup_{i \in I} \inf_{j \in J} d(x_i, x_j) \leq \mathbf{H}(d)(X_1, X_3) + \mathbf{H}(d)(X_3, X_2). \tag{2.6}$$

By definition of infimum, for each $\varepsilon_1 > 0$ we have that

$$\text{for each } i \in I \text{ there is an } h_i \in H \text{ s.t. } d(x_i, x_{h_i}) < \inf_{h \in H} d(x_i, x_h) + \varepsilon_1 \tag{2.7}$$

and, analogously, for each $\varepsilon_2 > 0$ we have that

$$\text{for each } h \in H \text{ there is an } j_h \in J \text{ s.t. } d(x_h, x_{j_h}) < \inf_{j \in J} d(x_h, x_j) + \varepsilon_2. \tag{2.8}$$

In particular given $i \in I$ let $h_i \in H$ be the index realizing Equation (2.7), with respect to ε_1, and let $j_{h_i} \in J$ be the index realizing Equation (2.8) with respect to h_i and ε_2. Then we have

$$d(x_i, x_{j_{h_i}})$$

$$\begin{aligned}
&\leq d(x_i, x_{h_i}) + d(x_{h_i}, x_{j_{h_i}}) && (d \text{ is metric})\\
&< \big(\inf_{h\in H} d(x_i, x_h) + \varepsilon_1\big) + \big(\inf_{j\in J} d(x_{h_i}, x_j) + \varepsilon_2\big) && (\text{by Eq. 2.7 and 2.8})\\
&\leq \big(\sup_{i\in I}\inf_{h\in H} d(x_i, x_h) + \varepsilon_1\big) + \big(\sup_{h\in H}\inf_{j\in J} d(x_h, x_j) + \varepsilon_2\big)
\end{aligned}$$

from which we gather

$$\inf_{j\in J} d(x_i, x_j) \leq d(x_i, x_{j_{h_i}}) < \sup_{i\in I}\inf_{h\in H} d(x_i, x_h) + \sup_{h\in H}\inf_{j\in J} d(x_h, x_j) + \varepsilon_1 + \varepsilon_2.$$

Thus, since i was arbitrary, we obtain

$$\sup_{i\in I}\inf_{j\in J} d(x_i, x_j) \leq \sup_{i\in I}\inf_{h\in H} d(x_i, x_h) + \sup_{h\in H}\inf_{j\in J} d(x_h, x_j) + \varepsilon_1 + \varepsilon_2$$

and since this relation holds for any ε_1 and ε_2 we can conclude that

$$\sup_{i\in I}\inf_{j\in J} d(x_i, x_j) \leq \sup_{i\in I}\inf_{h\in H} d(x_i, x_h) + \sup_{h\in H}\inf_{j\in J} d(x_h, x_i).$$

Then, by the inequalities in Equation (2.4) and Equation (2.5) we can conclude that

$$\sup_{i\in I}\inf_{j\in J} d(x_i, x_j) \leq \mathbf{H}(d)(X_1, X_3) + \mathbf{H}(d)(X_3, X_2)$$

and thus Equation (2.6) holds. An analogous reasoning, given by switching the roles of i and j in the steps above, allows us to infer

$$\sup_{j\in J}\inf_{i\in I} d(x_i, x_j) \leq \mathbf{H}(d)(X_2, X_3) + \mathbf{H}(d)(X_3, X_1). \tag{2.9}$$

Finally, we have

$$\begin{aligned}
&\mathbf{H}(d)(X_1, X_2)\\
=\ &\max\{\sup_{i\in I}\inf_{j\in J} d(x_i, x_j), \sup_{j\in J}\inf_{i\in I} d(x_j, x_i)\}\\
\leq\ &\mathbf{H}(d)(X_1, X_3) + \mathbf{H}(d)(X_3, X_2)
\end{aligned}$$

where the last relation follows by Equations (2.6) and (2.9).

■

2.2 The SOS framework

Structural Operational Semantics (SOS) was introduced in [138, 139] as "*an operational method of specifying semantics based on syntactic transformations of programs and simple operations on discrete data*". Due to its intuitive appeal and flexibility, SOS is nowadays the standard framework used to equip process algebras and specification languages with an

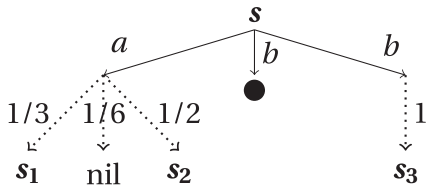


Figure 2.1: *An example of a nondeterministic probabilistic process.*

operational semantics. Roughly speaking, the behavior of a system is modeled as a process graph, mainly *transition systems* [111], whose vertices, called *processes*, are defined as closed terms over a proper algebra and whose edges, called *transitions*, are derived from a set of syntax-driven inference rules, the *transition system specification* [138].

In this Section we will recall only those notions and results necessary to our dissertation. We refer the interested reader to [3] for a complete presentation of the SOS framework.

Nondeterministic probabilistic labeled transition systems

As semantic model we consider that of *nondeterministic probabilistic transition systems* [145] which combine labeled transition systems (LTSs) [111] and discrete time Markov chains (MCs) [100, 150], allowing us to model reactive behavior, nondeterminism and probability.

As state space we take a set $\mathcal{S}$, whose elements are called *processes*, ranged over by $s, t, \dots$

Definition 2.8 (PTS, [145])**.** A *nondeterministic probabilistic labeled 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.*

We call $(s, a, \pi) \in \to$ a *transition*, and we write $s \xrightarrow{a} \pi$ for $(s, a, \pi) \in \to$ whose meaning is that process s can reach the probability distribution π by the execution of action a. We write $s \xrightarrow{a}$ if there is a $\pi \in \Delta(\mathcal{S})$ such that $s \xrightarrow{a} \pi$, and $s \not\xrightarrow{a}$ otherwise. Sometimes, we will refer to $s \xrightarrow{a}$ as an *a-move* of s.

We define the set of *initials* of process s as the set $\mathrm{init}(s) = \{a \in \mathcal{A} \mid s \xrightarrow{a}\}$ of the actions that can be performed by s. For each action $a \in \mathcal{A}$, the set of *a-derivatives* of process s is defined as the set $\mathrm{der}(s, a) = \{\pi \in \Delta(\mathcal{S}) \mid s \xrightarrow{a} \pi\}$ of distributions reachable from s through action a. We say that a process $s \in \mathcal{S}$ is *image-finite* if $\mathrm{der}(s, a)$ is finite for all $a \in \mathrm{init}(s)$ [104]. We say that a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$ is *image-finite* if all processes in $\mathcal{S}$ are.

Remark 2.2. We would like to point out that although also MCs allow to express both probability and nondeterminism, their combination with LTSs actually results into a more expressive model. In fact, in MCs probability and nondeterminism are independent from each other: probabilistic choices are determined by a transition probability function which depends solely on the current process in the graph, whereas nondeterminism is expressed as a set of labels assigned to each state. Conversely, in PTSs we have that each resolution of nondeterminism for a process in the graph leads to a particular probability distribution over processes. This means that all the results we will obtain for PTSs can be trivially adapted to the case of MCs, but the converse is not true in general.

Remark 2.3. To depict the dual nature of PTSs, in the figures we will use black arrows exiting from a process to represent its nondeterministic choices, and will use dotted black arrows to represent the probability distributions that are reached, as shown in Figure 2.1. As an example, we have represented a process s which can execute action a and then reach process s_1 with probability 1/3. Moreover, to simplify the figures, we represent the Dirac distribution on the process nil, namely the process which cannot execute any action, as ●.

Term algebras

In order to syntactically represent the configurations that are reachable by our systems, the notion of signature plays a central rôle.

Definition 2.9 (Signature, [3]). A *signature* is given by a countable set Σ of *operators* together with *rank function* $\mathrm{rk}\colon \Sigma \to \mathbb{N}$ which assigns to each operator the number of its arguments.

We let f range over Σ and $\mathfrak{n}$ range over the rank of f. Operators of rank 0 are called *constants*, while operators of rank 1 and 2 are called, resp., *unary* and *binary* operators.

Given a signature Σ, a Σ-*algebra* is a pair $\langle S, \Sigma_S \rangle$ where S is a set called *carrier* and Σ_S is a set of functions $\{f_S \colon S^{\mathfrak{n}} \to S \mid f \in \Sigma \text{ and } \mathrm{rk}(f) = \mathfrak{n}\}$ called *interpretations*. A special interpretation for a signature Σ is its *term algebra*. This algebra is a purely syntactic object: the carrier is the set of terms built over symbols in Σ and operators only syntactically manipulate them.

Definition 2.10 (Process terms, [3]). Assume a signature Σ and a countable set of (process) *variables* $\mathcal{V}_s$ disjoint from Σ. For a set of variables $V \subseteq \mathcal{V}_s$, the set $\mathsf{T}(\Sigma, V)$ of (process) *terms* over Σ and V is defined as the least set such that:

1. $x \in \mathsf{T}(\Sigma, V)$ for all $x \in V$, and

2. $f(t_1, \dots, t_{\mathfrak{n}}) \in \mathsf{T}(\Sigma, V)$ whenever $f \in \Sigma$ and $t_1, \dots, t_{\mathfrak{n}} \in \mathsf{T}(\Sigma, V)$.

By $\mathcal{T}(\Sigma)$ we denote the set of the *closed terms* $\mathsf{T}(\Sigma, \emptyset)$. By $\mathbb{T}(\Sigma)$ we denote the set of the *open terms* $\mathsf{T}(\Sigma, \mathcal{V}_s)$.

For a constant $c \in \Sigma$, the term $c()$ is abbreviated by c. For $f \in \Sigma$ and $\pi_i \in \Delta(\mathcal{T}(\Sigma))$, $f(\pi_1, \dots, \pi_{\mathfrak{n}})$ is the probability distribution defined by

$$f(\pi_1, \dots, \pi_{\mathfrak{n}})(t) = \begin{cases} \prod_{i=1}^{\mathfrak{n}} \pi_i(t_i) & \text{if } t = f(t_1, \dots, t_{\mathfrak{n}}) \\ 0 & \text{otherwise.} \end{cases}$$

Due to the dual nature of PTSs, we need also syntactic expressions that denote probability distributions.

Definition 2.11 (Distribution terms, [55]). Assume a countable set of *distribution variables* $\mathcal{V}_d$, let $\mathcal{V}$ denote the set of process and distribution variables $\mathcal{V} = \mathcal{V}_s \cup \mathcal{V}_d$, and let $\mu, \nu, \dots$ range over $\mathcal{V}_d$ and ζ range over $\mathcal{V}$. The set of *distribution terms* over Σ, $V_s \subseteq \mathcal{V}_s$ and $V_d \subseteq \mathcal{V}_d$, notation $\mathsf{DT}(\Sigma, V_s, V_d)$, is the least set satisfying:

1. $\{\delta_t \mid t \in \mathsf{T}(\Sigma, V_s)\} \subseteq \mathsf{DT}(\Sigma, V_s, V_d)$,

2. $V_d \subseteq \mathsf{DT}(\Sigma, V_s, V_d)$,

3. $f(\Theta_1, \dots, \Theta_n) \in \mathsf{DT}(\Sigma, V_s, V_d)$ whenever $f \in \Sigma$ and $\Theta_i \in \mathsf{DT}(\Sigma, V_s, V_d)$, and

4. $\sum_{i \in I} p_i \Theta_i \in \mathsf{DT}(\Sigma, V_s, V_d)$ whenever $\Theta_i \in \mathsf{DT}(\Sigma, V_s, V_d)$ and $p_i \in (0,1]$ with $\sum_{i \in I} p_i = 1$.

We write $\mathbb{DT}(\Sigma)$ for $\mathsf{DT}(\Sigma, \mathcal{V}_s, \mathcal{V}_d)$, i.e. the set of all *open distribution terms*, and $\mathcal{DT}(\Sigma)$ for $\mathsf{DT}(\Sigma, \emptyset, \emptyset)$, i.e. the set of all *closed distribution terms*.

Distribution terms have the following meaning. An *instantiable Dirac distribution* δ_t instantiates to $\delta_{t'}$ if t instantiates to t'. A *distribution variable* $\mu \in \mathcal{V}_d$ is a variable that takes values from $\Delta(\mathcal{T}(\Sigma))$. Case (3) lifts the structural inductive construction of terms to distribution terms. Case (4) allows us to construct convex combinations of distributions.

By $\mathrm{var}(t)$ (resp. $\mathrm{var}(\Theta)$) we denote the set of the variables occurring in term t (resp. distribution term Θ).

Probabilistic transitions system specifications

PTSs are defined by means of SOS rules, which are syntax-driven inference rules allowing us to infer the behavior of terms inductively with respect to their structure. Here we consider rules in the probabilistic GSOS format [25, 54], which allow for specifying most of probabilistic process algebras [92, 94].

A *positive (resp. negative) literal* is an expression of the form $t \xrightarrow{a} \Theta$ (resp. $t \stackrel{a}{\not\rightarrow}$) with $t \in \mathbb{T}(\Sigma)$, $a \in \mathcal{A}$ and $\Theta \in \mathbb{DT}(\Sigma)$. The literals $t \xrightarrow{a} \Theta$ and $t \stackrel{a}{\not\rightarrow}$ are said to *deny* each other.

Definition 2.12 (PGSOS rules, [54])**.** A *PGSOS rule r* has the form:

$$\frac{\{x_i \xrightarrow{a_{i,m}} \mu_{i,m} \mid i \in I, m \in M_i\} \qquad \{x_i \stackrel{a_{i,n}}{\not\longrightarrow} \mid i \in I, n \in N_i\}}{f(x_1, \dots, x_{\mathrm{n}}) \xrightarrow{a} \Theta}$$

where $f \in \Sigma$, $I = \{1, \dots, \mathrm{n}\}$, M_i, N_i are finite indexes sets, $a_{i,m}, a_{i,n}, a \in \mathcal{A}$ are actions, $x_i \in \mathcal{V}_s$ and $\mu_{i,m} \in \mathcal{V}_d$ are variables and $\Theta \in \mathbb{DT}(\Sigma)$ is a distribution term. Furthermore, it is required that (i) all $\mu_{i,m}$ for $i \in I$ and $m \in M_i$ are distinct, (ii) all $x_1, \dots, x_{\mathrm{n}}$ are distinct, and (iii) $\mathrm{var}(\Theta) \subseteq \{\mu_{i,m} \mid i \in I, m \in M_i\} \cup \{x_1, \dots, x_{\mathrm{n}}\}$.

A *PGSOS probabilistic transition system specification (PGSOS-PTSS)* is a tuple $P = (\Sigma, \mathcal{A}, R)$, with Σ a signature, $\mathcal{A}$ a countable set of actions and R a finite set of PGSOS rules.

For a PGSOS rule r, the positive (resp. negative) literals above the line are the *positive premises*, notation $\mathrm{pprem}(r)$ (resp. *negative premises*, notation $\mathrm{nprem}(r)$). The literal $f(x_1, \dots, x_{\mathrm{n}}) \xrightarrow{a} \Theta$ is called the *conclusion*, notation $\mathrm{conc}(r)$, the term $f(x_1, \dots, x_{\mathrm{n}})$ is called the *source*, notation $\mathrm{src}(r)$, and the distribution term Θ is called the *target*, notation $\mathrm{trg}(r)$. A PGSOS rule r is said to be *positive* if $\mathrm{nprem}(r) = \emptyset$. Then we say that a PGSOS-PTSS $P = (\Sigma, \mathcal{A}, R)$ is *positive* if all the PGSOS rules in R are positive.

We notice that the constraints (i)–(iii) in Definition 2.12 above, are exactly the constraints of the nondeterministic GSOS format [34] with the difference that we have distribution variables as right hand sides of positive literals.

***Example* 2.1.** The operators of synchronous parallel composition | and probabilistic alternative composition $+_p$, with $p \in (0,1]$, are specified by the following PGSOS rules:

$$\frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu}{x \mid y \xrightarrow{a} \mu \mid \nu} \qquad \frac{x \xrightarrow{a} \mu \quad y \overset{a}{\nrightarrow}}{x +_p y \xrightarrow{a} \mu} \qquad \frac{x \overset{a}{\nrightarrow} \quad y \xrightarrow{a} \nu}{x +_p y \xrightarrow{a} \nu} \qquad \frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu}{x +_p y \xrightarrow{a} p\mu + (1-p)\nu}\,.$$

◀

A PTS is derived from a PTSS through the notions of substitution and proof.

A *substitution* is a mapping $\sigma \colon \mathcal{V} \to \mathbb{T}(\Sigma) \cup \mathbb{DT}(\Sigma)$ such that $\sigma(x) \in \mathbb{T}(\Sigma)$ if $x \in \mathcal{V}_s$ and $\sigma(\mu) \in \mathbb{DT}(\Sigma)$ if $\mu \in \mathcal{V}_d$. It extends to terms, literals and rules by element-wise application. A substitution is *closed* if it maps variables to closed terms. A closed substitution instance of a literal (resp. PGSOS rule) is called a *closed literal* (resp. *closed PGSOS rule)*.

Definition 2.13 (Proof, [160])**.** A *proof* from a PTSS $P = (\Sigma, \mathcal{A}, R)$ of a closed literal α is a well-founded, upwardly branching tree, with nodes labeled by closed literals, such that the root is labeled α and, if β is the label of a node $\mathfrak{q}$ and $\mathcal{K}$ is the set of labels of the nodes directly above $\mathfrak{q}$, then:

- either β is positive and $\mathcal{K}/\beta$ is a closed substitution instance of a rule in R,
- or β is negative and for each closed substitution instance of a rule in R whose conclusion denies β, a literal in $\mathcal{K}$ denies one of its premises.

A literal α is *provable* from P, notation $P \vdash \alpha$, if there exists a proof from P of α.

We have that each PGSOS-PTSS P is *strictly stratifiable* [160] which implies that P *induces a unique model* corresponding to the PTS $(\mathcal{T}(\Sigma), \mathcal{A}, \to)$ whose transition relation $\to$ contains exactly the closed positive literals provable from P. Moreover, the existence of a stratification implies that P is also *complete* [160], thus giving that for any term $t \in \mathcal{T}(\Sigma)$ either $P \vdash t \xrightarrow{a} \pi$ for some $\pi \in \Delta(\mathcal{T}(\Sigma))$ or $P \vdash t \overset{a}{\nrightarrow}$, namely the PTS induced by P contains literals that do not deny each other [34]. In particular, the notion of provability in Definition 2.13 (which is called *supported* in [160]) subsumes the *negation as failure* principle of [50] for the derivation of negative literals: for each closed term t we have that $P \vdash t \overset{a}{\nrightarrow}$ if and only if $P \nvdash t \xrightarrow{a} \pi$ for any distribution $\pi \in \Delta(\mathcal{T}(\Sigma))$.

2.3 How to compare the behavior of processes?

Behavioral equivalences and preorders were introduced as a simple and elegant proof methodology for proving process equivalence and preorder resp., namely for establishing whether the behavior of two processes cannot be distinguished by an external observer.

In the previous Section 2.2, we have recalled that the behavior of processes is represented by a process graph, which, in our setting, is defined by the PTS model. It is then natural to obtain a comparison of the behavior of processes by comparing the respective process graphs. However, the knowledge of an external observer on the structure of a process graph should abstract away from the irrelevant information on the way processes compute and moreover is limited by the *observations* that they can make on it. Clearly, different kinds of observations lead to different kinds of behavioral relations (called accordingly *observational relations* in [102]) as (bi)simulations [10, 29, 103, 123, 126, 145, 146], (decorated) traces [29, 31, 144] and testing [29, 31, 51, 69, 95, 164]

A common feature to several notions of behavioral relation so obtained is that they relate processes that behave *exactly the same*. However, the values of the probability weights assigned in the PTS derive from statistical samplings or measures on physical systems and thus they are inevitably subject to errors and approximations. Consequently, one can be interested to know whether the behavior of two processes is similar *up-to* some tolerance or, more simply, *how far* the behavior of two processes is apart. For this reason, along with relations, the so called *behavioral metrics* [13, 14, 59, 61, 72, 73, 96, 114, 115, 122, 148, 157], have been introduced. They are 1-bounded metrics (as well as pseudometrics, hemimetrics,...) quantifying the behavioral distance between two processes.

In this Section, we recall some basic notions on probabilistic bisimulation [123], ready simulation and simulation [146], and the most studied behavioral (pseudo)metric: the *bisimilarity metric* [58, 64, 72, 91–94, 129, 157].

Probabilistic (bi)simulation

A probabilistic bisimulation is an equivalence over $\mathcal{S}$ that equates processes s, t if they can mimic each other's transitions and evolve to distributions related by the same relation. Hence, we need to lift relations over processes to relations over distributions.

Definition 2.14 (Relation lifting, [68])**.** The *lifting* of a relation $\mathcal{R} \subseteq \mathcal{S} \times \mathcal{S}$ is the relation $\mathcal{R}^\dagger \subseteq \Delta(\mathcal{S}) \times \Delta(\mathcal{S})$ with $\pi \mathcal{R}^\dagger \pi'$ whenever there is a set of indexes I with (i) $\pi = \sum_{i \in I} p_i \delta_{s_i}$, (ii) $\pi' = \sum_{i \in I} p_i \delta_{t_i}$ and (iii) $s_i \; \mathcal{R} \; t_i$ for all $i \in I$.

We recall some equivalent definitions to Definition 2.14 which will be useful in our proofs.

Proposition 2.3 ([66, Proposition 2.3])**.** *Consider a relation $\mathcal{R} \subseteq \mathcal{S} \times \mathcal{S}$. Then $\mathcal{R}^\dagger \subseteq \Delta(\mathcal{S}) \times \Delta(\mathcal{S})$ is the smallest relation satisfying*

1. *$s \mathcal{R} t$ implies $\delta_s \mathcal{R}^\dagger \delta_t$;*
2. *$\pi_i \mathcal{R}^\dagger \pi'_i$ for all $i \in I$, implies $(\sum_{i \in I} p_i \pi_i) \mathcal{R}^\dagger (\sum_{i \in I} p_i \pi'_i)$, for any set of indexes I with $\sum_{i \in I} p_i = 1$.*

Proposition 2.4 ([68, Proposition 1])**.** *Consider two sets X, Y, Let $\pi \in \Delta(X), \pi' \in \Delta(Y)$ and $\mathcal{R} \subseteq X \times Y$. Then $\pi \mathcal{R}^\dagger \pi'$ if and only if there is a matching $\mathfrak{w} \in \mathfrak{W}(\pi, \pi')$ (called* weight function *in [68]) such that*

(i) for each $x \in X$, $\sum_{y \in Y} \mathfrak{w}(x, y) = \pi(x)$;

(ii) for each $y \in Y$, $\sum_{x \in X} \mathfrak{w}(x, y) = \pi'(y)$;

(iii) for each $x \in X, y \in Y$, whenever $\mathfrak{w}(x, y) > 0$ then $x \mathcal{R} y$.

Moreover, we propose another alternative definition equivalent to Definition 2.14, as it will simplify the reasoning in some of the upcoming proofs.

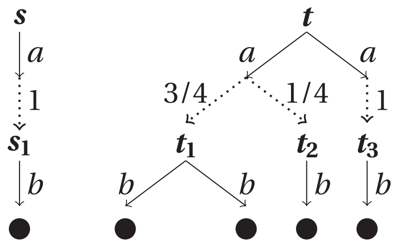


Figure 2.2: *Processes $s, t \in \mathcal{S}$ are probabilistic bisimilar.*

Definition 2.15. Let X be any set. Consider a relation $\mathcal{R} \subseteq X \times X$. Then the lifting of $\mathcal{R}$ is the relation $\mathcal{R}^\dagger \subseteq \Delta(X) \times \Delta(X)$ with $\pi \mathcal{R}^\dagger \pi'$ if whenever $\pi = \sum_{i \in I} p_i \delta_{x_i}$ then $\pi' = \sum_{i \in I, j_i \in J_i} p_{j_i} \delta_{y_{j_i}}$ with $\sum_{j_i \in J_i} p_{j_i} = p_i$ and $x_i \mathcal{R} y_{j_i}$ for all $j_i \in J_i$.

Proposition 2.5. *For any set X, the relation $\mathcal{R}^\dagger \subseteq \Delta(X) \times \Delta(X)$ defined in Definition 2.15 is equivalent to the lifting $\mathcal{R}^\dagger$ defined in Definition 2.14.*

Proof. Let us consider the probability distributions $\pi = \sum_{i \in I} p_i \delta_{x_i}$ and $\pi' = \sum_{j \in J} p_j \delta_{y_j}$. Assume first that $\pi \mathcal{R}^\dagger \pi'$ in the sense of Definition 2.15. Then from $\sum_{j_i \in J_i} p_{j_i} = p_i$ we can directly infer that $\pi = \sum_{i \in I, j_i \in J_i} p_{j_i} \delta_{x_i}$. Summarizing, we have (i) $\pi = \sum_{i \in I, j_i \in J_i} p_{j_i} \delta_{x_i}$, (ii) $\pi' = \sum_{i \in I, j_i \in J_i} p_{j_i} \delta_{y_{j_i}}$, (iii) $x_i \mathcal{R} y_{j_i}$ for all $j_i \in J_i, i \in I$, and therefore we can conclude that $\pi \mathcal{R}^\dagger \pi'$ in the sense of Definition 2.14.

Assume now that $\pi \mathcal{R} \pi'$ in the sense of Definition 2.14, namely there exists a set of indexes H such that $\pi = \sum_{h \in H} p_h \delta_{x_h}$, $\pi' = \sum_{h \in H} p_h \delta_{y_h}$ and $x_h \mathcal{R} y_h$ for all $h \in H$. Then Definition 2.15 is satisfied by taking $J_i = \{h \in H \mid x_h = x_i\}$ for all $i \in I$. ■

Definition 2.16 (Probabilistic (bi)simulations, [123, 145])**.** Assume a PTS $(\mathcal{S}, \mathcal{A}, \longrightarrow)$. Then:

1. A binary relation $\mathcal{R} \subseteq \mathcal{S} \times \mathcal{S}$ is a *probabilistic simulation* if whenever $s \mathcal{R} t$, for each $s \xrightarrow{a} \pi_s$ there is a transition $t \xrightarrow{a} \pi_t$ such that $\pi_s \mathcal{R}^\dagger \pi_t$.

2. A probabilistic simulation $\mathcal{R}$ is a *probabilistic ready simulation* if whenever $s \mathcal{R} t$, $s \overset{a}{\not\longrightarrow}$ implies $t \overset{a}{\not\longrightarrow}$.

3. 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), it is denoted by $\sqsubseteq$ (resp.: $\sqsubseteq_{\mathrm{r}}$, $\sim$), it is called *probabilistic similarity* (resp.: *ready similarity, bisimilarity*), and is a preorder (resp.: preorder, equivalence).

***Example* 2.2.** Consider the processes $s, t \in \mathcal{S}$ represented in Figure 2.2. We have that $s \sim t$. It is immediate to verify that processes t_1, t_2, t_3 are all bisimilar to process s_1, since they can execute only b-labeled transitions reaching with probability 1 the process nil, namely the process which cannot execute any action. As a consequence, we can directly conclude that $\delta_{s_1} \sim^\dagger \delta_{t_3}$. Likewise, it is quite easy to see that the Dirac distribution on s_1 can be rewritten as the convex combination $\delta_{s_1} = \frac{3}{4}\delta_{s_1} + \frac{1}{4}\delta_{s_1}$. Hence if we let $\pi = \frac{3}{4}\delta_{t_1} + \frac{1}{4}\delta_{t_2}$ be the probability distribution to which process t evolves by executing the leftmost a-labeled transition, from $s_1 \sim t_1$ and $s_1 \sim t_2$ we can conclude that $\delta_{s_1} \sim^{\dagger} \pi$ and thus $s \sim t$. ◀

These equivalences and preorders are approximated by relations that consider only the behavior of the first k transitions steps.

Definition 2.17 (Up-to-k (bi)simulations, [17, 104])**.** Assume a PTS $(\mathcal{S}, \mathcal{A}, \rightarrow)$. Then:

1. The family of the *up-to-k simulations* $\sqsubseteq_k$, for $k \in \mathbb{N}$, is defined inductively as:
   a. $\sqsubseteq_0 = \mathcal{S} \times \mathcal{S}$;
   b. $s \sqsubseteq_{k+1} t$ iff whenever $s \xrightarrow{a} \pi_s$ there is a transition $t \xrightarrow{a} \pi_t$ such that $\pi_s \sqsubseteq_k^{\dagger} \pi_t$.
2. The family of the *up-to-k ready simulations* $\sqsubseteq_k^r$, for $k \in \mathbb{N}$, is defined inductively as:
   a. $\sqsubseteq_0^{\mathrm{r}} = \mathcal{S} \times \mathcal{S}$;
   b. $s \sqsubseteq_{k+1}^{\mathrm{r}} t$ iff whenever $s \xrightarrow{a} \pi_s$ there is a transition $t \xrightarrow{a} \pi_t$ such that $\pi_s \sqsubseteq_k^{\mathrm{r}\,\dagger} \pi_t$, and whenever $s \not\xrightarrow{a}$ then $t \not\xrightarrow{a}$.
3. The *up-to-k bisimulation* $\sim_k$ is the kernel of $\sqsubseteq_k$.

Moreover, we define $\sqsubseteq_\omega = \bigcap_{k \geq 0} \sqsubseteq_k$, $\sqsubseteq_\omega^{\mathrm{r}} = \bigcap_{k \geq 0} \sqsubseteq_k^{\mathrm{r}}$, and $\sim_\omega = \bigcap_{k \geq 0} \sim_k$.

Proposition 2.6 ([104])**.** *On image-finite PTSs, the relation $\sqsubseteq_\omega$ (resp.: $\sqsubseteq_\omega^{\mathrm{r}}$, $\sim_\omega$), coincides with $\sqsubseteq$ (resp.: $\sqsubseteq_{\mathrm{r}}$, $\sim$).*

Bisimulation metrics

The quantitative analogue of the bisimulation game is defined by means of a functional $\mathbf{B}$ over the lattice $([0,1]^{\mathcal{S} \times \mathcal{S}}, \preceq)$. By means of a *discount factor* $\lambda \in (0,1]$, $\mathbf{B}$ allows us to specify how much the behavioral distance of future transitions is taken into account to determine the distance between two processes [60, 72]. The discount factor $\lambda = 1$ expresses no discount, meaning that the differences in the behavior between s and t are considered irrespective of after how many steps they can be observed.

Definition 2.18 (Bisimulation metric functional, [64])**.** Let $\mathbf{B} \colon [0,1]^{\mathcal{S} \times \mathcal{S}} \to [0,1]^{\mathcal{S} \times \mathcal{S}}$ be the function defined by

$$\mathbf{B}(d)(s,t) = \sup_{a \in \mathcal{A}} \{\mathbf{H}(\lambda \cdot \mathbf{K}(d))(\mathrm{der}(s,a), \mathrm{der}(t,a))\}$$

for $d \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ and $s, t \in \mathcal{S}$, with $(\lambda \cdot \mathbf{K}(d))(\pi, \pi') = \lambda \cdot \mathbf{K}(d)(\pi, \pi')$.

We remark that since the sets $\mathrm{der}(s,a)$ and $\mathrm{der}(t,a)$ are finite for all $a \in \mathcal{A}$, $s, t \in \mathcal{S}$, due to the image-finiteness assumption, the suprema and infima in the definition of the Hausdorff pseudometric are always achieved, thus becoming maxima and minima, respectively. Hence, considering that the lifting functional $\mathbf{K}$ is continuous, the continuity of the lifting functional $\mathbf{H}$ is guaranteed [154].

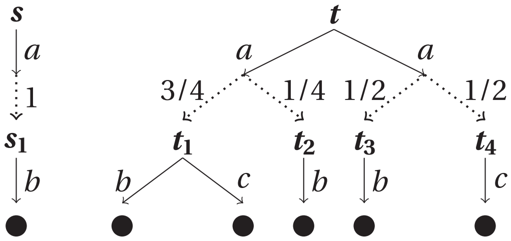


Figure 2.3: *The bisimilarity metric assigns distance* $\mathbf{d}_\lambda(s,t) = \frac{3}{4}\lambda$ *to processes* $s, t \in \mathcal{S}$.

It is not hard to show that $\mathbf{B}$ is monotone. Then, since $([0,1]^{\mathcal{S}\times\mathcal{S}}, \preceq)$ is a complete lattice, by the Knaster-Tarski theorem $\mathbf{B}$ has the least fixed point. Bisimulation metrics are the 1-bounded pseudometrics being prefixed points of $\mathbf{B}$ [64]. The bisimilarity metric is defined as the least fixed point of $\mathbf{B}$, which is also the least prefixed point, and is a 1-bounded pseudometric [64]. Hence, bisimilarity metric is the least bisimulation metric.

Definition 2.19 (Bisimulation metric, [64])**.** A 1-bounded pseudometric $d\colon \mathcal{S}\times\mathcal{S} \to [0,1]$ is a *bisimulation metric* if and only if $\mathbf{B}(d) \preceq d$. The least fixed point of $\mathbf{B}$ is denoted by $\mathbf{d}_\lambda$ and called the *bisimilarity metric.*

***Example* 2.3.** Consider the process $s \in \mathcal{S}$ from previous Example 2.2 and the process $t \in \mathcal{S}$, both represented in Figure 2.3. Assume a 1-bounded pseudometric d with $d(\text{nil},\text{nil}) = d(s_1,t_2) = d(s_1,t_3) = 0$. It is then immediate to see that $\mathbf{B}(d)(s_1,t_2) = \mathbf{B}(d)(s_1,t_3) = 0$ and $\mathbf{B}(d)(s_1,t_1) = \mathbf{B}(d)(s_1,t_4) = 1$. Furthermore, let $\text{der}(t,a) = \{\pi_1,\pi_2\}$ with $\pi_1 = \frac{3}{4}\delta_{t_1} + \frac{1}{4}\delta_{t_2}$ and $\pi_2 = \frac{1}{2}\delta_{t_3} + \frac{1}{2}\delta_{t_4}$. Then we have that $\mathbf{K}(d)(\delta_{s_1},\pi_1) = \frac{3}{4}$ by the matching $\mathfrak{w}_1 \in \mathfrak{W}(\delta_{s_1},\pi_1)$ defined by $\mathfrak{w}(s_1,t_1) = \frac{3}{4}$ and $\mathfrak{w}(s_1,t_2) = \frac{1}{4}$. Analogously, we obtain $\mathbf{K}(d)(\delta_{s_1},\pi_2) = \frac{1}{2}$ by the matching $\mathfrak{w}_2 \in \mathfrak{W}(\delta_{s_1},\pi_2)$ defined by $\mathfrak{w}(s_1,t_3) = \frac{1}{2}$ and $\mathfrak{w}(s_1,t_4) = \frac{1}{2}$. Then the Hausdorff lifting allows us to capture the distance between the nondeterministic choices in the sense that, since s has a unique choice, the nondeterministic evolution of t through the leftmost or the rightmost branch determines the distance between t and s, namely $\mathbf{B}(d)(s,t) = \max\{\frac{3}{4}\lambda, \frac{1}{2}\lambda\} = \frac{3}{4}\lambda$. Hence, the 1-bounded pseudometric d is a bisimulation metric if it satisfies $d(s_1,t_1) = d(s_1,t_4) = 1$ and $d(s,t) \geq \frac{3}{4}\lambda$. Furthermore the bisimilarity metric, as the fixed point of functional $\mathbf{B}$, assigns to processes s,t the distance $\mathbf{d}_\lambda(s,t) = \frac{3}{4}\lambda$. ◀

The kernel of $\mathbf{d}_\lambda$ is the probabilistic bisimulation, namely bisimilar processes are at distance 0.

Proposition 2.7 ([155])**.** *For processes* $s,t \in \mathcal{S}$, $\mathbf{d}_\lambda(s,t) = 0$ *if and only if* $s \sim t$.

The functional $\mathbf{B}$ allows us to define a notion of distance between processes that considers only the first k trasnsition steps.

Definition 2.20 (Up-to-k bisimilarity metric, [72])**.** We define the *up-to-k bisimilarity metric* $\mathbf{d}_\lambda^k$ for $k \in \mathbb{N}$ by $\mathbf{d}_\lambda^k = \mathbf{B}^k(\mathbf{0})$.

Due to the continuity of the lifting functionals $\mathbf{K}$ and $\mathbf{H}$, we can infer that also the functional $\mathbf{B}$ is continuous, besides monotone, thus ensuring that the closure ordinal of $\mathbf{B}$ is ω [154]. Hence, the up-to-k bisimilarity metrics converge to the bisimilarity metric when $k \to \infty$.

Proposition 2.8 ([154]). *Assume an image-finite PTS such that for each transition* $s \xrightarrow{a} \pi$ *we have that the probability distribution* π *has finite support. Then* $\mathbf{d}_\lambda = \lim_{k\to\infty} \mathbf{d}_\lambda^k$.

We also recall the following Lemma from [89], which will be useful in some proofs.

Lemma 2.9 ([89]). *Let* s *and* t *be two processes such that* $\mathrm{init}(s) \neq \mathrm{init}(t)$. *Then, for all* $k > 0$ *it holds that* $\mathbf{d}_\lambda^k(s,t) = 1$.

2.4 Logical characterizations

Modal logics are another tool that we can use to specify and compare the behavior of processes as, in particular, they allow for an immediate expression of the desired properties of processes. The *logical characterization* of a behavioral relation consists then in proving both: the fact that the logic is as expressive as the relation and that the relation preserves the properties of the processes, as expressed by the logic.

As discussed in Chapter 1, one of the contributions of this thesis consists in the definition of a general approach to the logical characterization of behavioral metrics. However, to obtain it we must first recall a few base notions on logical characterization of relations, that will be useful for our dissertation.

Adequate vs expressive characterizations

Since the seminal work [102] on the Hennessy-Milner logic, the strategy used to obtain logical characterizations of behavioral relations has been pretty much always the same: firstly we identify a process with the properties it enjoys and then we prove the relation between processes by checking whether they satisfy the same properties. Accordingly to which method we use to determine the set of properties satisfied by a process, we obtain two different kinds of logical characterization.

Using the terminology of [141], we say that a characterization is *adequate* for a behavioral relation $\mathcal{R}$ if we obtain that two processes s,t are related by $\mathcal{R}$ if and only if they satisfy the same formulae in the considered logic.

Conversely, a characterization is said to be *expressive* for $\mathcal{R}$ if for each process s we can construct a particular formula ϕ_s in the logic, called the *characteristic formula of* s *for* $\mathcal{R}$ [97], which is such that any process t is in relation $\mathcal{R}$ with s if and only if t satisfies ϕ_s. Roughly speaking, the characteristic formula of s for $\mathcal{R}$ can be considered as the representative of the equivalence (or preorder) class of s. Intuitively, the characteristic formula for $\mathcal{R}$ may not express all the properties satisfied by a process, but it subsumes all the properties whose satisfiability is discriminating with respect to $\mathcal{R}$.

Adequate characterizations are particularly useful in their contrapositive version: whenever there exists a formula that is satisfied by process s but not by process t, then we can immediately conclude that s,t cannot be equivalent (or in the considered preorder). However, an adequate characterization would potentially require to check for infinitely many formulae in order to establish the behavioral relation over processes. This is not the case of expressive characterizations, in which it is sufficient to check for the satisifiability of a single formula. Indeed, the construction of characteristic formulae is neither immediate nor always possible.

The modal logic $\mathcal{L}$

As a logic expressing behavioral properties over terms, we consider the modal logic $\mathcal{L}$ of [66], which extends the Hennessy-Milner Logic [102] with a probabilistic choice modality.

Definition 2.21 (Modal logic $\mathcal{L}$, [66])**.** The classes of *state formulae* $\mathcal{L}^{\mathrm{s}}$ and *distribution formulae* $\mathcal{L}^{\mathrm{d}}$ over $\mathcal{A}$ are defined by the following BNF-like grammar:

$$\begin{aligned}\mathcal{L}^{\mathrm{s}}: &\quad \varphi ::= \quad \top \mid \neg\varphi \mid \bigwedge_{j\in J}\varphi_j \mid \langle a\rangle\psi \\ \mathcal{L}^{\mathrm{d}}: &\quad \psi ::= \quad \bigoplus_{i\in I} r_i\varphi_i\end{aligned}$$

where: (i) $\varphi, \varphi_i, \varphi_j$ range over $\mathcal{L}^{\mathrm{s}}$, (ii) ψ ranges over $\mathcal{L}^{\mathrm{d}}$, (iii) $a \in \mathcal{A}$, (iv) J is an at most countable set of indexes with $J \neq \emptyset$, and (v) I is a finite set of indexes with $I \neq \emptyset$, $r_i \in (0,1]$ for each $i \in I$ and $\sum_{i\in I} r_i = 1$.

We shall write $\varphi_1 \wedge \varphi_2$ for $\bigwedge_{j\in J}\varphi_j$ with $J = \{1,2\}$, $r_1\varphi_1 \oplus r_2\varphi_2$ for $\bigoplus_{i\in I} r_i\varphi_i$ with $I = \{1,2\}$, and $\langle a\rangle\varphi$ for $\langle a\rangle\bigoplus_{i\in I} r_i\varphi_i$ with $I = \{i\}$, $r_i = 1$ and $\varphi_i = \varphi$. Notice that instead of using $\top$ we could use $\bigwedge_{\emptyset}$. We decided to use $\top$ to improve readability.

Formulae are interpreted over a PTS.

Definition 2.22 (Semantics of $\mathcal{L}$, [66])**.** Assume a PTS $(\mathcal{T}(\Sigma), \mathcal{A}, \rightarrow)$. The *satisfaction relation* $\models \subseteq (\mathcal{T}(\Sigma) \times \mathcal{L}^{\mathrm{s}}) \cup (\Delta(\mathcal{T}(\Sigma)) \times \mathcal{L}^{\mathrm{d}})$ is defined by structural induction on formulae by

- $t \models \top$ always;
- $t \models \neg\varphi$ iff $t \models \varphi$ does not hold;
- $t \models \bigwedge_{j\in J}\varphi_j$ iff $t \models \varphi_j$ for all $j \in J$;
- $t \models \langle a\rangle\psi$ iff $t \xrightarrow{a} \pi$ for a distribution $\pi \in \Delta(\mathcal{T}(\Sigma))$ with $\pi \models \psi$;
- $\pi \models \bigoplus_{i\in I} r_i\varphi_i$ iff $\pi = \sum_{i\in I} r_i\pi_i$ for a family $\{\pi_i\}_{i\in I}$ of distributions such that for each $i \in I$ whenever $t \in \mathrm{supp}(\pi_i)$ then $t \models \varphi_i$.

Dealing with $\mathcal{L}$ is motivated by its characterization of bisimilarity, proved in [66] (see Theorem 2.10 below), bisimilarity metric, proved in [41], and similarity and ready similarity, proved here (see Theorem 2.11 below).

Theorem 2.10 ([66])**.** *Assume an image finite PTS $(\mathcal{T}(\Sigma), \mathcal{A}, \rightarrow)$ and terms $s, t \in \mathcal{T}(\Sigma)$. Then, $s \sim t$ if and only if they satisfy the same formulae in $\mathcal{L}^{\mathrm{s}}$.*

The characterization of ready similarity and similarity requires two subclasses of $\mathcal{L}$.

Definition 2.23. The class of *ready formulae* $\mathcal{L}_{\mathrm{r}}$ is defined as

$$\begin{aligned}\mathcal{L}_{\mathrm{r}}^{\mathrm{s}}: &\quad \varphi ::= \quad \top \mid \bar{a} \mid \bigwedge_{j\in J} \varphi_j \mid \langle a \rangle \psi\\ \mathcal{L}_{\mathrm{r}}^{\mathrm{d}}: &\quad \psi ::= \quad \bigoplus_{i\in I} r_i \varphi_i\end{aligned}$$

where $\bar{a}$ stays for $\neg\langle a\rangle\top$, and the class of *positive formulae* $\mathcal{L}_{+}$ is defined as

$$\begin{aligned}\mathcal{L}_{+}^{\mathrm{s}}: &\quad \varphi ::= \quad \top \mid \bigwedge_{j\in J} \varphi_j \mid \langle a \rangle \psi\\ \mathcal{L}_{+}^{\mathrm{d}}: &\quad \psi ::= \quad \bigoplus_{i\in I} r_i \varphi_i.\end{aligned}$$

The classes $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ are strict sublogics of the one proposed in [69] for the characterization of failure similarity and forward similarity [145]. In particular, the logic used in [69] allows for arbitrary formulae to occur after the diamond modality.

We can show that our sublogics are powerful enough for the characterization of ready similarity and similarity.

Theorem 2.11. *Assume an image finite PTS $(\mathcal{T}(\Sigma), \mathcal{A}, \rightarrow)$ and terms $s, t \in \mathcal{T}(\Sigma)$. Then:*

1. *$s \sqsubseteq_{\mathrm{r}} t$ iff for any formula $\varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$, $s \models \varphi$ implies $t \models \varphi$.*
2. *$s \sqsubseteq t$ iff for any formula $\varphi \in \mathcal{L}_{+}^{\mathrm{s}}$, $s \models \varphi$ implies $t \models \varphi$.*

Proof. We prove only the first item, namely the characterization of the ready simulation preorder. The proof for simulation is analogous.

($\Rightarrow$) Let $\varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$. We aim to prove that

$$\text{whenever } s \sqsubseteq_{\mathrm{r}} t \text{ and } s \models \varphi, \text{ then } t \models \varphi. \tag{2.10}$$

We proceed by structural induction over φ.

- ★ Base case $\varphi = \top$. Then Equation (2.10) immediately follows.
- ★ Base case $\varphi = \bar{a}$. Then, by Definition 2.22, $s \models \bar{a}$ gives $s \overset{a}{\not\rightarrow}$. Since $s \sqsubseteq_{\mathrm{r}} t$, this implies that $t \overset{a}{\not\rightarrow}$ from which we draw $t \models \bar{a}$. Therefore, Equation (2.10) follows also in this case.
- ★ Inductive step $\varphi = \bigwedge_{j\in J} \varphi_j$. Then, by Definition 2.22, $s \models \bigwedge_{j\in J} \varphi_j$ gives that $s \models \varphi_j$ for each $j \in J$. Hence, by structural induction we obtain that $t \models \varphi_j$ for each $j \in J$, thus implying $t \models \bigwedge_{j\in J} \varphi_j$. Therefore, Equation (2.10) follows also in this case.
- ★ Inductive step $\varphi = \langle a \rangle \bigoplus_{i\in I} r_i \varphi_i$. Then, by Definition 2.22, $s \models \langle a \rangle \bigoplus_{i\in I} r_i \varphi_i$ gives that there exists a distribution π_s s.t. $s \xrightarrow{a} \pi_s$ and $\pi_s \models \bigoplus_{i\in I} r_i \varphi_i$. Since $s \sqsubseteq_{\mathrm{r}} t$, $s \xrightarrow{a} \pi_s$

implies the existence of a distribution π_t s.t. $t \xrightarrow{a} \pi_t$ and $\pi_s \sqsubseteq_{\mathrm{r}}^{\dagger} \pi_t$. Hence, to derive Equation (2.10) we need to prove that

$$\pi_t \models \bigoplus_{i \in I} r_i \varphi_i \,. \tag{2.11}$$

From $\pi_s \models \bigoplus_{i\in I} r_i \varphi_i$ we gather that $\pi_s = \sum_{i\in I} r_i \pi_i$ for some distributions π_i s.t. whenever $s' \in \mathsf{supp}(\pi_i)$ then $s' \models \varphi_i$ (Definition 2.22). Moreover, by Definition 2.14 and Proposition 2.3, $\pi_s \sqsubseteq_{\mathrm{r}}^{\dagger} \pi_t$ and $\pi_s = \sum_{i\in I} r_i \pi_i$ together imply the existence of distributions π'_i s.t. $\pi_t = \sum_{i\in I} r_i \pi'_i$ and for each $s' \in \mathsf{supp}(\pi_i)$ there is a $t' \in \mathsf{supp}(\pi'_i)$ s.t. $s' \sqsubseteq_{\mathrm{r}} t'$. Thus, from $s' \sqsubseteq_{\mathrm{r}} t'$ and $s' \models \varphi_i$, structural induction over φ_i gives $t' \models \varphi_i$. Hence, for each $t' \in \mathsf{supp}(\pi'_i)$ it holds that $t' \models \varphi_i$ thus giving Equation (2.11). Therefore, we can conclude that $t \models \langle a \rangle \bigoplus_{i\in I} r_i \varphi_i$ and Equation (2.10) follows also in this case.

($\Leftarrow$) Assume now that, for any $\varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$, $s \models \varphi$ implies $t \models \varphi$. We define the relation

$$\mathcal{R} = \{(s,t) \mid s \models \varphi \text{ implies } t \models \varphi \text{ for all } \varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}\}.$$

We aim to show that $\mathcal{R}$ is a probabilistic ready simulation.

Let $s \,\mathcal{R}\, t$. We aim to prove that

$$\text{whenever } s \stackrel{b}{\not\longrightarrow} \text{ then } t \stackrel{b}{\not\longrightarrow} \tag{2.12}$$

$$\text{whenever } s \xrightarrow{a} \pi_s \text{ then there is a transition } t \xrightarrow{a} \pi_t \text{ with } \pi_s \mathcal{R}^{\dagger} \pi_t. \tag{2.13}$$

Assume first that $s \stackrel{b}{\not\longrightarrow}$. Then, by Definition 2.22, we derive $s \models \bar{b}$. From $s \,\mathcal{R}\, t$ we gather $t \models \bar{b}$ thus giving $t \stackrel{b}{\not\longrightarrow}$ and Equation (2.12) follows.

Next, consider any transition $s \xrightarrow{a} \pi_s$. To prove Equation (2.13) we need to show that there exists a probability distribution π_t s.t. $t \xrightarrow{a} \pi_t$ and $\pi_s \mathcal{R}^{\dagger} \pi_t$. We recall that by definition of lifting of a relation (Definition 2.14) we have $\pi_s \mathcal{R}^{\dagger} \pi_t$ iff whenever $\pi_s = \sum_{i\in I} p_i \delta_{s_i}$, for some set of indexes I, then $\pi_t = \sum_{i\in I} p_i \delta_{t_i}$ for some processes t_i s.t. $s_i \,\mathcal{R}\, t_i$ for each $i \in I$. Since it is immediate to see that $\pi_s = \sum_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \delta_{s'}$, by Proposition 2.3 to prove Equation (2.13) we need to show that there exists a probability distribution π_t s.t. $\pi_t = \sum_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \pi_{s'}$ for a family of probability distributions $\{\pi_{s'}\}_{s' \in \mathsf{supp}(\pi_s)}$ s.t. whenever $t' \in \mathsf{supp}(\pi_{s'})$ then $s' \,\mathcal{R}\, t'$. Thus, let us consider the set

$$\Pi_{t,a} = \{\pi \mid t \xrightarrow{a} \pi \wedge \pi = \sum_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \pi_{s'} \wedge \exists\, s' \in \mathsf{supp}(\pi_s),\, t' \in \mathsf{supp}(\pi_{s'}) \colon s' \not\mathcal{R}\, t'\}.$$

Our aim is to prove that there is at least one probability distribution $\pi_t \in \mathrm{der}(t,a)$ which does not belong to the set $\Pi_{t,a}$.

By construction, for each $\pi \in \Pi_{t,a}$ there are some processes $s'_\pi \in \mathsf{supp}(\pi_s)$ and $t'_\pi \in \mathsf{supp}(\pi_{s'_\pi})$ and a ready state formula φ_π for which $s'_\pi \models \varphi_\pi$ but $t'_\pi \not\models \varphi_\pi$. Thus, for each $s' \in \mathsf{supp}(\pi_s)$ we have $s' \models \bigwedge_{\{\pi \in \Pi_{t,a} \mid s'_\pi = s'\}} \varphi_\pi$. Moreover, for each $\pi \in \Pi_{t,a}$ with $s'_\pi = s'$ there is some $t'_\pi \in \mathsf{supp}(\pi_{s'})$ s.t. $t'_\pi \not\models \bigwedge_{\{\pi \in \Pi_{t,a} \mid s'_\pi = s'\}} \varphi_\pi$.

Consider now that ready state formula

$$\varphi = \langle a \rangle \bigoplus_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \bigwedge_{\{\pi \in \Pi_{t,a} \mid s'_\pi = s'\}} \varphi_\pi .$$

Then, it is clear that $s \models \varphi$ thus implying $t \models \varphi$, as by hypothesis $s \mathcal{R} t$. From $t \models \varphi$ it follows that there exists a distribution π_t s.t. $t \xrightarrow{a} \pi_t$ and

$$\pi_t \models \bigoplus_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \bigwedge_{\{\pi \in \Pi_{t,a} \mid s'_\pi = s'\}} \varphi_\pi$$

namely $\pi_t = \sum_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \pi'_{s'}$ for some distributions $\pi'_{s'}$ s.t. whenever $t' \in \mathsf{supp}(\pi'_{s'})$ then $t' \models \bigwedge_{\{\pi \in \Pi_{t,a} \mid s'_\pi = s'\}} \varphi_\pi$. Consequently, $\pi_t \notin \Pi_{t,a}$ and hence for all $s' \in \mathsf{supp}(\pi_s)$ each $t' \in \mathsf{supp}(\pi'_{s'})$ is such that $s' \mathcal{R} t'$. Therefore, from Proposition 2.3 we obtain $\delta_{s'} \mathcal{R}^\dagger \pi'_{s'}$ and consequently (from the same Proposition 2.3) $\pi_s \mathcal{R}^\dagger \pi_t$, thus proving Equation (2.13). ■

We demand rigidly defined areas of doubt and uncertainty!

Douglas Adams,
The Hitchhiker's Guide to the Galaxy

CHAPTER 3

SOS-based Modal Decomposition on Nondeterministic Probabilistic Processes

One of the main concerns in the development of a meta-theory of process languages is to guarantee their compositionality, that is to prove the compatibility of the language operators with the behavioral relation, chosen for the application context. In algebraic terms, this compatibility is known as the *congruence* (resp. *precongruence*) *property* of the considered behavioral equivalence (resp. preorder) $\mathcal{R}$, which consists in verifying whether $f(t_1,\dots,t_\mathfrak{n})\,\mathcal{R}\,f(t'_1,\dots,t'_\mathfrak{n})$ whenever $t_i\,\mathcal{R}\,t'_i$ for all $i=1,\dots,\mathfrak{n}$, for any $\mathfrak{n}$-ary operator f.

The SOS framework plays a crucial rôle in supporting the compositional reasoning and verification: a *rule* (or *specification*) *format* [34, 63, 98, 99, 162], is a set of syntactical constraints over SOS-rules ensuring the desired semantic properties of the transition system derived from them. For instance, in [54] it is proved that probabilistic bisimilarity is a congruence for all operators defined by a transition system specification in PGSOS format [26] (cf. Chapter 2.2).

Moreover, as outlined in Chapter 1.1, we can combine the SOS framework with the logical characterization of a behavioral relation to favor the compositional reasoning: *modal decomposition* of formulae exploits the characterization of an equivalence to derive the compositional properties of the system. Roughly speaking, the definition of the semantic behavior of processes by means of the SOS framework allows for decomposing the satisfaction problem of a formula for a process into the verification of the satisfaction problem of certain formulae for its subprocesses (see [33, 80, 82–85, 125]). The decomposition of a formula ϕ with respect to a term t is defined by a set $t^{-1}(\phi)$ of *decomposition mappings* ξ assigning to each variable x in t a proper formula $\xi(x)$. These are obtained by means of the notion of *ruloid* [34], namely inference transition rules that are derived from the SOS specification and define the behavior of open processes in terms of the behavior of their variables. Then, in [33, 80, 82, 84, 85], the decomposition of modal formulae is used to systematically derive expressive (pre)congruence formats for several behavioral equivalences and preorders from

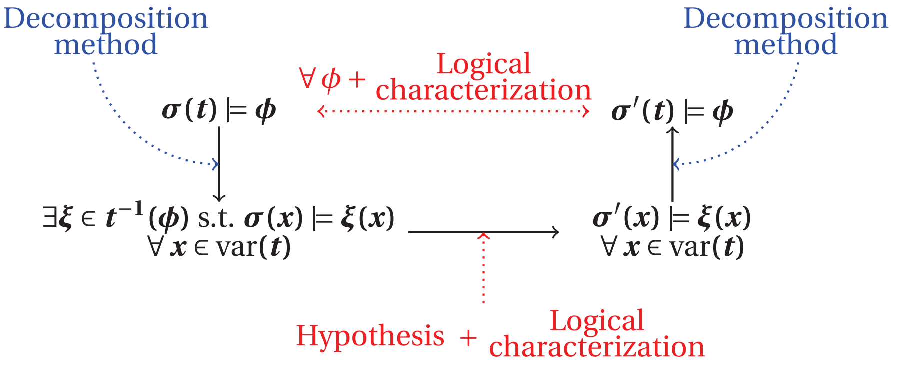

Decomposition method
Decomposition method
$\forall \phi$ + Logical characterization
$\boldsymbol{\sigma}(\boldsymbol{t}) \models \boldsymbol{\phi}$
$\boldsymbol{\sigma}'(\boldsymbol{t}) \models \boldsymbol{\phi}$
$\exists \boldsymbol{\xi} \in \boldsymbol{t}^{-1}(\boldsymbol{\phi})$ s.t. $\boldsymbol{\sigma}(\boldsymbol{x}) \models \boldsymbol{\xi}(\boldsymbol{x})$ $\forall \boldsymbol{x} \in \mathrm{var}(\boldsymbol{t})$
$\boldsymbol{\sigma}'(\boldsymbol{x}) \models \boldsymbol{\xi}(\boldsymbol{x})$ $\forall \boldsymbol{x} \in \mathrm{var}(\boldsymbol{t})$
Hypothesis + Logical characterization

Figure 3.1: *General schema to prove that $\sigma(x)\,\mathcal{R}\,\sigma'(x)$ for all $x \in \mathrm{var}(t)$ implies $\sigma(t)\,\mathcal{R}\,\sigma'(t)$, by combining a logical characterization of $\mathcal{R}$ with the related modal decomposition.*

their (adequate) modal characterizations. Informally, it is proved that the construction of the ruloids preserves the syntactic restrictions imposed by the considered rule format and thus that the decomposition of formulae in a certain class produces formulae in the same class. The (pre)congruence result then becomes an immediate consequence of the logical characterization of the considered behavioral relation, as schematized in Figure 3.1. Further, in [90] the semantic model of reactive probabilistic labeled transition systems [161] is considered and a method for decomposing formulae from a probabilistic version of HML [137] characterizing probabilistic bisimilarity with respect to a probabilistic transition system specification in the format of [116, 117] is proposed.

The purpose of this Chapter is to extend the SOS-driven decomposition approach to processes in the PTS model. All modal logics developed so far for the PTS model are equipped with modalities allowing for the specification of the quantitative properties of processes (cf. Chapter 2.4). In essence, this means that some modal formulae are (possibly indirectly) evaluated on distributions. In order to decompose this kind of formulae, we introduce an SOS-like machinery, called *distribution specification*, in which we syntactically represent open distribution terms as probability distributions over open terms. More precisely, our distribution specification, consisting in a set of *distribution rules* defined on a signature, will allow us to infer the expression $\Theta \xrightarrow{q} t$ whenever a closed distribution term Θ assigns probability weight q to a closed term t. Then, from these distribution rules we derive the *distribution ruloids*, which will play a fundamental rôle in the decomposition method. In fact, as happens for ruloids on terms, our distribution ruloids will allow us to derive expressions of the form $\Theta \xrightarrow{q} t$, for an arbitrary open distribution term Θ and open term t, by considering only the behavior of the variables occurring in Θ. Hence, they will allow us to decompose the formulae capturing the quantitative behavior of processes since through them we can relate the satisfaction problem of a formula of this kind for a closed distribution term to the satisfaction problem of certain derived formulae for its subterms. We stress that our distribution ruloids can support the decomposition of formulae in any modal logic for PTSs and moreover our distribution specification can be easily generalized to cover the case of models using sub-distributions in place of probability distributions (see for instance [115, 126, 127]).

We present the decomposition of formulae from the two-sorted boolean-valued modal logic $\mathcal{L}$ [66] and from its two subclasses of formulae $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ introduced in Chapter 2.4. Finally, to show the robustness of our approach we apply it to derive the congruence theorem for probabilistic bisimilarity with respect to the PGSOS format and the precongruence theorem for probabilistic ready similarity and similarity with respect to the PGSOS format and the positive PGSOS format, respectively.

The contribution of this Chapter can be summarized as follows:

1. We present new logical characterizations of probabilistic ready similarity and similarity obtained by means of two sublogics of $\mathcal{L}$, resp. $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ (Theorem 2.11).

2. We define an SOS machinery for the specification of the probabilistic behavior of processes, which can support the decomposition of any modal logic for PTSs.

3. We develop a method of decomposing formulae in $\mathcal{L}$ and in its sublogics $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ (Theorem 3.12 and Theorem 3.14).

4. We derive (pre)congruence formats for probabilistic bisimilarity, ready similarity and similarity by exploiting our decomposition method on the logics characterizing them (Theorem 3.15).

Organization of contents

In Section 3.1 we introduce the SOS-like machinery for the specification of the behavior of distribution terms and in Section 3.2 we define the two classes of ruloids: the *P-ruloids*, built on a PGSOS specification P, and the *distribution ruloids*, derived from a distribution specification. Section 3.3 is the core of this Chapter and provides our decomposition method which allows for the derivation of the (pre)congruence formats for probabilistic bisimilarity, ready similarity and similarity as shown in Section 3.4. In Section 3.5 we give an hint on how we can apply our decomposition method to the derivation of congruence formats for probabilistic weak semantics. Finally we end with some conclusions and discussion of related and future work in Section 3.6.

3.1 Distribution specifications

In this section we develop an SOS-like machinery consisting of a set of inference rules, called Σ-*distribution rules*, through which we syntactically represent open distribution terms as probability distributions over open terms. Informally, these rules allow us to infer the expression $\Theta \xrightarrow{q} t$ whenever a closed distribution term Θ assigns probability weight q to a closed term t. More precisely, the idea behind Σ-distribution rules is as follows: assuming that the distribution variable μ is characterized as the distribution $\{\mu \xrightarrow{q_i} x_i \mid i \in I\}$ and the distribution variable ν as the distribution $\{\nu \xrightarrow{q_j} x_j \mid j \in J\}$, then the Σ-distribution rule

$$\frac{\{\mu \xrightarrow{q_i} x_i \mid i \in I\} \quad \{\nu \xrightarrow{q_j} x_j \mid j \in J\}}{\{\mu \mid \nu \xrightarrow{q_i \cdot q_j} x_i \mid x_j \bigm| i \in I, j \in J\}}$$

allows us to describe the behavior of the distribution term $\mu \mid \nu$ as a probability distribution over the open terms $x_i \mid x_j$. As we will see in Definition 3.1 below the weights and the pattern of the target terms in the conclusion are chosen accordingly to the syntactic structure of the distribution term being the source. For this reason, the Σ-*distribution specification*, namely the set of Σ-distribution rules on a signature Σ, depends solely on the chosen signature. We also notice that for each possible interpretation of μ and ν as distributions we obtain a different Σ-distribution rule having $\mu \mid \nu$ as source. However, we will show that under a suitable notion of *provability*, the Σ-distribution specification correctly specifies the semantics of closed distribution terms. Moreover, our Σ-distribution specification will play a fundamental rôle in the decomposition method. In fact, in Section 3.2 from the Σ-distribution rules we will derive the Σ-distribution ruloids, which will allow us to derive expressions of the form $\Theta \xrightarrow{q} t$ for an arbitrary open distribution term Θ and open term t from the behavior of the variables occurring in Θ. We remark that our Σ-distribution specification can be exploited also to decompose formulae of any logic offering modalities for the specification of the probabilistic properties of processes. Moreover, it can be easily generalized to cover the case of sub-distributions, which are usually considered alongside a weak semantics for processes [126, 127].

Σ-DISTRIBUTION RULES

A *distribution literal* is an expression of the form $\Theta \xrightarrow{q} t$, with $\Theta \in \mathbb{DT}(\Sigma)$, $q \in (0,1]$ and $t \in \mathbb{T}(\Sigma)$. Given a set of (distribution) literals L we denote by $\mathrm{lhs}(L)$ the set of the left-hand sides of the (distribution) literals in L and by $\mathrm{rhs}(L)$ the set of right-hand sides of the (distribution) literals in L.

A set of distribution literals $\{\Theta \xrightarrow{q_i} t_i \mid i \in I\}$ is a *distribution over terms* if $\sum_{i \in I} q_i = 1$ and all terms t_i are pairwise distinct. This expresses that the possibly open distribution term $\Theta \in \mathbb{DT}(\Sigma)$ is the distribution over possibly open terms in $\mathbb{T}(\Sigma)$ giving weight q_i to t_i. Given an open distribution term $\Theta \in \mathbb{DT}(\Sigma)$ and a distribution over terms $L = \{\Theta \xrightarrow{q_i} t_i \mid i \in I\}$ we denote the set of terms in $\mathrm{rhs}(L)$ by $\mathsf{supp}(\Theta) = \{t_i \mid i \in I\} \subseteq \mathbb{T}(\Sigma)$.

Our target is to derive distributions over terms $\{\pi \xrightarrow{q_i} t_i \mid i \in I\}$ for a distribution $\pi \in \Delta(\mathcal{T}(\Sigma))$ (which coincides with a closed distribution term) and closed terms $t_i \in \mathcal{T}(\Sigma)$ such that: (i) $\{\pi \xrightarrow{q_i} t_i \mid i \in I\}$ if and only if $\pi(t_i) = q_i$ for all $i \in I$, and (ii) $\{\pi \xrightarrow{q_i} t_i \mid i \in I\}$ is obtained inductively with respect to the structure of π. To this aim, we introduce the Σ-*distribution rules* and the Σ-*distribution specification*.

Let $\delta_{\mathcal{V}_s} := \{\delta_x \mid x \in \mathcal{V}_s\}$ denote the set of all instantiable Dirac distributions with a variable as term, and $\vartheta, \vartheta_i, \ldots$ denote distribution terms in $\mathbb{DT}(\Sigma)$ ranging over $\mathcal{V}_d \cup \delta_{\mathcal{V}_s}$. Then, for arbitrary sets $S_1, \ldots, S_n$, we denote by $\times_{i=1}^{n} S_i$ the set of tuples $k = [s_1, \ldots, s_n]$ with $s_i \in S_i$. The i-th element of k is denoted by $k(i)$.

Informally, to define Σ-distribution rules we adopt a positive GSOS-like format: the source term can contain at most one operator symbol from Σ and the premises potentially bind the behavior of distribution variables and Dirac deltas occurring in the source. Accordingly to the term structure of the source, we will distinguish three types of rules:

1. Axioms for Dirac deltas over process variables. In this case, for each $x \in \mathcal{V}_s$, δ_x is characterized as the distribution assigning probability weight 1 to the process variable x.
2. Rules having as source a distribution term of the form $f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}})$ for $f \in \Sigma$ and $\vartheta_i \in \mathcal{V}_d \cup \delta_{\mathcal{V}_s}$ for each $i \in \{1,\dots,\mathfrak{n}\}$. In this case, given a particular characterization of each ϑ_i in the premises, $f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}})$ is characterized as the distribution assigning a positive probability weight q only to process terms of the form $f(x_1,\dots,x_{\mathfrak{n}})$ with $x_i \in \mathsf{supp}(\vartheta_i)$ for each $i \in \{1,\dots,\mathfrak{n}\}$. The weight q is obtained as the product over $i \in \{1,\dots,\mathfrak{n}\}$ of the probabilities q_i with $\vartheta_i \xrightarrow{q_i} x_i$ in the premises.
3. Rules having as source a convex combination $\sum_{i\in I} p_i\vartheta_i$, with $\vartheta_i \in \mathcal{V}_d \cup \delta_{\mathcal{V}_s}$ for each $i \in I$. In this case, given a particular characterization of each ϑ_i in the premises, $\sum_{i\in I} p_i\vartheta_i$ is characterized as their convex combination, namely the distribution that assigns to each $x \in \mathcal{V}_s$ the probability weight $q = \sum_{i\in I} p_i\vartheta_i(x)$.

We are now ready to formally define our Σ-distribution rules.

Definition 3.1 (Σ-distribution rules)**.** Assume a signature Σ. The set R_Σ of the Σ-*distribution rules* consists of the least set containing the following inference rules:

1. $$\frac{}{\{\delta_x \xrightarrow{1} x\}}$$
   for any state variable $x \in \mathcal{V}_s$;

2. $$\frac{\bigcup_{i=1,\dots,\mathfrak{n}} \left\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\right\}}{\left\{f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}}) \xrightarrow{q_k} f(x_{1,k(1)},\dots,x_{\mathfrak{n},k(\mathfrak{n})}) \;\middle|\; k \in \bigtimes_{i=1,\dots,\mathfrak{n}} J_i \text{ and } q_k = \prod_{i=1,\dots,\mathfrak{n}} q_{i,k(i)}\right\}}$$
   where:
   a. $f \in \Sigma$ and $\mathrm{rk}(f) = \mathfrak{n}$,
   b. the distribution terms $\vartheta_1,\dots,\vartheta_{\mathfrak{n}}$ are in $\mathcal{V}_d \cup \delta_{\mathcal{V}_s}$ and are all distinct,
   c. for each $i = 1,\dots,\mathfrak{n}$ the state variables $x_{i,j}$'s with $j \in J_i$ are all distinct,
   d. for each $i = 1,\dots,\mathfrak{n}$ we have $\sum_{j\in J_i} q_{i,j} = 1$;

3.
$$\frac{\bigcup_{i \in I} \left\{ \vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i \right\}}{\left\{ \sum_{i \in I} p_i \vartheta_i \xrightarrow{q_x} x \;\middle|\; x \in \{x_{i,j} \mid i \in I \wedge j \in J_i\} \text{ and } q_x = \sum_{i \in I, j \in J_i \text{ s.t. } x_{i,j}=x} p_i \cdot q_{i,j} \right\}}$$

where:

a. I is an at most countable set of indexes,

b. the distribution terms ϑ_i with $i \in I$ are in $\mathcal{V}_d \cup \delta_{\mathcal{V}_s}$ and are all distinct,

c. for each $i \in I$ the state variables $x_{i,j}$'s with $j \in J_i$ are all distinct,

d. for each $i \in I$ we have $\sum_{j \in J_i} q_{i,j} = 1$.

Then, the Σ-*distribution specification* (Σ-*DS*) is defined as the pair $D_\Sigma = (\Sigma, R_\Sigma)$.

For each Σ-distribution rule r_D, all sets above the line are called *premises*, notation $\mathrm{prem}(r_\mathrm{D})$, and the set below the line is called *conclusion*, notation $\mathrm{conc}(r_\mathrm{D})$. Then, we name the distribution term on the left side of all distribution literals in the conclusion of r_D as *source* of r_D, notation $\mathrm{src}(r_\mathrm{D})$, and the set of the terms in the right side of all distribution literals in the conclusion as target, notation $\mathrm{trg}(r_\mathrm{D})$.

***Example* 3.1.** An example of a Σ-distribution rule with source $\mu \mid \nu$ is the following:

$$\frac{\{\mu \xrightarrow{1/4} x_1, \quad \mu \xrightarrow{3/4} x_2\} \qquad \{\nu \xrightarrow{1/3} y_1, \quad \nu \xrightarrow{2/3} y_2\}}{\left\{\mu \mid \nu \xrightarrow{1/12} x_1 \mid y_1, \quad \mu \mid \nu \xrightarrow{1/6} x_1 | y_2, \quad \mu \mid \nu \xrightarrow{1/4} x_2 \mid y_1, \quad \mu \mid \nu \xrightarrow{1/2} x_2 \mid y_2\right\}}\,.$$

However, we remark that also

$$\frac{\{\mu \xrightarrow{1/2} x_1, \quad \mu \xrightarrow{1/2} x_2\} \qquad \{\nu \xrightarrow{1/2} z_1, \quad \nu \xrightarrow{1/2} z_2\}}{\left\{\mu \mid \nu \xrightarrow{1/4} x_1 \mid z_1, \quad \mu \mid \nu \xrightarrow{1/4} x_1 | z_2, \quad \mu \mid \nu \xrightarrow{1/4} x_2 \mid z_1, \quad \mu \mid \nu \xrightarrow{1/4} x_2 \mid z_2\right\}}$$

is a well defined Σ-distribution rule for $\mu | \nu$.

As another example, a Σ-distribution rule for the distribution term $1/3\mu + 1/2\nu + 1/6\delta_z$ is of the form

$$\frac{\{\mu \xrightarrow{1/2} x, \quad \mu \xrightarrow{1/2} z\} \qquad \{\nu \xrightarrow{2/3} y, \quad \nu \xrightarrow{1/3} z\} \qquad \{\delta_z \xrightarrow{1} z\}}{\left\{\frac{1}{3}\mu + \frac{1}{2}\nu + \frac{1}{6}\delta_z \xrightarrow{1/6} x, \quad \frac{1}{3}\mu + \frac{1}{2}\nu + \frac{1}{6}\delta_z \xrightarrow{1/3} y, \quad \frac{1}{3}\mu + \frac{1}{2}\nu + \frac{1}{6}\delta_z \xrightarrow{1/2} z\right\}}\,.$$

◀

Remark 3.1. We notice that by Definition 3.1.2, the only Σ-distribution rule for a constant function $c \in \Sigma$ is of the form

$$r_\mathrm{D} = \frac{}{c \xrightarrow{1} c}\,.$$

In fact, as the set of arguments of c is empty, we have $\mathrm{prem}(r_\mathrm{D}) = \emptyset$ and moreover, by convention, $\prod_\emptyset q = \sup_{(0,1]} q = 1$.

All premises in a Σ-distribution rule are distributions over terms. This is immediate for rules as in Definition 3.1.1, follows by constraints 2c and 2d for rules as in Definition 3.1.2 and follows by constraints 3c and 3d for rules as in Definition 3.1.3. We can show that also the conclusion is a distribution over terms.

Proposition 3.1. *The conclusion of any Σ-distribution rule is a distribution over terms.*

Proof. We proceed by a case analysis over the form of Σ-distribution rules.

- For Σ-distribution rules $r_{\mathrm{D}} = \{\delta_x \xrightarrow{1} x\}$, for some $x \in \mathcal{V}_s$, and $r_{\mathrm{D}} = \{c \xrightarrow{1} c\}$, for some constant function $c \in \Sigma$, the thesis is immediate.

- Consider a Σ-distribution rule r_{D} as in Definition 3.1.2. Then, to prove the thesis we need to show that $\sum_{k \in \times_{i=1}^{\mathfrak{n}} J_i} q_k = 1$. We have

$$\begin{aligned}
\sum_{k \in \times_{i=1}^{\mathfrak{n}} J_i} q_k &= \sum_{k \in \times_{i=1}^{\mathfrak{n}} J_i} \left(\prod_{i=1}^{\mathfrak{n}} q_{i,k(i)} \right) \\
&= \prod_{i=1}^{\mathfrak{n}} \left(\sum_{j \in J_i} q_{i,j} \right) \\
&= \prod_{i=1}^{\mathfrak{n}} (1) \\
&= 1
\end{aligned}$$

where $\sum_{k \in \times_{i=1}^{\mathfrak{n}} J_i} \left(\prod_{i=1}^{\mathfrak{n}} q_{i,k(i)} \right) = \prod_{i=1}^{\mathfrak{n}} \left(\sum_{j \in J_i} q_{i,j} \right)$ follows by the distributive property of the summation with respect to the product and can be formally proved by induction over $\mathfrak{n}$, with inductive step $\sum_{k \in \times_{i=1}^{\mathfrak{n}-1} J_i} \left(\prod_{i=1}^{\mathfrak{n}-1} q_{i,(i)} \right) = \prod_{i=1}^{\mathfrak{n}-1} \left(\sum_{j \in J_i} q_{i,j} \right)$, as follows:

$$\begin{aligned}
\sum_{k \in \times_{i=1}^{\mathfrak{n}} J_i} \left(\prod_{i=1}^{\mathfrak{n}} q_{i,k(i)} \right) &= \sum_{j \in J_{\mathfrak{n}}} q_{\mathfrak{n},j} \left(\sum_{k \in \times_{i=1}^{\mathfrak{n}-1} J_i} \left(\prod_{i=1}^{\mathfrak{n}-1} q_{i,k(i)} \right) \right) \\
&= \sum_{j \in J_{\mathfrak{n}}} q_{\mathfrak{n},j} \left(\prod_{i=1}^{\mathfrak{n}-1} \left(\sum_{j \in J_i} q_{i,j} \right) \right) && \text{(inductive step)} \\
&= \left(\sum_{j \in J_{\mathfrak{n}}} q_{\mathfrak{n},j} \right) \cdot \left(\prod_{i=1}^{\mathfrak{n}-1} \left(\sum_{j \in J_i} q_{i,j} \right) \right) \\
&= \prod_{i=1}^{\mathfrak{n}} \left(\sum_{j \in J_i} q_{i,j} \right).
\end{aligned}$$

- Finally, consider a Σ-distribution rule r_{D} as in Definition 3.1.3. Then, to prove the thesis we need to show that $\sum_{x \in \{x_{i,j} \mid j \in J_i, i \in I\}} q_x = 1$. We have

$$\sum_{x \in \{x_{i,j} \mid j \in J_i, i \in I\}} q_x = \sum_{x \in \{x_{i,j} \mid j \in J_i, i \in I\}} \left(\sum_{\substack{i \in I, j \in J_i \\ \text{s.t. } x_{i,j} = x}} p_i q_{i,j} \right)$$

$$
\begin{aligned}
&= \sum_{i \in I} p_i \Big(\sum_{\substack{x \in \{x_{i,j} \mid j \in J_i, i \in I\} \\ j \in J_i \text{ s.t. } x_{i,j} = x}} q_{i,j} \Big) \\
&= \sum_{i \in I} p_i \Big(\sum_{j \in J_i} q_{i,j} \Big) && (\forall\, i \in I \text{ the } x_{i,j} \text{ are distinct}) \\
&= \sum_{i \in I} p_i \\
&= 1.
\end{aligned}
$$

■

Reductions

The following notion of reduction with respect to a substitution allows us to extend the notion of substitution to distributions over terms and, then, to Σ-distribution rules. Roughly speaking, whenever a substitution maps the right-hand sides of two or more distribution literals in a set L to the same term, then L is *reduced* to the set L' in which those literals are substituted by a single distribution literal whose weight is given by the sum of their weights.

Definition 3.2 (Reduction with respect to a substitution)**.** Assume a substitution σ and a set of distribution literals $L = \{\Theta \xrightarrow{q_i} t_i \mid i \in I\}$. We say that σ *reduces* L to the set of distribution literals $L' = \{\sigma(\Theta) \xrightarrow{q_j} t_j \mid j \in J\}$, or that L' is the *reduction with respect to* σ of L, notation $\sigma(L) = L'$, if:

- ★ for each index $j \in J$ there is at least one index $i \in I$ with $\sigma(t_i) = t_j$;
- ★ the terms $\{t_j \mid j \in J\}$ are pairwise distinct;
- ★ for each index $j \in J$, we have $q_j = \sum_{\{i \in I \mid \sigma(t_i) = t_j\}} q_i$.

A reduction with respect to σ of a distribution over terms is, in turn, a distribution over terms.

Proposition 3.2. *For a substitution σ and a distribution over terms L, the set of distribution literals $\sigma(L)$ is a distribution over terms.*

Proof. The thesis follows directly by the definition of $\sigma(L)$. In fact, if we let $\sigma(L) = \{\sigma(\Theta) \xrightarrow{q_j} t_j \mid j \in J\}$, then the targets t_j are pairwise distinct by construction and moreover we have

$$
\begin{aligned}
\sum_{j \in J} q_j &= \sum_{j \in J} \Big(\sum_{\{i \in I \mid \sigma(t_i) = t_j\}} q_i \Big) \\
&= \sum_{i \in I} q_i \\
&= 1 && (L \text{ is a distribution over terms}).
\end{aligned}
$$

■

Definition 3.3 (Reduced instance of a Σ-distribution rule)**.** The *reduced instance* of a Σ-distribution rule r_{D} with respect to a substitution σ is the inference rule $\sigma(r_{\mathrm{D}})$ defined as follows:

1. If r_{D} is as in Definition 3.1.1 then $\sigma(r_{\mathrm{D}})$ is the Σ-distribution rule

$$\frac{}{\{\delta_{\sigma(x)} \xrightarrow{1} \sigma(x)\}}.$$

2. If r_{D} is as in Definition 3.1.2 then $\sigma(r_{\mathrm{D}})$ is the Σ-distribution rule

$$\frac{\bigcup_{i=1,\dots,\mathfrak{n}} \{\sigma(\vartheta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{ f(\sigma(\vartheta_1),\dots,\sigma(\vartheta_{\mathfrak{n}})) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) \,\middle|\, \kappa \in \bigtimes_{i=1,\dots\mathfrak{n}} H_i \text{ and } q_\kappa = \prod_{i=1,\dots,\mathfrak{n}} q_{i,\kappa(i)} \right\}}$$

where $\{\sigma(\vartheta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\} = \sigma(\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\})$.

3. If r_{D} is as in Definition 3.1.3 then $\sigma(r_{\mathrm{D}})$ is the Σ-distribution rule

$$\frac{\bigcup_{i \in I} \{\sigma(\vartheta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{ \sum_{i \in I} p_i \sigma(\vartheta_i) \xrightarrow{q_t} t \,\middle|\, t \in \{t_{i,h} \mid i \in I \wedge h \in H_i\} \text{ and } q_t = \sum_{i \in I \wedge h \in H_i \text{ s.t. } t_{i,h}=t} p_i \cdot q_{i,h} \right\}}$$

where $\{\sigma(\vartheta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\} = \sigma(\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\})$.

***Example* 3.2.** Consider the Σ-distribution rule r_{D} for the distribution term $\mu \mid \nu$ given in Example 3.1 and consider the substitution σ with

$$\sigma(x_1) = x \qquad \sigma(x_2) = x \qquad \sigma(y_1) = y \qquad \sigma(y_2) = \mathrm{nil}$$

where nil denotes the process that cannot perform any action. Then we have that the reduced instance of r_{D} with respect to σ is given by

$$\sigma(r_{\mathrm{D}}) = \frac{\{\sigma(\mu) \xrightarrow{1} x\} \qquad \{\sigma(\nu) \xrightarrow{1/3} y, \quad \sigma(\nu) \xrightarrow{2/3} \mathrm{nil}\}}{\{\sigma(\mu \mid \nu) \xrightarrow{1/3} x \mid y, \qquad \sigma(\mu \mid \nu) \xrightarrow{2/3} x \mid \mathrm{nil}\}}.$$

◀

Notice that Proposition 3.2 ensures that the premises of $\sigma(r_{\mathrm{D}})$ are distributions over terms. We can show that also the conclusion of $\sigma(r_{\mathrm{D}})$ is a distribution over terms.

Proposition 3.3. *Let D_Σ be the Σ-DS. The conclusion of a reduced instance of a Σ-distribution rule in D_Σ is a distribution over terms.*

Proof. The thesis immediately follows from the definition of reduced instance of a Σ-distribution rule (Definition 3.3), Proposition 3.1 and Proposition 3.2. ■

Semantics of distribution terms

We conclude this Section by showing that the Σ-distribution specification correctly defines the semantics of closed distribution terms as probability distributions over closed terms as outlined in Chapter 2.2.

Definition 3.4 (Proof from the Σ-DS)**.** A *proof* from the Σ-DS D_Σ of a closed distribution over terms L is a well-founded, upwardly branching tree, whose nodes are labeled by closed distributions over terms, such that the root is labeled L, and, if β is the label of a node $\mathfrak{q}$ and $\mathcal{K}$ is the set of labels of the nodes directly above $\mathfrak{q}$, then $\mathcal{K}/\beta$ is a closed reduced instance of a Σ-distribution rule in R_Σ.

A closed distribution over terms L is *provable* from D_Σ, notation $D_\Sigma \vdash L$, if there exists a proof from D_Σ for L.

***Example* 3.3.** Consider any signature Σ containing the operator of synchronous parallel composition | and let D_Σ be the Σ-DS built on it. We want to show that given a proper closed substitution σ, the distribution over terms

$$\begin{aligned}
L = \Big\{ &\frac{2}{5}\Big(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}\Big) + \frac{3}{5}\Big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}) \mid \delta_{t_5}\Big) \xrightarrow{\frac{1}{10}} t_1,\\
&\frac{2}{5}\Big(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}\Big) + \frac{3}{5}\Big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}) \mid \delta_{t_5}\Big) \xrightarrow{\frac{3}{10}} t_2,\\
&\frac{2}{5}\Big(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}\Big) + \frac{3}{5}\Big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}) \mid \delta_{t_5}\Big) \xrightarrow{\frac{1}{5}} t_3 \mid t_5,\\
&\frac{2}{5}\Big(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}\Big) + \frac{3}{5}\Big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}) \mid \delta_{t_5}\Big) \xrightarrow{\frac{2}{5}} t_4 \mid t_5 \Big\}
\end{aligned}$$

is provable from the Σ-DS. To this aim let us consider the following proof structure: the different instances of the Σ-distribution rules and the arrows between them constitute the proof tree, and the labels of its nodes are given by the closed substitution σ defined below. We decided to use as nodes the Σ-distribution rules instead of using solely the distributions over terms being their conclusions, to improve readability.

$$\begin{tikzcd}[column sep=small]
\{\delta_{x_1} \xrightarrow{1} x_1\} \arrow[d] & \{\delta_{x_2} \xrightarrow{1} x_2\} \arrow[dl] & \{\delta_{y_1} \xrightarrow{1} y_1\} \arrow[d] & \{\delta_{y_2} \xrightarrow{1} y_2\} \arrow[dl] & \{\delta_{z} \xrightarrow{1} z\} \arrow[ddl] \\
\dfrac{\{\delta_{x_1} \xrightarrow{1} x_1\}\ \{\delta_{x_2} \xrightarrow{1} x_2\}}{\left\{\begin{array}{l} \frac{1}{4}\delta_{x_1} + \frac{3}{4}\delta_{x_2} \xrightarrow{1/4} x_1, \\ \frac{1}{4}\delta_{x_1} + \frac{3}{4}\delta_{x_2} \xrightarrow{3/4} x_2 \end{array}\right\}} \arrow[ddrr] & & \dfrac{\{\delta_{y_1} \xrightarrow{1} y_1\}\ \{\delta_{y_2} \xrightarrow{1} y_2\}}{\left\{\begin{array}{l} \frac{1}{3}\delta_{y_1} + \frac{2}{3}\delta_{y_2} \xrightarrow{1/3} y_1, \\ \frac{1}{3}\delta_{y_1} + \frac{2}{3}\delta_{y_2} \xrightarrow{2/3} y_2 \end{array}\right\}} \arrow[dr] & & \\
 & & & \dfrac{\{\mu_1 \xrightarrow{1/3} y_1,\ \mu_1 \xrightarrow{2/3} y_2\}\ \{\nu_1 \xrightarrow{1} z\}}{\left\{\begin{array}{l} \mu_1 \mid \nu_1 \xrightarrow{1/3} y_1 \mid z, \\ \mu_1 \mid \nu_1 \xrightarrow{2/3} y_2 \mid z \end{array}\right\}} \arrow[dl] & \\
 & & \dfrac{\{\mu_2 \xrightarrow{1/4} x_1,\ \mu_2 \xrightarrow{3/4} x_2\} \quad \{\nu_2 \xrightarrow{1/3} w_1,\ \nu_2 \xrightarrow{2/3} w_2\}}{\left\{\begin{array}{l} \frac{2}{5}\mu_2 + \frac{3}{5}\nu_2 \xrightarrow{1/10} x_1,\ \frac{2}{5}\mu_2 + \frac{3}{5}\nu_2 \xrightarrow{3/10} x_2, \\ \frac{2}{5}\mu_2 + \frac{3}{5}\nu_2 \xrightarrow{1/5} w_1,\ \frac{2}{5}\mu_2 + \frac{3}{5}\nu_2 \xrightarrow{2/5} w_2 \end{array}\right\}} & &
\end{tikzcd}$$

Notice that we can assume, without loss of generality, that all the variables occurring in the Σ-distribution rules above are distinct. Then, we consider the closed substitution σ with

$$\sigma(x_1) = t_1 \qquad \sigma(x_2) = t_2 \qquad \sigma(y_1) = t_3 \qquad \sigma(y_2) = t_4 \qquad \sigma(z) = t_5$$
$$\sigma(\mu_1) = \tfrac{1}{3}\delta_{t_3} + \tfrac{2}{3}\delta_{t_4} \qquad\qquad \sigma(\nu_1) = \delta_{t_5}$$
$$\sigma(w_1) = t_3 \mid t_5 \quad \sigma(w_2) = t_4 \mid t_5 \quad \sigma(\mu_2) = \tfrac{1}{4}\delta_{t_1} + \tfrac{3}{4}\delta_{t_2} \quad \sigma(\nu_2) = (\tfrac{1}{3}\delta_{t_3} + \tfrac{2}{3}\delta_{t_4}) \mid \delta_{t_5}.$$

Therefore, we can conclude that $D_\Sigma \vdash L$. ◀

Since Σ-distribution rules have only positive premises, the set of the distribution over terms provable from the Σ-DS is unique [160]. The following result confirms that all probability distributions over $\mathcal{T}(\Sigma)$ can be inferred through the Σ-DS.

Theorem 3.4. *Assume a signature Σ. Let $\pi \in \mathcal{DT}(\Sigma)$ be a closed distribution term and $\{t_m\}_{m\in M} \subseteq \mathcal{T}(\Sigma)$ a set of pairwise distinct closed terms. Then*

$$D_\Sigma \vdash \{\pi \xrightarrow{q_m} t_m \mid m \in M\} \iff \textit{for all } m \in M \textit{ it holds } \pi(t_m) = q_m \textit{ and } \sum_{m\in M} q_m = 1.$$

Proof. (⇒) We aim to prove that

$$D_\Sigma \vdash \{\pi \xrightarrow{q_m} t_m \mid m \in M\} \text{ implies } \pi(t_m) = q_m \text{ for all } m \in M \text{ and } \sum_{m\in M} q_m = 1. \tag{3.1}$$

We proceed by induction over the length of a closed proof γ of $\{\pi \xrightarrow{q_m} t_m \mid m \in M\}$ from D_Σ.

- ★ Base case $|\gamma| = 1$. Since the only distributions over terms derivable in one step are the closed reduced substitution instances of distribution axioms, we have one of the following two cases:

  1. $\pi = \delta_t$ for some $t \in \mathcal{T}(\Sigma)$. The only Σ-distribution rule defining the instantiable Dirac function δ_t is the distribution axiom $r_{\mathrm{D}} = \frac{}{\{\delta_x \xrightarrow{1} x\}}$ (Definition 3.1.1), which should be reduced by a closed substitution σ such that $\sigma(x) = t$, thus giving $\sigma(r_{\mathrm{D}}) = \frac{}{\{\delta_t \xrightarrow{1} t\}}$ by Definition 3.3.1. Consequently the hypothesis $D_\Sigma \vdash \{\pi \xrightarrow{q_m} t_m \mid m \in M\}$ instantiates to $D_\Sigma \vdash \{\delta_t \xrightarrow{1} t\}$ for which Equation (3.1) is straightforward.
  2. $\pi = c$ for some constant operator $c \in \Sigma$. From Remark 3.1, the only Σ-distribution rule defining the behavior of constant operator c is the distribution axiom $r_{\mathrm{D}} = \frac{}{\{c \xrightarrow{1} c\}}$, which is reduced to $\sigma(r_{\mathrm{D}}) = \frac{}{\{c \xrightarrow{1} c\}}$ by Definition 3.3.2, independently on the substitution σ. Therefore, we can conclude that the hypothesis $D_\Sigma \vdash \{\pi \xrightarrow{q_m} t_m \mid m \in M\}$ instantiates to $D_\Sigma \vdash \{c \xrightarrow{1} c\}$ for which Equation (3.1) is straightforward.

- ★ Inductive step $|\gamma| > 1$. We can distinguish two cases, based on the structure of the closed distribution term π.

  1. $\pi = f(\pi_1,\dots,\pi_{\mathfrak{n}})$, for some $f \in \Sigma$ and $\pi_i \in \mathcal{DT}(\Sigma)$ for $i = 1,\dots,\mathfrak{n}$. Then, the bottom of the closed proof γ is constituted by the closed reduced instance of a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form

$$\frac{\bigcup_{i=1}^{\mathfrak{n}} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{ f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}}) \xrightarrow{q_k} f(x_{1,k(1)},\dots,x_{\mathfrak{n},k(\mathfrak{n})}) \;\middle|\; k \in \bigtimes_{i=1}^{\mathfrak{n}} J_i \text{ and } q_k = \prod_{i=1}^{\mathfrak{n}} q_{i,k(i)} \right\}}$$

  (see Definition 3.1.2) with respect to a closed substitution σ with $\sigma(\vartheta_i) = \pi_i$. By Definition 3.3.2 we get that $\sigma(r_{\mathrm{D}})$ has the form

$$\frac{\bigcup_{i=1}^{\mathfrak{n}} \{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{ f(\pi_1,\dots,\pi_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i \text{ and } q_\kappa = \prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)} \right\}}$$

  where

  - ∗ $t_{i,h}$ is a closed term in $\in \mathcal{T}(\Sigma)$ for all $i \in I$ and $h \in H_i$, since σ is closed;
  - ∗ for each $i = 1,\dots,\mathfrak{n}$, the closed terms $t_{i,h}$ are pairwise distinct for $h \in H_i$, since $\{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ is obtained as $\sigma(\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\})$ and we apply Proposition 3.3.

- there is a bijection $\mathfrak{f}\colon \bigtimes_{i=1}^{\mathfrak{n}} H_i \to M$ with $f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) = t_{\mathfrak{f}(\kappa)}$ and $q_\kappa = q_{\mathfrak{f}(\kappa)}$ for each $\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i$.

For each $i = 1,\dots,\mathfrak{n}$ there is a proof shorter than γ for $\{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ from D_Σ. By the inductive hypothesis, this implies that

$$q_{i,h} = \pi_i(t_{i,h}) \text{ for all } h \in H_i \text{ and } \sum_{h \in H_i} q_{i,h} = 1.$$

In particular, we have that for each $\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i$

$$q_{i,\kappa(i)} = \pi_i(t_{i,\kappa(i)}) \tag{3.2}$$

from which we draw

$$\begin{aligned}
q_{\mathfrak{f}(\kappa)} &= q_\kappa && \text{(by definition of } \mathfrak{f}\text{)}\\
&= \prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)} && \text{(by definition of } q_\kappa\text{)}\\
&= \prod_{i=1}^{\mathfrak{n}} \pi_i(t_{i,\kappa(i)}) && \text{(by Equation (3.2))}\\
&= \pi(f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})})) && (\pi = f(\pi_1,\dots,\pi_{\mathfrak{n}}))\\
&= \pi(t_{\mathfrak{f}(\kappa)}) && \text{(by definition of } \mathfrak{f}\text{).}
\end{aligned}$$

Summarizing, we have obtained that $q_m = \pi(t_m)$ for each $m \in M$. Moreover, we have that

$$\begin{aligned}
\sum_{m \in M} q_m &= \sum_{m \in M} q_{\mathfrak{f}^{-1}(m)}\\
&= \sum_{\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i} q_\kappa\\
&= 1 && \text{(by Proposition 3.3)}
\end{aligned}$$

thus giving Equation (3.1).

2. $\pi = \sum_{i \in I} p_i \pi_i$ for some $\pi_i \in \mathcal{D}T(\Sigma)$, $p_i \in (0,1]$ for each $i \in I$ and $\sum_{i \in I} p_i = 1$. Then, the bottom of the closed proof γ is constituted by the closed reduced instance of a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form

$$\frac{\bigcup_{i \in I} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{ \sum_{i \in I} p_i \vartheta_i \xrightarrow{q_x} x \;\middle|\; x \in \{x_{i,j} \mid i \in I \wedge j \in J_i\} \text{ and } q_x = \sum_{i \in I, j \in J_i \text{ s.t. } x_{i,j} = x} p_i q_{i,j} \right\}}$$

(see Definition 3.1.3) with respect to a closed substitution σ with $\sigma(\vartheta_i) = \pi_i$. By Definition 3.3.3 we get that $\sigma(r_{\mathrm{D}})$ is of the form

$$\frac{\bigcup_{i \in I} \{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{ \sum_{i \in I} p_i \pi_i \xrightarrow{q_u} u \;\middle|\; u \in \{t_{i,h} \mid i \in I \wedge h \in H_i\} \text{ and } q_u = \sum_{i \in I, h \in H_i \text{ s.t. } t_{i,h} = u} p_i q_{i,h} \right\}}$$

where

- $t_{i,h}$ is a closed term in $\mathcal{T}(\Sigma)$ for all $h \in H_i$, since σ is closed;
- for each $i \in I$ the closed terms $t_{i,h}$ are pairwise distinct for $h \in H_i$, since $\{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ is obtained as $\sigma(\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\})$ and we apply Proposition 3.3;
- there is a bijection $\mathfrak{f}\colon \{t_{i,h} \mid i \in I \wedge h \in H_i\} \to M$ with $u = t_{\mathfrak{f}(u)}$ and $q_u = q_{\mathfrak{f}(u)}$ for each $u \in \{t_{i,h} \mid i \in I \wedge h \in H_i\}$.

For each $i \in I$ there is a proof shorter than γ for $\{\pi_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ from D_Σ. By the inductive hypothesis, this implies that

$$q_{i,h} = \pi_i(t_{i,h}) \text{ for all } h \in H_i \text{ and } \sum_{h \in H_i} q_{i,h} = 1. \tag{3.3}$$

Then, we have

$$\begin{aligned}
q_{\mathfrak{f}(u)} &= q_u && \text{(by definition of } \mathfrak{f})\\
&= \sum_{i\in I, h\in H_i, \text{ s.t. } t_{i,h}=u} p_i\, q_{i,h} \\
&= \sum_{i\in I, h\in H_i, \text{ s.t. } t_{i,h}=u} p_i\, \pi_i(t_{i,h}) && \text{(by Equation (3.3))}\\
&= \sum_{i\in I, h\in H_i, \text{ s.t. } t_{i,h}=u} p_i\, \pi_i(u) \\
&= \sum_{i \in I} p_i\, \pi_i(u) \\
&= \pi(u) \\
&= \pi(t_{\mathfrak{f}(u)}) && \text{(by definition of } \mathfrak{f}).
\end{aligned}$$

Summarizing, we have obtained that $q_m = \pi(t_m)$ for each $m \in M$. Moreover, we have that

$$\begin{aligned}
\sum_{m\in M} q_m &= \sum_{m \in M} q_{\mathfrak{f}^{-1}(m)} \\
&= \sum_{u \in \{t_{i,h} \mid h \in H_i, i \in I\}} q_u \\
&= 1 && \text{(by Proposition 3.3)}
\end{aligned}$$

thus giving Equation (3.1).

($\Leftarrow$) We aim to prove that

$$\pi(t_m) = q_m \text{ for all } m \in M \text{ and } \sum_{m\in M} q_m = 1 \text{ imply } D_\Sigma \vdash \{\pi \xrightarrow{q_m} t_m \mid m \in M\}. \tag{3.4}$$

We proceed by structural induction over $\pi \in \mathcal{DT}(\Sigma)$.

★ Base case $\pi = \delta_t$ for some $t \in \mathcal{T}(\Sigma)$. Consider the Σ-distribution rule r_{D} $\dfrac{}{\{\delta_x \xrightarrow{1} x\}}$ (Definition 3.1.1) and a closed substitution σ such that $\sigma(x) = t$. By Definition 3.3.1 we get that $\sigma(r_{\mathrm{D}})$ is of the form $\dfrac{}{\{\delta_t \xrightarrow{1} t\}}$, from which we can directly conclude that $D_\Sigma \vdash \{\delta_t \xrightarrow{1} t\}$, thus giving Equation (3.4).

★ Inductive step $\pi = f(\pi_1, \dots, \pi_{\mathfrak{n}})$ for some $\pi_i \in \mathcal{DT}(\Sigma)$ for each $i = 1, \dots, \mathfrak{n}$ and $f \in \Sigma$. For each $i = 1, \dots, \mathfrak{n}$ there is a set of indexes M_i such that:

1. $\pi_i(t_{i,m}) = q_{i,m}$ for all $m \in M_i$,
2. $\sum_{m \in M_i} q_{i,m} = 1$ and
3. the terms $t_{i,m} \in \mathcal{T}(\Sigma)$ are pairwise distinct for each $m \in M_i$.

Let $M = \bigtimes_{i=1}^{\mathfrak{n}} M_i$. We have $\mathrm{supp}(\pi) = \{f(t_{1,\kappa(1)}, \dots, t_{\mathfrak{n},\kappa(\mathfrak{n})}) \mid \kappa \in M\}$ and

$$q_\kappa := \pi\big(f(t_{1,\kappa(1)}, \dots, t_{\mathfrak{n},\kappa(\mathfrak{n})})\big) = \prod_{i=1}^{\mathfrak{n}} \pi_i(t_{i,\kappa(i)}) = \prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)}$$

for each $\kappa \in M$. Hence, to prove Equation (3.4) we need to exhibit a proof of the distribution over terms $\{f(\pi_1, \dots, \pi_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)}, \dots, t_{\mathfrak{n},\kappa(\mathfrak{n})}) \mid \kappa \in M\}$ from D_Σ. By the inductive hypothesis, for each $i = 1, \dots, \mathfrak{n}$ from items (1)–(3) above we get

$$D_\Sigma \vdash \{\pi_i \xrightarrow{q_{i,m}} t_{i,m} \mid m \in M_i\}. \tag{3.5}$$

Consider the Σ-distribution rule r_{D}

$$\frac{\displaystyle\bigcup_{i=1}^{\mathfrak{n}} \{\vartheta_i \xrightarrow{q_{i,m}} x_{i,m} \mid m \in M_i\}}{\left\{f(\vartheta_1, \dots, \vartheta_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(x_{1,\kappa(1)}, \dots, x_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in M \text{ and } q_\kappa = \displaystyle\prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)}\right\}}$$

as in Definition 3.1.2 and a closed substitution σ with $\sigma(\vartheta_i) = \pi_i$ and $\sigma(x_{i,m}) = t_{i,m}$ for each $i = 1, \dots, \mathfrak{n}$ and $m \in M_i$ so that the closed reduced instance of r_{D} with respect to σ is of the form:

$$\frac{\displaystyle\bigcup_{i=1}^{\mathfrak{n}} \{\pi_i \xrightarrow{q_{i,m}} t_{i,m} \mid m \in M_i\}}{\left\{f(\pi_1, \dots, \pi_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)}, \dots, t_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in M \text{ and } q_\kappa = \displaystyle\prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)}\right\}}.$$

We observe that $\mathrm{trg}(\sigma(r_{\mathrm{D}})) = \mathrm{supp}(\pi)$ and since the premises of $\sigma(r_{\mathrm{D}})$ are provable from D_Σ (Equation (3.5)) we can conclude that

$$D_\Sigma \vdash \{f(\pi_1, \dots, \pi_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)}, \dots, t_{\mathfrak{n},\kappa(\mathfrak{n})}) \mid \kappa \in M\}$$

thus proving Equation (3.4).

- ⋆ Inductive step $\pi = \sum_{i \in I} p_i \pi_i$ for some $\pi_i \in \mathcal{DT}(\Sigma)$, $p_i \in (0,1]$ and $\sum_{i \in I} p_i = 1$. For each $i \in I$ there is a set of indexes M_i such that for each $m \in M_i$ such that

  1. $\pi_i(t_{i,m}) = q_{i,m}$,
  2. $\sum_{m \in M_i} q_{i,m} = 1$ and
  3. the terms $t_{i,m} \in \mathcal{T}(\Sigma)$ are pairwise distinct for each $m \in M_i$.

  Let $T = \{t_{i,m} \mid i \in I \text{ and } m \in M_i\}$. We have $\mathrm{supp}(\pi) = T$ and

  $$q_u := \pi(u) = \sum_{i \in I} p_i \pi_i(u) = \sum_{i \in I, m \in M_i, t_{i,m} = u} p_i q_{i,m}$$

  for each $u \in T$. Hence, to prove Equation (3.4) we need to exhibit a proof of $\{\pi \xrightarrow{q_u} u \mid u \in T\}$ from D_Σ.

  By the inductive hypothesis, for all $i \in I$ by items (1)–(3) above we get

  $$D_\Sigma \vdash \{\pi \xrightarrow{q_{i,m}} t_{i,m} \mid m \in M_i\}. \tag{3.6}$$

  Consider the Σ-distribution rule r_{D}

  $$\frac{\bigcup_{i \in I} \{\vartheta_i \xrightarrow{q_{i,m}} x_{i,m} \mid m \in M_i\}}{\left\{ \sum_{i \in I} p_i \vartheta_i \xrightarrow{q_x} x \;\middle|\; x \in \{x_{i,m} \mid i \in I \wedge m \in M_i\} \text{ and } q_x = \sum_{i \in I, m \in M_i \text{ s.t. } x_{i,m} = x} p_i q_{i,m} \right\}}$$

  as in Definition 3.1.3 and a closed substitution σ with $\sigma(\vartheta_i) = \pi_i$ and $\sigma(x_{i,m}) = t_{i,m}$ for each $i \in I$ and $m \in M_i$ so that the closed reduced instance of r_{D} with respect to σ is of the form:

  $$\frac{\bigcup_{i \in I} \{\pi_i \xrightarrow{q_{i,m}} t_{i,m} \mid m \in M_i\}}{\left\{ \sum_{i \in I} p_i \pi_i \xrightarrow{q_u} u \;\middle|\; u \in T \text{ and } q_u = \sum_{i \in I, m \in M_i \text{ s.t. } t_{i,m} = u} p_i q_{i,m} \right\}}$$

  We observe that $\mathrm{trg}(\sigma(r_{\mathrm{D}})) = \mathrm{supp}(\pi)$ and since the premises of $\sigma(r_{\mathrm{D}})$ are provable from D_Σ (Equation (3.6)) we can conclude that

  $$D_\Sigma \vdash \{\sum_{i \in I} p_i \pi_i \xrightarrow{q_u} u \mid u \in T \text{ and } q_u = \sum_{i \in I, m \in M_i \text{ s.t. } t_{i,m} = u} p_i q_{i,m}\}$$

  thus proving Equation (3.4).

■

3.2 Ruloids and distribution ruloids.

In this section we introduce the concept of *ruloid* [33, 34], namely a derived inference rule with an arbitrary term as source allowing us to deduce the behavior of that source term directly from the behavior of the variables occurring in it. This feature makes ruloids fundamental for the decomposition method. The characterization theorems seen in Chapter 2 (Theorem 2.10 and Theorem 2.11) assert that each formula satisfied by a process captures a different aspect of its behavior. Hence, the aim of a decomposition method, which we recall is to reduce the satisfaction problem of a formula for a process to the satisfaction problem of derived formulae for its subprocesses, can be restated by saying that we need to find a method to relate the behavior of a process to the behavior of its subprocesses. This is where ruloids play their rôle: they give us the constraints, expressed as premises of an inference rule, that the closed instances of the variables occurring in the source term of the ruloid must satisfy in order to guarantee that the closed instance of the source term behaves accordingly to the considered formula.

Formally, in this Section we introduce P-*ruloids*, namely the class of ruloids built from a PGSOS-PTSS P, and the Σ-*distribution ruloids*, namely derived Σ-distribution rules allowing us to infer the behavior of any distribution term directly from the behavior of the variables occurring in it. We prove that both classes of ruloids are *sound and specifically witnessing* [34], that is they completely define the behavior of all open (distribution) terms. More precisely, with Theorem 3.6 (resp. Theorem 3.10) we will prove that a closed literal α (resp. a distribution over terms L) is provable from a PGSOS-PTSS P (resp. the Σ-DS) if and only if α (resp. L) is a closed instance of the conclusion of a P-ruloid (resp. Σ-distribution ruloid).

Ruloids

Ruloids are a generalization of PGSOS rules that allow us to infer the behavior of all open terms directly from the behavior of their variables. A ruloid has an arbitrary open term as source, and positive and negative premises for the variables occurring in that term. Ruloids are defined by an inductive composition of PGSOS rules. In detail, from a rule r and a substitution σ, a ruloid ρ with conclusion $\sigma(\mathrm{conc}(r))$ is built as follows: 1. for each positive premise α in $\sigma(r)$, either we put α among the premises of ρ, if the left side of α is a variable, or, otherwise, we take any ruloid having α as conclusion and we put its premises among the premises of ρ; 2. for each negative premise α in $\sigma(r)$, either we put α among the premises of ρ, if the left side of α is a variable, or, otherwise, for each ruloid ρ' having any literal denying α as conclusion, we select any premise β of ρ', we take any literal β' denying β, and we put β' among the premises of ρ.

For a PGSOS-PTSS $P = (\Sigma, \mathcal{A}, R)$, let $\mathrm{Lit}(P)$ denote the set of literals that can be built with terms in $\mathbb{T}(\Sigma) \cup \mathbb{DT}(\Sigma)$ and actions in $\mathcal{A}$.

Definition 3.5 (Ruloids)**.** Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS. The set of P-*ruloids* $\Re^P$ is the smallest set such that:

- ★ $\dfrac{x \xrightarrow{a} \mu}{x \xrightarrow{a} \mu}$ is a P-ruloid for all $x \in \mathcal{V}_s$, $a \in \mathcal{A}$ and $\mu \in \mathcal{V}_d$;

- ★ For a PGSOS rule $r \in R$ of the form

$$\frac{\{x_i \xrightarrow{a_{i,m}} \mu_{i,m} \mid i \in I, m \in M_i\} \qquad \{x_i \not\xrightarrow{a_{i,n}} \mid i \in I, n \in N_i\}}{f(x_1,\dots,x_{\mathfrak{n}}) \xrightarrow{a} \Theta'}$$

  and a substitution σ with $\sigma(x_i) = t_i$ for $i = 1,\dots,\mathfrak{n}$ and $\sigma(\Theta') = \Theta$, the inference rule

$$\frac{\bigcup_{i \in I, m \in M_i} \mathcal{H}_{i,m} \cup \bigcup_{i \in I, n \in N_i} \mathcal{H}_{i,n}}{f(t_1,\dots,t_{\mathfrak{n}}) \xrightarrow{a} \Theta}$$

  is a P-ruloid if the following constraints are satisfied:

  - ∗ for every positive premise $x_i \xrightarrow{a_{i,m}} \mu_{i,m}$ of r
    - either $\sigma(x_i)$ is a variable and $\mathcal{H}_{i,m} = \{\sigma(x_i) \xrightarrow{a_{i,m}} \sigma(\mu_{i,m})\}$,
    - or there is a P-ruloid $\rho_{i,m} = \dfrac{\mathcal{H}_{i,m}}{\sigma(x_i) \xrightarrow{a_{i,m}} \sigma(\mu_{i,m})}$;
  - ∗ for every negative premise $x_i \not\xrightarrow{a_{i,n}}$ of r
    - either $\sigma(x_i)$ is a variable and $\mathcal{H}_{i,n} = \{\sigma(x_i) \not\xrightarrow{a_{i,n}}\}$,
    - or $\mathcal{H}_{i,n} = \mathrm{opp}(\mathrm{pick}(\Re^P_{\sigma(x_i),a_{i,n}}))$, where:

      i. $\Re^P_{\sigma(x_i),a_{i,n}} \in \mathcal{P}(\mathcal{P}(\mathrm{Lit}(P)))$ is the set containing the sets of premises of all P-ruloids with conclusion $\sigma(x_i) \xrightarrow{a_{i,n}} \theta$ for any distribution term $\theta \in \mathbb{DT}(\Sigma)$, formally

      $$\Re^P_{\sigma(x_i),a_{i,n}} = \{\mathrm{prem}(\rho) \mid \rho \in \Re^P \text{ and } \mathrm{conc}(\rho) = \sigma(x_i) \xrightarrow{a_{i,n}} \theta \text{ for } \theta \in \mathbb{DT}(\Sigma)\},$$

      ii. pick: $\mathcal{P}(\mathcal{P}(\mathrm{Lit}(P))) \to \mathcal{P}(\mathrm{Lit}(P))$ is any mapping such that, given any sets of literals L_k with $k \in K$, $\mathrm{pick}(\{L_k \mid k \in K\}) = \{l_k \mid k \in K \wedge l_k \in L_k\}$, namely pick selects exactly one literal from each set L_k,

      iii. opp: $\mathcal{P}(\mathrm{Lit}(P)) \to \mathcal{P}(\mathrm{Lit}(P))$ is any mapping satisfying $\mathrm{opp}(L) = \{\mathrm{opp}(l) \mid l \in L\}$ for all sets of literals L, where $\mathrm{opp}(t' \xrightarrow{a} \theta) = t' \not\xrightarrow{a}$, and $\mathrm{opp}(t' \not\xrightarrow{a}) = t' \xrightarrow{a} \theta$ for some fresh distribution term θ, namely opp applied to any literal returns a denying literal;
  - ∗ the sets of the right hand side variables in $\mathcal{H}_{i,m}$ and $\mathcal{H}_{i,n}$ are all pairwise disjoint, formally $\mathrm{rhs}(\mathcal{H}_{i,h}) \cap \mathrm{rhs}(\mathcal{H}_{j,k}) \neq \emptyset$ for any $h \in M_i \cup N_i$ and $k \in M_j \cup N_j$ implies $h = k$ and $i = j$.

Notice that P-ruloids having a process variable x as source (first item of Definition 3.5 above) state that the only way to derive $x \xrightarrow{a} \mu$ is to directly prove this literal from P.

Example **3.4.** From the rules in Example 2.1 specifying the synchronous parallel composition | and probabilistic alternative composition $+_p$, we derive the following ruloids for term

$x +_p (y \mid z)$:

$$\frac{x \xrightarrow{a} \mu \quad y \stackrel{a}{\nrightarrow}}{x +_p (y \mid z) \xrightarrow{a} \mu} \qquad \frac{x \xrightarrow{a} \mu \quad z \stackrel{a}{\nrightarrow}}{x +_p (y \mid z) \xrightarrow{a} \mu}$$

$$\frac{x \stackrel{a}{\nrightarrow} \quad y \xrightarrow{a} \nu \quad z \xrightarrow{a} \upsilon}{x +_p (y \mid z) \xrightarrow{a} \nu \mid \upsilon} \qquad \frac{x \xrightarrow{a} \mu \quad y \xrightarrow{a} \nu \quad z \xrightarrow{a} \upsilon}{x +_p (y \mid z) \xrightarrow{a} p\mu + (1-p)(\nu \mid \upsilon)} .$$

We describe the construction of the first ruloid.

$$\frac{x \xrightarrow{a} \mu \qquad \dfrac{y \stackrel{a}{\nrightarrow}}{y|z \stackrel{a}{\nrightarrow}}}{x +_p (y|z) \xrightarrow{a} \mu}$$

Accordingly to the second PGSOS rule in Example 2.1, whenever the term x makes an a-move to the distribution variable μ and the term $y|z$ cannot execute action a, then the term $x +_p (y|z)$ makes an a-move to μ. As the left-hand side of the positive premise $x \xrightarrow{a} \mu$ is already a variable, then there is nothing more to do. Conversely, the left-hand side of the negative premise $y|z \stackrel{a}{\nrightarrow}$ is a term. By Definition 3.5 we need to consider all the PGSOS rules having a literal $y|z \xrightarrow{a} \Theta$, for some Θ in $\mathbb{DT}(\Sigma)$, as conclusion, namely any proper instance of the first rule in Example 2.1. Then we need to choose one premise for each of those rules, for instance the one having y as left-hand side, and deny the ones we have selected. In our example, from this construction we obtain the single negative premise $y \stackrel{a}{\nrightarrow}$ whose left-hand side is a variable and thus concludes the construction of the first P-ruloid for the term $x +_p (y|z)$. ◀

We can show that if the PTSS is positive then also the derived ruloids are positive.

Lemma 3.5. *Let P be a positive PGSOS-PTSS. Then all the P-ruloids in $\Re^P$ are positive.*

Proof. The proof follows immediately from Definition 3.5 by noticing that since no rule in P contains negative premises, then the function opp is never applied. Therefore positive literals are never transformed into negative. ■

The following result states that ruloids completely define the behavior of all open terms.

Theorem 3.6 (Ruloid theorem)**.** *Assume a PGSOS-PTSS P, a closed substitution σ, a term $t \in \mathbb{T}(\Sigma)$ and a closed distribution term $\Theta' \in \mathcal{DT}(\Sigma)$. Then $P \vdash \sigma(t) \xrightarrow{a} \Theta'$ if and only if there are a P-ruloid $\frac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a closed substitution σ' with $\sigma'(t) = \sigma(t)$, $\sigma'(\Theta) = \Theta'$ and $P \vdash \sigma'(\mathcal{H})$.*

Proof. We proceed by structural induction on the term $t \in \mathbb{T}(\Sigma)$.

<u>*Base case :*</u> $t = x \in \mathcal{V}_s$.

($\Rightarrow$) The thesis follows immediately for the P-ruloid $\frac{x \xrightarrow{a} \mu}{x \xrightarrow{a} \mu}$ and any closed substitution σ' with $\sigma'(x) = \sigma(x)$ and $\sigma'(\mu) = \Theta'$.

($\Leftarrow$) Accordingly to Definition 3.5, a P-ruloid having x as source is of the form $\frac{x \xrightarrow{a} \mu}{x \xrightarrow{a} \mu}$. Thus, from $\sigma'(x) = \sigma(x)$, $\sigma'(\mu) = \Theta'$ and $P \vdash \sigma'(x) \xrightarrow{a} \sigma'(\mu)$ we can immediately infer that

$P \vdash \sigma(x) \xrightarrow{a} \Theta'$.

<u>*Inductive step*:</u> $t = f(t_1, \dots, t_{\mathfrak{n}}) \in \mathbb{T}(\Sigma)$ for some $\mathfrak{n}$-ary operator f.
($\Rightarrow$) We proceed by structural induction over a closed proof γ of $\sigma(t) \xrightarrow{a} \Theta'$ from P. The bottom of the closed proof γ is constituted by a PGSOS rule $r \in R$ of the form

$$\frac{\{x_i \xrightarrow{a_{i,m}} \mu_{i,m} \mid i \in I, m \in M_i\} \cup \{x_i \stackrel{a_{i,n}}{\not\longrightarrow} \mid i \in I, n \in N_i\}}{f(x_1, \dots, x_{\mathfrak{n}}) \xrightarrow{a} \upsilon}$$

together with a closed substitution ς such that:

1. $\varsigma(x_i) = \sigma(t_i)$ for each $i \in I$;
2. $\varsigma(\upsilon) = \Theta'$;
3. for all $i \in I$ and $m \in M_i$ there is a proof shorter than γ of $\varsigma(x_i) \xrightarrow{a_{i,m}} \varsigma(\mu_{i,m})$ from P;
4. for all $i \in I$ and $n \in N_i$ there is a proof shorter than γ of $\varsigma(x_i) \stackrel{a_{i,n}}{\not\longrightarrow}$ from P.

Let ς_0 be any substitution with $\varsigma_0(x_i) = t_i$ for each $i \in I$. Considering that $\varsigma(x_i) = \sigma(t_i) = \sigma(\varsigma_0(x_i))$, from items (3) and (4) above we get that $P \vdash \sigma(\varsigma_0(x_i)) \xrightarrow{a_{i,m}} \varsigma(\mu_{i,m})$, for $i \in I$ and $m \in M_i$, and $P \vdash \sigma(\varsigma_0(x_i)) \stackrel{a_{i,n}}{\not\longrightarrow}$, for $i \in I$ and $n \in N_i$.
Consider any $\sigma(\varsigma_0(x_i)) \xrightarrow{a_{i,m}} \varsigma(\mu_{i,m})$. By the inductive hypothesis, there are a P-ruloid

$$\frac{\mathcal{H}_{i,m}}{\varsigma_0(x_i) \xrightarrow{a_{i,m}} \Theta_{i,m}}$$

and a closed substitution $\sigma'_{i,m}$ with

- ⋆ $\sigma'_{i,m}(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$,
- ⋆ $\sigma'_{i,m}(\Theta_{i,m}) = \varsigma(\mu_{i,m})$, and
- ⋆ $P \vdash \sigma'_{i,m}(\mathcal{H}_{i,m})$.

Let us consider now any $\varsigma_0(x_i) \stackrel{a_{i,n}}{\not\longrightarrow}$. By definition, $P \vdash \sigma(\varsigma_0(x_i)) \stackrel{a_{i,n}}{\not\longrightarrow}$ only if $P \nvdash \sigma(\varsigma_0(x_i)) \xrightarrow{a_{i,n}} \pi$ for any $\pi \in \mathcal{DT}(\Sigma)$. By structural induction on $\varsigma_0(x_i) = t_i$, this implies that for all P-ruloids of the form

$$\frac{\mathcal{H}_{\Theta_{i,n}}}{\varsigma_0(x_i) \xrightarrow{a_{i,n}} \Theta_{i,n}}$$

and for all closed substitutions σ'' with $\sigma''(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$, it holds that $P \nvdash \sigma''(\mathcal{H}_{\Theta_{i,n}})$. We can distinguish two cases.

a) There is a negative literal $\alpha_{\Theta_{i,n}}$ in $\mathcal{H}_{\Theta_{i,n}}$ such that $P \nvdash \sigma''(\alpha_{\Theta_{i,n}})$ for any closed substitution σ'' with $\sigma''(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$. Then the completeness of P ensures that there are at least one positive literal $\beta_{\Theta_{i,n}}$ denying $\alpha_{\Theta_{i,n}}$ and one closed substitution $\sigma'_{i,n}$ with $\sigma'_{i,n}(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$ s.t. $P \vdash \sigma'_{i,n}(\beta_{\Theta_{i,n}})$.

b) The closed instances of negative literals possibly occurring in $\mathcal{H}_{\Theta_{i,n}}$, with respect to all closed substitutions σ'' with $\sigma''(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$, are provable from P. In this case, since the condition $P \nvdash \sigma''(\mathcal{H}_{\Theta_{i,n}})$ holds for all closed substitutions σ'' as above, we can infer that there is at least one positive literal in $\mathcal{H}_{\Theta_{i,n}}$, say $\alpha_{\Theta_{i,n}}$, s.t. $P \nvdash \sigma''(\alpha_{\Theta_{i,n}})$ for all such closed substitutions σ''. In detail, if we assume wlog. that $\alpha_{\Theta_{i,n}}$ is of the form $y \xrightarrow{a} v$ for some $y \in \mathrm{var}(\varsigma_0(x_i))$ and $v \in \mathcal{V}_d$, then we have obtained that given any closed substitution σ'', with $\sigma''(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$, we have $P \nvdash \sigma''(y) \xrightarrow{a} \pi$ for any $\pi \in \mathcal{DT}(\Sigma)$. By completeness of P, this implies that $P \vdash \sigma''(y) \stackrel{a}{\nrightarrow}$. In general, given a literal $\beta_{\Theta_{i,n}}$ denying $\alpha_{\Theta_{i,n}}$ and any closed substitution $\sigma'_{i,n}$ with $\sigma'_{i,n}(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$, we obtain that $P \vdash \sigma'_{i,n}(\beta_{\Theta_{i,n}})$.

Therefore, if we consider $\mathcal{H}_{i,n} = \bigcup_{\Theta_{i,n}} \beta_{\Theta_{i,n}}$ and we take a closed substitution $\sigma'_{i,n}$ as described in the two cases above, then we obtain

$$P \vdash \sigma'_{i,n}(\mathcal{H}_{i,n}).$$

We remark that since we are working with a countable set of variables, we can always assume that the variables in $\mathrm{rhs}(\mathcal{H}_{i,m})$ for $i \in I$ and $m \in M_i$ and the variables in $\mathrm{rhs}(\mathcal{H}_{i,n})$ for $i \in I$ and $n \in N_i$ are pairwise disjoint. Moreover, all those variables are disjoint from $\mathrm{var}(t)$. Therefore, we can define a closed substitution σ' as follows:

1. $\sigma'(y) = \sigma(y)$ for all $y \in \mathrm{var}(t)$;
2. $\sigma'(\mu) = \sigma'_{i,m}(\mu)$ for all $\mu \in \mathrm{rhs}(\mathcal{H}_{i,m})$, with $i \in I$ and $m \in M_i$;
3. $\sigma'(\mu) = \sigma'_{i,n}(\mu)$ for all $\mu \in \mathrm{rhs}(\mathcal{H}_{i,n})$, with $i \in I$ and $n \in N_i$.

Then define

$$\mathcal{H} = \bigcup_{i \in I, m \in M_i} \mathcal{H}_{i,m} \cup \bigcup_{i \in I, n \in N_i} \mathcal{H}_{i,n}.$$

Moreover, let ς_1 be a substitution with $\varsigma_1(x_i) = t_i$ and $\varsigma_1(\mu_{i,m}) = \Theta_{i,m}$ for all $i \in I$ and $m \in M_i$. We can show that the P-ruloid

$$\frac{\mathcal{H}}{f(t_1, \dots, t_n) \xrightarrow{a} \varsigma_1(v)}$$

together with the substitution σ' satisfies the required properties:

1. First we prove that $\sigma'(f(t_1, \dots, t_n)) = \sigma(f(t_1, \dots, t_n))$. This immediately follows from $\sigma'(y) = \sigma(y)$ for all $y \in \mathrm{var}(f(t_1, \dots, t_n))$.
2. Then we prove that $P \vdash \sigma'(\mathcal{H})$, which is derived from the following considerations:

   a. Substitutions σ' and $\sigma'_{i,m}$ agree on all variables occurring in $\dfrac{\mathcal{H}_{i,m}}{\varsigma_0(x_i) \xrightarrow{a_{i,m}} \Theta_{i,m}}$ for all $i \in I$ and $m \in M_i$. Indeed, assume any index $i \in I$ and $m \in M_i$. Since $\mathrm{var}(f(t_1, \dots, t_n)) = \bigcup_{i=1}^{n} \mathrm{var}(t_i) = \bigcup_{i=1}^{n} \mathrm{var}(\varsigma_0(x_i))$, and, moreover, σ and σ' agree on $\mathrm{var}(f(t_1, \dots, t_n))$ we obtain that $\sigma'(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$ for each $i \in I$. Moreover, by construction we have that $\sigma'_{i,m}(\varsigma_0(x_i)) = \sigma(\varsigma_0(x_i))$, thus giving $\sigma'(\varsigma_0(x_i)) = \sigma'_{i,m}(\varsigma_0(x_i))$, namely σ' and $\sigma'_{i,m}$ agree on $\mathrm{var}(\varsigma_0(x_i))$. Then, by definition σ' and $\sigma'_{i,m}$ agree on all variables in $\mathcal{H}_{i,m}$. Finally, as $\mathrm{var}(\Theta_{i,m}) \subseteq \mathrm{var}(\varsigma_0(x_i)) \cup \mathrm{rhs}(\mathcal{H}_{i,m})$ we can infer that σ' and $\sigma'_{i,m}$ agree also on $\mathrm{var}(\Theta_{i,m})$.

b. With a similar argument we obtain that σ' and $\sigma'_{i,n}$ agree on all variables occurring in $\dfrac{\mathcal{H}_{i,n}}{\varsigma_0(x_i) \overset{a_{i,n}}{\not\longrightarrow}}$ for all $i \in I$ and $n \in N_i$.

c. By item 2a above, for all $i \in I$ and $m \in M_i$ σ' agrees with $\sigma'_{i,m}$ on all variables in $\mathcal{H}_{i,m}$, hence $P \vdash \sigma'_{i,m}(\mathcal{H}_{i,m})$ implies $P \vdash \sigma'(\mathcal{H}_{i,m})$. Analogously, by item 2b above, for all $i \in I$ and $n \in N_i$ σ' agrees with $\sigma'_{i,n}$ on all variables in $\mathcal{H}_{i,n}$, hence $P \vdash \sigma'_{i,n}(\mathcal{H}_{i,n})$ implies $P \vdash \sigma'(\mathcal{H}_{i,n})$. Then, since $\mathcal{H} = \bigcup_{i \in I, m \in M_i} \mathcal{H}_{i,m} \cup \bigcup_{i \in I n \in N_i} \mathcal{H}_{i,n}$ we can conclude that $P \vdash \sigma'(\mathcal{H})$.

3. Finally, we prove that $\sigma'(\varsigma_1(v)) = \Theta'$. Notice that the substitutions ς_0 and ς_1 agree on $\mathrm{var}(f(t_1,\dots,t_{\mathfrak{n}}))$ thus giving $\sigma(\varsigma_0(x_i)) = \sigma(\varsigma_1(x_i))$ for all $i \in I$. Then we have that $\sigma'(\varsigma_1(x_j)) = \sigma'(t_j) = \sigma(t_j) = \varsigma(x_j)$ for $j = 1,\dots,\mathfrak{n}$. Moreover, since σ' and $\sigma'_{i,m}$ agree on $\mathrm{var}(\Theta_{i,m})$, we can infer that $\sigma'(\varsigma_1(\mu_{i,m})) = \sigma'(\Theta_{i,m}) = \sigma'_{i,m}(\Theta_{i,m}) = \varsigma(\mu_{i,m})$ for all $i \in I$ and $m \in M_i$. As $\mathrm{var}(v) \subseteq \{x_1,\dots,x_{\mathfrak{n}}\} \cup \{\mu_{i,m} \mid m \in M_i, i \in I\}$, it follows that $\sigma'(\varsigma_1(v)) = \varsigma(v) = \Theta'$.

($\Leftarrow$) Assume that there a P-ruloid $\rho = \dfrac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a closed substitution σ' with $P \vdash \sigma'(\mathcal{H})$, $\sigma'(t) = \sigma(t)$ and $\sigma'(\Theta) = \Theta'$. We note that the thesis $P \vdash \sigma(t) \xrightarrow{a} \Theta'$ is equivalent to $P \vdash \sigma'(t) \xrightarrow{a} \sigma'(\Theta)$.

Accordingly to Definition 3.5, let r and σ_0 be resp. the PGSOS rule and the substitution from which ρ is built, namely let r be of the form

$$r = \frac{\{x_i \xrightarrow{a_{i,m}} \mu_{i,m} \mid i \in I, m \in M_i\} \qquad \{x_i \overset{a_{i,n}}{\not\longrightarrow} \mid i \in I, n \in N_i\}}{f(x_1,\dots,x_{\mathfrak{n}}) \xrightarrow{a} \Theta''}$$

for $I = \{1,\dots,\mathfrak{n}\}$, and σ_0 be such that $\sigma_0(x_i) = t_i$ and $\sigma_0(\Theta'') = \Theta$. Then ρ is of the form

$$\rho = \frac{\bigcup\limits_{i \in I, m \in M_i} \mathcal{H}_{i,m} \cup \bigcup\limits_{i \in I, n \in N_i} \mathcal{H}_{i,n}}{f(t_1,\dots,t_{\mathfrak{n}}) \xrightarrow{a} \Theta}$$

where:

★ For every positive premise $x_i \xrightarrow{a_{i,m}} \mu_{i,m}$ of r:

∗ Either $\sigma_0(x_i)$ is a variable and $\mathcal{H}_{i,m} = \{\sigma_0(x_i) \xrightarrow{a_{i,m}} \sigma_0(\mu_{i,m})\} = \{t_i \xrightarrow{a_{i,m}} \sigma_0(\mu_{i,m})\}$. Hence from $P \vdash \sigma'(\mathcal{H})$ we can directly infer that $P \vdash \sigma'(t_i) \xrightarrow{a_{i,m}} \sigma'(\sigma_0(\mu_{i,m}))$.

    - Or there is a P-ruloid $\rho_{i,m} = \dfrac{\mathcal{H}_{i,m}}{\sigma_0(x_i) \xrightarrow{a_{i,m}} \sigma_0(\mu_{i,m})} = \dfrac{\mathcal{H}_{i,m}}{t_i \xrightarrow{a,m} \sigma_0(\mu_{i,m})}$. Since $P \vdash \sigma'(\mathcal{H})$ implies $P \vdash \sigma'(\mathcal{H}_{i,m})$, by structural induction on t_i we can infer that $P \vdash \sigma'(t_i) \xrightarrow{a_{i,m}} \sigma'(\sigma_0(\mu_{i,m}))$.

  We can therefore conclude that the closed instances with respect to $\sigma' \circ \sigma_0$ of the positive premises of r are provable from P.

- For every negative premise $x_i \overset{a_{i,n}}{\nrightarrow}$ of r:

    - Either $\sigma_0(x_i)$ is a variable and $\mathcal{H}_{i,n} = \{\sigma_0(x_i) \overset{a_{i,n}}{\nrightarrow}\} = \{t_i \overset{a_{i,n}}{\nrightarrow}\}$. Hence from $P \vdash \sigma'(\mathcal{H})$ we can immediately infer that $P \vdash \sigma'(t_i) \overset{a_{i,n}}{\nrightarrow}$.
    - Or $\mathcal{H}_{i,n} = \mathrm{opp}(\mathrm{pick}(\Re^P_{\sigma_0(x_i),a_{i,n}}))$, namely (see Definition 3.5) for each P-ruloid ρ' such that $\mathrm{conc}(\rho') = \sigma_0(x_i) \xrightarrow{a_{i,n}} \theta$, for any $\theta \in \mathbb{DT}(\Sigma)$, we have that $\mathcal{H}_{i,n}$ contains at least one literal denying a literal in $\mathrm{prem}(\rho')$. Hence, since $P \vdash \sigma'(\mathcal{H})$ implies $P \vdash \sigma'(\mathcal{H}_{i,n})$, we can infer that $P \nvdash \sigma'(\mathrm{prem}(\rho'))$. Hence, the structural induction on $\sigma_0(x_i) = t_i$ (case ($\Rightarrow$)) gives that $P \nvdash \sigma'(t_i) \xrightarrow{a_{i,n}} \sigma'(\sigma_0(\theta))$, for any $\theta \in \mathbb{DT}(\Sigma)$, thus implying $P \vdash \sigma'(t_i) \overset{a_{i,n}}{\nrightarrow}$.

  We can therefore conclude that the closed instances with respect to $\sigma' \circ \sigma_0$ of the negative premises of r are provable from P.

We have obtained that all the closed instances with respect to $\sigma' \circ \sigma_0$ of the premises of r are provable from P and therefore we can infer that there is a proof from P of $\sigma'(t) \xrightarrow{a} \sigma'(\Theta)$, which concludes the proof. ■

Clearly, an analogous result holds if we restrict our attention to positive PGSOS-PTSSs.

Corollary 3.7. *Let P be a positive PGSOS-PTSS. Then $P \vdash \sigma(t) \xrightarrow{a} \Theta'$ for $t \in \mathbb{T}(\Sigma)$, $\Theta' \in \mathcal{DT}(\Sigma)$ and σ a closed substitution, iff there are a positive P-ruloid $\dfrac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a closed substitution σ' with $P \vdash \sigma'(\mathcal{H})$, $\sigma'(t) = \sigma(t)$ and $\sigma'(\Theta) = \Theta'$.*

Proof. The proof follows immediately from Lemma 3.5 and Theorem 3.6. ■

Distribution ruloids

Σ-distribution ruloids are a generalization of Σ-distribution rules and define the behavior of arbitrary open distribution terms. More precisely, they allow us to infer the behavior of a distribution term as a probability distribution over terms from the distribution over terms that characterize the behavior of the variables occurring in it. Similarly to P-ruloids, a Σ-distribution ruloid is defined by an inductive composition of Σ-distribution rules and the left-hand sides of its premises are the variables occurring in the source, which is an arbitrary open distribution term. As the Σ-DS is positive, the definition of Σ-distribution ruloids results technically simpler than that of P-ruloids.

Definition 3.6 (Σ-distribution ruloids)**.** Let $D_\Sigma = (\Sigma, R_\Sigma)$ be the Σ-DS. The set of Σ-*distribution ruloids* $\mathfrak{R}^\Sigma$ is the smallest set such that:

- The inference rule
$$\frac{\{\delta_x \xrightarrow{1} x\}}{\{\delta_x \xrightarrow{1} x\}}$$
is a Σ-distribution ruloid for any $x \in \mathcal{V}_s$;

- The inference rule
$$\frac{\{\mu \xrightarrow{q_i} x_i \mid i \in I\}}{\{\mu \xrightarrow{q_i} x_i \mid i \in I\}}$$
is a Σ-distribution ruloid for any $\mu \in \mathcal{V}_d$, provided that $\sum_{i \in I} q_i = 1$ and all variables x_i with $i \in I$ are distinct;

- For a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form
$$\frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{ f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}}) \xrightarrow{q_k} f(x_{1,k(1)},\dots,x_{\mathfrak{n},k(\mathfrak{n})}) \;\middle|\; k \in \mathop{\times}\limits_{i=1,\dots,\mathfrak{n}} J_i \text{ and } q_k = \prod\limits_{i=1,\dots,\mathfrak{n}} q_{i,k(i)} \right\}}$$
as in Definition 3.1.2 and a substitution σ with $\sigma(r_{\mathrm{D}})$ of the form
$$\frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{ f(\Theta_1,\dots,\Theta_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in \mathop{\times}\limits_{i=1,\dots,\mathfrak{n}} H_i \text{ and } q_\kappa = \prod\limits_{i=1,\dots,\mathfrak{n}} q_{i,\kappa(i)} \right\}}$$
(see Definition 3.3.2), the inference rule
$$\frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \mathcal{H}_i}{\left\{ f(\Theta_1,\dots,\Theta_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in \mathop{\times}\limits_{i=1,\dots,\mathfrak{n}} H_i \text{ and } q_\kappa = \prod\limits_{i=1,\dots,\mathfrak{n}} q_{i,\kappa(i)} \right\}}$$
is a Σ-distribution ruloid if for each $i = 1,\dots,\mathfrak{n}$ we have that:
  - either Θ_i is a variable or a Dirac distribution and $\mathcal{H}_i = \{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$,
  - or there is a Σ-distribution ruloid $\rho_i^{\mathrm{D}} = \dfrac{\mathcal{H}_i}{\{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}$;

- For a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form
$$\frac{\bigcup\limits_{i \in I} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{ \sum\limits_{i \in I} p_i \vartheta_i \xrightarrow{q_x} x \;\middle|\; x \in \{x_{i,j} \mid i \in I \wedge j \in J_i\} \text{ and } q_x = \sum\limits_{i \in I, j \in J_i \text{ s.t. } x_{i,j} = x} p_i \cdot q_{i,j} \right\}}$$

as in Definition 3.1.3 and a substitution σ with $\sigma(r_{\mathrm{D}})$ of the form

$$\frac{\bigcup_{i\in I}\{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{\sum_{i\in I} p_i\Theta_i \xrightarrow{q_u} u \;\middle|\; u \in \{t_{i,h} \mid i\in I \wedge h \in H_i\} \text{ and } q_u = \sum_{i\in I, h\in H_i \text{ s.t. } t_{i,h}=u} p_i \cdot q_{i,h}\right\}}$$

(see Definition 3.3.3), the inference rule

$$\frac{\bigcup_{i\in I}\mathcal{H}_i}{\left\{\sum_{i\in I} p_i\Theta_i \xrightarrow{q_u} u \;\middle|\; u \in \{t_{i,h} \mid i\in I \wedge h \in H_i\} \text{ and } q_u = \sum_{i\in I, h\in H_i \text{ s.t. } t_{i,h}=u} p_i \cdot q_{i,h}\right\}}$$

is a Σ-distribution ruloid if for every $i \in I$ we have that:

- ∗ either Θ_i is a variable or a Dirac distribution and $\mathcal{H}_i = \{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$,
- ∗ or there is a Σ-distribution ruloid $\rho_i^{\mathrm{D}} = \dfrac{\mathcal{H}_i}{\{\Theta_i \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}$.

Similarly to the case of process variables, the Σ-distribution ruloids having a Dirac delta or a distribution variable as source (resp. first and second item of Definition 3.6 above) state that to derive $\{\delta_x \xrightarrow{1} x\}$ and $\{\mu \xrightarrow{q_i} x_i \mid i \in I\}$ these distributions over terms have to be directly provable from the Σ-DS.

***Example* 3.5.** Consider the distribution term $\frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon)$ (which is an instance of the target of the fourth P-ruloid in Example 3.4). Then, we can build the following Σ-distribution ruloid:

$$\frac{\{\mu \xrightarrow{1/4} x_1 \quad \mu \xrightarrow{3/4} x_2\} \qquad \dfrac{\{\nu \xrightarrow{1/3} y_1,\ \ \nu \xrightarrow{2/3} y_2\} \quad \{\upsilon \xrightarrow{1} z\}}{\{\nu|\upsilon \xrightarrow{1/3} y_1|z \quad \nu|\upsilon \xrightarrow{2/3} y_2|z\}}}{\left\{\frac{2}{5}\mu+\frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{1}{10}} x_1, \frac{2}{5}\mu+\frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{3}{10}} x_2, \frac{2}{5}\mu+\frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{1}{5}} y_1|z, \frac{2}{5}\mu+\frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{2}{5}} y_2|z\right\}}.$$

◀

Proposition 3.8. *The conclusion of a Σ-distribution ruloid is a distribution over terms.*

Proof. As the conclusion of a Σ-distribution ruloid coincides with the conclusion of a reduced instance of the Σ-distribution rule on which the Σ-distribution ruloid is built, the thesis follows immediately from Proposition 3.3. ■

The inductive construction of ruloids with respect to the structure of distribution terms, gives the following Lemma.

Lemma 3.9. *Any Σ-distribution ruloid $\dfrac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} t_m \mid m \in M\}}$ is such that:*

1. *for all $\mu \in \mathcal{V}_d$, $\mu \in \mathrm{var}(\Theta)$ iff μ is the left-hand side of a premise in $\mathcal{H}$;*

2. *for all $x \in \mathcal{V}_s$, $x \in \mathrm{var}(\Theta)$ iff δ_x is the left-hand side of a premise in $\mathcal{H}$;*
3. $\bigcup_{m \in M} \mathrm{var}(t_m) = \mathrm{rhs}(\mathcal{H})$.

Proof. The proof follows by structural induction over the source term $\Theta \in \mathbb{DT}(\Sigma)$. ■

The following result states that Σ-distribution ruloids define completely the behavior of all open distribution terms.

Theorem 3.10 (Distribution ruloid theorem)**.** *Assume the Σ-DS D_Σ, a closed substitution σ, a distribution term $\Theta \in \mathbb{DT}(\Sigma)$ and closed terms $t_m \in \mathcal{T}(\Sigma)$ with $m \in M$ pairwise distinct. Then $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ if and only if there are a Σ-distribution ruloid $\dfrac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} u_m \mid m \in M\}}$ and a closed substitution σ' with $\sigma'(\Theta) = \sigma(\Theta)$, $\sigma'(u_m) = t_m$ for each $m \in M$ and $D_\Sigma \vdash \sigma'(\mathcal{H})$.*

Proof. We proceed by structural induction over $\Theta \in \mathbb{DT}(\Sigma)$.

1. Base case: Θ is a Dirac distribution $\Theta = \delta_x$ for some $x \in \mathcal{V}_s$.

   ($\Rightarrow$) The thesis follows immediately for the Σ-distribution ruloid

   $$\frac{\{\delta_x \xrightarrow{1} x\}}{\{\delta_x \xrightarrow{1} x\}}$$

   and the closed substitution $\sigma' = \sigma$.

   ($\Leftarrow$) By Definition 3.6 the only possible Σ-distribution ruloid for Θ has the form

   $$\frac{\{\delta_x \xrightarrow{1} x\}}{\{\delta_x \xrightarrow{1} x\}}.$$

   Thus the thesis follows immediately from $D_\Sigma \vdash \sigma'(\{\delta_x \xrightarrow{1} x\})$ and the choice of σ'.

2. Base case: Θ is a variable $\mu \in \mathcal{V}_d$.

   ($\Rightarrow$) The thesis immediately follows for the Σ-distribution ruloid

   $$\frac{\{\mu \xrightarrow{q_m} x_m \mid m \in M\}}{\{\mu \xrightarrow{q_m} x_m \mid m \in M\}}$$

   and the closed substitution σ' with $\sigma'(\mu) = \sigma(\mu)$ and $\sigma'(x_m) = t_m$ for each $m \in M$.

   ($\Leftarrow$) By Definition 3.6 the considered Σ-distribution ruloid for Θ has the form

   $$\frac{\{\mu \xrightarrow{q_m} x_m \mid m \in M\}}{\{\mu \xrightarrow{q_m} x_m \mid m \in M\}}.$$

   Thus the thesis follows immediately from $D_\Sigma \vdash \sigma'(\{\mu \xrightarrow{q_m} x_m \mid m \in M\})$ and the choice of σ'.

3. Inductive step $\Theta = f(\Theta_1,\dots,\Theta_{\mathfrak{n}})$ for some $f \in \Sigma$ and $\Theta_i \in \mathbb{DT}(\Sigma)$ for $i = 1,\dots,\mathfrak{n}$.

   ($\Rightarrow$) First of all, we recall that by Theorem 3.4 we have $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ iff $\sigma(\Theta)(t_m) = q_m$ for each $m \in M$ and $\sum_{m\in M} q_m = 1$. Thus, for the particular choice of $\sigma(\Theta)$, we have that the closed terms t_m are of the form $t_m = f(t_{1,m},\dots,t_{\mathfrak{n},m})$ for some $\{t_{i,m} \mid i = 1,\dots,\mathfrak{n}\} \subseteq \mathcal{T}(\Sigma)$, for $m \in M$, so that $t_{i,m} \in \mathsf{supp}(\sigma(\Theta_i))$ for each $m \in M$. Next, let us consider a closed proof γ of $\{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ from D_Σ. The bottom of γ is constituted by the closed reduced instance of a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form

$$\frac{\bigcup_{i=1}^{\mathfrak{n}} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{f(\vartheta_1,\dots,\vartheta_{\mathfrak{n}}) \xrightarrow{q_k} f(x_{1,k(1)},\dots,x_{\mathfrak{n},k(\mathfrak{n})}) \;\middle|\; k \in \bigtimes_{i=1}^{\mathfrak{n}} J_i \text{ and } q_k = \prod_{i=1}^{\mathfrak{n}} q_{i,k(i)}\right\}}$$

   with respect to a closed substitution ς with $\varsigma(\vartheta_i) = \sigma(\Theta_i)$ for $i = 1,\dots,\mathfrak{n}$. More precisely, let $\varsigma(r_{\mathrm{D}})$ be the inference rule of the form

$$\frac{\bigcup_{i=1}^{\mathfrak{n}} \{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{f(\sigma(\Theta_1),\dots,\sigma(\Theta_{\mathfrak{n}})) \xrightarrow{q_\kappa} f(t_{1,\kappa(1)},\dots,t_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i \text{ and } q_\kappa = \prod_{i=1}^{\mathfrak{n}} q_{i,\kappa(i)}\right\}}$$

   where

   ★ each set $\{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ is the reduction with respect to σ of the corresponding set $\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}$,

   ★ there is bijection $\mathfrak{f}\colon \bigtimes_{i=1}^{\mathfrak{n}} H_i \to M$ with $t_{i,\kappa(i)} = t_{i,\mathfrak{f}(\kappa)}$ for each $i = 1,\dots,\mathfrak{n}$,

   ★ for all $i = 1,\dots,\mathfrak{n}$ there is a proof shorter than γ of $\{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ from D_Σ.

   Let ς_0 be a substitution with $\varsigma_0(\vartheta_i) = \Theta_i$ for $i = 1,\dots,\mathfrak{n}$. Considering that $\varsigma(\vartheta_i) = \sigma(\Theta_i) = \sigma(\varsigma_0(\vartheta_i))$, we have $\varsigma(\vartheta_i) = \sigma(\varsigma_0(\vartheta_i))$ for $i = 1,\dots,\mathfrak{n}$. As a consequence, $\{\sigma(\varsigma_0(\vartheta_i)) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ for $i = 1,\dots,\mathfrak{n}$, is provable from D_Σ with a proof shorter than γ. Hence, by structural induction over each $\Theta_i = \varsigma_0(\vartheta_i)$, for each $i = 1,\dots,\mathfrak{n}$ there are a Σ-distribution ruloid

$$\frac{\mathcal{H}_i}{\{\varsigma_0(\vartheta_i) \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}$$

   and a closed substitution σ_i with

   a. $\sigma_i(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\zeta_i))$,

   b. $\sigma_i(u_{i,h}) = t_{i,h}$, and

   c. $D_\Sigma \vdash \sigma_i(\mathcal{H}_i)$.

   Consider a closed substitution σ' with

- ⋆ $\sigma'(\zeta) = \sigma(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$,
- ⋆ $\sigma'(\mathrm{rhs}(\mathcal{H}_i)) = \sigma_i(\mathrm{rhs}(\mathcal{H}_i))$ for all $i = 1, \dots, \mathfrak{n}$

and let $\mathcal{H} = \bigcup_{i=1}^{\mathfrak{n}} \mathcal{H}_i$. Moreover, let ς_1 be a substitution with $\varsigma_1(\vartheta_i) = \Theta_i$ and $\varsigma_1(x_{i,j}) = u_{i,h}$ for some $h \in H_i$ accordingly to the reduced instance $\varsigma(r_{\mathrm{D}})$, for all $i = 1, \dots, \mathfrak{n}$ and $j \in J_i$. We recall that $\sigma_i(u_{i,\kappa(i)}) = t_{i,\kappa(i)} = t_{i,\mathfrak{f}(\kappa)}$ for each $i = 1, \dots, \mathfrak{n}$ and we show that the Σ-distribution ruloid

$$\frac{\mathcal{H}}{\{f(\Theta_1, \dots, \Theta_{\mathfrak{n}}) \xrightarrow{q_\kappa} f(u_{1,\kappa(1)}, \dots, u_{\mathfrak{n},\kappa(\mathfrak{n})}) \mid \kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i\}}$$

together with the substitution σ' satisfies the required properties:

a. First we prove that $\sigma'(\Theta) = \sigma(\Theta)$. This immediately follows from $\sigma'(\zeta) = \sigma(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$.

b. Then we show that $D_\Sigma \vdash \sigma'(\mathcal{H})$, which is derived from the following considerations:

   i. Notice that $\mathrm{var}(\Theta) = \bigcup_{i=1}^{\mathfrak{n}} \mathrm{var}(\Theta_i) = \bigcup_{i=1}^{\mathfrak{n}} \mathrm{var}(\varsigma_0(\vartheta_i))$. Thus, since σ and σ' agree on $\mathrm{var}(\Theta)$ we obtain that $\sigma'(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\vartheta_i))$ for each $i = 1, \dots, \mathfrak{n}$. Moreover, by construction we have that $\sigma_i(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\vartheta_i))$ for each $i = 1, \dots, \mathfrak{n}$, thus giving $\sigma'(\varsigma_0(\vartheta_i)) = \sigma_i(\varsigma_0(\vartheta_i))$ for each $i = 1, \dots, \mathfrak{n}$. Further, by definition σ' and σ_i agree on all variables in $\mathrm{rhs}(\mathcal{H}_i)$. As by Lemma 3.9.3, $\mathrm{rhs}(\mathcal{H}_i) = \bigcup_{h \in H_i} \mathrm{var}(u_{i,h})$, we can conclude that σ' and σ_i agree on all variables occurring in $\dfrac{\mathcal{H}_i}{\{\varsigma_0(\vartheta_i) \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}$ for each $i = 1, \dots, \mathfrak{n}$.

   ii. As by the previous item we know that σ' agrees with σ_i on all variables in $\mathcal{H}_i$ and $D_\Sigma \vdash \sigma_i(\mathcal{H}_i)$, we infer that $D_\Sigma \vdash \sigma'(\mathcal{H}_i)$, for $i = 1, \dots, \mathfrak{n}$. Then, from $\mathcal{H} = \bigcup_{i=1}^{\mathfrak{n}} \mathcal{H}_i$, we can immediately conclude that $D_\Sigma \vdash \sigma'(\mathcal{H})$.

c. Finally, we prove that $\sigma'(f(u_{1,\kappa(1)}, \dots, u_{\mathfrak{n},\kappa(\mathfrak{n})})) = t_{\mathfrak{f}(\kappa)}$ for each $\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i$. By Lemma 3.9.3 we have that $\mathrm{var}(f(u_{1,\kappa(1)}, \dots, u_{\mathfrak{n},\kappa(\mathfrak{n})})) \subseteq \mathrm{rhs}(\mathcal{H})$. In addition, we have

   - ⋆ $\mathrm{var}(u_{i,\kappa(i)}) \subseteq \mathrm{rhs}(\mathcal{H}_i)$;
   - ⋆ σ' agrees with σ_i on all variables in $\mathrm{rhs}(\mathcal{H}_i)$, for all $i = 1, \dots, \mathfrak{n}$;
   - ⋆ $\mathrm{rhs}(\mathcal{H}) = \bigcup_{i=1}^{\mathfrak{n}} \mathrm{rhs}(\mathcal{H}_i)$.

   Therefore, we have that $\sigma'(u_{i,\kappa(i)}) = \sigma_i(u_{i,\kappa(i)}) = t_{i,\kappa(i)} = t_{i,\mathfrak{f}(\kappa)}$ for each $i = 1, \dots, \mathfrak{n}$ and for each $\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i$. Hence, we can conclude that for each $\kappa \in \bigtimes_{i=1}^{\mathfrak{n}} H_i$ we have $\sigma'\big(f(u_{1,\kappa(1)}, \dots, u_{\mathfrak{n},\kappa(\mathfrak{n})})\big) = t_{\mathfrak{f}(\kappa)}$.

($\Leftarrow$) We aim to show that $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$. To this aim it is enough to show that $D_\Sigma \vdash \{\sigma'(\Theta) \xrightarrow{q_m} \sigma'(u_m) \mid m \in M\}$ which, since the closed terms t_m are pairwise distinct by the hypothesis, by the choice of σ' is equivalent to $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$.

Notice that by the choice of Θ we have that the open terms u_m are of the form $u_m = f(u_{1,m},\dots,u_{\mathfrak{n},m})$ for some $\{u_{i,m} \mid i = 1,\dots,\mathfrak{n}\} \subseteq \mathcal{T}(\Sigma)$ for $m \in M$, so that $u_{i,m} \in \mathsf{supp}(\Theta_i)$ for each $m \in M$.

Accordingly to Definition 3.6, let r_{D} and σ_0 be resp. the Σ-distribution rule and the substitution from which ρ^{D} is built, namely let r_{D} be of the form

$$\frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{f(\vartheta_1,\dots,\vartheta_\mathfrak{n}) \xrightarrow{q_k} f(x_{1,k(1)},\dots,x_{\mathfrak{n},k(\mathfrak{n})}) \;\middle|\; k \in \bigtimes\limits_{i=1,\dots,\mathfrak{n}} J_i \text{ and } q_k = \prod\limits_{i=1,\dots,\mathfrak{n}} q_{i,k(i)}\right\}}$$

as in Definition 3.1.2 and σ_0 be such that $\sigma_0(r_{\mathsf{D}})$ is of the form

$$\frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}{\left\{f(\Theta_1,\dots,\Theta_\mathfrak{n}) \xrightarrow{q_\kappa} f(u_{1,\kappa(1)},\dots,u_{\mathfrak{n},\kappa(\mathfrak{n})}) \;\middle|\; \kappa \in \bigtimes\limits_{i=1,\dots,\mathfrak{n}} H_i \text{ and } q_\kappa = \prod\limits_{i=1,\dots,\mathfrak{n}} q_{i,\kappa(i)}\right\}}$$

(see Definition 3.3.2) and there is a bijection $\mathfrak{f}\colon \bigtimes_{i=1,\dots,\mathfrak{n}} H_i \to M$ so that $u_{i,\kappa(i)} = u_{i,\mathfrak{f}(\kappa)}$ for each $i = 1,\dots,\mathfrak{n}$, and $q_\kappa = q_{\mathfrak{f}(\kappa)}$ for each $\kappa \in \bigtimes_{i=1,\dots,\mathfrak{n}} H_i$.

Then ρ^{D} is of the form

$$\rho^{\mathsf{D}} = \frac{\bigcup\limits_{i=1,\dots,\mathfrak{n}} \mathcal{H}_i}{\{f(\Theta_1,\dots,\Theta_\mathfrak{n}) \xrightarrow{q_m} u_m \mid m \in M\}}$$

where for each $i = 1,\dots,\mathfrak{n}$ we have that:

- ★ Either $\sigma_0(\vartheta_i) = \Theta_i$ is a variable or a Dirac distribution and $\mathcal{H}_i = \{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}$. Hence from $D_\Sigma \vdash \sigma'(\mathcal{H})$ we can immediately infer that $D_\Sigma \vdash \sigma'(\{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\})$.
- ★ Or there is a Σ-distribution ruloid $\rho_i^{\mathsf{D}} = \dfrac{\mathcal{H}_i}{\{\sigma_0(\vartheta) \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}$. Since $D_\Sigma \vdash \sigma'(\mathcal{H})$ implies $D_\Sigma \vdash \sigma'(\mathcal{H}_i)$, by structural induction on Θ_i we can infer that $D_\Sigma \vdash \sigma'(\{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\})$.

Hence, we have obtained that the closed instances with respect to $\sigma' \circ \sigma_0$ of the premises of r_{D} are provable from D_Σ and therefore we can infer that there is a proof from D_Σ of $\{\sigma'(\Theta) \xrightarrow{q_m} \sigma'(u_m) \mid m \in M\}$. By the choice of σ', we can conclude that $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$.

4. Inductive step $\Theta = \sum_{i\in I} p_i\Theta_i$ for some $\Theta_i \in \mathbb{DT}(\Sigma)$, $p_i \in [0,1]$ for $i \in I$ and $\sum_{i\in I} p_i = 1$.

   ($\Rightarrow$) First of all, we recall that by Theorem 3.4 $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ iff $\sigma(\Theta)(t_m) = q_m$ and $\sum_{m\in M} q_m = 1$. Thus, for the particular choice of $\sigma(\Theta)$, we have that the closed terms t_m are such that $\{t_m \mid m \in M\} = \bigcup_{i\in I}\mathsf{supp}(\sigma(\Theta_i))$. Next, let us consider a closed proof γ of $\{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ from D_Σ. The bottom of γ is constituted by the closed reduced instance of a Σ-distribution rule $r_{\mathrm{D}} \in R_\Sigma$ of the form

   $$\frac{\bigcup_{i\in I}\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j \in J_i\}}{\left\{\sum_{i\in I} p_i\vartheta_i \xrightarrow{q_x} x \,\middle|\, x \in \{x_{i,j} \mid i \in I \wedge j \in J_i\} \text{ and } q_x = \sum_{i\in I, j\in J_i \text{ s.t. } x_{i,j}=x} p_i q_{i,j}\right\}}$$

   with respect to a closed substitution ς with $\varsigma(\vartheta_i) = \sigma(\Theta_i)$ for $i \in I$. More precisely, let $\varsigma(r_{\mathrm{D}})$ be the inference rule of the form

   $$\frac{\bigcup_{i\in I}\{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}}{\left\{\sum_{i\in I} p_i\sigma(\Theta_i) \xrightarrow{q_u} u \,\middle|\, u \in \{t_{i,h} \mid i \in I \wedge h \in H_i\} \text{ and } q_u = \sum_{i\in I, h\in H_i \text{ s.t. } t_{i,h}=u} p_i q_{i,h}\right\}}$$

   where

   - ★ each set $\{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ is the reduction with respect to σ of the corresponding set $\{\varsigma(\vartheta_i) \xrightarrow{q_{i,j}} \varsigma(x_{i,j}) \mid j \in J_i\}$,
   - ★ there is bijection $\mathfrak{f}\colon \{t_{i,h} \mid h \in H_i, i \in I\} \to M$ so that $u = t_{\mathfrak{f}(u)}$ for each $u \in \{t_{i,h} \mid h \in H_i, i \in I\}$ and
   - ★ for each $i \in I$ there is a proof shorter than γ of $\{\sigma(\Theta_i) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ from D_Σ.

   Let ς_0 be a substitution with $\varsigma_0(\vartheta_i) = \Theta_i$ for each $i \in I$. Considering that $\varsigma(\vartheta_i) = \sigma(\Theta_i) = \sigma(\varsigma_0(\vartheta_i))$, we have $\varsigma(\vartheta_i) = \sigma(\varsigma_0(\vartheta_i))$ for each $i \in I$. As a consequence, $\{\sigma(\varsigma_0(\vartheta_i)) \xrightarrow{q_{i,h}} t_{i,h} \mid h \in H_i\}$ for each $i \in I$, is provable from D_Σ with a proof shorter than γ. Hence, by structural induction over each $\Theta_i = \varsigma_0(\vartheta_i)$, for each $i \in I$ there are a Σ-distribution ruloid

   $$\frac{\mathcal{H}_i}{\{\varsigma_0(\vartheta_i) \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}$$

   and a closed substitution σ_i with

   a. $\sigma_i(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\vartheta_i))$,
   b. $\sigma_i(u_{i,h}) = t_{i,h}$, and
   c. $D_\Sigma \vdash \sigma_i(\mathcal{H}_i)$.

   So let us consider a closed substitution σ' with

   - ★ $\sigma'(\zeta) = \sigma(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$,

- ★ $\sigma'(\mathrm{rhs}(\mathcal{H}_i)) = \sigma_i(\mathrm{rhs}(\mathcal{H}_i))$ for all $i \in I$

and let $\mathcal{H} = \bigcup_{i\in I} \mathcal{H}_i$. Moreover, let ς_1 be a substitution with $\varsigma_1(\vartheta_i) = \Theta_i$ and $\varsigma_1(x_{i,j}) = u_{i,h}$ for some $h \in H_i$ accordingly to the reduced instance $\varsigma(r_{\mathrm{D}})$, for all $j \in J_i, i \in I$. We recall that $\sigma_i(u_{i,h}) = t_{i,h}$ for each $h \in H_i, i \in I$. and we prove that the Σ-distribution ruloid

$$\frac{\mathcal{H}}{\left\{\sum_{i\in I} p_i\Theta_i \xrightarrow{q_u} u \mid u \in \{t_{i,h} \mid h \in H_i, i \in I\}\right\}}$$

together with the substitution σ' satisfies the required properties:

a. First we prove that $\sigma'(\Theta) = \sigma(\Theta)$. This immediately follows from $\sigma'(\zeta) = \sigma(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$.

b. Then we prove that $D_\Sigma \vdash \sigma'(\mathcal{H})$, which is derived from the following considerations:

   i. Notice that $\mathrm{var}(\Theta) = \bigcup_{i\in I} \mathrm{var}(\Theta_i) = \bigcup_{i\in I} \mathrm{var}(\varsigma_0(\vartheta_i))$. Thus, since σ and σ' agree on $\mathrm{var}(\Theta)$ we obtain that $\sigma'(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\vartheta_i))$ for each $i \in I$. Moreover, by construction we have that $\sigma_i(\varsigma_0(\vartheta_i)) = \sigma(\varsigma_0(\vartheta_i))$ for each $i \in I$, thus giving $\sigma'(\varsigma_0(\vartheta_i)) = \sigma_i(\varsigma_0(\vartheta_i))$ for each $i \in I$. Furthermore, by definition σ' and σ_i agree on all variables in $\mathrm{rhs}(\mathcal{H}_i)$. As by Lemma 3.9.3, $\mathrm{rhs}(\mathcal{H}_i) = \bigcup_{h\in H_i} \mathrm{var}(u_{i,h})$, we can conclude that σ' and σ_i agree on all variables occurring in $\dfrac{\mathcal{H}_i}{\{\varsigma_0(\vartheta_i) \xrightarrow{q_{i,h}} u_{i,h} \mid h \in H_i\}}$ for each $i \in I$.

   ii. As by the previous item σ' agrees with σ_i on all variables in $\mathcal{H}_i$ and $D_\Sigma \vdash \sigma_i(\mathcal{H}_i)$, we infer $D_\Sigma \vdash \sigma'(\mathcal{H}_i)$, for each $i \in I$. Then, from $\mathcal{H} = \bigcup_{i\in I} \mathcal{H}_i$, we can immediately conclude that $D_\Sigma \vdash \sigma'(\mathcal{H})$.

c. Finally, we prove that $\sigma'(u) = t_{\mathfrak{f}(u)}$ for each $u \in \{t_{i,h} \mid h \in H_i, i \in I\}$. By Lemma 3.9.3 we have that $\mathrm{var}(u) \subseteq \mathrm{rhs}(\mathcal{H})$. Furthermore, we have that

   - ★ $\mathrm{var}(u_{i,h}) \subseteq \mathrm{rhs}(\mathcal{H}_i)$;
   - ★ σ' agrees with σ_i on all variables in $\mathrm{rhs}(\mathcal{H}_i)$, for all $i \in I$;
   - ★ $\mathrm{rhs}(\mathcal{H}) = \bigcup_{i\in I} \mathrm{rhs}(\mathcal{H}_i)$.

   Therefore, we have that $\sigma'(u_{i,h}) = \sigma_i(u_{i,h}) = t_{i,h}$ for each $h \in H_i, i \in I$. Hence, we can conclude that $\sigma'(u) = t_{\mathfrak{f}(u)}$ for each $u \in \{t_{i,h} \mid h \in H_i, i \in I\}$.

($\Leftarrow$) We aim to show that $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$. To this aim it is enough to show that $D_\Sigma \vdash \{\sigma'(\Theta) \xrightarrow{q_m} \sigma'(u_m) \mid m \in M\}$ which, since the closed terms t_m are pairwise distinct by hypothesis, by the choice of σ' is equivalent to $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$.

Accordingly to Definition 3.6, let r_{D} and σ_0 be resp. the Σ-distribution rule and the substitution from which ρ^{D} is built, namely let r_{D} be of the form

$$\frac{\bigcup_{i\in}\{\vartheta_i \xrightarrow{q_{i,j}} x_{i,j} \mid j\in J_i\}}{\left\{\sum_{i\in I} p_i\vartheta_i \xrightarrow{q_x} x \,\middle|\, x\in\{x_{i,j}\mid i\in I\wedge j\in J_i\} \text{ and } q_x = \sum_{i\in I, j\in J_i \text{ s.t. } x_{i,j}=x} p_i\cdot q_{i,j}\right\}}$$

as in Definition 3.1.3 and σ_0 be such that $\sigma_0(r_{\mathrm{D}})$ is of the form

$$\frac{\bigcup_{i\in}\{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h\in H_i\}}{\left\{\sum_{i\in I} p_i\Theta_i \xrightarrow{q_m} u_m \,\middle|\, u_m\in\{u_{i,h}\mid i\in I\wedge h\in H_i\} \text{ and } q_m = \sum_{i\in I, h\in H_i \text{ s.t. } u_{i,h}=u_m} p_i\cdot q_{i,h}\right\}}$$

(see Definition 3.3.3). Then ρ^{D} is of the form

$$\rho^{\mathrm{D}} = \frac{\bigcup_{i\in I}\mathcal{H}_i}{\{\sum_{i\in I} p_i\Theta_i \xrightarrow{q_m} u_m \mid m\in M\}}$$

where for each $i\in I$ we have that:

- ★ Either $\sigma_0(\vartheta_i) = \Theta_i$ is a variable or a Dirac distribution and $\mathcal{H}_i = \{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h\in H_i\}$. Hence from $D_\Sigma \vdash \sigma'(\mathcal{H})$ we can immediately infer that $D_\Sigma \vdash \sigma'(\{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h\in H_i\})$.
- ★ Or there is a Σ-distribution ruloid $\rho_i^{\mathrm{D}} = \dfrac{\mathcal{H}_i}{\{\sigma_0(\vartheta) \xrightarrow{q_{i,h}} u_{i,h} \mid h\in H_i\}}$. Since $D_\Sigma \vdash \sigma'(\mathcal{H})$ implies $D_\Sigma \vdash \sigma'(\mathcal{H}_i)$, by structural induction on Θ_i we can infer that $D_\Sigma \vdash \sigma'(\{\Theta_i \xrightarrow{q_{i,h}} u_{i,h} \mid h\in H_i\})$.

Hence, we have obtained that the closed instances with respect to $\sigma'\circ\sigma_0$ of the premises of r_{D} are provable from D_Σ and therefore we can infer that there is a proof from D_Σ of $\{\sigma'(\Theta) \xrightarrow{q_m} \sigma'(u_m) \mid m\in M\}$. By the choice of σ', we can conclude that $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m\in M\}$.

■

***Example* 3.6.** Consider the distribution term $\Theta = \frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon)$ and the closed substitution σ with $\sigma(\Theta) = \frac{2}{5}(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}) + \frac{3}{5}\big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}) \mid \delta_{t_5}\big)$. Notice that $\sigma(\Theta)$ is the source term of the distribution over terms L in Example 3.3. Thus, we know that

$$D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{1/10} t_1,\ \sigma(\Theta) \xrightarrow{3/10} t_2,\ \sigma(\Theta) \xrightarrow{1/5} t_3|t_5,\ \sigma(\Theta) \xrightarrow{2/5} t_4|t_5\}.$$

Consider the Σ-distribution ruloid ρ^{D} for Θ given in Example 3.5

$$\frac{\{\mu \xrightarrow{1/4} x_1 \quad \mu \xrightarrow{3/4} x_2\} \qquad \{\nu \xrightarrow{1/3} y_1, \quad \nu \xrightarrow{2/3} y_2\} \qquad \{\upsilon \xrightarrow{1} z\}}{\left\{\frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{1}{10}} x_1, \frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{3}{10}} x_2, \frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{1}{5}} y_1|z, \frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon) \xrightarrow{\frac{2}{5}} y_2|z\right\}}.$$

We want to exhibit a proper closed substitution σ' such that ρ^{D} and σ' satisfy Theorem 3.10 with respect to $\sigma(\Theta)$. Let

$$\begin{array}{ccccc} \sigma'(x_1) = t_1 & \sigma'(x_2) = t_2 & \sigma'(y_1) = t_3 & \sigma'(y_2) = t_4 & \sigma'(z) = t_5 \end{array}$$
$$\sigma'(\mu) = \tfrac{1}{4}\delta_{t_1} + \tfrac{3}{4}\delta_{t_2} \qquad \sigma'(\nu) = \tfrac{1}{3}\delta_{t_3} + \tfrac{2}{3}\delta_{t_4} \qquad \sigma'(\upsilon) = \delta_{t_5}.$$

Then we have

$$\sigma'(\Theta) = \frac{2}{5}\sigma'(\mu) + \frac{3}{5}\sigma'(\nu|\upsilon) = \frac{2}{5}(\frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}) + \frac{3}{5}\Big((\frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4})\,|\,\delta_{t_5}\Big).$$

Moreover

$$\sigma'(y_1|z) = t_3|t_5 \qquad \sigma'(y_2|z) = t_4|t_5$$

thus giving that $\sigma'(\mathrm{trg}(\rho^{\mathrm{D}})) = \mathrm{rhs}(L)$. Finally, we remark that

- ⋆ the proof presented for $\{\sigma(\mu_2) \xrightarrow{1/4} t_1, \sigma(\mu_2) \xrightarrow{3/4} t_2\}$ with $\sigma(\mu_2) = \frac{1}{4}\delta_{t_1} + \frac{3}{4}\delta_{t_2}$ in Example 3.3 gives us $D_\Sigma \vdash \{\sigma'(\mu) \xrightarrow{1/4} t_1, \sigma'(\mu) \xrightarrow{3/4} t_2\}$;
- ⋆ the proof presented for $\{\sigma(\mu_1) \xrightarrow{1/3} t_3, \sigma(\mu_1) \xrightarrow{2/3} t_4\}$ with $\sigma(\mu_1) = \frac{1}{3}\delta_{t_3} + \frac{2}{3}\delta_{t_4}$ in Example 3.3 gives us $D_\Sigma \vdash \{\sigma'(\nu) \xrightarrow{1/3} t_3, \sigma'(\nu) \xrightarrow{2/3} t_4\}$;
- ⋆ the proof presented for $\{\sigma(\nu_1) \xrightarrow{1} t_5\}$ with $\sigma(\nu_1) = \delta_{t_5}$ in Example 3.3 gives us $D_\Sigma \vdash \{\sigma'(\upsilon) \xrightarrow{1} t_5\}$.

We have therefore obtained that $D_\Sigma \vdash \sigma'(\mathrm{prem}(\rho^{\mathrm{D}}))$ and thus that ρ^{D} and σ' satisfy Theorem 3.10 with respect to $\sigma(\Theta)$. ◀

3.3 The decomposition method

In this section we present our method for decomposing formulae in the classes $\mathcal{L}$, $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ introduced in Chapter 2.4. To this purpose we exploit the two classes of ruloids introduced in Section 3.2. In fact, the idea behind the decomposition of state (resp. distribution) formulae is to establish which constraints the closed instances of the variables occurring in a (distribution) term must satisfy to guarantee that the closed instance of that (distribution) term satisfies the chosen state (resp. distribution) formula. Thus, since (Σ-distribution) ruloids derive the behavior of a (distribution) term directly from the behavior of the variables occurring in it, the decomposition method is firmly related to them.

Formally, starting from the class $\mathcal{L}$, the decomposition of state formulae follows those in [33, 80, 82–85, 90] and consists in assigning to each term $t \in \mathbb{T}(\Sigma)$ and state formula $\varphi \in \mathcal{L}^{\mathrm{s}}$, a set of functions $\xi \colon \mathcal{V}_s \to \mathcal{L}^{\mathrm{s}}$, called *decomposition mappings*, assigning to each variable x in t a proper formula in $\mathcal{L}^{\mathrm{s}}$ such that for any closed substitution σ it holds that $\sigma(t) \models \varphi$ if and only if $\sigma(x) \models \xi(x)$ for each $x \in \mathrm{var}(t)$ (Theorem 3.12). Each mapping ξ will be defined on a P-ruloid having t as source, P being the considered PGSOS-PTSS.

Similarly, the decomposition of distribution formulae consists in assigning to each distribution term $\Theta \in \mathbb{DT}(\Sigma)$ and distribution formula $\psi \in \mathcal{L}^{\mathrm{d}}$ a set of decomposition mappings $\eta \colon \mathcal{V} \to \mathcal{L}^{\mathrm{d}} \cup \mathcal{L}^{\mathrm{s}}$ such that for any closed substitution σ we get that $\sigma(\Theta) \models \psi$ if and only if $\sigma(\zeta) \models \eta(\zeta)$ for each $\zeta \in \mathrm{var}(\Theta)$ (Theorem 3.12). Each mapping η will be defined on a Σ-distribution ruloid having Θ as source.

Finally, as $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ are subclasses of $\mathcal{L}$, we will show how we can easily derive the decomposition method for them from the one proposed for $\mathcal{L}$ (Theorem 3.14).

Decomposition of $\mathcal{L}$

First we need to introduce the notion of *matching* for a distribution over terms and a distribution formula, seen as a probability distribution over state formulae [41, 66].

Definition 3.7 (Matching)**.** Consider a distribution over terms $L = \{\Theta \xrightarrow{q_m} t_m \mid m \in M\}$ and a distribution formula $\psi = \bigoplus_{i \in I} r_i \varphi_i \in \mathcal{L}^{\mathrm{d}}$. A *matching* for L and ψ is a distribution over the product space $\mathfrak{w} \in \Delta(\mathbb{T}(\Sigma) \times \mathcal{L}^{\mathrm{s}})$ having L and ψ as left and right marginals respectively, that is $\sum_{i \in I} \mathfrak{w}(t_m, \varphi_i) = q_m$ for all $m \in M$ and $\sum_{m \in M} \mathfrak{w}(t_m, \varphi_i) = r_i$ for all $i \in I$. We denote by $\mathfrak{W}(L, \psi)$ the set of all matchings for L and ψ.

Definition 3.8 (Decomposition of $\mathcal{L}$)**.** Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS and let D_Σ be the Σ-DS. We define the mappings

- ⋆ $\cdot^{-1} \colon \mathbb{T}(\Sigma) \to (\mathcal{L}^{\mathrm{s}} \to \mathcal{P}(\mathcal{V}_s \to \mathcal{L}^{\mathrm{s}}))$, and
- ⋆ $\cdot^{-1} \colon \mathbb{DT}(\Sigma) \to (\mathcal{L}^{\mathrm{d}} \to \mathcal{P}(\mathcal{V} \to \mathcal{L}))$

as follows. For each term $t \in \mathbb{T}(\Sigma)$ and state formula $\varphi \in \mathcal{L}^{\mathrm{s}}$, $t^{-1}(\varphi) \in \mathcal{P}(\mathcal{V}_s \to \mathcal{L}^{\mathrm{s}})$ is the set of *decomposition mappings* $\xi \colon \mathcal{V}_s \to \mathcal{L}^{\mathrm{s}}$ such that for any univariate term t we have:

1. $\xi \in t^{-1}(\top)$ iff $\xi(x) = \top$ for all $x \in \mathcal{V}_s$;

2. $\xi \in t^{-1}(\neg\varphi)$ iff there is a function $\mathfrak{f} \colon t^{-1}(\varphi) \to \mathrm{var}(t)$ such that

$$\xi(x) = \begin{cases} \bigwedge\limits_{\xi' \in \mathfrak{f}^{-1}(x)} \neg \xi'(x) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise;} \end{cases}$$

3. $\xi \in t^{-1}(\bigwedge_{j \in J} \varphi_j)$ iff there exist decomposition mappings $\xi_j \in t^{-1}(\varphi_j)$ for all $j \in J$ such that

$$\xi(x) = \bigwedge_{j \in J} \xi_j(x) \text{ for all } x \in \mathcal{V}_s;$$

4. $\xi \in t^{-1}(\langle a\rangle\psi)$ iff there exist a P-ruloid $\frac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a decomposition mapping $\eta \in \Theta^{-1}(\psi)$ such that:

$$\xi(x) = \begin{cases} \displaystyle\bigwedge_{x \xrightarrow{b} \mu \in \mathcal{H}} \langle b\rangle \eta(\mu) \quad \wedge \quad \bigwedge_{x \not\xrightarrow{c} \in \mathcal{H}} \neg\langle c\rangle\top \quad \wedge \quad \eta(x) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise;} \end{cases}$$

5. $\xi \in (\sigma(t))^{-1}(\varphi)$ for a non injective substitution $\sigma\colon \mathrm{var}(t) \to \mathcal{V}_s$ iff there is a decomposition mapping $\xi' \in t^{-1}(\varphi)$ such that

$$\xi(x) = \begin{cases} \displaystyle\bigwedge_{y \in \sigma^{-1}(x)} \xi'(y) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise.} \end{cases}$$

Then, for each distribution term $\Theta \in \mathbb{DT}(\Sigma)$ and distribution formula $\psi \in \mathcal{L}^{\mathrm{d}}$, $\Theta^{-1}(\psi) \in \mathcal{P}(\mathcal{V} \to \mathcal{L})$ is the set of *decomposition mappings* $\eta\colon \mathcal{V} \to \mathcal{L}$ such that for any univariate distribution term Θ we have:

6. $\eta \in \Theta^{-1}(\bigoplus_{i\in I} r_i\varphi_i)$ iff there are a Σ-distribution ruloid $\frac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} t_m \mid m \in M\}}$ and a matching $\mathfrak{w} \in \mathfrak{W}(\{\Theta \xrightarrow{q_m} t_m \mid m \in M\}, \bigoplus_{i\in I} r_i\varphi_i)$ such that for all $m \in M$ and $i \in I$ there is a decomposition mapping $\xi_{m,i}$ with $\begin{cases} \xi_{m,i} \in t_m^{-1}(\varphi_i) & \text{if } \mathfrak{w}(t_m,\varphi_i) > 0 \\ \xi_{m,i} \in t_m^{-1}(\top) & \text{otherwise} \end{cases}$ and we have

   a. for all $\mu \in \mathcal{V}_d$, $\eta(\mu) = \begin{cases} \displaystyle\bigoplus_{\mu \xrightarrow{q_j} x_j \in \mathcal{H}} q_j \bigwedge_{\substack{i\in I \\ m \in M}} \xi_{m,i}(x_j) & \text{if } \mu \in \mathrm{var}(\Theta) \\ 1\top & \text{otherwise} \end{cases}$

   b. for all $x \in \mathcal{V}_s$, $\eta(x) = \begin{cases} \displaystyle\bigwedge_{\substack{i\in I \\ m \in M}} \xi_{m,i}(x) & \text{if } x \in \mathrm{var}(\Theta) \\ \top & \text{otherwise.} \end{cases}$

7. $\eta \in (\sigma(\Theta))^{-1}(\psi)$ for a non injective substitution $\sigma\colon \mathrm{var}(\Theta) \to \mathcal{V}$ iff there is a decomposition mapping $\eta' \in \Theta^{-1}(\psi)$ such that for all $\zeta \in \mathrm{var}(\sigma(\Theta))$ it holds $\eta'(z) = \eta'(z')$ for all $z, z' \in \sigma^{-1}(\zeta)$ and

$$\eta(\zeta) = \begin{cases} \eta'(\tilde{z}) & \text{if } \zeta \in \mathrm{var}(\sigma(\Theta)) \text{ and } \tilde{z} \in \sigma^{-1}(\zeta) \\ \top & \text{if } \zeta \not\in \mathrm{var}(\sigma(\Theta)). \end{cases}$$

We explain our decomposition method for the diamond modality for state formulae and for distribution formulae. For the other modalities on state formulae, which do not directly involve the quantitative properties of processes, we refer to [83].

We discuss first the decomposition of a state formula $\varphi = \langle a \rangle \psi \in \mathcal{L}^{\mathrm{s}}$. Given any term $t \in \mathbb{T}(\Sigma)$ and closed substitution σ, we need to identify in $\xi \in t^{-1}(\varphi)$ which properties each $\sigma(x)$ with $x \in \mathrm{var}(t)$ has to satisfy in order to guarantee $\sigma(t) \models \varphi$. By Definition 2.22 we have that $\sigma(t) \models \varphi$ if and only if $P \vdash \sigma(t) \xrightarrow{a} \pi$ for some probability distribution π such that $\pi \models \psi$. By Theorem 3.6 there is such a transition if and only if there are a P-ruloid $\mathcal{H}/t \xrightarrow{a} \Theta$ and a closed substitution σ' with $\sigma'(t) = \sigma(t)$ and (i) $P \vdash \sigma'(\mathcal{H})$ and (ii) $\sigma'(\Theta) \models \psi$. The validity of condition (i) follows if, for each $x \in \mathrm{var}(t)$, the literals in $\mathcal{H}$ having x as left hand side test only the provable behavior of $\sigma'(x)$. More precisely, we need that $\sigma'(x) \models \langle b \rangle \eta(\mu)$ for each $x \xrightarrow{b} \mu \in \mathcal{H}$, for a chosen decomposition mapping $\eta \in \Theta^{-1}(\psi)$ with $\sigma'(\mu) \models \eta(\mu)$ for each $\mu \in \mathrm{var}(\Theta)$, and that $\sigma'(x) \models \neg\langle c \rangle \top$ for each $x \stackrel{c}{\nrightarrow} \in \mathcal{H}$. The decomposed formula $\xi(x)$ is then defined as the conjunction of such formulae. Moreover, we also add in $\xi(x)$ a conjunct $\eta(x)$ to capture the potential behavior of x as a subterm of the target term Θ. Further, the choice of η and its use in ξ also guarantees that condition (ii) holds.

We discuss now the decomposition of a distribution formula $\psi = \bigoplus_{i \in I} r_i \varphi_i \in \mathcal{L}^{\mathrm{d}}$. Given any distribution term $\Theta \in \mathbb{DT}(\Sigma)$ and a closed substitution σ, we need to identify in $\eta \in \Theta^{-1}(\psi)$ which properties each $\sigma(\zeta)$ with $\zeta \in \mathrm{var}(\Theta)$ has to satisfy in order to guarantee $\sigma(\Theta) \models \psi$. By Definition 2.22 we have that $\sigma(\Theta) \models \psi$ if and only if $\sigma(\Theta) = \sum_{i \in I} r_i \pi_i$ with $t \models \varphi_i$ for all $t \in \mathrm{supp}(\pi_i)$. Assume $\mathrm{supp}(\sigma(\Theta)) = \{t_m \mid m \in M\}$ and $\sigma(\Theta)(t_m) = q_m$. By Theorem 3.4, this is equivalent to have $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in \mathcal{M}\}$ which, by Theorem 3.10, is equivalent to say that there are a Σ-distribution ruloid $\mathcal{H}/\{\Theta \xrightarrow{q_m} u_m \mid m \in M\}$ and a closed substitution σ' with $\sigma'(\Theta) = \sigma(\Theta)$ and (i) $D_\Sigma \vdash \sigma'(\mathcal{H})$ and (ii) $\sigma'(u_m) \models \varphi_i$ whenever $\sigma'(u_m) \in \mathrm{supp}(\pi_i)$. Since the weights q_m are univocally determined by the distributions over terms in $\mathcal{H}$ and moreover they already represent the exact probability weights of $\sigma(\Theta)$, we define, for each $\mu \in \mathrm{var}(\Theta) \cap \mathcal{V}_d$, the decomposition mapping $\eta(\mu)$ using as weights the q_j in the distributions over terms $\{\mu \xrightarrow{q_j} x_j\} \in \mathcal{H}$. Then, to guarantee condition (ii), we define $\mathfrak{w}(u_m, \varphi_i)$ to be positive if $\sigma'(u_m) \in \mathrm{supp}(\pi_i)$ so that we can assign the proper decomposed formula $\xi_{m,i}(x)$ to each $x \in \mathrm{var}(u_m)$ such that $\sigma'(x) \models \xi_{m,i}(x)$. Moreover, since each $\sigma'(u_m)$ may occur in the support of more than one π_i, we impose that each $x \in \mathrm{var}(u_m)$ satisfies the conjunction of all the decomposed formulae $\xi_{m,i}(x)$. Therefore, also condition (i) follows.

***Example* 3.7.** We exemplify two decomposition mappings in the set $t^{-1}(\varphi)$ for term $t = x +_{2/5} (y|z)$, which is the term considered in Example 3.4 with $p = 2/5$, and the formula $\varphi = \langle a \rangle \psi$, with $\psi = \frac{1}{2}\langle a \rangle \top \oplus \frac{1}{2}\neg\langle a \rangle \top$. As this example is aimed at providing a deeper insight on the mechanism of our decomposition method, we will choose arbitrarily the ruloids and the matching for the considered terms and formulae in order to minimize the number of the mappings involved in the decomposition and improve readability. Let ρ be the last ruloid for t in Example 3.4, $\Theta = \frac{2}{5}\mu + \frac{3}{5}(\nu|\upsilon)$ denote its target, and ρ^{D} be the Σ-distribution ruloid for Θ showed in Example 3.5. By Definition 3.8.4, the decomposition mappings $\xi \in t^{-1}(\varphi)$ built over ρ are such that:

$$\xi(x) = \langle a \rangle \eta(\mu) \qquad \xi(y) = \langle a \rangle \eta(\nu) \qquad \xi(z) = \langle a \rangle \eta(\upsilon) \tag{3.7}$$

where $\eta \in \Theta^{-1}(\psi)$. Consider the matching $\mathfrak{w} \in \mathfrak{W}(\mathrm{conc}(\rho^{\mathrm{D}}), \psi)$ for $\mathrm{conc}(\rho^{\mathrm{D}})$ and ψ defined by

$$\mathfrak{w}(x_1, \langle a\rangle\top) = \frac{1}{10} \quad \mathfrak{w}(x_2, \neg\langle a\rangle\top) = \frac{3}{10} \quad \mathfrak{w}(y_1|z, \neg\langle a\rangle\top) = \frac{1}{5} \quad \mathfrak{w}(y_2|z, \langle a\rangle\top) = \frac{2}{5}.$$

For the terms and the formulae to which $\mathfrak{w}$ gives a positive weight, we obtain the decomposition mappings in Table 3.1, where ξ_3 and ξ_4 derive from Definition 3.8.2.

$x_1^{-1}(\langle a\rangle\top) = \{\xi_1\}$	$\xi_1(x_1) = \langle a\rangle\top$, $\xi_1(x) = \top$ for all other $x \in \mathcal{V}_s$
$x_2^{-1}(\neg\langle a\rangle\top) = \{\xi_2\}$	$\xi_2(x_2) = \neg\langle a\rangle\top$, $\xi_2(x) = \top$ for all other $x \in \mathcal{V}_s$
$(y_1|z)^{-1}(\neg\langle a\rangle\top) = \{\xi_3, \xi_4\}$	$\xi_3(y_1) = \neg\langle a\rangle\top$, $\xi_3(z) = \top$, $\xi_3(x) = \top$ for all other $x \in \mathcal{V}_s$ $\xi_4(y_1) = \top$, $\xi_4(z) = \neg\langle a\rangle\top$, $\xi_4(x) = \top$ for all other $x \in \mathcal{V}_s$
$(y_2|z)^{-1}(\langle a\rangle\top) = \{\xi_5\}$	$\xi_5(y_2) = \langle a\rangle\top$, $\xi_5(z) = \langle a\rangle\top$, $\xi_5(x) = \top$ for all other $x \in \mathcal{V}_s$

Table 3.1: *Derived decomposition mappings*

Next, we construct the decomposition mappings for the variable v in Θ with respect to ρ^{D} and $\mathfrak{w}$. By Definition 3.8.6a we consider the weights of the premises of ρ^D having v as left-hand side, namely $\mathcal{H}_v = \{v \xrightarrow{1/3} y_1,\quad v \xrightarrow{2/3} y_2\}$, and use them as weights of the $\bigoplus$ operator. Then for each of the variables y_1, y_2 in the right side of $\mathcal{H}_v$, we consider the conjunction of the formulae assigned to it by one decomposition mapping from each set in the first column of Table 3.1. In detail, by omitting multiple occurrences of the $\top$ formulae in conjunctions, for y_1 we consider $\xi_1(y_1) \wedge \xi_2(y_1) \wedge \xi_3(y_1) \wedge \xi_5(y_1) = \neg\langle a\rangle\top$ and $\xi_1(y_1) \wedge \xi_2(y_1) \wedge \xi_4(y_1) \wedge \xi_5(y_1) = \top$, and for y_2 we consider $\xi_1(y_2) \wedge \xi_2(y_2) \wedge \xi_3(y_2) \wedge \xi_5(y_2) = \langle a\rangle\top$ and $\xi_1(y_1) \wedge \xi_2(y_1) \wedge \xi_4(y_1) \wedge \xi_5(y_1) = \langle a\rangle\top$. Hence the choice between ξ_3 or ξ_4 generates two different decomposition mappings in $\Theta^{-1}(\psi)$: by ξ_3 we obtain the decomposition mapping $\eta_1 \in \Theta^{-1}(\psi)$ with $\eta_1(v) = \frac{1}{3}\neg\langle a\rangle\top \oplus \frac{2}{3}\langle a\rangle\top$ and by ξ_4 we obtain the decomposition mapping $\eta_2 \in \Theta^{-1}(\psi)$ with $\eta_2(v) = \frac{1}{3}\top \oplus \frac{2}{3}\langle a\rangle\top$. By applying the same reasoning to μ and υ we obtain

$$\eta_1(\mu) = \frac{1}{4}\langle a\rangle\top \oplus \frac{3}{4}\neg\langle a\rangle\top \quad \eta_1(v) = \frac{1}{3}\neg\langle a\rangle\top \oplus \frac{2}{3}\langle a\rangle\top \quad \eta_1(\upsilon) = 1(\top \wedge \langle a\rangle\top)$$

$$\eta_2(\mu) = \frac{1}{4}\langle a\rangle\top \oplus \frac{3}{4}\neg\langle a\rangle\top \quad \eta_2(v) = \frac{1}{3}\top \oplus \frac{2}{3}\langle a\rangle\top \quad \eta_2(\upsilon) = 1(\neg\langle a\rangle\top \wedge \langle a\rangle\top)$$

where we have omitted multiple occurrences of the $\top$ formulae in conjunctions. Finally, we obtain two decomposition mappings in $t^{-1}(\varphi)$ by substituting η with either η_1 or η_2 in Equation (3.7), obtaining respectively

$$\xi^1(x) = \langle a\rangle\Big(\frac{1}{4}\langle a\rangle\top \oplus \frac{3}{4}\neg\langle a\rangle\top\Big) \quad \xi^1(y) = \langle a\rangle\Big(\frac{1}{3}\neg\langle a\rangle\top \oplus \frac{2}{3}\langle a\rangle\top\Big) \quad \xi^1(z) = \langle a\rangle\Big(1(\langle a\rangle\top \wedge \top)\Big)$$

$$\xi^2(x) = \langle a\rangle\Big(\frac{1}{4}\langle a\rangle\top \oplus \frac{3}{4}\neg\langle a\rangle\top\Big) \quad \xi^2(y) = \langle a\rangle\Big(\frac{1}{3}\top \oplus \frac{2}{3}\langle a\rangle\top\Big) \quad \xi^2(z) = \langle a\rangle\Big(1(\neg\langle a\rangle\top \wedge \langle a\rangle\top)\Big).$$

◀

The following Lemma proves that our decomposition method preserves the syntactic restrictions of the considered modal class, namely that by decomposing formulae in $\mathcal{L}$ we get formulae in $\mathcal{L}$ thus preserving the logical characterization of Theorem 2.10.

Lemma 3.11. *Assume the terms $t \in \mathbb{T}(\Sigma)$ and $\Theta \in \mathbb{DT}(\Sigma)$ and the formulae $\varphi \in \mathcal{L}^{\mathrm{s}}$ and $\psi \in \mathcal{L}^{\mathrm{d}}$.*

1. *For all $x \in \mathcal{V}_s$ we have $\xi(x) \in \mathcal{L}^{\mathrm{s}}$ for each $\xi \in t^{-1}(\varphi)$.*
2. *For all $\zeta \in \mathcal{V}_d$ we have $\eta(\zeta) \in \mathcal{L}^{\mathrm{d}}$ for each $\eta \in \Theta^{-1}(\psi)$.*
3. *For all $\zeta \in \mathcal{V}_s$ we have $\eta(\zeta) \in \mathcal{L}^{\mathrm{s}}$ for each $\eta \in \Theta^{-1}(\psi)$.*

Proof. The proof follows immediately from Definition 3.8. ■

The following result confirms that our decomposition method is correct.

Theorem 3.12 (Decomposition theorem)**.** *Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS and let D_Σ be the Σ-DS. For any term $t \in \mathbb{T}(\Sigma)$, closed substitution σ and state formula $\varphi \in \mathcal{L}^{\mathrm{s}}$ we have*

$$\sigma(t) \models \varphi \Leftrightarrow \exists \xi \in t^{-1}(\varphi) \textit{ such that for all } x \in \mathrm{var}(t) \textit{ it holds } \sigma(x) \models \xi(x)$$

and for any distribution term $\Theta \in \mathbb{DT}(\Sigma)$, closed substitution σ and distribution formula $\psi \in \mathcal{L}^{\mathrm{d}}$ we have

$$\sigma(\Theta) \models \psi \Leftrightarrow \exists \eta \in \Theta^{-1}(\psi) \textit{ such that for all } \zeta \in \mathrm{var}(\Theta) \textit{ it holds } \sigma(\zeta) \models \eta(\zeta).$$

Proof. We start with univariate terms. We proceed by structural induction over $\phi \in \mathcal{L}$ to prove that for any univariate $t \in \mathbb{T}(\Sigma)$, closed substitution σ and $\phi = \varphi \in \mathcal{L}^{\mathrm{s}}$ we have

$$\sigma(t) \models \varphi \Leftrightarrow \exists \xi \in t^{-1}(\varphi) \text{ such that } \forall x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \xi(x) \tag{3.8}$$

and for any univariate $\Theta \in \mathbb{DT}(\Sigma)$, closed substitution σ and $\phi = \psi \in \mathcal{L}^{\mathrm{d}}$ we have

$$\sigma(\Theta) \models \psi \Leftrightarrow \exists \eta \in \Theta^{-1}(\psi) \text{ such that } \forall \zeta \in \mathrm{var}(\Theta) \text{ it holds } \sigma(\zeta) \models \eta(\zeta). \tag{3.9}$$

- Base case $\phi = \top$. Then by Definition 3.8.1 we have that $\xi \in t^{-1}(\top)$ iff $\xi(x) = \top$ for all $x \in \mathcal{V}_s$. Then Equation (3.8) directly follows from the definition of $\models$ (Definition 2.22).

- Inductive step $\phi = \neg\varphi$ for some $\varphi \in \mathcal{L}^{\mathrm{s}}$. We have

$$\begin{aligned}
& \sigma(t) \models \neg\varphi \\
\Leftrightarrow\ & \sigma(t) \not\models \varphi \\
\Leftrightarrow\ & \forall \xi \in t^{-1}(\varphi)\ \exists x \in \mathrm{var}(t) \text{ s.t. } \sigma(x) \not\models \xi(x) \\
\Leftrightarrow\ & \exists \mathfrak{f} \colon t^{-1}(\varphi) \to \mathrm{var}(t) \text{ s.t. } \forall \xi' \in t^{-1}(\varphi) \text{ it holds } \sigma(\mathfrak{f}(\xi')) \not\models \xi'(\mathfrak{f}(\xi')) \\
\Leftrightarrow\ & \exists \mathfrak{f} \colon t^{-1}(\varphi) \to \mathrm{var}(t) \text{ s.t. } \forall x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \bigwedge_{\xi' \in \mathfrak{f}^{-1}(x)} \neg\xi'(x)
\end{aligned}$$

$$\Leftrightarrow \exists \xi \in t^{-1}(\neg\varphi) \text{ s.t. } \forall\, x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \xi(x)$$

where the second relation follows by the inductive hypothesis and the last relation follows by construction of $t^{-1}(\neg\varphi)$ (Definition 3.8.2). Hence, Equation (3.8) holds also in this case.

★ Inductive step $\phi = \bigwedge_{j\in J} \varphi_j$ for some index set J and $\varphi_j \in \mathcal{L}^{\mathrm{s}}$ for all $\jmath i n J$. We have

$$\begin{aligned}
&\sigma(t) \models \bigwedge_{j\in J} \varphi_j \\
\Leftrightarrow\ &\sigma(t) \models \varphi_j, \text{ for all } j \in J \\
\Leftrightarrow\ &\exists \xi_j \in t^{-1}(\varphi_j) \text{ s.t. } \forall\, x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \xi_j(x), \text{ for all } j \in J \\
\Leftrightarrow\ &\exists \xi_j \in t^{-1}(\varphi_j) \text{ for all } j \in J \text{ s.t. } \forall\, x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \bigwedge_{j\in J} \xi_j(x) \\
\Leftrightarrow\ &\exists \xi \in t^{-1}(\bigwedge_{j\in J} \varphi_j) \text{ s.t. } \forall\, x \in \mathrm{var}(t) \text{ it holds } \sigma(x) \models \xi(x)
\end{aligned}$$

where the second relation follows by the inductive hypothesis and the last relation follows by construction of $t^{-1}(\bigwedge_{j\in J}\varphi_j)$ (Definition 3.8.3). Hence, Equation (3.8) holds also in this case.

★ Inductive step $\phi = \bigoplus_{i\in I} r_i\varphi_i$ for some $\varphi_i \in \mathcal{L}^{\mathrm{s}}$, with $r_i \in (0,1]$ for $i \in I$ and $\sum_{i\in I} r_i = 1$. Notice that in this case we have $\phi \in \mathcal{L}^{\mathrm{d}}$ and therefore we need to show Equation (3.9). To this aim, we prove the two implications separately.

($\Rightarrow$) Assume first that $\sigma(\Theta) \models \bigoplus_{i\in I} r_i\varphi_i$. Then, by definition of $\models$ (Definition 2.22), this implies that there exists a family of probability distributions $\{\pi_i\}_{i\in I} \subseteq \Delta(\mathcal{T}(\Sigma))$ with $\sigma(\Theta) = \sum_{i\in I} r_i\pi_i$ and whenever $t \in \mathsf{supp}(\pi_i)$ for some $t \in \mathcal{T}(\Sigma)$, then $t \models \varphi_i$. Notice that $\mathsf{supp}(\sigma(\Theta)) = \bigcup_{i\in I} \mathsf{supp}(\pi_i)$. Let us order the elements of the support of the distribution $\sigma(\Theta)$ through indexes in a suitable set M, namely $\mathsf{supp}(\sigma(\Theta)) = \{t_m \mid m \in M\}$, with $t_m, t_{m'}$ pairwise distinct for all $m, m' \in M$ with $m \neq m'$. We have $\sigma(\Theta) = \sum_{m\in M} q_m\delta_{t_m}$, for some $q_m \in (0,1]$ such that $\sum_{m\in M} q_m = 1$. In particular, this gives $q_m = \sigma(\Theta)(t_m)$, which, by Theorem 3.4, implies that $D \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$. By Theorem 3.10, $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_m} t_m \mid m \in M\}$ implies that there are a Σ-distribution ruloid $\rho^{\mathsf{D}} = \dfrac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} u_m \mid m \in M\}}$ and a closed substitution σ' with $D_\Sigma \vdash \sigma'(\mathcal{H})$, $\sigma'(\Theta) = \sigma(\Theta)$ and $\sigma'(u_m) = t_m$ for each $m \in M$. Let us show that the rewriting of $\sigma'(\Theta)$ as convex combination of the $\{\pi_i\}_{i\in I}$ gives rise to a matching for $\mathrm{conc}(\rho^{\mathsf{D}})$ and $\bigoplus_{i\in I} r_i\varphi_i$. Define $\mathfrak{w} \in \mathfrak{W}(\mathrm{conc}(\rho^{\mathsf{D}}), \bigoplus_{i\in I} r_i\varphi_i)$ as $\mathfrak{w}(u_m, \varphi_i) = r_i\pi_i(\sigma'(u_m))$, then $\mathfrak{w}$ is a matching with left marginal $\mathrm{conc}(\rho^{\mathsf{D}})$, and right marginal the distribution formula $\bigoplus_{i\in I} r_i\varphi_i$. More precisely, we have

$$\begin{aligned}
&q_m \\
&= \sigma(\Theta)(t_m) && \text{(by construction of } \sigma(\Theta)) \\
&= \sum_{i\in I} r_i\pi_i(t_m) && \text{(by definition of convex combination of distributions)}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i\in I} r_i \pi_i(\sigma'(u_m)) && \text{(by construction of } \sigma'\text{)}\\
&= \sum_{i\in I} \mathfrak{w}(u_m,\varphi_i) && \text{(by definition of } \mathfrak{w}\text{)}
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{m\in M} \mathfrak{w}(u_m,\varphi_i) &= \sum_{m\in M} r_i\pi_i(\sigma'(u_m)) && \text{(by definition of } \mathfrak{w}\text{)}\\
&= \sum_{m\in M} r_i\pi_i(t_m) && \text{(by construction of } \sigma'\text{)}\\
&= r_i \sum_{t\in\mathsf{supp}(\sigma(\Theta))} \pi_i(t) && \text{(by the choice of } M\text{)}\\
&= r_i \sum_{t\in\mathsf{supp}(\pi_i)} \pi_i(t)\\
&= r_i.
\end{aligned}
$$

We derive that:

1. from $\sigma'(\Theta) = \sigma(\Theta)$ we obtain that $\sigma'(\zeta) = \sigma(\zeta)$ for all variables $\zeta \in \mathrm{var}(\Theta)$;
2. whenever $\mathfrak{w}(u_m,\varphi_i) > 0$ it holds that $\sigma'(u_m) \in \mathsf{supp}(\pi_i)$ and, therefore, we infer $\sigma'(u_m) \models \varphi_i$. By the inductive hypothesis we derive that there is a decomposition mapping $\xi_{m,i} \in u_m^{-1}(\varphi_i)$ such that $\sigma'(x) \models \xi_{m,i}(x)$ for all $x \in \mathrm{var}(u_m)$;
3. from $D_\Sigma \vdash \sigma'(\mathcal{H})$ we obtain that for all premises $\{\zeta \xrightarrow{q_j} x_j \mid j \in J\} \in \mathcal{H}$ we have $D_\Sigma \vdash \{\sigma'(\zeta) \xrightarrow{q_h} t'_h \mid h \in H\}$, where $\{\sigma'(\zeta) \xrightarrow{q_h} t'_h \mid h \in H\}$ is $\sigma'(\{\zeta \xrightarrow{q_j} x_j \mid j \in J\})$, for a suitable set of indexes H and proper terms t'_h. By Theorem 3.4, $D_\Sigma \vdash \{\sigma'(\zeta) \xrightarrow{q_h} t'_h \mid h \in H\}$ iff $\sigma'(\zeta)(t'_h) = q_h$ and $\sum_{h\in H} q_h = 1$. Hence we have that

$$
\begin{aligned}
\sigma'(\zeta) &= \sum_{h\in H} q_h\delta_{t'_h}\\
&= \sum_{h\in H}\Big(\sum_{j\in J,\sigma'(x_j)=t'_h} q_j\Big)\delta_{t'_h} && \text{(by Definition 3.2)}\\
&= \sum_{h\in H}\Big(\sum_{j\in J,\sigma'(x_j)=t'_h} q_j\delta_{\sigma'(x_j)}\Big)\\
&= \sum_{j\in J} q_j\delta_{\sigma'(x_j)} && \text{(the } t'_h \text{ are pairwise distinct).}
\end{aligned}
$$

We remark that this reasoning holds since we assumed that Θ is univariate, and therefore there is only one set of distribution premises in $\mathcal{H}$ with left-hand side ζ, for each $\zeta \in \mathrm{var}(\Theta)$.

Let $\eta \in \Theta^{-1}(\bigoplus_{i\in I} r_i\varphi_i)$ be the decomposition mapping defined as in Definition 3.8.6 by means of the Σ-distribution ruloid $\dfrac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} u_m \mid m \in M\}}$ and the decomposition mappings $\xi_{m,i}$ as in item (2) above for each $m \in M$ and $i \in I$ such that $\mathfrak{w}(u_m,\varphi_i) > 0$, and $\xi_{m,i}$ defined by $\xi_{m,i}(x) = \top$ for all $x \in \mathcal{V}_s$ for those m, i such that $\mathfrak{w}(u_m, \varphi_i) = 0$. We aim to show that for this η it holds that $\sigma'(\zeta) \models \eta(\zeta)$ for each $\zeta \in \mathrm{var}(\Theta)$. By construction,

$$\eta(\zeta) = \begin{cases} \bigoplus\limits_{\{\zeta \xrightarrow{q_j} x_j \mid j \in J\} \in \mathcal{H}} q_j \bigwedge\limits_{\substack{m \in M \\ i \in I}} \xi_{m,i}(x_j) & \text{if } \zeta \in \mathcal{V}_d \\ \bigwedge\limits_{\substack{m \in M \\ i \in I}} \xi_{m,i}(x) & \text{if } \zeta = x \in \mathcal{V}_s. \end{cases}$$

For each variable $y \in \{x_j \mid j \in J\} \cup \{x\}$ and for each $m \in M$ and $i \in I$, we can distinguish three cases:

4. $y \in \mathrm{var}(u_m)$ and $\mathfrak{w}(u_m, \varphi_i) > 0$. Then, by item (2) above, we have $\sigma'(y) \models \xi_{m,i}(y)$.
5. $y \in \mathrm{var}(u_m)$ and $\mathfrak{w}(u_m, \varphi_i) = 0$. Then by construction $\xi_{m,i}(y) = \top$, thus giving that $\sigma'(y) \models \xi_{m,i}(y)$ holds trivially also in this case.
6. $y \not\in \mathrm{var}(u_m)$. Then, whichever is the value of $\mathfrak{w}(u_m, \varphi_i)$, we have $\xi_{m,i}(y) = \top$ (see Definition 3.8) and consequently $\sigma'(y) \models \xi_{m,i}(y)$ holds trivially also in this case.

Since these considerations apply to each $m \in M$ and $i \in I$ we can conclude that if $\zeta \in \mathcal{V}_d$ then for all $\{\zeta \xrightarrow{q_j} x_j \mid j \in J\} \in \mathcal{H}$ it holds that for each x_j with $j \in J$ we have $\sigma'(x_j) \models \bigwedge_{m \in M, i \in I} \xi_{m,i}(x_j)$. Furthermore, by item (3) above, if $\{\zeta \xrightarrow{q_j} x_j \mid j \in J\} \in \mathcal{H}$ then $D_\Sigma \vdash \sigma'(\mathcal{H})$ gives $\sigma'(\zeta) = \sum_{j \in J} q_j \delta_{\sigma'(x_j)}$, from which we can conclude that

$$\sigma'(\zeta) \models \bigoplus_{j \in J} q_j \bigwedge_{i \in I, m \in M} \xi_{m,i}(x_j), \text{ namely } \sigma'(\zeta) \models \eta(\zeta).$$

Similarly, if $\zeta = x \in \mathcal{V}_s$ then

$$\sigma'(x) \models \bigwedge_{m \in M, i \in I} \xi_{m,i}(x), \text{ namely } \sigma'(x) \models \eta(x).$$

Thus, we can conclude that for each $\zeta \in \mathrm{var}(\Theta)$ it holds that $\sigma'(\zeta) \models \eta(\zeta)$. Since moreover $\sigma(\zeta) = \sigma'(\zeta)$ (item (1) above), we can conclude that $\sigma(\zeta) \models \eta(\zeta)$ as required.

($\Leftarrow$) Assume now that there is a decomposition mapping $\eta \in \Theta^{-1}(\bigoplus_{i \in I} r_i \varphi_i)$ such that $\sigma(\zeta) \models \eta(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$. Following Definition 3.8.6, the existence of such a decomposition mapping η entails the existence of a Σ-distribution ruloid $\rho^{\mathrm{D}} = \dfrac{\mathcal{H}}{\{\Theta \xrightarrow{q_m} t_m \mid m \in M\}}$ with $\sum_{m \in M} q_m = 1$ (Proposition 3.8) and of a matching $\mathfrak{w}$ for $\mathrm{conc}(\rho^{\mathrm{D}})$ and $\bigoplus_{i \in I} r_i \varphi_i$ from which we can build the following decomposition mappings:

$$\begin{cases} \xi_{m,i} \in t_m^{-1}(\varphi_i) & \text{if } \mathfrak{w}(t_m, \varphi_i) > 0 \\ \xi_{m,i} \in t_m^{-1}(\top) & \text{otherwise.} \end{cases}$$

In particular, we have that for each $\mu \in \mathrm{var}(\Theta)$

$$\eta(\mu) = \bigoplus_{\{\mu \xrightarrow{q_j} x_j \mid \sum_{j \in J} q_j = 1\} \in \mathcal{H}} q_j \bigwedge_{i \in I, m \in M} \xi_{m,i}(x_j)$$

and for each $x \in \mathrm{var}(\Theta)$

$$\eta(x) = \bigwedge_{i \in I, m \in M} \xi_{m,i}(x).$$

We define a closed substitution σ' such that $\sigma'(\zeta) = \sigma(\zeta)$ for each $\zeta \in \mathrm{var}(\Theta)$ and $\sigma'(x) = \sigma(x)$ for each $x \in \mathrm{rhs}(\mathcal{H})$. Then, the following properties hold:

(a) From $\sigma'(\zeta) = \sigma(\zeta)$ and $\sigma(\zeta) \models \eta(\zeta)$ we derive $\sigma'(\zeta) \models \eta(\zeta)$. In particular we obtain that $\sigma'(x) \models \bigwedge_{i \in I, m \in M} \xi_{m,i}(x)$ for each $x \in \mathrm{var}(\Theta)$.

(b) As $\sigma'(\mu) \models \eta(\mu)$ for each $\mu \in \mathrm{var}(\Theta)$, by previous item (a), we derive that there are probability distributions π_j such that $\sigma'(\mu) = \sum_{j \in J} q_j \pi_j$ and whenever $t \in \mathrm{supp}(\pi_j)$, for some $t \in \mathcal{T}(\Sigma)$, then $t \models \bigwedge_{i \in I, m \in M} \xi_{m,i}(x_j)$. By Definition 3.8.6a, the weights of the distribution formula $\eta(\mu)$ coincide with the weights of the distribution literals in $\{\mu \xrightarrow{q_j} x_j \mid \sum_{j \in J} q_j = 1\} \in \mathcal{H}$. Therefore, we have that $\sigma'(\mu) = \sum_{j \in J} q_j \delta_{\sigma'(x_j)}$ from which we gather $\sigma'(x_j) \models \bigwedge_{i \in I, m \in M} \xi_{m,i}(x_j)$, for each $j \in J$.

(c) From $\sigma(\zeta) = \sigma'(\zeta)$ for each $\zeta \in \mathrm{var}(\Theta)$ we infer that $\sigma'(\Theta) = \sigma(\Theta)$. Moreover, by Lemma 3.9.3 we have that $\mathrm{rhs}(\mathcal{H}) = \bigcup_{m \in M} \mathrm{var}(t_m)$, so that $\sigma'(x) = \sigma(x)$ for each $x \in \mathrm{rhs}(\mathcal{H})$ implies $\sigma'(t_m) = \sigma(t_m)$ for each $m \in M$.

From items (a), (b) above and by structural induction we gather $\sigma'(t_m) \models \varphi_i$ for each $m \in M, i \in I$ with $\mathfrak{w}(t_m, \varphi_i) > 0$. Moreover, from $\sigma'(\zeta) \models \eta(\zeta)$ for each $\zeta \in \mathrm{var}(\Theta)$, item (a) above, we obtain that $D_\Sigma \vdash \sigma'(\mathcal{H})$, namely D_Σ proves the reduced instance w.r.t, σ' of each set of distribution premises $\{\zeta \xrightarrow{q_j} x_j \mid \sum_{j \in J} q_j = 1\} \in \mathcal{H}$. This fact taken together with item (c) above and Theorem 3.10 gives that D_Σ proves the reduced instance of $\{\Theta \xrightarrow{q_m} t_m \mid m \in M\}$ with respect to σ, that is $D_\Sigma \vdash \{\sigma(\Theta) \xrightarrow{q_h} t'_h \mid h \in H\}$ for a suitable set of indexes H and a proper set of closed terms t'_h such that for each $h \in H$ there is at least one $m \in M$ such that $t'_h = \sigma'(t_m)$ and moreover $q_h = \sum_{\{m \in M \mid \sigma'(t_m) = t'_h\}} q_m$ (Definition 3.2). In addition, by Theorem 3.4 it follows that $q_h = \sigma(\Theta)(t'_h)$ for each $h \in H$ and $\sum_{h \in H} q_h = 1$. Since moreover $q_h \in (0,1]$ for each $h \in H$, this is equivalent to say that $\sigma(\Theta) = \sum_{h \in H} q_h \delta_{\sigma'(t_h)}$. Finally, we notice that

$$\begin{aligned}
\sigma(\Theta) &= \sum_{h \in H} q_h \delta_{t'_h} \\
&= \sum_{h \in H} \Big(\sum_{\{m \in M \mid \sigma'(t_m) = t'_h\}} q_m \Big) \delta_{t'_h} \\
&= \sum_{m \in M} q_m \delta_{\sigma'(t_m)} && (t'_h \text{ pairwise distinct}) \\
&= \sum_{m \in M} \Big(\sum_{i \in I} \mathfrak{w}(t_m, \varphi_i) \Big) \delta_{\sigma'(t_m)} && \Big(\sum_{i \in I} \mathfrak{w}(t_m, \varphi_i) = q_m\Big) \\
&= \sum_{i \in I} \Big(\sum_{m \in M} \mathfrak{w}(t_m, \varphi_i) \delta_{\sigma'(t_m)} \Big) \\
&= \sum_{i \in I} \Big(\sum_{m \in M} r_i \frac{\mathfrak{w}(t_m, \varphi_i)}{r_i} \delta_{\sigma'(t_m)} \Big) \\
&= \sum_{i \in I} r_i \Big(\sum_{m \in M} \frac{\mathfrak{w}(t_m, \varphi_i)}{r_i} \delta_{\sigma'(t_m)} \Big)
\end{aligned}$$

$$= \sum_{i \in I} r_i \pi_i$$

where for each $i \in I$, $\pi_i = \sum_{m \in M} \frac{\mathfrak{w}(t_m,\varphi_i)}{r_i} \delta_{\sigma'(t_m)}$ is a probability distribution as it is obtained as a convex combination of probability distributions ($\sum_{m \in M} \frac{\mathfrak{w}(t_m,\varphi_i)}{r_i} = 1$). Moreover, the π_i are such that whenever $\sigma'(t) \in \mathsf{supp}(\pi_i)$ it holds that $\sigma'(t) \models \varphi_i$. In fact, we have that whenever $\mathfrak{w}(t_m,\varphi_i) > 0$, then the only closed term in the support of $\delta_{\sigma'(t_m)}$ is indeed $\sigma'(t_m)$. Furthermore, whenever $\sigma'(t_m) \not\models \varphi_i$ we are granted that $\mathfrak{w}(t_m,\varphi_i) = 0$, thus giving that $\sigma'(t_m)$ is not in the support of π_i. Therefore, we can conclude that $\sigma(\Theta) \models \bigoplus_{i \in I} r_i \varphi_i$ as requested.

Hence, Equation (3.9) follows from the two implications.

★ Inductive step $\phi = \langle a \rangle \psi$ for some $\psi \in \mathcal{L}^{\mathrm{d}}$ and $a \in \mathcal{A}$. Notice that in this case we have $\phi \in \mathcal{L}^{\mathrm{s}}$ and therefore we need to show Equation (3.8). To this aim, we prove the two implications separately.

(⇒) Assume first that $\sigma(t) \models \langle a \rangle \psi$. Then, by definition of relation $\models$ (Definition 2.22), there exists a probability distribution $\pi \in \Delta(\mathcal{T}(\Sigma))$ with $P \vdash \sigma(t) \xrightarrow{a} \pi$ and $\pi \models \psi$. By Theorem 3.6, $P \vdash \sigma(t) \xrightarrow{a} \pi$ implies that there are a P-ruloid $\frac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a closed substitution σ' with $P \vdash \sigma'(\mathcal{H})$, $\sigma'(t) = \sigma(t)$ and $\sigma'(\Theta) = \pi$. We infer the following facts:

1. from $\sigma'(t) = \sigma(t)$ we obtain that $\sigma'(x) = \sigma(x)$ for all $x \in \mathrm{var}(t)$;
2. from $\sigma'(\Theta) = \pi$ and $\pi \models \psi$, we gather $\sigma'(\Theta) \models \psi$ and by the inductive hypothesis we obtain that there exists a $\eta \in \Theta^{-1}(\psi)$ such that $\sigma'(\zeta) \models \eta(\zeta)$ for all $\zeta \in \mathrm{var}(\Theta)$;
3. from $P \vdash \sigma'(\mathcal{H})$ we obtain that whenever $x \xrightarrow{b} \mu \in \mathcal{H}$ we have $P \vdash \sigma'(x) \xrightarrow{b} \sigma'(\mu)$. Then, if $\mu \in \mathrm{var}(\Theta)$, by previous item (2), we get $\sigma'(\mu) \models \eta(\mu)$. Otherwise, if $\mu \not\in \mathrm{var}(\Theta)$, we have $\eta(\mu) = \top$ thus giving $\sigma'(\mu) \models \eta(\mu)$ also in this case. Hence, $\sigma'(\mu) \models \eta(\mu)$ and $\sigma'(x) \models \langle b \rangle \eta(\mu)$ in all cases.
4. from $P \vdash \sigma'(\mathcal{H})$ we obtain that whenever $x \overset{c}{\not\rightarrow} \in \mathcal{H}$ we have $P \vdash \sigma'(x) \overset{c}{\not\rightarrow}$, namely $P \not\vdash \sigma'(x) \xrightarrow{c} v$ for any $v \in \mathcal{DT}(\Sigma)$, giving $\sigma'(x) \models \neg\langle c \rangle \top$.

Let $\xi \in t^{-1}(\langle a \rangle \psi)$ be defined as in Definition 3.8.4 by means of the P-ruloid $\frac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and the decomposition mapping η introduced in item (2) above. We aim to show that for this ξ it holds that $\sigma'(x) \models \xi(x)$ for each $x \in \mathrm{var}(t)$. By construction,

$$\xi(x) = \bigwedge_{x \xrightarrow{b} \mu \in \mathcal{H}} \langle b \rangle \eta(\mu) \wedge \bigwedge_{x \overset{c}{\not\rightarrow} \in \mathcal{H}} \neg\langle c \rangle \top \wedge \eta(x).$$

By item (3) above we have $\sigma'(x) \models \langle b \rangle \eta(\mu)$ for each $x \xrightarrow{b} \mu \in \mathcal{H}$. By item (4) above we have $\sigma'(x) \models \neg\langle c \rangle \top$ for each $x \overset{c}{\not\rightarrow} \in \mathcal{H}$. Finally, if $x \in \mathrm{var}(\Theta)$ by item (2) above we get $\sigma'(x) \models \eta(x)$. If $x \not\in \mathrm{var}(\Theta)$ then we have $\eta(x) = \top$ (Definition 3.8.6b) thus giving $\sigma'(x) \models \eta(x)$ also in this case. Hence, $\sigma'(x) \models \eta(x)$ in all cases. Thus, we can conclude that $\sigma'(x) \models \xi(x)$. Since, by item (1) above, $\sigma(x) = \sigma'(x)$ we can conclude that $\sigma(x) \models \xi(x)$ as required.

($\Leftarrow$) Assume now that there is a $\xi \in t^{-1}(\langle a \rangle \psi)$ such that $\sigma(x) \models \xi(x)$ for all $x \in \mathrm{var}(t)$. Following Definition 3.8.4, we construct ξ in terms of some P-ruloid $\frac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and decomposition mapping $\eta \in \Theta^{-1}(\psi)$. In particular, we have that for each $x \in \mathrm{var}(t)$

$$\xi(x) = \bigwedge_{x \xrightarrow{b} \mu \in \mathcal{H}} \langle b \rangle \eta(\mu) \wedge \bigwedge_{x \overset{c}{\not\to} \in \mathcal{H}} \neg \langle c \rangle \top \wedge \eta(x).$$

We define a closed substitution σ' such that the following properties hold:

(a) $\sigma'(x) = \sigma(x)$ for all $x \in \mathrm{var}(t)$. As a consequence, from $\sigma(x) \models \xi(x)$ we derive $\sigma'(x) \models \xi(x)$.

(b) As $\sigma'(x) \models \xi(x)$, by previous item (a), we derive that $\sigma'(x) \models \langle b \rangle \eta(\mu)$ for each $x \xrightarrow{b} \mu \in \mathcal{H}$. This implies that for each positive premise in $\mathcal{H}$ there exists a probability distribution $\pi_{b,\mu}$ such that $P \vdash \sigma'(x) \xrightarrow{b} \pi_{b,\mu}$ and $\pi_{b,\mu} \models \eta(\mu)$. We define $\sigma'(\mu) = \pi_{b,\mu}$ thus obtaining that for each $x \xrightarrow{b} \mu \in \mathcal{H}$ we have $P \vdash \sigma'(x) \xrightarrow{b} \sigma'(\mu)$ and $\sigma'(\mu) \models \eta(\mu)$.

(c) As $\sigma'(x) \models \xi(x)$, by previous item (a), we derive that $\sigma'(x) \models \neg \langle c \rangle \top$ for each $x \overset{c}{\not\to} \in \mathcal{H}$. Therefore, we obtain that $P \vdash \sigma'(x) \overset{c}{\not\to}$ for each $x \overset{c}{\not\to} \in \mathcal{H}$.

(d) Since $\mathrm{var}(\Theta) \subseteq \mathrm{var}(t) \cup \mathrm{rhs}(\mathcal{H})$, previous items (b) and (c) we obtain that $\sigma'(\mu) \models \eta(\mu)$ for each $\mu \in \mathcal{V}_d$.

(e) $\sigma'(x) \models \eta(x)$ for each $x \in \mathrm{var}(\Theta)$.

From items (d), (e) and structural induction, we gather $\sigma'(\Theta) \models \psi$. Moreover, items (b) and (c) give $P \vdash \sigma'(\mathcal{H})$. Hence, by Theorem 3.6 we obtain $P \vdash \sigma'(t) \xrightarrow{a} \sigma'(\Theta)$. From item (a) we have that $\sigma'(t) = \sigma(t)$ and, therefore, we can conclude that $\sigma(t) \models \langle a \rangle \psi$.

Hence, Equation (3.8) follows from the two implications.

Finally, let us deal with terms that are not univariate.

Assume first that t is not univariate, namely $t = \varsigma(s)$ for some univariate s and non-injective substitution $\varsigma \colon \mathrm{var}(s) \to \mathcal{V}_s$. Then, $\sigma(\varsigma(s)) \models \varphi$ iff there exists a decomposition mapping $\xi' \in s^{-1}(\varphi)$ such that $\sigma(\varsigma(y)) \models \xi'(y)$, which by Definition 3.8.5 is equivalent to require that there exists a decomposition mapping $\xi' \in s^{-1}(\varphi)$ such that for each $x \in \mathrm{var}(t)$ we have $\sigma(x) \models \bigwedge_{y \in \varsigma^{-1}(x)} \xi'(y)$. By defining the decomposition mapping $\xi \in t^{-1}(\varphi)$ as $\xi(x) = \bigwedge_{y \in \varsigma^{-1}(x)} \xi'(y)$, we obtain the thesis.

Assume now that Θ is not univariate, namely $\Theta = \varsigma(\Theta_1)$ for some univariate Θ_1 and non-injective substitution $\varsigma \colon \mathrm{var}(\Theta_1) \to \mathcal{V}_d \cup \delta_{\mathcal{V}_s}$. Then, $\sigma(\varsigma(\Theta_1)) \models \psi$ iff there exists a decomposition function $\eta_1 \in \Theta_1^{-1}(\psi)$ such that $\sigma(\varsigma(z)) \models \eta_1(z)$, which by Definition 3.8.7 is equivalent to require that there exists a decomposition mapping $\eta' \in \Theta_1^{-1}(\psi)$ such that for each $\zeta \in \mathrm{var}(\Theta)$ we have $\eta'(z) = \eta'(z')$ for all $z, z' \in \varsigma^{-1}(\zeta)$ and, for a chosen $\tilde{z} \in \varsigma^{-1}(\zeta)$, $\sigma(\zeta) \models \eta'(\tilde{z})$. By defining the decomposition mapping $\eta \in \Theta^{-1}(\psi)$ as $\eta(\zeta) = \eta'(\tilde{z})$, for $\tilde{z} \in \varsigma^{-1}(\zeta)$, we obtain the thesis. ■

Decomposition of $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$

The decomposition of formulae in $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$ can be derived from the one for $\mathcal{L}$.

Definition 3.9 (Decomposition of $\mathcal{L}_{\mathrm{r}}$ and $\mathcal{L}_{+}$)**.** Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS and D_Σ be the Σ-DS. The mappings $\cdot^{-1} \colon \mathbb{T}(\Sigma) \to (\mathcal{L}_{\mathrm{r}}^{\mathrm{s}} \to \mathcal{P}(\mathcal{V}_s \to \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}))$ and $\cdot^{-1} \colon \mathbb{DT}(\Sigma) \to (\mathcal{L}_{\mathrm{r}}^{\mathrm{d}} \to \mathcal{P}(\mathcal{V} \to \mathcal{L}_{\mathrm{r}}))$ are obtained as in Definition 3.8 by rewriting Definition 3.8.2 and Definition 3.8.4, respectively, by

2'. $\xi \in t^{-1}(\bar{a})$ iff there is a function $\mathfrak{f} \colon t^{-1}(\langle a \rangle \top) \to \mathrm{var}(t)$ such that

$$\xi(x) = \begin{cases} \displaystyle\bigwedge_{\xi' \in \mathfrak{f}^{-1}(x)} \neg \xi'(x) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise;} \end{cases}$$

4'. $\xi \in t^{-1}(\langle a \rangle \psi)$ iff there are a ruloid $\dfrac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a decomposition mapping $\eta \in \Theta^{-1}(\psi)$ such that

$$\xi(x) = \begin{cases} \displaystyle\bigwedge_{x \xrightarrow{b} \mu \in \mathcal{H}} \langle b \rangle \eta(\mu) \wedge \bigwedge_{x \overset{c}{\nrightarrow} \in \mathcal{H}} \bar{c} \wedge \eta(x) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise.} \end{cases}$$

If P is positive, the mappings $\cdot^{-1} \colon \mathbb{T}(\Sigma) \to (\mathcal{L}_{+}^{\mathrm{s}} \to \mathcal{P}(\mathcal{V}_s \to \mathcal{L}_{+}^{\mathrm{s}}))$ and $\cdot^{-1} \colon \mathbb{DT}(\Sigma) \to (\mathcal{L}_{+}^{\mathrm{d}} \to \mathcal{P}(\mathcal{V} \to \mathcal{L}_{+}))$ are obtained as in Definition 3.8 by removing Definition 3.8.2 and by rewriting Definition 3.8.4 by

4''. $\xi \in t^{-1}(\langle a \rangle \psi)$ iff there are a positive P-ruloid $\dfrac{\mathcal{H}}{t \xrightarrow{a} \Theta}$ and a decomposition mapping $\eta \in \Theta^{-1}(\psi)$ such that

$$\xi(x) = \begin{cases} \displaystyle\bigwedge_{x \xrightarrow{b} \mu \in \mathcal{H}} \langle b \rangle \eta(\mu) \wedge \eta(x) & \text{if } x \in \mathrm{var}(t) \\ \top & \text{otherwise.} \end{cases}$$

The following Lemma shows that our decomposition method preserves the syntactic restrictions of the considered modal classes, namely that by decomposing formulae in $\mathcal{L}_{\mathrm{r}}$ (resp. $\mathcal{L}_{+}$) we get formulae in $\mathcal{L}_{\mathrm{r}}$ (resp. $\mathcal{L}_{+}$) thus preserving the logical characterization of Theorem 2.11.

Lemma 3.13. *Let P be a PGSOS-PTSS and consider the term $t \in \mathbb{T}(\Sigma)$ and the formulae $\varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$, $\psi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{d}}$, $\varphi' \in \mathcal{L}_{+}^{\mathrm{s}}$ and $\psi' \in \mathcal{L}_{+}^{\mathrm{d}}$.*

1. ⋆ *For all $x \in \mathcal{V}_s$ we have $\xi(x) \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$ for each $\xi \in t^{-1}(\varphi)$.*

⋆ *For all $\zeta \in \mathcal{V}_d$ we have $\eta(\zeta) \in \mathcal{L}_{\mathrm{r}}^{\mathrm{d}}$ for each $\eta \in \Theta^{-1}(\psi)$.*

- ★ *For all $\zeta \in \mathcal{V}_s$ we have $\eta(\zeta) \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$ for each $\eta \in \Theta^{-1}(\psi)$.*

2. *If P is positive, then*

   - ★ *For all $x \in \mathcal{V}_s$ we have $\xi(x) \in \mathcal{L}_{+}^{\mathrm{s}}$ for each $\xi \in t^{-1}(\varphi')$.*
   - ★ *For all $\zeta \in \mathcal{V}_d$ we have $\eta(\zeta) \in \mathcal{L}_{+}^{\mathrm{d}}$ for each $\eta \in \Theta^{-1}(\psi')$.*
   - ★ *For all $\zeta \in \mathcal{V}_s$ we have $\eta(\zeta) \in \mathcal{L}_{+}^{\mathrm{s}}$ for each $\eta \in \Theta^{-1}(\psi')$.*

Proof. The proofs of items (1) and (2) follow immediately from Definition 3.9. ■

Theorem 3.14 (Decomposition theorem II)**.** *Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS and D_Σ be the Σ-DS. Assume the decomposition mappings as in Definition 3.9. Then:*

- ★ *The results in Theorem 3.12 hold for $\varphi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{s}}$ and $\psi \in \mathcal{L}_{\mathrm{r}}^{\mathrm{d}}$.*

- ★ *Moreover, if P is positive, then the results in Theorem 3.12 hold for $\varphi \in \mathcal{L}_{+}^{\mathrm{s}}$ and $\psi \in \mathcal{L}_{+}^{\mathrm{d}}$.*

Proof. The proof of both items can be obtained by following the one of Theorem 3.12 with respect to the decompositions of the two logics (Definition 3.9). In particular, we remark that in the proof for the diamond modality in $\mathcal{L}_{+}$, we use Corollary 3.7 in place of Theorem 3.6. ■

3.4 Congruence theorems

To support the compositional reasoning, the congruence (resp. precongruence) property is required for any behavioral equivalence (resp. preorder) $\mathcal{R}$. It consists in verifying whether $f(t_1,\dots,t_{\mathfrak{n}})\ \mathcal{R}\ f(t_1',\dots,t_{\mathfrak{n}}')$ whenever $t_i\,\mathcal{R}\,t_i'$ for $i = 1,\dots,\mathfrak{n}$. In [54] it is proved that probabilistic bisimilarity is a congruence for all operators defined by a PGSOS-PTSS. We can restate this result as a direct consequence of the characterization result of [66] (Theorem 2.10) combined with our first decomposition result in Theorem 3.12 schematized in Figure 3.1. Then, by our characterization results in Theorem 2.11 and our decomposition results in Theorem 3.14 we can derive precongruence formats for both ready similarity and similarity.

Theorem 3.15. *Let $P = (\Sigma, \mathcal{A}, R)$ be a PGSOS-PTSS. Then:*

1. *Probabilistic bisimilarity is a congruence for all operators defined by P;*

2. *Probabilistic ready similarity is a precongruence for all operators defined by P;*

3. *If P is positive, probabilistic similarity is a precongruence for all operators defined by P.*

Proof.

1. Let $t \in \mathbb{T}(\Sigma)$ and let σ, σ' be two closed substitutions. We aim to show that

$$\text{whenever } \sigma(x) \sim \sigma'(x) \text{ for each } x \in \mathrm{var}(t) \text{ then it holds that } \sigma(t) \sim \sigma'(t). \qquad (3.10)$$

Considering the characterization result of $\mathcal{L}$ for probabilistic bisimilarity (Theorem 2.10), to prove Equation (3.10) we simply have to show that $\sigma(t)$ and $\sigma'(t)$ satisfy the same formulae in $\mathcal{L}$. Assume that $\sigma(t) \models \varphi$, for some state formula $\varphi \in \mathcal{L}$. By Theorem 3.12, there is a decomposition mapping $\xi \in t^{-1}(\varphi)$ such that $\sigma(x) \models \xi(x)$ for each $x \in \mathrm{var}(t)$. From Lemma 3.11 we gather that $\xi(x) \in \mathcal{L}^{\mathrm{s}}$ and moreover by Theorem 2.10 from $\sigma(x) \sim \sigma'(x)$ we obtain that $\sigma'(x) \models \xi(x)$ for each $x \in \mathrm{var}(t)$. By applying Theorem 3.12 once again, we obtain that $\sigma'(t) \models \varphi$, thus proving Equation (3.10).

2. The proof for probabilistic ready simulation is analogous to the one for item 1 by exploiting Theorem 2.11.1 in place of Theorem 2.10, Theorem 3.14.1 in place of Theorem 3.12 and Lemma 3.13.1 in place of Lemma 3.11.

3. Under the assumption of P positive, the proof for probabilistic simulation is analogous to the one for item 1 by exploiting Theorem 2.11.2 in place of Theorem 2.10, Theorem 3.14.2 in place of Theorem 3.12 and Lemma 3.13.2 in place of Lemma 3.11.

■

3.5 Generalization

We have proposed a decomposition method that allowed us to derive congruence formats for probabilistic strong (bi)similarities directly from their modal characterizations. Due to the presence of probabilistic choice modalities in the characterizing classes of formulae, the decomposition was made possible by the introduction of an SOS-like machinery for the specification of the behavior of distribution terms. We claim that our method can be extended to the other semantics in the probabilistic strong and weak linear time - branching time spectra (see Chapter 4 for a presentation of the strong spectrum) thus obtaining a class of '*Probabilistic Divide and Congruence*' results in the line of [33, 80, 82, 84, 85], proposed in the fully-nondeterministic setting. To give an intuition on how this can actually be done, in this Section we sketch the reasoning that would lead to the definition of a congruence format for probabilistic (rooted) branching bisimilarity starting from specifications in the PGSOS format. As we will outline, this can be obtained by combining the decomposition method proposed in this Chapter with the format for (rooted) branching bisimilarity, the RBB format, defined in [84]. We recall that the RBB format is built on the predicates Λ and $\aleph$, where the former marks running processes, namely processes that have already started their execution, and the latter marks the ones that can start their execution immediately. These predicates are a refined version of the *tame/wild* labeling of operators from [34, 81].

Decomposition method

The first step in the definition of a decomposition method is to identify the class of modal formulae characterizing probabilistic branching bisimilarity and its rooted version. It's not difficult to figure out that this class can be obtained by extending the modal logic characterizing these equivalences in the fully-nondeterministic case with the probabilistic choice modality $\bigoplus$. For simplicity let $\mathbb{L}$ be such a class of formulae. Then the decomposition of formulae in $\mathbb{L}$ is then defined on our (Σ-distribution) ruloids by combining the decomposition in [84] with the one we have proposed for distribution formulae. We stress that our notions of ruloids and Σ-distribution ruloids would not need to be changed to obtain the decomposition theorem.

Format

Interestingly, the RBB format proposed in [84] holds even in our probabilistic setting. More precisely, the syntactic constraints imposed by the format are enough to obtain, in combination the decomposition method described above, that probabilistic (rooted) branching bisimilarity is a congruence with respect to all operators defined by PGSOS rules satisfying them. This, however, should not be surprising. First of all we notice that both the RBB format and the PGSOS format do not allow look-ahead. In particular, in the case of the PGSOS format, the impossibility of testing for two consecutive moves of a process implies that probability is never involved in the derivation of nondeterministic transitions. Equivalently, we are guaranteed that Σ-distribution rules (and ruloids) are never used to determine the provability of a closed literal. Moreover, the constraints on the probability weights in the definition of behavioral relations do not depend on the syntactical definition of processes and thus they are independent from the constraints of the rule format. Therefore, to obtain our *probabilistic RBB format* we simply need to lift the definition of the predicates $\aleph$ and Λ on arguments of operators to the arguments of distribution terms and impose on PGSOS rules the same constraints of the RBB format (see [84, Definition 14]).

The congruence theorem

Once we have the decomposition method and the probabilistic RBB format, we can simply proceed in the classic way to obtain the congruence result. Firstly, since the decomposition method is not defined in terms of PGSOS rules but of ruloids, we need to guarantee that the syntactic constraints imposed by the PRBB format are preserved in the construction of ruloids from PGSOS rules fitting the format. Secondly, to ensure that the decomposition of formulae in a chosen class preserves the syntactic restrictions of that class, and thus the logical characterization, we need to show that a formula in $\mathbb{L}$ is decomposed into a formula in $\mathbb{L}$ or at least to a formula equivalent to a formula in $\mathbb{L}$. Finally, by applying the same reasoning schematized in Figure 3.1 we can conclude that probabilistic (rooted) branching bisimilarity in a congruence with respect to all operators defined by a PGSOS PTSS in probabilistic RBB format.

3.6 Concluding remarks

In this Chapter we developed a modal decomposition of formulae in $\mathcal{L}$ and its subclasses $\mathcal{L}_{\mathrm{r}}, \mathcal{L}_{+}$ presented in Chapter 2.4 as adequate logics for, respectively, probabilistic bisimilarity, ready similarity and similarity. Our decomposition method is novel with respect to the ones existing in the literature (see for instance [33, 80, 82–85, 90]) as it is based on the structural operational semantics of nondeterministic probabilistic processes in the PTS model.

The dual nature of these processes, and of the classes of formulae characterizing them, enforced the introduction of an SOS framework tailored for the specification of distribution terms, namely the Σ-*distribution specification* in which we have syntactically represented open distribution terms as probability distributions over open terms. Moreover, the Σ-*distribution ruloids*, built from it, provide a general tool that can be used to support the decomposition of any logic with modalities specifying quantitative properties for the PTS model and they can be easily adapted to models admitting subdistributions [115, 126, 127].

To prove the robustness of our decomposition method we have showed how the congruence theorems for probabilistic bisimilarity, ready similarity and similarity with respect to the PGSOS format can be restated as an application of our decomposition theorems. Moreover, we sketched how our method can be generalized to derive congruence formats for other relations in the probabilistic strong and weak spectra.

To the best of our knowledge, [90] is the only other paper dealing with ruloids for the specification of probabilistic process calculi. As previously outlined, [90] deals with reactive transition systems, which are less expressive than PTSs as they do not admit internal nondeterminism. Transitions are of the form $t \xrightarrow{a,p} t'$, denoting that t evolves by a to t' with probability p. Informally, our P-ruloids generalize those in [90] in the same way PTSSs generalize reactive systems. In fact, to deal with the quadruple $t \xrightarrow{a,p} t'$, ruloids in [90] are defined by keeping track of rules and ruloids used in their construction, in order to assign a proper probability weight to their conclusion. In detail, to guarantee the property of semi-stochasticity, stating that the sum of the probabilities of all transitions for an action from a term is either 0 or 1, a partitioning over ruloids is needed in [90]: given a term t the ruloids in the partition for t related to action a allow one to derive a-labeled transitions from t whose total probability is 1. To do so, one also has to constantly keep track of the rules and ruloids used in the construction of the ruloids in a partition, because the exact probability weight of a transition depends on this construction. This technical expedient was introduced in [117], in which the SOS framework on which [90] builds was defined.

Here we do not need this technicality, since probabilities are directly managed by Σ-distribution ruloids and we can use P-ruloids to derive the transitions leading to probability distributions. More precisely, we should say that given a term t, all ruloids in one partition for t of [90] are captured by one of our P-ruloids and one Σ-distribution ruloid. The P-ruloid captures all the requirements that the subterms of t must satisfy to derive the transition to the desired probability distribution over terms. The proper probability weights are then automatically assigned to terms by the Σ-distribution ruloid, without necessity of keeping track of all the rules and ruloids used in the construction.

'Is there any point to which you would wish to draw my attention?'
'To the curious incident of the dog in the night-time.'
'The dog did nothing in the night-time.'
'That was the curious incident,' remarked Sherlock Holmes.

Sir Arthur Conan Doyle,
Silver Blaze

CHAPTER 4

A Quantitative Spectrum for Nondeterministic Probabilistic Processes

With this Chapter we start our studies on behavioral metrics for nondeterministic probabilistic processes. In particular, we propose a quantitative analogue to ready similarity and similarity and moreover we introduce novel distances measuring the disparities of processes with respect to the testing and (decorated) trace semantics.

The definition of (ready) similarity metric follows the quantitative characterization of bisimilarity: we identify a suitable functional expressing the differences related to probability and nondeterminism of processes that are relevant with respect to the considered semantics, and we define the desired metric as the least fixed point of this functional.

Indeed, to obtain proper behavioral distances for linear semantics, as those of (decorated) traces and testing, we will follow a different approach. Intuitively, these metrics should measure the differences in the probabilities that the processes assign to *semantic-specific events*, namely sequences of events aimed at capturing the considered semantics. For instance, we will consider sequences of actions for the trace semantics and sequences of actions leading to success for the testing semantics. In the literature we can find a wealth of behavioral equivalences and preorders for theses semantics, based on the class of schedulers chosen to resolve nondeterminism and on how the probabilities are compared [29–31, 51, 69, 95, 108, 109, 144, 146, 147, 164, 166]. Conversely, little has been studied of their quantitative analogues. We can find a few proposals for trace metrics [14, 43, 53, 59, 148], but no metric for testing and decorated traces has been proposed so far. One of the main contributions of this Chapter is to provide those metrics. To this purpose we consider the resolutions of nondeterminism given by *deterministic schedulers* [144] that select exactly one transition among the possible ones. Then, to compare the probabilities of semantic-specific events we will follow the *trace-by-trace* approach of [29]: each event is tested on all possible resolutions of nondeterminism for the two processes and we evaluate the difference between the best cases, namely between the suprema of the probabilities of performing the considered event. The distance between two processes is then given by the suprema of these differences with respect to all events. By means of this technique we obtain original metrics for decorated trace and testing semantics and a novel notion of trace metric with respect to to the literature. In fact, our trace metric generalizes the total variation distance used in [14, 53] on Markov Chains by capturing the interaction between nondeterminism and probability proper of PTSs. Moreover, differently from the distance in [43, 148], based on the *trace-distribution* approach of [144] and obtained by combining the Kantorovich and Hausdorff metrics, the trace metric proposed here induces an equivalence relation that is coarser than bisimilarity (see Section 4.5 for a detailed discussion on this issue).

In the nonprobabilistic case, [159] classified behavioral relations with respect to their discriminating power and mutual relationships, the idea being to help in determining the most suitable semantics for a given application and also to stress the similarities and differences among those semantics. Here we obtain a *quantitative* analogue to the *linear time-branching time spectrum* of [159] for the proposed behavioral metrics that will be classified with respect to their discriminating power on processes in the PTS model. More precisely, we consider (cf. upper part of Figure 4.1) the bisimilarity metric ($\mathbf{d}_\lambda$) of [64, 72, 157] together with the novel notions of: 1. *ready simulation hemimetric* ($\mathbf{d}_{\mathrm{r},\lambda}$); 2. *simulation hemimetric* ($\mathbf{d}_{\mathrm{s},\lambda}$); 3. *ready trace hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}$); 4. *ready trace metric* ($\mathbf{d}_{\mathrm{TrR},\lambda}$); 5. *readiness hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}$); 6. *readiness metric* ($\mathbf{d}_{\mathrm{R},\lambda}$); 7. *failure trace hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}$); 8. *failure trace metric* ($\mathbf{d}_{\mathrm{TrF},\lambda}$); 9. *failure hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}$); 10. *failure metric* ($\mathbf{d}_{\mathrm{F},\lambda}$); 11. *completed trace hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}$); 12. *completed trace metric* ($\mathbf{d}_{\mathrm{TrC},\lambda}$); 13. *trace hemimetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$); 14. *trace metric* ($\mathbf{d}_{\mathrm{Tr},\lambda}$); 15. *testing premetric* ($\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}$); 16. *testing semimetric* ($\mathbf{d}_{\mathrm{test},\lambda}$). Then we order these metrics by the relation '*makes processes farther than*', represented in the upper part of Figure 4.1. Here, a blue arrow $\mathbf{d} \to \mathbf{d}'$ means that $\mathbf{d}(s,t) \geq \mathbf{d}'(s,t)$ for all processes s,t and, moreover, there are processes s,t for which $\mathbf{d}(s,t) > \mathbf{d}'(s,t)$. As far as we know, this is the first proposal of a *quantitative spectrum* on the PTS model and, moreover, it comes with the first definition of metrics capturing the probabilistic testing, decorated traces and ready simulation semantics, and a novel notion of trace metric.

Another interesting feature of our metrics is in that the equivalences and preorders induced by them, namely their kernels, satisfy a lot of desirable properties. In Figure 4.1 red dotted arrows connect each distance, on the upper side, with its kernel, on the lower side, namely the equivalence or preorder that relates precisely the processes at distance 0. In detail, we have probabilistic bisimulation equivalence ($\sim$) and the kernels of our behavioral distances: 1. *ready simulation* preorder ($\sqsubseteq_{\mathrm{r}}$); 2. *simulation* preorder ($\sqsubseteq$); 3. *ready trace preorder* ($\sqsubseteq_{\mathrm{TrR}}$); 4. *ready trace equivalence* ($\sim_{\mathrm{TrR}}$); 5. *readiness preorder* ($\sqsubseteq_{\mathrm{R}}$); 6. *readiness equivalence* ($\sim_{\mathrm{R}}$); 7. *failure trace preorder* ($\sqsubseteq_{\mathrm{TrF}}$); 8. *failure trace equivalence* ($\sim_{\mathrm{TrF}}$); 9. *failure preorder* ($\sqsubseteq_{\mathrm{F}}$); 10. *failure equivalence* ($\sim_{\mathrm{F}}$); 11. *completed trace preorder* ($\sqsubseteq_{\mathrm{TrC}}$); 12. *completed trace equivalence* ($\sim_{\mathrm{TrC}}$); 13. *trace preorder* ($\sqsubseteq_{\mathrm{Tr}}$); 14. *trace equivalence* ($\sim_{\mathrm{Tr}}$); 15. *testing preorder* ($\sqsubseteq_{\mathrm{test}}$); 16. *testing equivalence* ($\sim_{\mathrm{test}}$). Interestingly, the spectrum on metrics in the upper part of Figure 4.1 together with the kernel properties of these relations, ensure that each black arrow on the lower side takes a relation to a larger one, giving a *spectrum of probabilistic equivalences and preorders* with respect to the relation '*makes strictly less identification than*' which is perfectly consistent with the spectrum on metrics. The relations obtained from the kernels of our metrics are a slightly coarser version of the ones

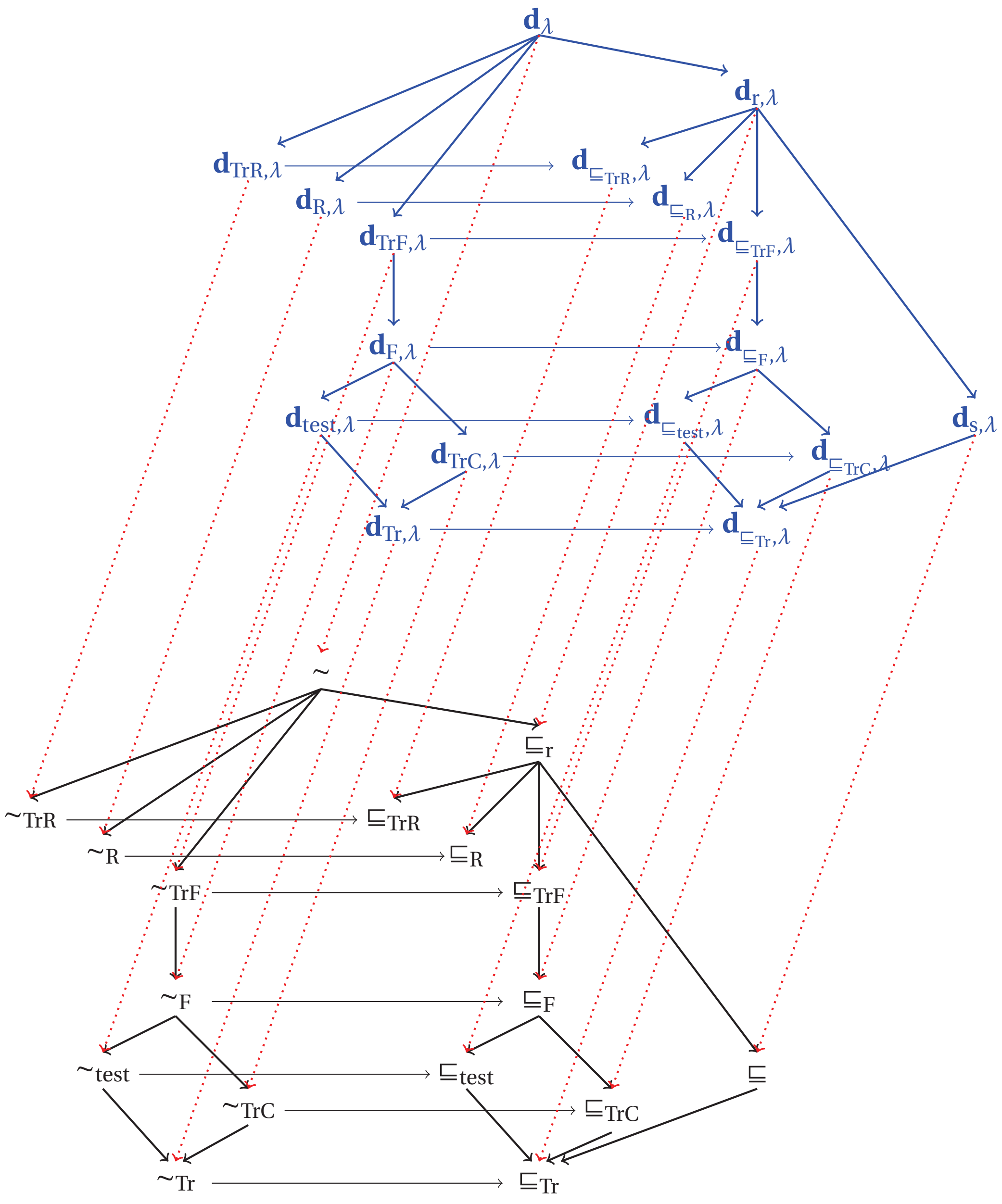

$\mathbf{d}_{\lambda}$
$\mathbf{d}_{\mathrm{r},\lambda}$
$\mathbf{d}_{\mathrm{TrR},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}$
$\mathbf{d}_{\mathrm{R},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}$
$\mathbf{d}_{\mathrm{TrF},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}$
$\mathbf{d}_{\mathrm{F},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}$
$\mathbf{d}_{\mathrm{test},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}$
$\mathbf{d}_{\mathrm{s},\lambda}$
$\mathbf{d}_{\mathrm{TrC},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}$
$\mathbf{d}_{\mathrm{Tr},\lambda}$
$\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$
$\sim$
$\sqsubseteq_{\mathrm{r}}$
$\sim_{\mathrm{TrR}}$
$\sqsubseteq_{\mathrm{TrR}}$
$\sim_{\mathrm{R}}$
$\sqsubseteq_{\mathrm{R}}$
$\sim_{\mathrm{TrF}}$
$\sqsubseteq_{\mathrm{TrF}}$
$\sim_{\mathrm{F}}$
$\sqsubseteq_{\mathrm{F}}$
$\sim_{\mathrm{test}}$
$\sqsubseteq_{\mathrm{test}}$
$\sqsubseteq$
$\sim_{\mathrm{TrC}}$
$\sqsubseteq_{\mathrm{TrC}}$
$\sim_{\mathrm{Tr}}$
$\sqsubseteq_{\mathrm{Tr}}$

Figure 4.1: *The spectrum of metrics (top) and the spectrum of probabilistic relations (bottom). An arrow $d \to d'$ between two distances (top) stands for $d(s,t) \geq d'(s,t)$ for all processes s, t, and $d(s,t) > d'(s,t)$ for some processes s, t. An arrow $\mathcal{R} \to \mathcal{R}'$ between two relations (bottom) stands for $\mathcal{R} \subset \mathcal{R}'$. A dotted arrow $d \dashrightarrow \mathcal{R}$ between a distance d and a relation $\mathcal{R}$ means that $\mathcal{R}$ is the kernel of d.*

proposed in the spectrum in [29,30] and thus they share some important properties with them along with some new important feature. We will show that our probabilistic relations satisfy: (i) compositionality; (ii) full backward compatibility with the fully-nondeterministic case; (iii) full backward compatibility with the fully-probabilistic case; (iv) they are all coarser than bisimilarity.

Organization of contents

In Section 4.1 we briefly recall some well know notions on traces and we justify our choice of dealing with deterministic schedulers. Then we proceed to define the behavioral metrics: the ones for ready similarity and similarity in Section 4.2 and those expressing (decorated) trace and testing semantics in Section 4.3. These metric are then ordered in the spectrum presented in Section 4.4. In Section 4.5 we study the kernels of the metrics introduced in Section 4.3 and the relations so obtained are then ordered with (bi)similarities in the spectrum in Section 4.6. We conclude discussing related work in Section 4.7.

4.1 Preliminary notions

We delay the discussion of our results to recall first some basic notions necessary to reason about the (decorated) trace and testing semantics. As the main term of comparison for our results is the work in [29–31], we decided to keep our notation as much closer as possible to theirs.

We start with the notion of *computation* which expresses a weighted sequence a process-to-process action-labeled transitions for processes in a PTS.

Definition 4.1 (Computation)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS and $s, s' \in \mathcal{S}$. We say that

$$c := s_0 \xrightarrow{a_1}\!\!\!\!\rightarrow s_1 \xrightarrow{a_2}\!\!\!\!\rightarrow s_2 \dots s_{n-1} \xrightarrow{a_n}\!\!\!\!\rightarrow s_n$$

is a *computation from* $s = s_0$ *to* $s' = s_n$, notation $\mathrm{first}(c) = s_0$ and $\mathrm{last}(c) = s_n$, if and only if for all $i = 1, \dots, n$ there exists a transition $s_{i-1} \xrightarrow{a_i} \pi_i$ such that $s_i \in \mathsf{supp}(\pi_i)$.

Note that $\pi_i(s_i)$ is the *execution probability* of step $s_{i-1} \overset{a_i}{\twoheadrightarrow} s_i$ conditioned on the selection of the transition $s_{i-1} \xrightarrow{a_i} \pi_i$ at s_{i-1}. We denote by $\Pr(c) = \prod_{i=1}^{n} \pi_i(s_i)$ the product of the execution probabilities of the steps in c.

We say that c is a *computation from* s if c is a computation from s to some process s'. Then, c 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. Given 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.

We denote by $\mathcal{A}^\star$ the set of *finite traces* in $\mathcal{A}$ and we denote the *empty trace* with the special symbol $\mathfrak{e}$. We say that a computation is *compatible* with the trace $\alpha \in \mathcal{A}^\star$ if and only 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 of s which are compatible with α, and by $\mathcal{C}_{\max}(s, \alpha)$ the set $\mathcal{C}_{\max}(s, \alpha) = \mathcal{C}_{\max}(s) \cap \mathcal{C}(s, \alpha)$.

As one can expect, to capture the (decorated) trace semantics we will need to evaluate and compare the probability of a particular sequence of *events* to occur. However, in the PTS model, this probability highly depends also on internal nondeterministic choices. For this reason a fundamental decision we need to make is in the choice of *schedulers* (or *adversaries*), namely the functions resolving the nondeterminism for processes. In the literature we can find several proposals for schedulers (see [95, 144, 164] and the references therein), but we can regroup them into two main classes: *deterministic schedulers* and *randomized schedulers* [144]. We say that a scheduler is deterministic if for each process it selects exactly one transition among the possible ones or none of them. Thus, internal nondeterministic choices are always treated as distinct by this class of schedulers. Conversely, randomized schedulers allow for a convex combination of the equally labeled transitions. Exemplifying, if we consider a PTS such that $s \xrightarrow{a} \pi_1$ and $s \xrightarrow{a} \pi_2$ are both valid transitions, then a randomized scheduler can assign to s a transition $s \xrightarrow{a} \pi$ with $\pi = p\pi_1 + (1-p)\pi_2$ for any value of $p \in [0,1]$. Clearly, each resolution of nondeterminism induced by a deterministic scheduler can be also induced by a randomized one. Still, the result of the interaction of a process with a deterministic scheduler is a fully probabilistic process (as formalized in Definition 4.2 below), whereas when randomized schedulers are involved we obtain a fully probabilistic process with *combined transitions* [146]. Considering that the main purpose of this Chapter is to introduce novel notions of behavioral metrics and to study the relations among them, we decided to consider the resolutions of nondeterminism induced by deterministic schedulers. In this way, we can reason on classic PTSs and the equivalences and metrics on them, like the (bi)simulations in Definition 2.16 and the bisimilarity metric in Definition 2.19. We leave as future work the investigation of a spectrum of metrics and relations on processes with combined transitions, on which randomized schedulers can be naturally applied.

Definition 4.2 (Resolution)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS and $s \in \mathcal{S}$. We say that a PTS $\mathcal{Z} = (Z, \mathcal{A}, \to_{\mathcal{Z}})$ is a *resolution* for s if and only if there exists a state correspondence function $\mathrm{corr}_{\mathcal{Z}} \colon Z \to \mathcal{S}$ such that $s = \mathrm{corr}_{\mathcal{Z}}(z_s)$ for some $z_s \in Z$, called the *initial state* of $\mathcal{Z}$, and moreover it holds that:

1. $z_s \not\in \mathsf{supp}(\pi)$ for any $\pi \in \bigcup_{z \in Z, a \in \mathcal{A}} \mathrm{der}(z,a)$.
2. Each $z \in Z \setminus \{z_s\}$ is such that $z \in \mathsf{supp}(\pi)$ for some $\pi \in \bigcup_{z' \in Z \setminus \{z\}, a \in \mathcal{A}} \mathrm{der}(z',a)$.
3. Whenever $z \xrightarrow{a}_{\mathcal{Z}} \pi$, then $\mathrm{corr}_{\mathcal{Z}}(z) \xrightarrow{a} \pi'$ with $\pi(z') = \pi'(\mathrm{corr}_{\mathcal{Z}}(z'))$ for all $z' \in Z$.
4. Whenever $z \xrightarrow{a_1}_{\mathcal{Z}} \pi_1$ and $z \xrightarrow{a_2}_{\mathcal{Z}} \pi_2$ then $a_1 = a_2$ and $\pi_1 = \pi_2$.

Then, $\mathcal{Z}$ is *maximal* if and only if it cannot be further extended in accordance with the graph structure of P and the constraints above. We denote by $\mathrm{Res}(s)$ the set of resolutions for s and by $\mathrm{Res}_{\max}(s)$ the subset of maximal resolutions for s.

Finally, we recall a notion of CSP-like [106] fully synchronous parallel composition for PTSs.

Definition 4.3 (Parallel composition)**.** Let $P_1 = (S_1, \mathcal{A}, \to_1)$ and $P_2 = (S_2, \mathcal{A}, \to_2)$ be two PTSs. The *synchronous parallel composition of* P_1 *and* P_2 is the PTS $P_1 \parallel P_2 = (S_1 \times S_2, \mathcal{A}, \to)$, where

$\rightarrow \subseteq (S_1 \times S_2) \times \mathcal{A} \times \Delta(S_1 \times S_2)$ is such that $(s_1, s_2) \xrightarrow{a} \pi$ 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 $(s_1', s_2') \in S_1 \times S_2$.

Parallel composition of PTSs naturally induces the parallel composition of processes.

Finally, we recall the notion of *premetric* and *semimetric* which will be useful to reason about the quantitative testing semantics.

Definition 4.4. For a set X, a non-negative function $d \colon X \times X \to \mathbb{R}^+$ is said to be a *premetric* on X whenever $d(x,x) = 0$ for all $x \in X$. The non-negative function d is then said to be a *pseudosemimetric* if it is a symmetric premetric, namely if it satisfies also the condition $d(x,y) = d(y,x)$ for all $x, y \in X$.

For simplicity, we will call a pseudosemimetric d on X a *semimetric*. Hence a semimetric is a pseudometric that does not necessarily satisfy the triangular inequality.

4.2 Branching hemimetrics

In this Section we present the hemimetrics for ready similarity and similarity, whose construction is analogous to that of bisimulation metrics. More precisely, the quantitative analogues of the ready simulation and simulation game are defined resp. by means of functionals **R** and **S** over the lattice $([0,1]^{\mathcal{S}\times\mathcal{S}}, \preceq)$, the idea being that whenever $s \in \mathcal{S}$ is at some given distance d from $t \in \mathcal{S}$, then t can mimic s transitions and evolve into distributions that are at distance not greater than d.

We remark that since preorders are asymmetrical relations, their quantitative analogous should share this property, and thus our distance for (ready) simulation will be actually a hemimetric. However, in accordance with the usual conventions in the related literature, we will use the term (ready) simulation *metric* in place of (ready) simulation hemimetric.

Definition 4.5 ((Ready) simulation metric functional)**.** Let $\mathbf{R}, \mathbf{S} \colon [0,1]^{\mathcal{S}\times\mathcal{S}} \to [0,1]^{\mathcal{S}\times\mathcal{S}}$ be the functions defined for all functions $d \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ and processes $s, t \in \mathcal{S}$ by

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

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

Notice that, due to the image-finiteness assumption, maxima and minima in Definition 4.5 are well-defined. It is not hard to show that **R** and **S** are monotone. Then, since $([0,1]^{\mathcal{S}\times\mathcal{S}}, \preceq)$ is a complete lattice, by the Knaster-Tarski theorem **R** and **S** have the least fixed point. Ready simulation metrics (resp. simulation metrics) are the 1-bounded hemimetrics being prefixed points of **R** (resp. **S**). We define the *ready similarity metric* (resp. *similarity metric*) as the least fixed point of **R** (resp. **S**).

Definition 4.6 ((Ready) simulation metric.)**.** A 1-bounded hemimetric $d \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is a *ready simulation metric* if and only if $\mathbf{R}(d) \preceq d$. The least fixed point of **R** is denoted by $\mathbf{d}_{\mathrm{r},\lambda}$ and called the *ready similarity metric*. Analogously, a 1-bounded hemimetric $d \colon \mathcal{S}\times\mathcal{S} \to [0,1]$ is a *simulation metric* if and only if $\mathbf{S}(d) \preceq d$. The least fixed point of **S** is denoted by $\mathbf{d}_{\mathrm{s},\lambda}$ and called the *similarity metric*.

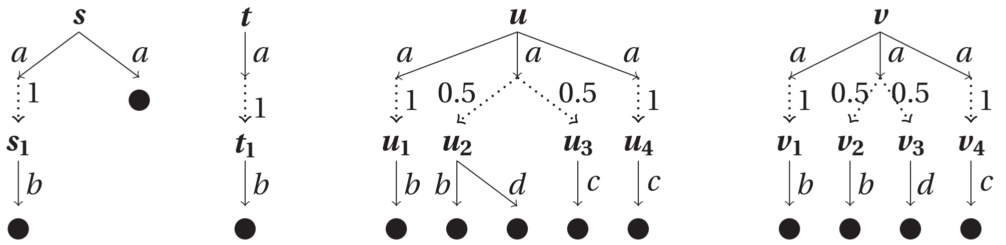


Figure 4.2: *Processes s, t, u, v are such that $\mathbf{d}_{\mathrm{r},\lambda}(s,t) = \lambda$, $\mathbf{d}_{\mathrm{s},\lambda}(s,t) = 0$, and $\mathbf{d}_{\mathrm{r},\lambda}(u,v) = \mathbf{d}_{\mathrm{s},\lambda}(u,v) = \frac{1}{2}\lambda$.*

Moreover, as in the case of the bisimilarity metric, this fixed-point characterization of the (ready) similarity metric allows us to define a notion of distance between processes that considers only the first k transition steps.

Definition 4.7 (Up-to-k (ready) similarity metric)**.** We define the *up-to-k ready similarity metric* $\mathbf{d}^k_{\mathrm{r},\lambda}$ for $k \in \mathbb{N}$ by $\mathbf{d}^k_{\mathrm{r},\lambda} = \mathbf{R}^k(\mathbf{0})$. Analogously, we define the *up-to-k ready similarity metric* $\mathbf{d}^k_{\mathrm{s},\lambda}$ for $k \in \mathbb{N}$ by $\mathbf{d}^k_{\mathrm{s},\lambda} = \mathbf{S}^k(\mathbf{0})$.

Due to the continuity of the lifting functional **K** we can infer that also the functional **R** (resp. **S**) is continuous, besides monotone, thus ensuring that the closure ordinal of **R** (resp. **S**) is ω [154]. Hence, the up-to-k (ready) similarity metrics converge to the (ready) similarity metric when $k \to \infty$.

Proposition 4.1. *Assume an image-finite PTS such that for each transition $s \xrightarrow{a} \pi$ we have that the probability distribution π has finite support. Then*

1. $\mathbf{d}_{\mathrm{r},\lambda} = \lim_{k\to\infty} \mathbf{d}^k_{\mathrm{r},\lambda}$ *and*
2. $\mathbf{d}_{\mathrm{s},\lambda} = \lim_{k\to\infty} \mathbf{d}^k_{\mathrm{s},\lambda}$.

Proof. The proof of both items follows by applying the same arguments used in the proof of Proposition 2.8. ■

***Example* 4.1.** Consider processes s, t in Figure 4.2. We aim to evaluate $\mathbf{d}_{\mathrm{r},\lambda}(t,s)$ and $\mathbf{d}_{\mathrm{s},\lambda}(t,s)$. We start with $\mathbf{d}_{\mathrm{r},\lambda}(t,s)$. We have $\mathbf{d}_{\mathrm{r},\lambda}(t_1,s_1) = 0$ and $\mathbf{d}_{\mathrm{r},\lambda}(t_1,\mathrm{nil}) = 1$, thus giving

$$\mathbf{d}_{\mathrm{r},\lambda}(t,s) = \min\{\lambda \cdot \mathbf{d}_{\mathrm{r},\lambda}(t_1,s_1), \lambda \cdot \mathbf{d}_{\mathrm{r},\lambda}(t_1,\mathrm{nil})\} = \min\{\lambda \cdot 0, \lambda \cdot 1\} = 0.$$

Similarly, we obtain that $\mathbf{d}_{\mathrm{s},\lambda}(t,s) = 0$. Let us evaluate now $\mathbf{d}_{\mathrm{r},\lambda}(s,t)$ and $\mathbf{d}_{\mathrm{s},\lambda}(s,t)$. Clearly, $\mathbf{d}_{\mathrm{r},\lambda}(s_1,t_1) = 0$ and $\mathbf{d}_{\mathrm{r},\lambda}(\mathrm{nil},t_1) = 1$, thus giving

$$\mathbf{d}_{\mathrm{r},\lambda}(s,t) = \max\{\lambda \cdot \mathbf{d}_{\mathrm{r},\lambda}(s_1,t_1), \lambda \cdot \mathbf{d}_{\mathrm{r},\lambda}(\mathrm{nil},t_1)\} = \max\{\lambda \cdot 0, \lambda \cdot 1\} = \lambda.$$

Interestingly, the evaluation of the similarity distance between s and t is different. In fact we have $\mathbf{d}_{\mathrm{s},\lambda}(s_1,t_1) = 0$ and $\mathbf{d}_{\mathrm{s},\lambda}(\mathrm{nil},t_1) = 0$ as well, since nil cannot execute any action and thus the supremum over the distributions reachable by it trivially becomes 0. Therefore, we get

$$\mathbf{d}_{\mathrm{s},\lambda}(s,t) = \max\{\lambda \cdot \mathbf{d}_{\mathrm{s},\lambda}(s_1,t_1), \lambda \cdot \mathbf{d}_{\mathrm{s},\lambda}(\mathrm{nil},t_1)\} = \max\{\lambda \cdot 0, \lambda \cdot 0\} = 0.$$

Consider now processes u, v from the same Figure 4.2. We have that $\mathbf{d}_{\mathrm{r},\lambda}(u_1, v_1) = \mathbf{d}_{\mathrm{r},\lambda}(u_1, v_2) = \mathbf{d}_{\mathrm{r},\lambda}(u_3, v_4) = \mathbf{d}_{\mathrm{r},\lambda}(u_4, v_4) = 0$ whereas $\mathbf{d}_{\mathrm{r},\lambda}(u_i, v_j) = 1$ for all other combinations for $i, j \in \{1, \dots, 4\}$. Clearly, the leftmost and the rightmost a-branches of u can be matched resp. by the leftmost and rightmost a-branches of v. Therefore, to evaluate $\mathbf{d}_{\mathrm{r},\lambda}(u, v)$ we need to match the distribution $\pi_1 = \frac{1}{2}\delta_{u_2} + \frac{1}{2}\delta_{u_3}$ with the distribution in $\mathrm{der}(v, a)$ that minimizes the Kantorovich distance from it. Let $\pi_2 = \delta_{v_1}$, $\pi_3 = \frac{1}{2}\delta_{v_2} + \frac{1}{2}\delta_{v_3}$ and $\pi_4 = \delta_{v_4}$. We have

$$\begin{aligned}
\mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_1, \pi_2) &= \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_2, v_1) + \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_3, v_1) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 = 1\\
\mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_1, \pi_3) &= \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_2, v_2) + \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_3, v_3) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 = 1\\
\mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_1, \pi_4) &= \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_2, v_4) + \frac{1}{2}\mathbf{d}_{\mathrm{r},\lambda}(u_3, v_4) = \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 = \frac{1}{2}
\end{aligned}$$

from which we can conclude that

$$\mathbf{d}_{\mathrm{r},\lambda}(u, v) = \max\left\{0, \lambda\cdot\min\{1, \frac{1}{2}\}\right\} = \frac{1}{2}\cdot\lambda.$$

Notice that $\mathbf{d}_{\mathrm{r},\lambda}(v, u) = \frac{1}{2}\cdot\lambda$ as well, which is obtained by $\mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_{v_2}, \delta_{u_1}) = \frac{1}{2}$.

By similar calculations we get that $\mathbf{d}_{\mathrm{s},\lambda}(u, v) = \mathbf{d}_{\mathrm{s},\lambda}(v, u) = \frac{1}{2}\cdot\lambda$. We simply remark that in this case it holds that $\mathbf{d}_{\mathrm{s},\lambda}(u_2, v_2) = \mathbf{d}_{\mathrm{s},\lambda}(u_2, v_3) = 1$ but $\mathbf{d}_{\mathrm{s},\lambda}(v_2, u_2) = \mathbf{d}_{\mathrm{s},\lambda}(v_3, u_2) = 0$. The value of $\mathbf{d}_{\mathrm{s},\lambda}(v, u)$ is then due to the fact that the weight assigned to u_2 is 0.5, and thus it is not enough to simulate both v_2 and v_3, having weight 0.5 each. ◀

We show now that $\mathbf{d}_{\mathrm{r},\lambda}$ is a 1-bounded hemimetric, thus implying that it is the least ready simulation metric.

Theorem 4.2. *Function $\mathbf{d}_{\mathrm{r},\lambda}$ is a 1-bounded hemimetric on $\mathcal{S}$.*

Proof. Firstly, we show that $\mathbf{d}_{\mathrm{r},\lambda}$ is a hemimetric. To this aim we show the stronger property that

$$\text{for each } k \in \mathbb{N} \text{ we have that } \mathbf{d}^k_{\mathrm{r},\lambda} \text{ is a hemimetric.} \tag{4.1}$$

The thesis will then follow by $\mathbf{d}_{\mathrm{r},\lambda} = \lim_{k\to\infty}\mathbf{d}^k_{\mathrm{r},\lambda}$ and the linearity of the limit. We proceed by induction over $k \in \mathbb{N}$ to prove Equation (4.1), namely that for each $k \in \mathbb{N}$ it holds that

1. $\mathbf{d}^k_{\mathrm{r},\lambda}(s, s) = 0$, for all $s \in \mathcal{S}$, and
2. $\mathbf{d}^k_{\mathrm{r},\lambda}(s, t) \le \mathbf{d}^k_{\mathrm{r},\lambda}(s, u) + \mathbf{d}^k_{\mathrm{r},\lambda}(u, t)$, for all $s, t, u \in \mathcal{S}$.

The base case $k = 0$ is trivial since $\mathbf{d}^0_{\mathrm{r},\lambda}(s, t) = 0$ for all $s, t \in \mathcal{S}$.

Consider the base case $k > 0$. The proof of item 1 is immediate from the definition of $\mathbf{d}^k_{\mathrm{r},\lambda}$. Let us prove item 2, namely the triangular inequality. We can distinguish two cases.

(a) $\mathrm{init}(s) \neq \mathrm{init}(t)$ and thus $\mathbf{d}^k_{\mathrm{r},\lambda}(s, t) = 1$. Given any process u we have that

- ⋆ either $\mathrm{init}(s) \neq \mathrm{init}(u)$, thus implying $\mathbf{d}^k_{\mathrm{r},\lambda}(s, u) = 1$,

⋆ or $\mathrm{init}(t) \neq \mathrm{init}(u)$, thus implying $\mathbf{d}^k_{\mathrm{r},\lambda}(u,t) = 1$.

In both cases we obtain that $\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \le \mathbf{d}^k_{\mathrm{r},\lambda}(s,u) + \mathbf{d}^k_{\mathrm{r},\lambda}(u,t)$ as requested.

(b) $\mathrm{init}(s) = \mathrm{init}(t)$, thus giving

$$\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) = \sup_{a \in \mathcal{A}} \left\{ \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s, \pi_t) \right\}$$

By induction over $k-1$, we have that $\mathbf{d}^{k-1}_{\mathrm{r},\lambda}$ is a hemimetric. Thus, by Proposition 2.1 we directly gather that

$$\mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_1,\pi_2) \le \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_1,\pi_3) + \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_3,\pi_2). \tag{4.2}$$

Then, by definition of supremum we have that for each $\varepsilon > 0$ there is an action $a_\varepsilon \in \mathcal{A}$ such that

$$\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) < \max_{\pi_s \in \mathrm{der}(s,a_\varepsilon)} \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s,\pi_t) + \varepsilon \tag{4.3}$$

Let $\tilde{\pi}_s \in \mathrm{der}(s,a_\varepsilon)$ be the distribution realizing the maximum in Equation (4.3). Given any process p let

$$\begin{aligned} \tilde{\pi}_u &= \arg \min_{\pi_u \in \mathrm{der}(u,a_\varepsilon)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\pi_u) \\ \tilde{\pi}_t &= \arg \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_u,\pi_t). \end{aligned}$$

Then we have

$$\begin{aligned} & \max_{\pi_s \in \mathrm{der}(s,a_\varepsilon)} \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s,\pi_t) + \varepsilon \\ = & \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\pi_t) + \varepsilon \\ \le & \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\tilde{\pi}_t) + \varepsilon \\ \le & \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\tilde{\pi}_u) + \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_u,\tilde{\pi}_t) + \varepsilon \qquad \text{(by Equation (4.2))} \\ = & \min_{\pi_u \in \mathrm{der}(u,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\pi_u) + \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_u,\pi_t) + \varepsilon \\ \le & \max_{\pi_s \in \mathrm{der}(s,a_\varepsilon)} \min_{\pi_u \in \mathrm{der}(u,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s,\pi_u) + \\ & + \max_{\pi_u \in \mathrm{der}(u,a_\varepsilon)} \min_{\pi_t \in \mathrm{der}(t,a_\varepsilon)} \lambda\, \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_u,\pi_t) + \varepsilon \\ \le & \mathbf{d}^k_{\mathrm{r},\lambda}(s,u) + \mathbf{d}^k_{\mathrm{r},\lambda}(u,t) + \varepsilon \end{aligned}$$

and since this holds for all $\varepsilon > 0$, it concludes the proof of Equation (4.1).

To conclude, we need to show that $\mathbf{d}_{\mathrm{r},\lambda}$ is 1-bounded. This follows by showing that

$$\text{for each } k \in \mathbb{N} \text{ we have } \mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \le 1 \text{ for all } s,t \in \mathcal{S} \tag{4.4}$$

and the monotonicity of the limit. Equation (4.4) follows by an easy induction over $k \in \mathbb{N}$. ■

Next, we show that ready simulation is the kernel of $\mathbf{d}_{\mathrm{r},\lambda}$.

Theorem 4.3. *For processes $s,t \in \mathcal{S}$, we have $\mathbf{d}_{\mathrm{r},\lambda}(s,t)=0$ iff $s \sqsubseteq_{\mathrm{r}} t$.*

Proof. ($\Rightarrow$) We aim to show that the relation

$$\mathcal{R} = \{(s,t) \mid \mathbf{d}_{\mathrm{r},\lambda}(s,t)=0\}$$

is a probabilistic ready simulation.

Assume that $s \mathcal{R} t$, namely $\mathbf{d}_{\mathrm{r},\lambda}(s,t)=0$. By Definition 4.6, this implies that $\mathrm{init}(s)=\mathrm{init}(t)$ from which we can immediately infer that whenever $s \stackrel{a}{\not\rightarrow}$ then also $t \stackrel{a}{\not\rightarrow}$. Then we notice that

$$\begin{aligned}
\mathbf{d}_{\mathrm{r},\lambda}(s,t) = & \sup_{a \in \mathrm{init}(s)} \left\{ \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_s,\pi_t) \right\} \\
& = 0 \\
& \text{iff for all } a \in \mathrm{init}(s) \\
& \quad \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_s,\pi_t) = 0 \\
& \text{iff for all } a \in \mathrm{init}(s), \text{ for each } \pi_s \in \mathrm{der}(s,a) \text{ there is } \pi_t \in \mathrm{der}(t,a) \\
& \quad \mathbf{K}(\mathbf{d}_{\mathrm{r},\lambda})(\pi_s,\pi_t) = \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s,\pi_t)} \sum_{s' \in \mathrm{supp}(\pi_s),\, t' \in \mathrm{supp}(\pi_t)} \mathfrak{w}(s',t')\mathbf{d}_{\mathrm{r},\lambda}(s',t') = 0 \\
& \qquad \text{iff for } \tilde{\mathfrak{w}} \in \mathfrak{W}(\pi_s,\pi_t) \text{ optimal} \\
& \qquad\quad \text{whenever } \tilde{\mathfrak{w}}(s',t') > 0 \text{ then } \mathbf{d}_{\mathrm{r},\lambda}(s',t') = 0.
\end{aligned}$$

More precisely, we have obtained that for each $a \in \mathrm{init}(s)$, for each $\pi_s \in \mathrm{der}(s,a)$ there are a distribution $\pi_t \in \mathrm{der}(t,a)$ and a weight function $\tilde{\mathfrak{w}}$ such that

- ★ for each $s' \in \mathrm{supp}(\pi_s)$ we have $\sum_{t' \in \mathrm{supp}(\pi_t)} \tilde{w}(s',t') = \pi_s(s')$,
- ★ for each $t' \in \mathrm{supp}(\pi_t)$ we have $\sum_{s' \in \mathrm{supp}(\pi_s)} \tilde{w}(s',t') = \pi_t(t')$,
- ★ for each $s' \in \mathrm{supp}(\pi_s)$, $t' \in \mathrm{supp}(\pi_t)$ whenever $\tilde{w}(s',t') > 0$ then $s' \mathcal{R} t'$.

Thus, from Proposition 2.4, we can conclude that for each $a \in \mathrm{init}(s)$, for each $\pi_s \in \mathrm{der}(s,a)$ there is a distribution $\pi_t \in \mathrm{der}(t,a)$ such that $\pi_s \mathcal{R}^{\dagger} \pi_t$.

Summarizing, we have obtained that whenever $s \mathcal{R} t$ then

$$\begin{aligned}
& \text{for each } s \xrightarrow{a} \pi_s \text{ there is a } \pi_t \text{ such that } t \xrightarrow{a} \pi_t \text{ and } \pi_s \mathcal{R}^{\dagger} \pi_t, \text{ and} \\
& \text{whenever } s \stackrel{a}{\not\rightarrow} \text{ then } t \stackrel{a}{\not\rightarrow}.
\end{aligned}$$

Therefore, we can conclude that the relation $\mathcal{R}$ is a ready simulation equivalence.

($\Leftarrow$) Assume now that $s \sqsubseteq_{\mathrm{r}} t$. We aim to show that $\mathbf{d}_{\mathrm{r},\lambda}(s,t)=0$. To this aim, we prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ s \sqsubseteq^{\mathrm{r}}_{k} t \text{ implies } \mathbf{d}^{k}_{\mathrm{r},\lambda}(s,t) = 0. \tag{4.5}$$

The thesis will then follow by observing that $\sqsubseteq_{\mathrm{r}} = \lim_{k\to\infty} \sqsubseteq^{\mathrm{r}}_{k}$ and $\mathbf{d}_{\mathrm{r},\lambda} = \lim_{k\to\infty} \mathbf{d}^{k}_{\mathrm{r},\lambda}$. We proceed by induction over $k \in \mathbb{N}$ to prove Equation (4.5).

★ The base case $k = 0$ is immediate as $\sqsubseteq^{\mathrm{r}}_{0} = \mathcal{S} \times \mathcal{S}$ and $\mathbf{d}^{0}_{\mathrm{r},\lambda}(s,t) = 0$ for all $s, t \in \mathcal{S}$.

★ Consider now the inductive step $k > 0$. By definition we have that $s \sqsubseteq^{\mathrm{r}}_{k} t$ if and only if whenever $s \xrightarrow{a} \pi_s$ then there is a π_t with $t \xrightarrow{a} \pi_t$ and $\pi_s \sqsubseteq^{\mathrm{r}^\dagger}_{k-1} \pi_t$ and whenever $s \overset{a}{\nrightarrow}$ then $t \overset{a}{\nrightarrow}$. Clearly these two conditions imply that $\mathrm{init}(s) = \mathrm{init}(t)$ and thus

$$\mathbf{d}^{k}_{\mathrm{r},\lambda}(s,t) = \sup_{a \in \mathrm{init}(s)} \left\{ \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s, \pi_t) \right\}.$$

Let $s \xrightarrow{a} \pi_s$ and let π_t be any distribution such that $t \xrightarrow{a} \pi_t$ and $\pi_s \sqsubseteq^{\mathrm{r}^\dagger}_{k-1} t$. By Proposition 2.4 $\pi_s \sqsubseteq^{\mathrm{r}^\dagger}_{k-1} \pi_t$ implies the existence of a matching $\tilde{\mathfrak{w}} \in \mathfrak{W}(\pi_s, \pi_t)$ such that for each $s' \in \mathrm{supp}(\pi_s)$, $t' \in \mathrm{supp}(\pi_t)$ whenever $\tilde{\mathfrak{w}}(s',t') > 0$ then $s' \sqsubseteq^{\mathrm{r}}_{k-1} t'$. By induction over $k-1$, $s' \sqsubseteq^{\mathrm{r}}_{k-1} t$ implies $\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') = 0$. Thus, we have obtained that there is a matching $\tilde{\mathfrak{w}} \in \mathfrak{W}(\pi_s, \pi_t)$ such that whenever $\tilde{\mathfrak{w}}(s',t') > 0$ then $\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') = 0$. Therefore, we can infer that

$$\begin{aligned} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s, \pi_t) &= \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s,\pi_t)} \sum_{s' \in \mathrm{supp}(\pi_s),\, t' \in \mathrm{supp}(\pi_t)} \mathfrak{w}(s',t') \mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') \\ &\leq \sum_{s' \in \mathrm{supp}(\pi_s),\, t' \in \mathrm{supp}(\pi_t)} \tilde{\mathfrak{w}}(s',t') \mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') \\ &= 0. \end{aligned}$$

Hence, we have obtained that for each $\pi_s \in \mathrm{der}(s,a)$ there is a $\pi_t \in \mathrm{der}(t,a)$ such that $\lambda \cdot \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s, \pi_t) = 0$. Thus, we have

$$\begin{aligned} \mathbf{d}^{k}_{\mathrm{r},\lambda}(s,t) &= \sup_{a \in \mathrm{init}(s)} \left\{ \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \cdot \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s, \pi_t) \right\} \\ &= \sup_{a \in \mathrm{init}(s)} \left\{ \max_{\pi_s \in \mathrm{der}(s,a)} 0 \right\} \\ &= \sup_{a \in \mathrm{init}(s)} \{0\} \\ &= 0. \end{aligned}$$

■

The results for $\mathbf{d}_{\mathrm{s},\lambda}$ are analogous.

Theorem 4.4. *Function $\mathbf{d}_{\mathrm{s},\lambda}$ is a 1-bounded hemimetric on $\mathcal{S}$.*

Proof. The thesis follows by applying the same arguments used in the proof of Theorem 4.2. ■

Theorem 4.5. *For processes $s, t \in \mathcal{S}$, we have $\mathbf{d}_{\mathrm{s},\lambda}(s,t) = 0$ if and only if $s \sqsubseteq t$.*

Proof. The thesis follows by applying the same arguments used in the proof of Theorem 4.3. ■

4.3 Linear (hemi)metrics

In this Section we introduce the behavioral metrics capturing linear semantics as (decorated) traces and testing.

The trace metric

In the fully-nondeterministic case, the trace semantics is mainly based on the capability of processes to execute particular sequences of *events*, called *traces*. Clearly, the same principle should hold when also probability is taken into account: probabilistic trace semantics should be based on the evaluation of the probability of a particular trace to be executed by a process. Intuitively, this should result into a metric semantics expressing the distance between two processes by quantifying the difference in the execution probabilities of the traces. However, as we are combining nondeterminism and probability, a few more considerations are required in order to define a robust trace metric semantics.

First of all, we need to establish what are the *events* we are considering. Accordingly to the original idea in [144] an event consists on the execution of an action followed by a probability distribution over events. Traces are then seen as *trace distributions*, namely probability distributions over sequences of actions. In [148] a metric for this semantics was proposed and in [43] we proposed a logical characterization for it. Besides, this semantics is neither fully backward compatible with the fully-nondeterministic case [31] nor compositional [144]. More importantly, the distance from [148], due to an overpowered discriminating capability of schedulers, is incompatible with the bisimilarity metric, that is, denoting the distance from [148] by d_{tr}, there are processes s, t such that $d_{\mathrm{tr}}(s,t) > \mathbf{d}_{\lambda}(s,t)$. For all these reasons we decided to look for an alternative notion of trace metric that, together with its kernel, would satisfy these desirable properties. So, we switch to a standard notion of event in the trace semantics, namely the execution of a certain action, so that a trace is no more than a sequence of actions, as in the fully nondeterministic case.

Next, we need to deal with nondeterminism. Clearly, a process may execute a given trace with different probabilities, accordingly to which resolution of nondeterminism for it we are considering. For instance, process v in Figure 4.2 can execute the same trace ab with probability 1, 0.5 or 0 with respect to the choice of the leftmost, central or rightmost a-branch of v by the scheduler. As our trace metric has to quantify the differences in those executions, we need to establish how they will be compared. We let the fully-nondeterministic case guide us in this choice: when we compare two fully-nondeterministic processes we simply check that whenever a trace is executable by a process then also the other process can execute it. We shall say that only *positive information* about the execution is considered: if there is a resolution of nondeterminism for a process in which it can execute a certain trace, then this information is used in the comparison; conversely if there is a resolution in which the same process cannot execute such a trace, then this resolution is not taken into account. The same principle should hold in the PTS model. So when we consider the resolution of nondeterminism for process v in Figure 4.2 corresponding its central a-branch what we obtain is that v executes trace ab with *at least* probability 0.5.

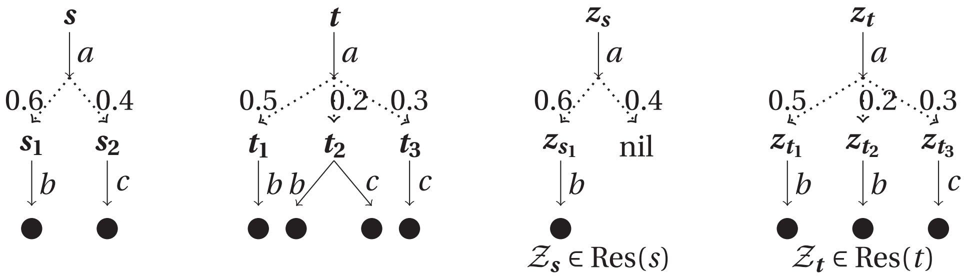


Figure 4.3: *Processes s, t are such that $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0$ and $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = \mathbf{d}_{\mathrm{Tr},\lambda}(s,t) = \frac{1}{10} \cdot \lambda$.*

Therefore, our *trace metric* will express the difference in the execution probabilities of traces by two processes at their best, namely it will be the difference between the suprema execution probabilities with respect to all resolutions of nondeterminism for the two processes.

Definition 4.8 (Trace metric)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS and $\lambda \in (0,1]$. For each trace $\alpha \in \mathcal{A}^\star$ we consider the function $\mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all processes $s,t \in \mathcal{S}$ by

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

The *trace hemimetric* and the *trace metric* are the functions $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}, \mathbf{d}_{\mathrm{Tr},\lambda} \colon \mathcal{S}\times\mathcal{S} \to [0,1]$ defined for all processes $s,t \in \mathcal{S}$ by

$$\begin{aligned}\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) &= \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)\\ \mathbf{d}_{\mathrm{Tr},\lambda}(s,t) &= \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\right\}.\end{aligned}$$

Notice that we also make use of a factor $\lambda \in (0,1]$ which discounts the final distance with respect to the length of the observed trace minus 1 since the first computation step is not discounted. This follows the same principle of the discount factor introduced in the branching metrics "*the longer the trace the less the distance should weight*" and it will allow us to properly compose the spectrum of metrics.

***Example* 4.2.** We aim to evaluate the trace distance $\mathbf{d}_{\mathrm{Tr},\lambda}(s,t)$ for processes s,t in Figure 4.3. Clearly, we have that $\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,a)) = \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,a)) = 1$. Consider now trace ab and the resolutions $\mathcal{Z}_s$ and $\mathcal{Z}_t$ for resp. s and t in Figure 4.3. We have

$$\begin{aligned}\sup_{\mathcal{Z}'_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z'_s,ab)) &= \Pr(\mathcal{C}(z_s,ab)) = 0.6\\ \sup_{\mathcal{Z}'_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z'_t,ab)) &= \Pr(\mathcal{C}(z_t,ab)) = 0.7\end{aligned}$$

thus giving $\mathbf{d}^{ab}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0$ and $\mathbf{d}^{ab}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = 0.1 \cdot \lambda$. Similarly, considering trace ac we get $\mathbf{d}^{ac}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0$ and $\mathbf{d}^{ac}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = 0.1 \cdot \lambda$. Since there are no other traces executable by the two processes we can conclude that

$$\begin{aligned}
&\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0\\
&\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = 0.1\cdot\lambda,\\
&\mathbf{d}_{\mathrm{Tr},\lambda}(s,t) = \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} = 0.1\cdot\lambda.
\end{aligned}$$

◀

We show that $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$ (resp. $\mathbf{d}_{\mathrm{Tr},\lambda}$) is a 1-bounded hemimetric (resp. pseudometric) on $\mathcal{S}$. We delay the discussion on the kernels of these distances and their properties to Section 4.5.

Theorem 4.6. *1. Function $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$ is a 1-bounded hemimetric on $\mathcal{S}$. 2. Function $\mathbf{d}_{\mathrm{Tr},\lambda}$ is a 1-bounded pseudometric on $\mathcal{S}$.*

Proof.

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

   a. $\mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,s) = 0$ for each $s\in\mathcal{S}$.

   b. $\mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_2) \le \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_3) + \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_3,s_2)$ for each $s_1,s_2,s_3\in\mathcal{S}$.

   The first item is immediate by Definition 4.8. Let us prove the triangular inequality. We can distinguish two cases.

   ★ $\sup_{\mathcal{Z}_1\in\mathrm{Res}(s_1)}\Pr(\mathcal{C}(z_1,\alpha)) \le \sup_{\mathcal{Z}_2\in\mathrm{Res}(s_2)}\Pr(\mathcal{C}(z_2,\alpha))$. Hence we have

   $$\mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_2) = 0 \le \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_3) + \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_3,s_2)$$

   for all $s_3\in\mathcal{S}$.

   ★ $\sup_{\mathcal{Z}_1\in\mathrm{Res}(s_1)}\Pr(\mathcal{C}(z_1,\alpha)) > \sup_{\mathcal{Z}_2\in\mathrm{Res}(s_2)}\Pr(\mathcal{C}(z_2,\alpha))$. Hence we have

   $$\begin{aligned}
   &\mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_2)\\
   &= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1\in\mathrm{Res}(s_1)}\Pr(\mathcal{C}(z_1,\alpha)) - \sup_{\mathcal{Z}_2\in\mathrm{Res}(s_2)}\Pr(\mathcal{C}(z_2,\alpha))\Big)\\
   &= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1\in\mathrm{Res}(s_1)}\Pr(\mathcal{C}(z_1,\alpha)) - \sup_{\mathcal{Z}_2\in\mathrm{Res}(s_2)}\Pr(\mathcal{C}(z_2,\alpha)) \pm \sup_{\mathcal{Z}_3\in\mathrm{Res}(s_3)}\Pr(\mathcal{C}(z_3,\alpha))\Big)\\
   &= \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_1\in\mathrm{Res}(s_1)}\Pr(\mathcal{C}(z_1,\alpha)) - \sup_{\mathcal{Z}_3\in\mathrm{Res}(s_3)}\Pr(\mathcal{C}(z_3,\alpha))\Big) +\\
   &\quad \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_3\in\mathrm{Res}(s_3)}\Pr(\mathcal{C}(z_3,\alpha)) - \sup_{\mathcal{Z}_2\in\mathrm{Res}(s_2)}\Pr(\mathcal{C}(z_2,\alpha))\Big)\\
   &\le \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_1,s_3) + \mathbf{d}^\alpha_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s_3,s_2).
   \end{aligned}$$

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

   $$\begin{aligned}
   &\Pr(\mathcal{C}(z_i,\alpha)) \le 1\\
   \Rightarrow\ &\sup_{\mathcal{Z}_i\in\mathrm{Res}(s_i)}\Pr(\mathcal{C}(z_i,\alpha)) \le 1\\
   \Rightarrow\ &\sup_{\mathcal{Z}_i\in\mathrm{Res}(s_i)}\Pr(\mathcal{C}(z_i,\alpha)) - \sup_{\mathcal{Z}_j\in\mathrm{Res}(s_j)}\Pr(\mathcal{C}(z_j,\alpha)) \le 1.
   \end{aligned}$$

2. $\mathbf{d}_{\mathrm{Tr},\lambda}$ being a 1-bounded (pseudo)metric follows by the fact that it is defined as the maximum between two 1-bounded (hemi)metrics (Theorem 4.6.1).

■

The decorated trace metrics

The idea behind the definition of metrics for decorated trace semantics is the same of the trace metric: we evaluate the difference in the suprema of the execution probabilities of a particular event over all possible resolutions of nondeterminism for processes, where an event changes from a sequence of actions to a sequence of actions and *decorations* accordingly to the considered semantics.

The completed trace metric

We start by considering the *completed trace metric*. Assume a process $s \in \mathcal{S}$ and consider any resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ for s. We denote by $\mathcal{CC}(z_s, \alpha)$ the set of *completed* α*-compatible computations* from z_s, namely $\mathcal{CC}(z_s, \alpha) = \{c \in \mathcal{C}(z_s, \alpha) \mid \mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(\mathrm{last}(c))) = \emptyset\}$. For sake of readability, we say that $\alpha \in \mathcal{A}^\star$ is a *completed trace* of process s if there exists a completed α-compatible computation from z_s, for some resolution $\mathcal{Z}_s$ for s.

Definition 4.9 (Completed trace metric)**.** Let $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ be a PTS and $\lambda \in (0, 1]$. For each trace $\alpha \in \mathcal{A}^\star$ we consider the function $\mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{TrC}},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) = \max\left\{0, \lambda^{|\alpha|}\left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{CC}(z_s,\alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{CC}(z_t,\alpha))\right)\right\}.$$

The *completed trace hemimetric* and the *completed trace metric* are the functions $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}$, $\mathbf{d}_{\mathrm{TrC},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0, 1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\ \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t)\right\}$$

$$\mathbf{d}_{\mathrm{TrC},\lambda}(s,t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(t,s)\right\}.$$

Notice that differently from the trace metric, the discount factor on completed traces is considered with respect to the total length of the considered completed trace. Informally, this is due to the fact that to establish whether a trace α is a completed trace or not we also need to check process behavior at step $|\alpha| + 1$, thus making the exponent of the discount factor to be $(|\alpha| + 1) - 1 = |\alpha|$, where the minus 1 is related to the first computation step which is not discounted.

***Example* 4.3.** Consider processes s, t in Figure 4.4. Firstly notice that $\mathbf{d}_{\mathrm{Tr},\lambda}(s, t) = 0$ as the only interesting traces for this case are a,ab and ac and clearly $\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) = \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t, \alpha))$ for all $\alpha \in \{a, ab, ac\}$. Hence, the completed trace distance between s and t will be obtained by comparing the probabilities of executing the completed traces.

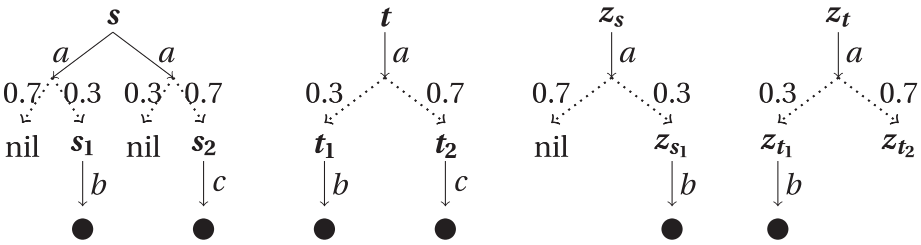


Figure 4.4: *Processes s, t are such that $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(t,s) = 0$ and $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(s,t) = \mathbf{d}_{\text{TrC},\lambda}(s,t) = 0.7 \cdot \lambda$.*

We start by evaluating $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(t,s)$. Clearly, the completed traces for t are ab and ac for which $\sup_{\mathcal{Z}_t \in \text{Res}(t)} \Pr(\mathcal{CC}(z_t, ab)) = 0.3$ and $\sup_{\mathcal{Z}_t \in \text{Res}(t)} \Pr(\mathcal{CC}(z_t, ac)) = 0.7$. These values are matched by the analogous suprema over the resolutions for s: the one corresponding to the leftmost a-branch for trace ab and the one corresponding to the rightmost a-branch for trace ac. Thus we get $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(t,s) = 0$.

Let us evaluate now $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(s,t)$. Notice that the trace $\alpha = a$ is a completed trace for s. Consider the resolutions $\mathcal{Z}_s \in \text{Res}(s)$ and $\mathcal{Z}_t \in \text{Res}(t)$ represented in Figure 4.4. We have that $\Pr(\mathcal{CC}(z_s, \alpha)) = 0.7$ whereas $\Pr(\mathcal{CC}(z_t, \alpha)) = 0$, since $\text{init}(\text{corr}_{\mathcal{Z}_t}(z_{t_2})) \neq \emptyset$. It is easy to see that $\mathbf{d}_{\sqsubseteq_{\text{TrC}},\lambda}(s,t) = \sup_{\alpha \in \{a,ab,ac\}} \mathbf{d}^{\alpha}_{\sqsubseteq_{\text{TrC}},\lambda}(s,t) = 0.7 \cdot \lambda$ and thus

$$\mathbf{d}_{\text{TrC},\lambda}(s,t) = \max\{0, 0.7 \cdot \lambda\} = 0.7 \cdot \lambda.$$

◀

The failure metric

Next we consider the *failure semantics* [37, 109] which expresses the *safety* properties of processes: it ensures that whenever a particular sequence of events takes place then the process will refuse with a positive probability to execute the actions from a given set. More formally, an element $\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ is called a *failure pair* and it is constituted by a trace α and a set F called *failure set* (sometimes called *refusal* set) containing the actions that have to be refused. Given a process $s \in \mathcal{S}$ and a resolution $\mathcal{Z}_s \in \text{Res}(s)$ for s, we denote by $\mathcal{FC}(z_s, \mathfrak{f})$ the set of $\mathfrak{f}$-*compatible* computations from z_s: for $\mathfrak{f} = (\alpha, F)$, we let $\mathcal{FC}(z_s, \mathfrak{f}) = \{c \in \mathcal{C}(z_s, \alpha) \mid \text{init}(\text{corr}_{\mathcal{Z}_s}(\text{last}(c))) \cap F = \emptyset\}$ and we say that a process s *admits* the failure set F if $\text{init}(s) \cap F = \emptyset$. This notion is lifted to resolutions via the correspondence function.

For simplicity of notation, we denote a failure pair $\mathfrak{f} = (\alpha, F)$ by $\mathfrak{f} = \alpha F$. Moreover, we define the *length of the failure pair* $\mathfrak{f} = \alpha F$ as

$$|\mathfrak{f}| = \begin{cases} |\alpha| & \text{if } F \neq \emptyset \\ |\alpha| - 1 & \text{otherwise.} \end{cases}$$

Clearly, the metric capturing the failure semantics will quantify the difference in the probabilities of satisfying the same safety properties.

Definition 4.10 (Failure metric)**.** Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS and $\lambda \in (0,1]$. For each failure pair $\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ we consider the function $\mathbf{d}^{\mathfrak{f}}_{\sqsubseteq_{\mathrm{F}},\lambda} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0,1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}^{\mathfrak{f}}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) = \max\left\{0, \lambda^{|\mathfrak{f}|}\left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f})) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}))\right)\right\}.$$

The *failure hemimetric* and the *failure metric* are the functions $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}, \mathbf{d}_{\mathrm{F},\lambda} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0,1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) = \sup_{\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})} \mathbf{d}^{\mathfrak{f}}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t)$$

$$\mathbf{d}_{\mathrm{F},\lambda}(s,t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s)\right\}.$$

The choice for the exponent of the discount factor follows, as for the three other upcoming decorated trace metrics, from the same reasoning applied to the case of completed traces. In fact, notice that given any failure pair $\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ we have that the distance with respect to $\mathfrak{f} = \alpha F$ is discounted with respect to the actual number of computation steps that are investigated, that is $|\mathfrak{f}|$. Notice that whenever $F = \emptyset$, namely we do not investigate the behavior of processes reached by executing α, then $|\mathfrak{f}| = |\alpha| - 1$ as happens for traces. Conversely, if $F \neq \emptyset$, namely we impose some constraint on the behavior of the processes reached via α, then $|\mathfrak{f}| = |\alpha|$ as for completed traces.

***Example* 4.4.** Consider again processes s, t in Figure 4.4. Consider the failure pairs $\mathfrak{f}_1 = a\{b\}$, $\mathfrak{f}_2 = a\{c\}$ and $\mathfrak{f}_3 = a\{b,c\}$ for which, considering process t and its resolutions, we have respectively

$$\sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}_1)) = 0.7 \qquad \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}_2)) = 0.3 \qquad \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}_3)) = 0.$$

These values can be evaluated on the resolution $\mathcal{Z}_t$ represented in Figure 4.4. By executing a process z_t reaches z_{t_1} with probability 0.3 and z_{t_2} with probability 0.7. Thus, to evaluate for instance $\Pr(\mathcal{FC}(z_t,\mathfrak{f}_2))$ we need to evaluate the probability of z_t to reach by executing a a process whose correspondent in t cannot execute action c. As $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_t}(z_{t_1})) = \{b\}$ and $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_t}(z_{t_2})) = \{c\}$, we have that $\Pr(\mathcal{FC}(z_t,\mathfrak{f}_2))$ corresponds to the probability of reaching z_{t_1}, namely 0.3. Consider now the resolution $\mathcal{Z}_s$ for s represented in Figure 4.4. We have that $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(\mathrm{nil})) = \emptyset$ and $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(z_{s_1})) = \{b\}$ and thus both processes admit the failure set $\{c\}$. Hence, $\Pr(\mathcal{FC}(z_s,\mathfrak{f}_2)) = 1$. By applying a similar argument to process s and its resolutions, we get

$$\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}_1)) = 1 \qquad \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}_2)) = 1 \qquad \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}_3)) = 0.7.$$

As the considered failure pairs are the only interesting ones for this particular case, the suprema for s are always greater than those for t. Then, since $|\mathfrak{f}_1| = |\mathfrak{f}_2| = |\mathfrak{f}_3| = 1$, we get that

$$\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s) = 0 \qquad \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) = \lambda^1 \cdot \max\{1-0.7,\ 1-0.3,\ 0.7-0\} = 0.7 \cdot \lambda.$$

Therefore, we can conclude that $\mathbf{d}_{\mathrm{F},\lambda}(s,t) = \max\{0.7 \cdot \lambda,\ 0\} = 0.7 \cdot \lambda$. ◀

The failure trace metric

We can extend the failure semantics to traces obtaining the *failure trace semantics* in which the safety properties of processes are tested in a step-by-step fashion. It is formalized by means of *failure traces*, namely sequences $\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ of pairs (a, F) of an action and a failure set or the empty trace and a failure set $(\mathfrak{e}, F)$. Given a process $s \in \mathcal{S}$ and a resolution $\mathcal{Z}_s \in \text{Res}(s)$ for s, we denote by $\mathcal{FC}(z_s, \mathfrak{F})$ the set of $\mathfrak{F}$-*compatible* computations from z_s: for $\mathfrak{F} = (a_1, F_1)\dots(a_n, F_n)$, we let $\alpha = a_1 \dots a_n$ and $\mathcal{FC}(z_s, \mathfrak{F}) = \{c \in \mathcal{C}(z_s, \alpha) \mid \text{init}(\text{corr}_{\mathcal{Z}_s}(z_i)) \cap F_i = \emptyset \text{ for all } i = 1, \dots, n\}$, where for each $i \in \{1, \dots, n\}$ we let z_i denote the state reached by computation c after i steps.

For simplicity of notation, we denote a failure trace $\mathfrak{F} = (a_1, F_1)\dots(a_n, F_n)$ simply by $\mathfrak{F} = a_1 F_1 \dots a_n F_n$. Moreover, we define the *length of the failure trace* $\mathfrak{F} = a_1 F_1 \dots a_n F_n$ as

$$|\mathfrak{F}| = \begin{cases} |a_1 \dots a_n| & \text{if } F_n \neq \emptyset \\ |a_1 \dots a_n| - 1 & \text{otherwise.} \end{cases}$$

We remark that although we are using the same notation $\mathcal{FC}(z_s, \_)$ to denote the set of _-compatible computations for both the failure pairs and failure traces, the meaning will always be clear from the context.

The *failure trace metric* refines the failure metric by quantifying the disparities in the probabilities of satisfying the same step-by-step safety properties.

Definition 4.11 (Failure trace metric)**.** Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS and $\lambda \in (0, 1]$. For each failure trace $\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ we consider the function $\mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\text{TrF}}, \lambda} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\text{TrF}}, \lambda}(s, t) = \max\left\{0, \lambda^{|\mathfrak{F}|}\left(\sup_{\mathcal{Z}_s \in \text{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F})) - \sup_{\mathcal{Z}_t \in \text{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{F}))\right)\right\}.$$

The *failure trace hemimetric* and the *failure trace metric* are the functions $\mathbf{d}_{\sqsubseteq_{\text{TrF}}, \lambda}, \mathbf{d}_{\text{TrF}, \lambda} \colon \mathcal{S} \times \mathcal{S} \rightarrow [0, 1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}_{\sqsubseteq_{\text{TrF}}, \lambda}(s, t) = \sup_{\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))} \mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\text{TrF}}, \lambda}(s, t)$$

$$\mathbf{d}_{\text{TrF}, \lambda}(s, t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\text{TrF}}, \lambda}(s, t), \mathbf{d}_{\sqsubseteq_{\text{TrF}}, \lambda}(t, s)\right\}.$$

***Example* 4.5.** Consider processes s, t in Figure 4.5. Firstly we evaluate $\mathbf{d}_{\sqsubseteq_{\text{TrF}}, \lambda}(s, t)$. Consider the failure trace $\mathfrak{F} = a\{d\}c\{f\}$. We have that

$$\sup_{\mathcal{Z}_s \in \text{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F})) = 1$$

given by the resolution $\mathcal{Z}_s \in \text{Res}(s)$ represented in the same Figure, whereas

$$\sup_{\mathcal{Z}_t \in \text{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{F})) = 0$$

since t_3 can execute f and if we consider the resolution $\mathcal{Z}_t \in \text{Res}(t)$ in the same Figure we have that $\text{init}(\text{corr}_{\mathcal{Z}_t}(z_{t_2})) = \{c, d\}$ and thus z_{t_2} does not admit the failure set $\{d\}$. This implies

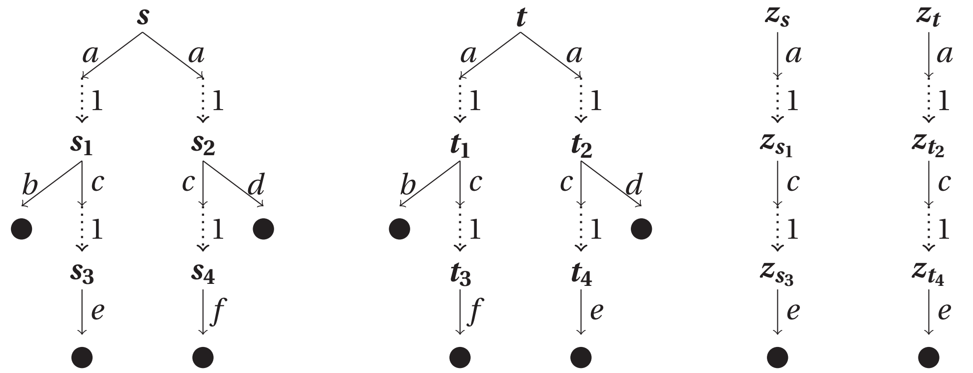


Figure 4.5: *Processes s, t are such that* $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s) = \mathbf{d}_{\mathrm{TrF},\lambda}(s,t) = \lambda^2$.

that $\mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \lambda^{|\mathfrak{F}|}\cdot(1-0) = \lambda^2$. As the distance over all failure traces of length less than 2 is 0, we can immediately infer that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \lambda^2$. From similar calculations on the failure trace $\mathfrak{F}' = a\{d\}c\{e\}$ we get that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s) = \lambda^2$ as well, and thus we can conclude

$$\mathbf{d}_{\mathrm{TrF},\lambda}(s,t) = \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s)\} = \lambda^2.$$

◀

The readiness metric

Almost complementary to failure, we have the *readiness semantics* expressing the *liveness* properties of processes: it ensures that whenever a particular sequence of events takes place then the process will have a positive probability to execute the actions from a given set. To define such a semantics we make use of *ready pairs*, namely elements $\mathfrak{r} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ constituted by a trace α and a set R called *ready set*, which is the set of the actions that have to be executed. More precisely, we require that the process reached by the trace α, executes exactly the actions specified in the ready set. Formally, given a process $s \in \mathcal{S}$ and a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ for s, we denote by $\mathcal{RC}(z_s,\mathfrak{r})$ the set of $\mathfrak{r}$-*compatible* computations from z_s: for $\mathfrak{r} = (\alpha, R)$, we let $\mathcal{RC}(z_s,\mathfrak{r}) = \{c \in \mathcal{C}(z_s,\alpha) \mid \mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(\mathrm{last}(c))) = R\}$ and we say that a process s *admits* the ready set R if $\mathrm{init}(s) = R$. This notion is lifted to resolutions via the correspondence function.

For simplicity of notation, we denote a ready pair $\mathfrak{r} = (\alpha, R)$ by $\mathfrak{r} = \alpha R$. Moreover, we define the *length of the ready pair* $\mathfrak{r} = \alpha R$ as the length of the trace α, namely $|\mathfrak{r}| = |\alpha|$.

Dually to failure metrics, the *readiness metric* will quantify the difference in the probabilities of satisfying the same liveness properties.

Definition 4.12 (Readiness metric)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS and $\lambda \in (0,1]$. For each ready pair $\mathfrak{r} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ we consider the function $\mathbf{d}^{\mathfrak{r}}_{\sqsubseteq_{\mathrm{R}},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all $s, t \in \mathcal{S}$ by

$$\mathbf{d}^{\mathfrak{r}}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t) = \max\left\{0, \lambda^{|\mathfrak{r}|}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{r})) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{r}))\Big)\right\}.$$

The *readiness hemimetric* and the *readiness metric* are the functions $\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}, \mathbf{d}_{\mathrm{R},\lambda} \colon \mathcal{S}\times\mathcal{S} \to [0,1]$ defined for all $s,t \in \mathcal{S}$ by

$$\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t) = \sup_{\mathfrak{r}\in\mathcal{A}^\star\times\mathcal{P}(\mathcal{A})} \mathbf{d}^{\mathfrak{r}}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t)$$

$$\mathbf{d}_{\mathrm{R},\lambda}(s,t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(t,s)\right\}.$$

Notice that, differently from failure semantics, testing a ready pair $\mathfrak{r} = \alpha R$ always subsumes that R is tested on the processes reached by performing α. For this reason we let $|\mathfrak{r}| = |\alpha|$ for all ready sets R.

***Example* 4.6.** Consider again processes s,t in Figure 4.4. Consider the ready pairs $\mathfrak{r}_1 = a\{b\}$, $\mathfrak{r}_2 = a\{c\}$ and $\mathfrak{r}_3 = a\{\emptyset\}$ for which, considering process t and its resolutions, we have respectively

$$\sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{r}_1)) = 0.3 \qquad \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{r}_2)) = 0.7 \qquad \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{r}_3)) = 0.$$

By executing a process z_t reaches z_{t_1} with probability 0.3 and z_{t_2} with probability 0.7. Thus, to evaluate for instance $\Pr(\mathcal{RC}(z_t,\mathfrak{r}_1))$ we need to evaluate the probability of z_t to reach by executing a a process whose correspondent in t can execute only action b. As $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_t}(z_{t_1})) = \{b\}$ and $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_t}(z_{t_2})) = \{c\}$, we have that $\Pr(\mathcal{RC}(z_t,\mathfrak{r}_1))$ corresponds to the probability of reaching z_{t_1}, namely 0.3. Consider now the resolution $\mathcal{Z}_s$ for s represented in Figure 4.4. We have that $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(\mathrm{nil})) = \emptyset$ and $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(z_{s_1})) = \{b\}$ and thus process z_{s_1} admits the ready set $\{b\}$. Hence, $\Pr(\mathcal{RC}(z_s,\mathfrak{r}_1)) = 0.3$. By applying a similar argument to process s and its resolutions, we get

$$\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{r}_1)) = 0.3 \qquad \sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{r}_2)) = 0.7 \qquad \sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{r}_3)) = 0.7.$$

As the considered ready pairs are the only interesting ones for this particular case and $|\mathfrak{r}_1| = |\mathfrak{r}_2| = |\mathfrak{r}_3| = 1$, we get that

$$\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(t,s) = 0 \qquad \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t) = \lambda^1 \cdot \max\{0.3-0.3,\ 0.7-0.7,\ 0.7-0\} = 0.7\cdot\lambda.$$

Therefore, we can conclude that $\mathbf{d}_{\mathrm{R},\lambda}(s,t) = \max\{0.7\cdot\lambda,\ 0\} = 0.7\cdot\lambda$. ◀

The ready trace metric

By considering the liveness properties in a step-by-step fashion, we obtain the *ready trace semantics*. A *ready trace* is a sequence $\mathfrak{R} \in (\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e}\times\mathcal{P}(\mathcal{A}))$ of pairs (a,R) of an action and a ready set or the empty trace and a ready set $(\mathfrak{e},F)$. Given a process $s\in\mathcal{S}$ and a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ for s, we denote by $\mathcal{RC}(z_s,\mathfrak{R})$ the set of $\mathfrak{R}$-*compatible* computations from z_s: for $\mathfrak{R} = a_1R_1\dots a_nR_n$, we let $\alpha = a_1\dots a_n$ and $\mathcal{RC}(z_s,\mathfrak{R}) = \{c \in \mathcal{C}(z_s,\alpha) \mid \mathrm{init}(\mathrm{corr}_{\mathcal{Z}_s}(z_i)) = R_i \text{ for all } i = 1,\dots,n\}$, where for each $i \in \{1,\dots,n\}$ we let z_i denote the state reached by computation c after i steps.

For simplicity of notation, we denote a ready trace $\mathfrak{R} = (a_1,R_1)\dots(a_n,R_n)$ simply by $\mathfrak{R} = a_1F_1\dots a_nF_n$. Moreover, we define the *length of the ready trace* $\mathfrak{R} = a_1R_1\dots a_nR_n$ as the length of the trace constituting it, namely $|\mathfrak{R}| = |a_1\dots a_n| = n$.

Also in this case, despite we are using the same notation $\mathcal{RC}(z_s,\_)$ to denote the set of _-compatible computations for both the ready pairs and ready traces, the meaning will always be clear from the context.

The *ready trace metric* refines the readiness metric by quantifying the disparities in the probabilities of satisfying the same step-by-step liveness properties.

Definition 4.13 (Ready trace metric)**.** Let $P = (\mathcal{S},\mathcal{A},\longrightarrow)$ be a PTS and $\lambda \in (0,1]$. For each ready trace $\mathfrak{R} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ we consider the function $\mathbf{d}^{\mathfrak{R}}_{\sqsubseteq_{\mathrm{TrR}},\lambda} : \mathcal{S}\times\mathcal{S} \to [0,1]$ defined for all $s,t \in \mathcal{S}$ by

$$\mathbf{d}^{\mathfrak{R}}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t) = \max\left\{0, \lambda^{|\mathfrak{R}|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{R})) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{R}))\Big)\right\}.$$

The *ready trace hemimetric* and the *ready trace metric* are the functions $\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}, \mathbf{d}_{\mathrm{TrR},\lambda} : \mathcal{S}\times\mathcal{S} \to [0,1]$ defined for all $s,t\in\mathcal{S}$ by

$$\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t) = \sup_{\mathfrak{R}\in(\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star\cup(\mathfrak{e}\times\mathcal{P}(\mathcal{A}))} \mathbf{d}^{\mathfrak{R}}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)$$

$$\mathbf{d}_{\mathrm{TrR},\lambda}(s,t) = \max\left\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(t,s)\right\}.$$

***Example* 4.7.** Consider again processes s,t in Figure 4.5. Firstly we evaluate $\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)$- Consider the ready trace $\mathfrak{R} = a\{b,c\}c\{e\}$. We have that

$$\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{R})) = 1$$

given by the resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ represented in the same Figure, whereas

$$\sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{RC}(z_t,\mathfrak{R})) = 0$$

since t_3 can execute f and if we consider the resolution $\mathcal{Z}_t \in \mathrm{Res}(t)$ in the same Figure we have that $\mathrm{init}(\mathrm{corr}_{\mathcal{Z}_t}(z_{t_2})) = \{c,d\}$ and thus z_{t_2} does not admit the ready set $\{b,c\}$. This implies that $\mathbf{d}^{\mathfrak{R}}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t) = \lambda^{|\mathfrak{R}|}\cdot(1-0) = \lambda^2$. As the distance over all ready traces of length less than 2 is 0, we can immediately infer that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t) = \lambda^2$. From similar calculations on the failure trace $\mathfrak{R}' = a\{c,d\}c\{e\}$ we get that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(t,s) = \lambda^2$ as well, and thus we can conclude

$$\mathbf{d}_{\mathrm{TrR},\lambda}(s,t) = \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(t,s)\} = \lambda^2.$$

◀

Well-defined distances

The following Theorem formalizes our definitions by proving that all the functions proposed so far to quantify decorated trace semantics are actually hemimetrics and pseudometrics.

Theorem 4.7. *Let* $x \in \{\mathrm{TrC},\mathrm{F},\mathrm{TrF},\mathrm{R},\mathrm{TrR}\}$.

1. *Function* $\mathbf{d}_{\sqsubseteq_x,\lambda}$ *is a* 1*-bounded hemimetric on* $\mathcal{S}$.

2. *Function* $\mathbf{d}_{x,\lambda}$ *is a* 1*-bounded pseudometric on* $\mathcal{S}$*.*

Proof. The thesis follows by applying the same arguments used in the proof of Theorem 4.6. ■

The testing metric

We conclude this Section by defining a metric expressing the probabilistic testing semantics. To the best of our knowledge, ours is the first proposal for a quantitative version of this semantics. Since the work in [62], testing semantics expresses the reliability of a process in different environments, which are modeled by the notion of *test*.

Definition 4.14 (Test)**.** A *nondeterministic probabilistic test transition systems* (NPT) is a finite PTS $O = (\mathcal{O}, \mathcal{A}, \rightarrow)$ where $\mathcal{O}$ is a set of processes, called *tests*, containing a distinguished *success process* denoted by $\surd$ with no outgoing transitions. We say that a computation from $o \in \mathcal{O}$ is *successful* if and only if its last state is $\surd$.

Given a test $o \in \mathcal{O}$, we define the depth of o to be the length of the longest executable sequence of transitions in o, namely

$$\mathrm{dpt}(o) = \begin{cases} 0 & \text{if } \mathrm{init}(o) = \emptyset \\ 1 + \sup\limits_{a \in \mathrm{init}(o), \pi \in \mathrm{der}(o,a), o' \in \mathrm{supp}(\pi)} \mathrm{dpt}(o') & \text{otherwise.} \end{cases}$$

The behavior of a process in an environment is then determined by means of the *interaction system* of the process and the test modeling the given environment.

Definition 4.15 (Interaction system)**.** Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS and $O = (\mathcal{O}, \mathcal{A}, \rightarrow_O)$ be an NPT. The *interaction system* of P and O is the PTS $P \parallel O$ where

- every $(s, o) \in \mathcal{S} \times \mathcal{O}$ is called a *configuration* and is said to be *successful* if and only if $o = \surd$;
- a computation from $(s, o) \in P \parallel O$ is *successful* if and only if its last configuration is successful.

Notice that since O is finite, then also $P \parallel O$ is finite.

The notion of interaction systems of a PTS and an NPT naturally induces the interaction systems of a process and a test.

Hence, the behavior of a process is given by the probability it has to *pass a test*, that is the probability of reaching a successful process in the interaction system. Given an interaction system $s \parallel o$ and a resolution $\mathcal{Z}_{s,o} \in \mathrm{Res}(s, o)$, we denote by $\mathcal{SC}(z_{s,o})$ the set of *successful computations* from the state $z_{s,o}$. Moreover, given any trace $\alpha \in \mathcal{A}^\star$ we let $\mathcal{SC}(z_{s,o}, \alpha)$ denote the set of α-compatible successful computations from $z_{s,o}$. For sake of readability, we say that $\alpha \in \mathcal{A}^\star$ is successful for an interaction system $s \parallel o$ if there is a successful α-compatible computation from $z_{s,o}$ for some resolution $z_{s,o}$ for (s, o).

Our aim is to define a metric suitable to express the testing semantics. Intuitively, we shall quantify the disparities in the success probabilities of processes with respect to all tests. However, a simple comparison of this kind would not be sufficient. As previously pointed out in the discussion of trace metric, to avoid unexpected evaluations of success probabilities, we need to limit the power of schedulers. Thus, inspired by [29], we will reason in a *trace-by-trace* fashion also on testing: our testing metric will quantify the differences in the probabilities of processes to reach success by executing a given trace, with respect to all traces and tests.

Although only maximal computations may lead to success, considering them alone is not sufficient to guarantee full backward compatibility of the kernel of our metric with the fully-nondeterministic case. In [62] a distinction between *may testing* and *must testing* is made: a process may pass a test if at least one of the computations of the interaction system is successful. Conversely, a process must pass a test if all the computations of the interaction system are successful. In the literature, we can find several proposals of probabilistic testing semantics [51, 69, 95, 166] and in particular of may and must testing obtained as a probabilistic generalization of those in [62]. Briefly, two processes are probabilistic may (resp. must) testing equivalent if the suprema (resp. infima) of their success probabilities with respect to all resolutions of nondeterminism are the same. In [31], it has been shown that these notions are not fully backward compatible with the fully-nondeterministic case. To guarantee the compatibility for the trace-to-trace approach we need to impose the same restrictions given by the must testing from [62] on our resolutions. Our testing metric should quantify the disparities in the trace-by-trace success probabilities and also guarantee that for each trace $\alpha \in \mathcal{A}^\star$, whenever all α-compatible maximal computations of a process lead to failure, then also the other process should fail.

For this reason we introduce a particular subset of resolutions on which we will evaluate the suprema of success probabilities. Given a process $s \in \mathcal{S}$, a test $o \in \mathcal{O}$ and a trace $\alpha \in \mathcal{A}^\star$, we denote by $\mathrm{Res}_{\max,\alpha}(s,o)$ the subset of maximal resolutions $\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max}(s,o)$ such that $\mathcal{C}_{\max}(z_{s,o},\alpha) \neq \emptyset$, namely the subset of maximal resolutions for $s \parallel o$ having at least one maximal computation compatible with α.

Interestingly, for any trace $\alpha \in \mathcal{A}^\star$, parallel composition being synchronous guarantees that whenever two interaction systems have at least one maximal α-compatible computation then either both systems will reach a successful process by executing α or they will both fail.

Lemma 4.8. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ and an NPT $O = (\mathcal{O}, \mathcal{A}, \rightarrow_O)$. Then for all $s, t \in \mathcal{S}$, $o \in \mathcal{O}$ and $\alpha \in \mathcal{A}^\star$ we have that whenever both $\mathrm{Res}_{\max,\alpha}(s,o), \mathrm{Res}_{\max,\alpha}(t,o) \neq \emptyset$, then*

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) > 0 \quad \textit{iff} \quad \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) > 0$$

Proof. The proof follows by noticing that either there is a successful computation from o which is compatible with α, and thus $\mathrm{Res}_{\max,\alpha}(s,o), \mathrm{Res}_{\max,\alpha}(t,o) \neq \emptyset$ implies that there is also one such computation in both (s,o) and (t,o), or there is no computation like that, and thus no successful maximal computation from (s,o) and (t,o) can be compatible with α. ■

Finally, we remark that the execution probabilities of processes are inevitably modified by the interaction with a test. Thus, to obtain a metric comparable with the ones proposed so far, we ought to introduce a *normalization factor*. To this purpose, notice that for a process $s \in \mathcal{S}$ and a test $o \in \mathcal{O}$, we have that $\mathcal{Z}_{s,o}$ is a resolution in $\mathrm{Res}(s,o)$ if and only if

$\mathcal{Z}_{s,o} = \mathcal{Z}_s \parallel \mathcal{Z}_o$ for some resolutions $\mathcal{Z}_s \in \mathrm{Res}(s)$ and $\mathcal{Z}_o \in \mathrm{Res}(o)$. Consequently, for each resolution $\mathcal{Z}_{s,o} = \mathcal{Z}_s \parallel \mathcal{Z}_o$ for (s,o) we have that for all traces $\alpha \in \mathcal{A}^\star$

$$\begin{aligned}
\Pr(\mathcal{C}(z_{s,o},\alpha)) &= \sum_{c \in \mathcal{C}(z_{s,o},\alpha)} \Pr(c) \\
&= \sum_{c \in \{c_s \parallel c_o \mid c_s \in \mathcal{C}(z_s,\alpha) \wedge c_o \in \mathcal{C}(z_o,\alpha)\}} \Pr(c_s)\Pr(c_o) \\
&= \sum_{c_s \in \mathcal{C}(z_s,\alpha)} \Pr(c_s) \cdot \sum_{c_o \in \mathcal{C}(z_o,\alpha)} \Pr(c_o) \\
&= \Pr(\mathcal{C}(z_s,\alpha)) \cdot \Pr(\mathcal{C}(z_o,\alpha))
\end{aligned}$$

where $c = c_s \parallel c_o$ denotes that the projection on the first component of the configurations in c gives $c_s \in \mathcal{C}(z_s)$ and the projection on their second component gives $c_o \in \mathcal{C}(z_o)$. This implies that

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) \cdot \sup_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha))$$

where the supremum of the success probabilities for the test o is evaluated over maximal resolutions, since no computation that is not maximal can reach success. Conversely, the analogous supremum for the process s is evaluated over all possible resolutions for s since success can be given only by the test and thus we only need to quantify the probability of executing the trace α.

Now, we have all the ingredients necessary to define our *testing metric*.

Definition 4.16 (Testing metric)**.** Let $P = (\mathcal{S},\mathcal{A},\to)$ be a PTS and $O = (\mathcal{O},\mathcal{A},\to_O)$ an NPT. Let $\lambda \in (0,1]$. For each test $o \in \mathcal{O}$ and trace $\alpha \in \mathcal{A}^\star$ define

$$d(s,t,o,\alpha) = \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) - \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha))$$

for all $s,t \in \mathcal{S}$. Then we define the function $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ for all $s,t \in \mathcal{S}$ as

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \begin{cases}
\lambda^{\mathrm{dpt}(o)-1} \dfrac{d(s,t,o,\alpha)}{\sup_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha))} & \text{if } d(s,t,o,\alpha) > 0 \\[2ex]
\lambda^{\mathrm{dpt}(o)-1} \dfrac{\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))} & \text{if } \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0 \wedge \\
& \wedge \, \mathrm{Res}_{\max,\alpha}(t,o) = \emptyset \\[1ex]
0 & \text{otherwise.}
\end{cases}$$

The *testing premetric* and the *testing metric* are the functions $\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}, \mathbf{d}_{\text{test},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ defined for all processes $s,t \in \mathcal{S}$ by

$$\begin{aligned}
\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t) &= \sup_{o \in \mathcal{O}} \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) \\
\mathbf{d}_{\text{test},\lambda}(s,t) &= \max\left\{\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(t,s)\right\}.
\end{aligned}$$

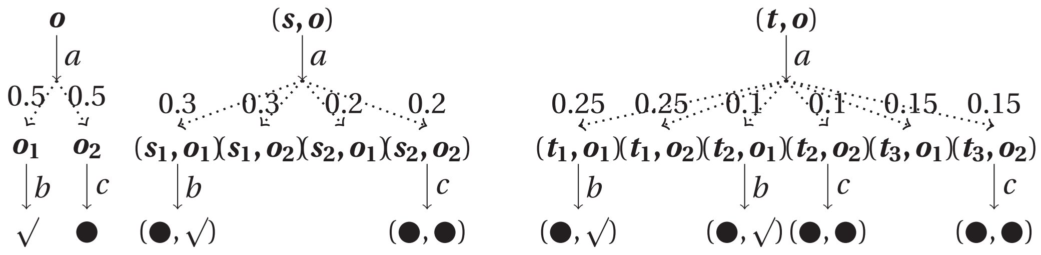


Figure 4.6: *A test showing that processes s, t from Figure 4.3 are such that* $\mathbf{d}_{\text{test},\lambda}(s,t) \geq 0.1 \cdot \lambda$.

The choice for the exponent of the discount factor follows from the same reasoning applied to the case of trace metric.

Now that we have formally introduced our testing metric, we can discuss in detail the technical choices that we have made to define it. In particular, given a test o and a trace α, we will focus on the definition of $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}$, which we will explain with the help of a few examples.

We remark that for all processes s, t, $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$ is an asymmetric distance and thus it is mainly constructed on the properties of process s.

The first case in the definition of $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$ should be the most intuitive: whenever s has a positive probability of reaching success by executing α in the interaction with o, then we consider the same success probability for t and we quantify their difference. If such a difference is positive, then we normalize it with respect to the success probability of α in o (which is necessarily non null), and we assign it as value to $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$.

***Example* 4.8.** We represent the probability distribution $\delta_{\surd}$, namely the distribution assigning probability 1 to the process $\surd$, simply as $\surd$.

Consider processes s, t in Figure 4.3 and their interaction systems with the test o represented in Figure 4.6. Let $\alpha = ab$. We aim to evaluate $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(t,s)$. Clearly,

$$\sup_{\mathcal{Z}_{t,o} \in \text{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) = 0.35 \qquad \sup_{\mathcal{Z}_{s,o} \in \text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0.3$$

from which we gather $d(t,s,o,\alpha) = 0.35 - 0.3 = 0.05$. As $\sup_{\mathcal{Z}_o \in \text{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha)) = 0.5$, we can conclude that

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = \lambda^{\text{dpt}(o)-1} \frac{0.05}{0.5} = 0.1 \cdot \lambda.$$

Let us evaluate now $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$. We have that $\sup_{\mathcal{Z}_{s,o} \in \text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) > 0$ and $d(s,t,o,\alpha) < 0$. Hence, $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0$. Notice that for $\beta \in \{a, ac\}$ we have $\text{Res}_{\max,\beta}(s,o) \neq \emptyset$, $\sup_{\text{Res}_{\max,\beta}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\beta)) = 0$ but $\text{Res}_{\max,\beta}(t,o) \neq \emptyset$ as well and thus $\mathbf{d}^{o,\beta}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0$. Similarly, we derive $\mathbf{d}^{o,\beta}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = 0$. ◄

Let us focus now on the second case in the definition of $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$. Assume that process s cannot reach success in the interaction with the test o by executing α and that $t \parallel o$ has no α-compatible maximal computations, namely $\text{Res}_{\max,\alpha}(t,o) = \emptyset$. In this case we need to check whether s may fail by executing α. This is to guarantee the fully backward

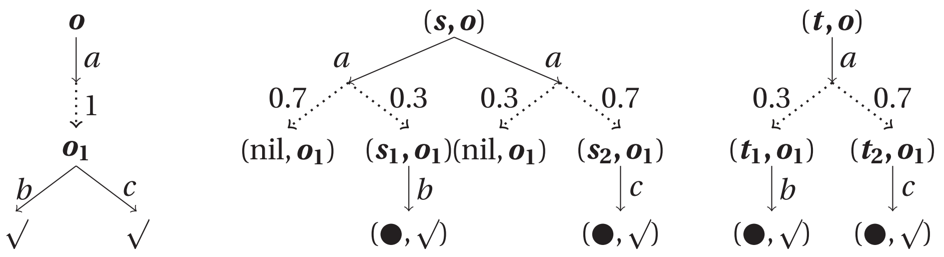


Figure 4.7: *A test showing that processes s, t from Figure 4.4 are such that* $\mathbf{d}_{\text{test},\lambda}(s,t) \geq 0.7 \cdot \lambda$.

compatibility of the kernel of our metric with the fully-nondeterministic case. In fact, as previously outlined, testing semantics expresses non only success probabilities but also the possibility of failure. Therefore the metric for testing should also emphasize the differences of processes related to failure. However, we do not quantify the disparities in the probabilities of failing a test by executing a particular trace, but simply the possibility of a process to fail against the impossibility for the other one to fail. Our aim is simply to guarantee that whenever a process fails then so does the other one, without imposing any constraint on the probabilities. Those are considered only in quantifying the differences on success.

The probability of s to fail wrt. the execution of α is given by $\sup_{\text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))$, where maximal computations are considered as success and failure can only be established on them. This probability is then normalized by $\sup_{\text{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))$, where instead all α-compatible computations are considered to avoid the denominator to be 0. In fact, $s \parallel o$ can execute α only if both s and o can execute it (synchronization is full) but since α is not successful ($\sup_{\text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0$) we cannot establish whether the α-compatible computations from o are maximal or not. In the following example we show that the restriction to $\text{Res}_{\max,\alpha}(\_)$ and the second case in the definition of $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}$ are necessary to guarantee full backward compatibility with the fully-nondeterministic case.

***Example* 4.9.** Consider processes s, t in Figure 4.4 and their interaction systems with the test o represented in Figure 4.7. Clearly,

$$\sup_{\mathcal{Z}_{s,o} \in \text{Res}_{\max,ab}(s,o)} \Pr(\mathcal{SC}(z_{s,o},ab)) = \sup_{\mathcal{Z}_{t,o} \in \text{Res}_{\max,ab}(t,o)} \Pr(\mathcal{SC}(z_{t,o},ab)) = 0.3$$

$$\sup_{\mathcal{Z}_{s,o} \in \text{Res}_{\max,ac}(s,o)} \Pr(\mathcal{SC}(z_{s,o},ac)) = \sup_{\mathcal{Z}_{t,o} \in \text{Res}_{\max,ac}(t,o)} \Pr(\mathcal{SC}(z_{t,o},ac)) = 0.7$$

and thus $\mathbf{d}^{o,ab}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \mathbf{d}^{o,ab}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = \mathbf{d}^{o,ac}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \mathbf{d}^{o,ac}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = 0$. Let us consider the trace $\alpha = a$. We aim to evaluate $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(t,s)$. Since $\text{Res}_{\max,\alpha}(t,o) = \emptyset$, we can immediately conclude that $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = 0$. Consider now $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$. Notice that $\text{Res}_{\max,\alpha}(s,o) \neq \emptyset$, $\sup_{\text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0$ and, as already noticed, $\text{Res}_{\max,\alpha}(t,o) = \emptyset$. Therefore, the conditions of the second case of the definition of $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}$ are satisfied and we can infer that

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \lambda^{\text{dpt}(o)-1} \frac{\sup\limits_{\text{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup\limits_{\text{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))} = \lambda \cdot \frac{0.7}{1} = 0.7 \cdot \lambda.$$

◀

We conclude by showing that $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}$ (resp. $\mathbf{d}_{\mathrm{test},\lambda}$) is a 1-bounded premetric (resp. semimetric) on $\mathcal{S}$.

Theorem 4.9. *1. Function* $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}$ *is a* 1*-bounded premetric on* $\mathcal{S}$*. 2. Function* $\mathbf{d}_{\mathrm{test},\lambda}$ *is a* 1*-bounded semimetric on* $\mathcal{S}$*.*

Proof.

1. To prove that $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}$ is a 1-bounded premetric it is enough to show that for each test $o \in \mathcal{O}$ and for each trace $\alpha \in \mathcal{A}^{\star}$, the function $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}$ is a 1-bounded premetric, that is we need to show that $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,s) = 0$ for each $s \in \mathcal{S}$, which is immediate by Definition 4.16.

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

   $$\frac{d(s,t,o,\alpha)}{\sup\limits_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha))} \le 1 \qquad \frac{\sup\limits_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup\limits_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))} \le 1.$$

2. $\mathbf{d}_{\mathrm{test},\lambda}$ being a 1-bounded semimetric follows by the fact that it is defined as the symmetrization of the 1-bounded premetric.

■

4.4 A spectrum of behavioral metrics

In this Section we show that the behavioral distances discussed so far can be partially ordered in a spectrum by the relation '*makes processes farther than*', as represented by the blue arrows in the upper part of Figure 4.1.

More formally, the purpose of this Section is to prove the following Theorem.

Theorem 4.10. *The following relations among the proposed behavioral metrics hold:*

1. $\mathbf{d}_{\lambda} > \mathbf{d}_{\mathrm{r},\lambda} > \mathbf{d}_{\mathrm{s},\lambda}$.
2. $\mathbf{d}_{\mathrm{r},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
3. $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
4. $\mathbf{d}_{\mathrm{s},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
5. $\mathbf{d}_{\mathrm{r},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}$ *and* $\mathbf{d}_{\mathrm{r},\lambda} > \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}$.
6. $\mathbf{d}_{\lambda} > \mathbf{d}_{\mathrm{TrF},\lambda} > \mathbf{d}_{\mathrm{F},\lambda} > \mathbf{d}_{\mathrm{test},\lambda} > \mathbf{d}_{\mathrm{Tr},\lambda}$.
7. $\mathbf{d}_{\mathrm{F},\lambda} > \mathbf{d}_{\mathrm{TrC},\lambda} > \mathbf{d}_{\mathrm{Tr},\lambda}$.
8. $\mathbf{d}_{\lambda} > \mathbf{d}_{\mathrm{TrR},\lambda}$ *and* $\mathbf{d}_{\lambda} > \mathbf{d}_{\mathrm{R},\lambda}$.

It is worth noticing the isolation of the metrics for the readiness semantics in the spectrum as well as the mutual incomparability of the metrics for testing, completed traces and similarity. These results may seem odd, especially in comparison with the linear time branching time spectrum of [159] on fully-nondeterministic processes. However, as we will detail in Section 4.6, they are due to the close interaction of nondeterminism and probability in the PTS model.

For each relation $d > d'$ in Theorem 4.10 we will prove, in the upcoming Theorems 4.11–4.19, the non strict version $d \geq d'$. Then, $d > d'$ follows from (i) relation $d \geq d'$, (ii) the results in Section 4.5 showing that the kernel of d (resp. d') is the target of the red dotted arrow in Figure 4.1 originating from d (resp. d'), (iii) the strict inclusion between the kernels of d and d' that we will prove in Section 4.6. Indeed, (i)–(iii) ensure the existence of processes $s,t \in \mathcal{S}$ with $d(s,t) > 0 = d'(s,t)$. Moreover, the examples in Section 4.6 showing that certain kernels are incomparable automatically give that also the corresponding metrics are incomparable.

Theorem 4.11. *For each $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_\lambda(s,t) \geq \mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{s},\lambda}(s,t)$.*

Proof. First of all we notice that whenever $\mathrm{init}(s) \neq \mathrm{init}(t)$ then $\mathbf{d}_\lambda(s,t) = \mathbf{d}_{\mathrm{r},\lambda}(s,t) = 1$ and $\mathbf{d}_{\mathrm{s},\lambda}(s,t) \leq 1$ an the thesis holds.

Consider now the case of $\mathrm{init}(s) = \mathrm{init}(t)$. We proceed by induction over $k \in \mathbb{N}$ to prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ \mathbf{d}^k_\lambda(s,t) \geq \mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}^k_{\mathrm{s},\lambda}(s,t). \tag{4.6}$$

The thesis will then follow by the monotonicity of the limit.

The base case $k = 0$ is immediate since by definition $\mathbf{d}^0_\lambda(s,t) = \mathbf{d}^0_{\mathrm{r},\lambda}(s,t) = \mathbf{d}^0_{\mathrm{s},\lambda}(s,t) = 0$.

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

$$\begin{aligned}
\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) &= \sup_{a \in \mathcal{A}^\star} \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s,\pi_t)\\
&\geq \sup_{a \in \mathcal{A}^\star} \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\pi_s,\pi_t)\\
&= \mathbf{d}^k_{\mathrm{s},\lambda}(s,t)
\end{aligned}$$

where the inequality follows by induction over $k-1$ and the monotonicity of **K**, inf and sup.

Similarly, we have

$$\begin{aligned}
&\mathbf{d}^k_\lambda(s,t)\\
&= \sup_{a \in \mathcal{A}} \max\left\{ \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_\lambda)(\pi_s,\pi_t),\ \max_{\pi_t \in \mathrm{der}(t,a)} \min_{\pi_s \in \mathrm{der}(s,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_\lambda)(\pi_s,\pi_t)\right\}\\
&\geq \sup_{a \in \mathcal{A}} \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_\lambda)(\pi_s,\pi_t)\\
&\geq \sup_{a \in \mathcal{A}} \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \lambda \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\pi_s,\pi_t)\\
&= \mathbf{d}^k_{\mathrm{r},\lambda}(s,t)
\end{aligned}$$

where the second inequality follows by induction over $k-1$ and the monotonicity of **K**, inf and sup. ■

Theorem 4.12. *For all $s,t \in \mathcal{S}$ it holds that* $\mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)$.

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

$$\mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \sup_{\mathfrak{F} \in (\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e}\times\mathcal{P}(\mathcal{A})), |\mathfrak{F}|\leq k} \mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t).$$

Notice that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \lim_{k\to\infty} \mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)$. Therefore, to prove the thesis, we prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ \mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t). \tag{4.7}$$

The thesis will the follow by Proposition 4.1 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

$$\mathbf{d}^1_{\mathrm{r},\lambda}(s,t) = \mathbf{d}^1_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \begin{cases} 1 & \text{if } \mathrm{init}(s) \neq \mathrm{init}(t) \\ 0 & \text{otherwise} \end{cases}$$

and thus Equation (4.7) directly follows.

Consider now the inductive step $k > 0$. If $\mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = 0$, then there is nothing to prove. Hence assume that $\mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) > 0$. Notice that by definition of supremum we have that for each $\varepsilon > 0$ there is a failure trace $\mathfrak{F}_\varepsilon \in (\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e}\times\mathcal{P}(\mathcal{A}))$, with $|\mathfrak{F}_\varepsilon| \leq k$ s.t.

$$\mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) = \sup_{\mathfrak{F} \in (\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e}\times\mathcal{P}(\mathcal{A})), |\mathfrak{F}|\leq k} \mathbf{d}^{\mathfrak{F}}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) < \mathbf{d}^{\mathfrak{F}_\varepsilon}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) + \varepsilon.$$

In the following we will prove that $\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}^{\mathfrak{F}_\varepsilon}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)$ and that such a result does not depend on the choice of ε. Therefore, we will get that for all $\varepsilon > 0$ it holds $\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) + \varepsilon \geq \mathbf{d}^k_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)$ from which Equation (4.7) directly follows.

We ca assume, without loss of generality, that $\mathfrak{F}_\varepsilon = aF\mathfrak{F}'$ for some $a \in \mathcal{A}, F \in \mathcal{P}(\mathcal{A})$ and $\mathfrak{F}' \in (\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e}\times\mathcal{P}(\mathcal{A}))$ with $|\mathfrak{F}'| \leq k-1$. Then we have

$$\begin{aligned}
&\mathbf{d}^{\mathfrak{F}_\varepsilon}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) \\
&= \lambda^{|\mathfrak{F}_\varepsilon|}\Bigg(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{F}_\varepsilon)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\mathfrak{F}_\varepsilon))\Bigg) \\
&= \lambda^{|\mathfrak{F}_\varepsilon|}\Bigg(\max_{\pi_s \in \mathrm{der}(s,a)} \sum_{\substack{s' \in \mathrm{supp}(\pi_s) \\ \text{s.t. } \mathrm{init}(s')\cap F = \emptyset}} \pi_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) + \\
&\quad - \max_{\pi_t \in \mathrm{der}(t,a)} \sum_{\substack{t' \in \mathrm{supp}(\pi_t) \\ \text{s.t. } \mathrm{init}(t')\cap F = \emptyset}} \pi_t(t') \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}))\Bigg).
\end{aligned}$$

Let

$$\tilde{\pi}_s = \arg\max_{\pi_s \in \mathrm{der}(s,a)} \sum_{\substack{s' \in \mathrm{supp}(\pi_s) \\ \text{s.t. } \mathrm{init}(s')\cap F = \emptyset}} \pi_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')).$$

Moreover, for sake of readability, for any distribution π we let $F(\pi) = \{s \in \mathrm{supp}(\pi) \mid \mathrm{init}(s) \cap F = \emptyset\}$ and we let $\neg F(\pi) = \{s \in \mathrm{supp}(\pi) \mid \mathrm{init}(s) \cap F \neq \emptyset\}$ be its complementary. Then we have

$$
\begin{aligned}
&\mathbf{d}^{k}_{\mathrm{r},\lambda}(s,t)\\
&= \sup_{a \in \mathcal{A}} \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s,\pi_t)} \lambda \sum_{\substack{s' \in \mathrm{supp}(\pi_s)\\ t' \in \mathrm{supp}(\pi_t)}} \mathfrak{w}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t')\\
&\geq \max_{\pi_s \in \mathrm{der}(s,a)} \min_{\pi_t \in \mathrm{der}(t,a)} \min_{\mathfrak{w} \in \mathfrak{W}(\pi_s,\pi_t)} \lambda \sum_{\substack{s' \in \mathrm{supp}(\pi_s)\\ t' \in \mathrm{supp}(\pi_t)}} \mathfrak{w}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t')\\
&\geq \min_{\pi_t \in \mathrm{der}(t,a)} \min_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s,\pi_t)} \lambda \sum_{\substack{s' \in \mathrm{supp}(\tilde{\pi}_s)\\ t' \in \mathrm{supp}(\pi_t)}} \mathfrak{w}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t')\\
&\geq \lambda \sum_{\substack{s' \in \mathrm{supp}(\tilde{\pi}_s)\\ t' \in \mathrm{supp}(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t')\\
&= \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t'') +\\
&\quad + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s'',t') + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t'')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s'',t'') \Bigg]\\
&\geq \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t') + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'') + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t') \Bigg]\\
&\geq \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\mathbf{d}^{k-1}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s',t') + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'') + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t') \Bigg]\\
&\geq \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t), \mathrm{init}(t') \cap F = \emptyset}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|}\Bigg(\sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg) +\\
&\quad + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'') + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t') \Bigg]\\
&= \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}')) +\\
&\quad + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'') + \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t') \Bigg]\\
&\geq \lambda \Bigg[\sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) + \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t'' \in \neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t'')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) +\\
&\quad - \sum_{\substack{s' \in F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}')) - \sum_{\substack{s'' \in \neg F(\tilde{\pi}_s)\\ t' \in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}')) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda\Bigg[\sum_{\substack{s'\in F(\tilde{\pi}_s)\\ t'\in \mathrm{supp}(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sum_{\substack{s'\in \mathrm{supp}(\tilde{\pi}_s)\\ t'\in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg]\\
&= \lambda\Bigg[\sum_{s'\in F(\tilde{\pi}_s)} \tilde{\pi}_s(s')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sum_{t'\in F(\tilde{\pi}_t)} \tilde{\pi}_t(t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg]\\
&\geq \lambda^{|\mathfrak{F}_\varepsilon|}\Bigg[\sum_{s'\in F(\tilde{\pi}_s)} \tilde{\pi}_s(s') \sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \max_{\pi_t\in \mathrm{der}(t,a)} \sum_{t'\in F(\pi_t)} \pi_t(t') \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg]\\
&= \lambda^{|\mathfrak{F}_\varepsilon|}\Bigg[\max_{\pi_s\in \mathrm{der}(s,a)} \sum_{s'\in F(\pi_s)} \pi_s(s') \sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \max_{\pi_t\in \mathrm{der}(t,a)} \sum_{t'\in F(\pi_t)} \pi_t(t') \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg]\\
&= \lambda^{|\mathfrak{F}_\varepsilon|}\Bigg[\sup_{\mathcal{Z}_{s}\in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_{s},\mathfrak{F}_\varepsilon)) - \sup_{\mathcal{Z}_{t}\in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_{t},\mathfrak{F}_\varepsilon))\Bigg]\\
&= \mathbf{d}^{\mathfrak{F}_\varepsilon}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)
\end{aligned}
$$

where:

- ★ The second step follows by evaluating the ready simulation distance on a particular action, namely the action a of $\mathfrak{F}_\varepsilon$.
- ★ The third step follows by choosing $\tilde{\pi}_s$ among all distributions in $\mathrm{der}(s,a)$.
- ★ The fourth step follows letting

$$
\begin{aligned}
\tilde{\pi}_t &= \arg\min_{\pi_t\in \mathrm{der}(t,a)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\pi_t)\\
\mathfrak{w} &= \arg\min_{\mathfrak{w}\in \mathfrak{W}(\tilde{\pi}_s,\tilde{\pi}_t)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{r},\lambda})(\tilde{\pi}_s,\tilde{\pi}_t).
\end{aligned}
$$

- ★ The sixth step follows by noticing that $\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s',t'')=1$ whenever $\mathrm{init}(s')\cap F=\emptyset$ and $\mathrm{init}(t'')\cap F\neq\emptyset$ or viceversa. Moreover we delete the non negative quantity

$$
\sum_{\substack{s''\in\neg F(\tilde{\pi}_s)\\ t''\in\neg F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s'',t'')\mathbf{d}^{k-1}_{\mathrm{r},\lambda}(s'',t'').
$$

- ★ The seventh step follows by induction over $k-1$.
- ★ The eighth step follows by $|\mathfrak{F}'|\leq k-1$ and the definition of supremum.
- ★ the ninth step follows by

$$
\begin{aligned}
&\sum_{\substack{s'\in F(\tilde{\pi}_s)\\ t'\in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|}\Bigg(\sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg)\\
&= \sum_{\substack{s'\in F(\tilde{\pi}_s)\\ t'\in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_{s'}\in \mathrm{Res}(s')} \Pr(\mathcal{FC}(z_{s'},\mathfrak{F}')) - \sum_{\substack{s'\in F(\tilde{\pi}_s)\\ t'\in F(\tilde{\pi}_t)}} \tilde{\mathfrak{w}}(s',t')\lambda^{\mathfrak{F}'} \sup_{\mathcal{Z}_{t'}\in \mathrm{Res}(t')} \Pr(\mathcal{FC}(z_{t'},\mathfrak{F}'))\Bigg)
\end{aligned}
$$

which can be proved by applying the same arguments used in the proof of Theorem 4.15.

- ⋆ The tenth step follows by $\lambda^{|\mathfrak{F}'|} \sup_{\mathcal{Z}_p \in \mathrm{Res}(p)} \Pr(\mathcal{FC}(z_p, \mathfrak{F}')) \le 1$ for all processes $p \in \mathcal{S}$.
- ⋆ The twelfth step follows by the choice of $\mathfrak{w}$.
- ⋆ The thirteenth step follows by the choice of $\tilde{\pi}_t$.
- ⋆ The fourteenth step follows by the choice of $\tilde{\pi}_s$.

This gives that $\mathbf{d}^k_{\mathrm{r},\lambda}(s,t) \ge \mathbf{d}^{\mathfrak{F}_\varepsilon}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)$ and thus concludes the proof. ■

Theorem 4.13. *For all $s, t \in \mathcal{S}$ it holds that $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$.*

Proof. We start with $\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t)$. This directly follows by noticing that each failure pair $\mathfrak{f} = a_1 a_2 \dots a_n F$ can be seen as the failure trace $\mathfrak{F} = a_1 \emptyset a_2 \emptyset \dots a_n F$, for which moreover we have $\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{f})) = \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F}))$ for any process $s \in \mathcal{S}$. Thus we have

$$
\begin{aligned}
\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) &= \sup_{\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})} \max \left\{ 0, \lambda^{|\mathfrak{f}|} \Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{f})) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{f})) \Big) \right\} \\
&< \max \left\{ 0, \lambda^{|\mathfrak{f}_\varepsilon|} \Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{f}_\varepsilon)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{f}_\varepsilon)) \Big) \right\} + \varepsilon \\
&= \max \left\{ 0, \lambda^{|\mathfrak{F}_\varepsilon|} \Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F}_\varepsilon)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{F}_\varepsilon)) \Big) \right\} + \varepsilon \\
&\le \sup_{\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))} \max \left\{ 0, \lambda^{|\mathfrak{F}|} \Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F})) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t, \mathfrak{F})) \Big) \right\} + \varepsilon \\
&= \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) + \varepsilon
\end{aligned}
$$

where

- ⋆ The second step follows by definition of supremum which guarantees that for each $\varepsilon > 0$ there is a failure trace $\mathfrak{f}_\varepsilon \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ s.t $\mathbf{d}^{\mathfrak{f}_\varepsilon}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) > \sup_{\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})} \mathbf{d}^{\mathfrak{f}}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) - \varepsilon$.
- ⋆ The third step follows by choosing the failure trace $\mathfrak{F}_\varepsilon$ as described above.
- ⋆ The forth step follows by definition of supremum.

Since the inequality $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) < \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that the thesis for this case follows.

We proceed to $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t)$. If $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0$, then there is nothing to prove. Hence assume that $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) > 0$. By definition of supremum, given any $\varepsilon > 0$ there are $o_\varepsilon \in \mathcal{O}$ and $\alpha_\varepsilon \in \mathcal{A}^\star$ such that

$$
\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \sup_{o \in \mathcal{O}} \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) < \mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) + \varepsilon.
$$

In the following, we will prove that $\mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\text{test}},\lambda}(s,t) \le \mathbf{d}_{\sqsubseteq_{\text{F}},\lambda}(s,t)$ and that this result does not depend on the choice of $\varepsilon > 0$. Therefore we will get that for all $\varepsilon > 0$ it holds $\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t) \le \mathbf{d}_{\sqsubseteq_{\text{F}},\lambda}(s,t) + \varepsilon$ from the thesis for this case directly follows.

For simplicity of notation, we let $o_\varepsilon = o$ and $\alpha_\varepsilon = \alpha$. We can distinguish two cases.

1. Either

$$\begin{aligned}
\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) &= \lambda^{\operatorname{dpt}(o)-1} \frac{\sup_{\mathcal{Z}_{s,o}\in\operatorname{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) - \sup_{\mathcal{Z}_{t,o}\in\operatorname{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha))}{\sup_{\mathcal{Z}_o\in\operatorname{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha))} \\
&= \lambda^{\operatorname{dpt}(o)-1}\Big(\sup_{\mathcal{Z}_s\in\operatorname{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t\in\operatorname{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha))\Big) \\
&= \lambda^{\operatorname{dpt}(o)-1}\Big(\sup_{\mathcal{Z}_s\in\operatorname{Res}(s)} \Pr(\mathcal{FC}(z_s,\alpha\emptyset)) - \sup_{\mathcal{Z}_t\in\operatorname{Res}(t)} \Pr(\mathcal{FC}(z_t,\alpha\emptyset))\Big) \\
&\le \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_s\in\operatorname{Res}(s)} \Pr(\mathcal{FC}(z_s,\alpha\emptyset)) - \sup_{\mathcal{Z}_t\in\operatorname{Res}(t)} \Pr(\mathcal{FC}(z_t,\alpha\emptyset))\Big) \\
&\le \mathbf{d}_{\sqsubseteq_{\text{F}},\lambda}(s,t)
\end{aligned}$$

where the second step follows by

$$\sup_{\mathcal{Z}_{s,o}\in\operatorname{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = \sup_{\mathcal{Z}_s\in\operatorname{Res}(s)} \Pr(C(z_s,\alpha)) \cdot \sup_{\mathcal{Z}_o\in\operatorname{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha))$$

for all $s \in \mathcal{S}$ and the forth step follows by the choice of the failure pair and the forth step follows from $|\alpha| \le \operatorname{dpt}(o)$ and $\lambda \le 1$.

2. Or

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \lambda^{\operatorname{dpt}(o)-1} \frac{\sup_{\mathcal{Z}_{s,o}\in\operatorname{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup_{\mathcal{Z}_o\in\operatorname{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))}$$

with $\operatorname{Res}_{\max,\alpha}(t,o) = \emptyset$. Notice that as $\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) > 0$ implies $\operatorname{Res}_{\max,\alpha}(s,o) \neq \emptyset$, we have that at least one process reached by s through the execution of α cannot synchronize with the processes reached by o through α. Now $\operatorname{Res}_{\max,\alpha}(t,o) = \emptyset$ is due to the fact that either t cannot perform α at all, or there is a set of actions F such that the processes reached by t through α can always synchronize with processes reached by o through α on at least one action in F. In the first case let $\mathfrak{f} = \alpha\emptyset$, whereas in the second case we let $\mathfrak{f} = \alpha F$. Notice that in both cases we are guaranteed that $|\mathfrak{f}| \le \operatorname{dpt}(o) - 1$ and moreover we have that $\sup_{\mathcal{Z}_t\in\operatorname{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f})) = 0$. Then,

$$\begin{aligned}
\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\text{test}},\lambda}(s,t) &= \lambda^{\operatorname{dpt}(o)-1} \frac{\sup_{\mathcal{Z}_{s,o}\in\operatorname{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup_{\mathcal{Z}_o\in\operatorname{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))} \\
&< \lambda^{\operatorname{dpt}(o)-1} \frac{\Pr(\mathcal{C}_{\max}(z^{\varepsilon'}_{s,o},\alpha)) + \varepsilon'}{\sup_{\mathcal{Z}_o\in\operatorname{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))}
\end{aligned}$$

$$
\begin{aligned}
&\le \lambda^{|\mathfrak{f}|}\Pr(\mathcal{C}_{\max}(z_s^{\varepsilon'},\alpha)) + \varepsilon'' \\
&= \lambda^{|\mathfrak{f}|}\Pr(\mathcal{FC}(z_s^{\varepsilon'},\mathfrak{f})) + \varepsilon'' \\
&\le \lambda^{|\mathfrak{f}|} \sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f})) + \varepsilon'' \\
&= \lambda^{|\mathfrak{f}|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f})) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}))\Big) + \varepsilon'' \\
&\le \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) + \varepsilon''
\end{aligned}
$$

where:

- ⋆ The second step follows by definition of supremum which guarantees that for each $\varepsilon' > 0$ there is a resolution $\mathcal{Z}_{s,o}^{\varepsilon'} \in \mathrm{Res}_{\max,\alpha}(s,o)$ such that $\Pr(\mathcal{C}_{\max}(z_{s,o}^{\varepsilon'},\alpha)) > \sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha)) - \varepsilon'$.
- ⋆ The third step follows by considering $\mathcal{Z}_{s,o}^{\varepsilon'} = \mathcal{Z}_s^{\varepsilon'} \parallel \mathcal{Z}_o^{\varepsilon'}$ with $\mathcal{Z}_s^{\varepsilon'}$ being such that each maximal computation of $\mathcal{Z}_{s,o}^{\varepsilon'}$ is projected on a maximal computation of $\mathcal{Z}_s^{\varepsilon'}$, and $\varepsilon'' = \lambda^{|\mathfrak{f}|}\varepsilon' / \sup_{\mathcal{Z}_o\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_o,\alpha))$.
- ⋆ The forth and sixth steps follow by the choice of $\mathfrak{f}$.

Since this kind of reasoning holds for all $\varepsilon',\varepsilon'' > 0$, we obtain that the thesis follows also in this case.

Finally, we show that $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$. First of all we notice that for each test $o \in \mathcal{O}$ and for each trace $\alpha \in \mathcal{A}^\star$ we have that

$$
\sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = \sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) \cdot \sup_{\mathcal{Z}_o\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha)).
$$

For each trace $\alpha \in \mathcal{A}^\star$ we let o_α be the test consisting in a single successful computation compatible with a α. Then, we let $\mathcal{O}_{\mathcal{A}^\star} = \{o_\alpha \in \mathcal{O} \mid \alpha \in \mathcal{A}^\star\}$. Notice that for each $\alpha \in \mathcal{A}^\star$ we have that $\mathrm{dpt}(o_\alpha) = |\alpha|$. Therefore, we have

$$
\begin{aligned}
\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) &= \sup_{o\in\mathcal{O}} \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \\
&\ge \sup_{o\in\mathcal{O}_{\mathcal{A}^\star}} \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \\
&= \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{o_\alpha,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \\
&\ge \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) \\
&= \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)
\end{aligned}
$$

where

- ⋆ the second step follows by the fact that we are evaluating the supremum over a the smaller set of tests $\mathcal{O}_{\mathcal{A}^\star}$;

- the third step follows by construction of $\mathcal{O}_{\mathcal{A}^\star}$ and by noticing that given a test $o_\alpha \in \mathcal{O}_{\mathcal{A}^\star}$, for each $\beta \in \mathcal{A}^\star$ we have that

$$\sup_{\mathcal{Z}_o \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{SC}(z_o,\alpha)) = \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{otherwise.} \end{cases}$$

- the fourth step follows by noticing that given a test $o_\alpha \in \mathcal{O}_{\mathcal{A}^\star}$, for each $\beta \in \mathcal{A}^\star$ we have

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\beta}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\beta)) = \begin{cases} \sup\limits_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\beta)) & \text{if } \beta = \alpha \\ 0 & \text{otherwise.} \end{cases}$$

Therefore we get that for each $\alpha \in \mathcal{A}^\star$

$$\mathbf{d}^{o_\alpha,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \begin{cases} \lambda^{|\alpha|-1} \sup\limits_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) & \text{if } \mathcal{C}(s,\alpha) \neq \emptyset \wedge \mathcal{C}(t,\alpha) = \emptyset \\ \lambda^{|\alpha|-1}\Big(\sup\limits_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup\limits_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha))\Big) & \text{if } \sup\limits_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) > \\ & \quad \sup\limits_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha)) \\ 0 & \text{otherwise.} \end{cases}$$

which gives $\lambda^{|\alpha|-1} \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) \geq \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$.

■

Theorem 4.14. *For all $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$.*

Proof. We start with $\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t)$. Accordingly to Definition 4.9 we can distinguish two cases.

- $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) = \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$. In this case the thesis follows by noticing that each trace $\alpha \in \mathcal{A}^\star$ can be seen as the failure pair $\mathfrak{f} = \alpha\emptyset$, for which we have $\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) = \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}))$ for any process $s \in \mathcal{S}$. Thus we have

$$\begin{aligned} \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) &= \sup_{\alpha \in \mathcal{A}^\star} \max\left\{0, \lambda^{|\alpha|-1}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha))\Big)\right\} \\ &< \max\left\{0, \lambda^{|\alpha_\varepsilon|-1}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha_\varepsilon))\Big)\right\} + \varepsilon \\ &= \max\left\{0, \lambda^{|\mathfrak{f}_\varepsilon|-1}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}_\varepsilon)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}_\varepsilon))\Big)\right\} + \varepsilon \\ &\leq \sup_{\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})} \max\left\{0, \lambda^{|\mathfrak{f}|}\Big(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f})) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}))\Big)\right\} + \varepsilon \\ &= \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) + \varepsilon \end{aligned}$$

where

  - The second step follows by definition of supremum which guarantees that for each $\varepsilon > 0$ there is a failure trace $\alpha_\varepsilon \in \mathcal{A}^\star$ s.t $\mathbf{d}^{\alpha_\varepsilon}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) > \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) - \varepsilon$.
  - The third step follows by choosing the failure pair $\mathfrak{f}_\varepsilon$ as described above.
  - The forth step follows by definition of supremum.

  Since the inequality $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) < \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that the thesis for this case follows.

- ★ $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) = \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t)$. In this case the thesis follows by noticing that each completed trace $\alpha \in \mathcal{A}^\star$ can be seen as the failure pair $\mathfrak{f} = \alpha\mathcal{A}$, for which we have $\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{CC}(z_s,\alpha)) = \sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}))$ for any process $s \in \mathcal{S}$. Thus we have

$$\begin{aligned}
\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) &= \sup_{\alpha\in\mathcal{A}^\star} \max\left\{0, \lambda^{|\alpha|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{CC}(z_s,\alpha)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{CC}(z_t,\alpha))\Big)\right\}\\
&< \max\left\{0, \lambda^{|\alpha_\varepsilon|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{CC}(z_s,\alpha_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{CC}(z_t,\alpha_\varepsilon))\Big)\right\} + \varepsilon\\
&= \max\left\{0, \lambda^{|\mathfrak{f}_\varepsilon|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f}_\varepsilon)) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}_\varepsilon))\Big)\right\} + \varepsilon\\
&\le \sup_{\mathfrak{f}\in\mathcal{A}^\star\times\mathcal{P}(\mathcal{A})} \max\left\{0, \lambda^{|\mathfrak{f}|}\Big(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s,\mathfrak{f})) - \sup_{\mathcal{Z}_t\in\mathrm{Res}(t)} \Pr(\mathcal{FC}(z_t,\mathfrak{f}))\Big)\right\} + \varepsilon\\
&= \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) + \varepsilon
\end{aligned}$$

  where

  - The second step follows by definition of supremum which guarantees that for each $\varepsilon > 0$ there is a failure trace $\alpha_\varepsilon \in \mathcal{A}^\star$ s.t $\mathbf{d}^{\alpha_\varepsilon}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) > \sup_{\alpha\in\mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) - \varepsilon$.
  - The third step follows by choosing the failure pair $\mathfrak{f}_\varepsilon$ as described above.
  - The forth step follows by definition of supremum.

  Since the inequality $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) < \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that the thesis for this case follows.

The relation $\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$ follows directly from Definition 4.9. ■

Theorem 4.15. *For all $s, t \in \mathcal{S}$ it holds that $\mathbf{d}_{\mathrm{s},\lambda}(s,t) \ge \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$.*

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

$$\mathbf{d}^{k}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = \sup_{\alpha\in\mathcal{A}^\star, |\alpha|\le k} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t).$$

Notice that $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = \lim_{k\to\infty} \mathbf{d}^{k}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)$. Therefore, to prove the thesis, we prove the stronger property that

$$\text{for each } k \in \mathbb{N},\ \mathbf{d}^{k}_{\mathrm{s},\lambda}(s,t) \ge \mathbf{d}^{k}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t). \tag{4.8}$$

The thesis will the follow by Proposition 4.1 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

$$\mathbf{d}^1_{\mathrm{s},\lambda}(s,t) = \mathbf{d}^1_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = \begin{cases} 1 & \text{if } \mathrm{init}(s) \not\subseteq \mathrm{init}(t) \\ 0 & \text{otherwise} \end{cases}$$

and thus Equation (4.8) directly follows.

Consider now the inductive step $k > 0$. If $\mathbf{d}^k_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0$, then there is nothing to prove. Hence assume that $\mathbf{d}^k_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) > 0$. We have

$$\begin{aligned}
&\mathbf{d}^k_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) \\
&= \sup_{\alpha \in \mathcal{A}^\star, |\alpha| \le k} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) \\
&< \mathbf{d}^{\alpha_\varepsilon}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \Bigg(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s,\alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t,\alpha)) \Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \Bigg(\max_{\pi_s \in \mathrm{der}(s,a)} \sum_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) + \\
&\quad - \max_{\pi_t \in \mathrm{der}(t,a)} \sum_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) \Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \Bigg(\sum_{s' \in \mathrm{supp}(\tilde{\pi}_s)} \tilde{\pi}_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) + \\
&\quad - \max_{\pi_t \in \mathrm{der}(t,a)} \sum_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) \Bigg) + \varepsilon \\
&\le \lambda^{|\alpha_\varepsilon|-1} \Bigg(\sum_{s' \in \mathrm{supp}(\tilde{\pi}_s)} \tilde{\pi}_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) + \\
&\quad - \sum_{t' \in \mathrm{supp}(\tilde{\pi}_t)} \tilde{\pi}_t(t') \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) \Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \Bigg(\sum_{s' \in \mathrm{supp}(\tilde{\pi}_s)} \Bigg(\sum_{t' \in \mathrm{supp}(\tilde{\pi}_t)} \mathfrak{w}(s',t') \Bigg) \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) + \\
&\quad - \sum_{t' \in \mathrm{supp}(\tilde{\pi}_t)} \Bigg(\sum_{s' \in \mathrm{supp}(\tilde{\pi}_s)} \mathfrak{w}(s',t') \Bigg) \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) \Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1} \Bigg(\sum_{\substack{s' \in \mathrm{supp}(\tilde{\pi}_s) \\ t' \in \mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) + \\
&\quad - \sum_{\substack{s' \in \mathrm{supp}(\tilde{\pi}_s) \\ t' \in \mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'} \in \mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) \Bigg) + \varepsilon
\end{aligned}$$

$$
\begin{aligned}
&= \lambda^{|\alpha_\varepsilon|-1}\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t')\Big(\sup_{\mathcal{Z}_{s'}\in\mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) \pm \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big) + \\
&\quad - \sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1}\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \Big(\mathfrak{w}(s',t')\Big(\sup_{\mathcal{Z}_{s'}\in\mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) - \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big) + \\
&\qquad\qquad\qquad + \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big) + \\
&\quad - \sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Bigg) + \varepsilon \\
&= \lambda^{|\alpha_\varepsilon|-1}\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t')\Big(\sup_{\mathcal{Z}_{s'}\in\mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) - \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big) + \\
&\quad + \sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta)) - \sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Bigg) + \varepsilon \\
&= \lambda\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t')\lambda^{|\beta|-1}\Big(\sup_{\mathcal{Z}_{s'}\in\mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) - \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big)\Bigg) + \varepsilon \\
&\le \lambda\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\beta\in\mathcal{A}^\star, |\beta|\le k-1} \lambda^{|\beta|-1}\Big(\sup_{\mathcal{Z}_{s'}\in\mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)) - \sup_{\mathcal{Z}_{t'}\in\mathrm{Res}(t')} \Pr(\mathcal{C}(z_{t'},\beta))\Big)\Bigg) + \varepsilon \\
&= \lambda\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \sup_{\beta\in\mathcal{A}^\star, |\beta|\le k-1} \mathbf{d}^{\beta}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)\Bigg) + \varepsilon \\
&= \lambda\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \mathbf{d}^{k-1}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)\Bigg) + \varepsilon \\
&\le \lambda\Bigg(\sum_{\substack{s'\in\mathrm{supp}(\tilde{\pi}_s)\\ t'\in\mathrm{supp}(\tilde{\pi}_t)}} \mathfrak{w}(s',t') \mathbf{d}^{k-1}_{\mathrm{s},\lambda}(s,t)\Bigg) + \varepsilon \\
&= \lambda\,\mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\tilde{\pi}_s,\tilde{\pi}_t) + \varepsilon \\
&= \min_{\pi_t\in\mathrm{der}(t,a)} \lambda\,\mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\tilde{\pi}_s,\pi_t) + \varepsilon \\
&\le \max_{\pi_s\in\mathrm{der}(s,a)} \min_{\pi_t\in\mathrm{der}(t,a)} \lambda\,\mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\pi_s,\pi_t) + \varepsilon \\
&\le \sup_{a\in\mathcal{A}} \max_{\pi_s\in\mathrm{der}(s,a)} \min_{\pi_t\in\mathrm{der}(t,a)} \lambda\,\mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\pi_s,\pi_t) + \varepsilon \\
&= \mathbf{d}^{k}_{\mathrm{s},\lambda}(s,t) + \varepsilon
\end{aligned}
$$

where:

- ⋆ The second step follows by definition of supremum which guarantees that for each $\varepsilon > 0$ there is a trace $\alpha_\varepsilon \in \mathcal{A}^\star$, with $|\alpha_\varepsilon| \leq k$ s.t $\mathbf{d}^{\alpha_\varepsilon}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) > \sup_{\alpha\in\mathcal{A}^\star,|\alpha|\leq k} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) - \varepsilon$.
- ⋆ The fourth step follows since we can assume wlog that $\alpha_\varepsilon = a\beta$ for some $\beta \in \mathcal{A}^\star$, with $|\beta| \leq k-1$, and by construction of resolutions.
- ⋆ The fifth step follows by choosing
$$\tilde{\pi}_s = \operatorname*{arg\,max}_{\pi_s \in \mathrm{der}(s,a)} \sum_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') \sup_{\mathcal{Z}_{s'} \in \mathrm{Res}(s')} \Pr(\mathcal{C}(z_{s'},\beta)).$$
- ⋆ The sixth step follows by choosing
$$\tilde{\pi}_t = \operatorname*{arg\,min}_{\pi_t \in \mathrm{der}(t,a)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\tilde{\pi}_s,\pi_t).$$
- ⋆ The seventh step follows by choosing
$$\mathfrak{w} = \operatorname*{arg\,min}_{\mathfrak{w} \in \mathfrak{W}(\tilde{\pi}_s,\tilde{\pi}_t)} \mathbf{K}(\mathbf{d}^{k-1}_{\mathrm{s},\lambda})(\tilde{\pi}_s,\tilde{\pi}_t).$$
- ⋆ The sixteenth step follows by induction over $k-1$.
- ⋆ The seventeenth step follows by the choice of $\mathfrak{w}$.
- ⋆ The eighteenth step follows by the choice of $\tilde{\pi}_t$.

Since the inequality $\mathbf{d}^k_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) < \mathbf{d}^k_{\mathrm{s},\lambda}(s,t) + \varepsilon$ holds for all $\varepsilon > 0$, we can conclude that Equation (4.8) holds. ■

Theorem 4.16. *For all $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)$ and $\mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t)$.*

Proof. The proof of $\mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)$ follows by applying similar arguments to the ones used in the proof of Theorem 4.12.

The proof of $\mathbf{d}_{\mathrm{r},\lambda}(s,t) \geq \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t)$ follows by applying similar arguments to the ones used in the proof of Theorem 4.15. ■

Theorem 4.17. *For all $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_{\lambda}(s,t) \geq \mathbf{d}_{\mathrm{TrF},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{F},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{test},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{Tr},\lambda}(s,t)$.*

Proof. We have that

$$\begin{aligned}
\mathbf{d}_{\lambda}(s,t) &\geq \max\{\mathbf{d}_{\mathrm{r},\lambda}(s,t),\, \mathbf{d}_{\mathrm{r},\lambda}(t,s)\} && \text{(by Theorem 4.11)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s)\} && \text{(by Theorem 4.12)}\\
&= \mathbf{d}_{\mathrm{TrF},\lambda}(s,t) && \text{(by Definition 4.11).}
\end{aligned}$$

Then, we have

$$\begin{aligned}
\mathbf{d}_{\mathrm{TrF},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s)\} && \text{(by Definition 4.11)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s)\} && \text{(by Theorem 4.13)}\\
&= \mathbf{d}_{\mathrm{F},\lambda}(s,t) && \text{(by Definition 4.10).}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\mathbf{d}_{\mathrm{F},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s)\} && \text{(by Definition 4.10)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(t,s)\} && \text{(by Theorem 4.13)}\\
&= \mathbf{d}_{\mathrm{test},\lambda}(s,t) && \text{(by Definition 4.16).}
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathbf{d}_{\mathrm{test},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(t,s)\} && \text{(by Definition 4.16)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} && \text{(by Theorem 4.13)}\\
&= \mathbf{d}_{\mathrm{Tr},\lambda}(s,t) && \text{(by Definition 4.8).}
\end{aligned}$$

■

Theorem 4.18. *For all $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_{\mathrm{F},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{TrC},\lambda}(s,t) \geq \mathbf{d}_{\mathrm{Tr},\lambda}(s,t)$.*

Proof. We have that

$$\begin{aligned}
\mathbf{d}_{\mathrm{F},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s)\} && \text{(by Definition 4.10)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(t,s)\} && \text{(by Theorem 4.14)}\\
&= \mathbf{d}_{\mathrm{TrC},\lambda}(s,t) && \text{(by Definition 4.9).}
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathbf{d}_{\mathrm{TrC},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(t,s)\} && \text{(by Definition 4.9)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} && \text{(by Theorem 4.14)}\\
&= \mathbf{d}_{\mathrm{Tr},\lambda}(s,t) && \text{(by Definition 4.8).}
\end{aligned}$$

■

Theorem 4.19. *For all $s,t \in \mathcal{S}$ it holds that $\mathbf{d}_{\lambda}(s,t) \geq \mathbf{d}_{\mathrm{TrR},\lambda}(s,t)$ and $\mathbf{d}_{\lambda}(s,t) \geq \mathbf{d}_{\mathrm{R},\lambda}(s,t)$.*

Proof. We have that

$$\begin{aligned}
\mathbf{d}_{\lambda}(s,t) &\geq \max\{\mathbf{d}_{\mathrm{r},\lambda}(s,t),\, \mathbf{d}_{\mathrm{r},\lambda}(t,s)\} && \text{(by Theorem 4.11)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(t,s)\} && \text{(by Theorem 4.16)}\\
&= \mathbf{d}_{\mathrm{TrR},\lambda}(s,t) && \text{(by Definition 4.13).}
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\mathbf{d}_{\lambda}(s,t) &\geq \max\{\mathbf{d}_{\mathrm{r},\lambda}(s,t),\, \mathbf{d}_{\mathrm{r},\lambda}(t,s)\} && \text{(by Theorem 4.11)}\\
&\geq \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t),\, \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(t,s)\} && \text{(by Theorem 4.16)}\\
&= \mathbf{d}_{\mathrm{R},\lambda}(s,t) && \text{(by Definition 4.12).}
\end{aligned}$$

■

4.5 Probabilistic linear relations

In this Section we characterize the kernels of the metrics introduced in Section 4.3. We will show that the kernels of our metrics define novel notions of (decorated) trace and testing relations, which result into coarser versions of the ones presented in [29]. More precisely, our relations can be considered as the *may* part of the relations presented in the *max-min partially matching resolution approach* in [29]. Roughly speaking, this means that the kernel of our trace (resp: decorated trace; testing) metric will relate processes s and t if and only if for each trace (resp: decorated trace; test and trace) the suprema of the probabilities of executing that trace (resp: decorated trace; trace in the interaction with the test) with respect to all resolutions of nondeterminism for s and t are the same.

Although the relations derived from our metrics are coarser than those studied in the literature (see [29–31] and the references therein), they enjoy a lot of desirable properties that their finer versions may in part lack. First of all, as already noticed in [29, 95], to avoid questionable estimations of the execution probabilities we need to limit the power of schedulers. This is obtained by the *partially matching resolutions* approach, allowing to match any resolution for a process with different resolutions for the other process depending on the considered semantics-specific event, and by a *trace-by-trace* analysis of these semantics. Moreover, this approach also allows us to obtain full backward compatibility with the fully-nondeterministic and fully-probabilistic cases, also with respect to the testing semantics. Finally, comparing only the suprema of the execution probabilities instead of the extremal ones, namely suprema and infima, results into relations that are congruences with respect to parallel composition (see Example 4.11) and that are, more importantly, coarser than (bi)similarities.

Summarizing, we will show that our relations satisfy the following desirable properties:

1. compositionality;
2. full backward compatibility with the fully-nondeterministic case;
3. full backward compatibility with the fully-probabilistic case;
4. they are coarser than (bi)similarities.

Finally, in Section 4.6 we will show that these relations can be ordered with respect to the ordering '*makes strictly less identifications than*' in the spectrum in the lower part of Figure 4.1.

Probabilistic trace equivalence

We start by characterizing the kernels of the trace hemimetric and the trace metric. Our proposal is that two processes s and t are related by the *trace preorder*, $s \sqsubseteq_{\mathrm{Tr}} t$, if and only if for each trace $\alpha \in \mathcal{A}^\star$ the probability that s performs α is not greater than the probability that t performs α. Then, the symmetric closure of trace preorder gives the *trace equivalence*.

Definition 4.17 (Probabilistic trace equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{Tr}} t$ if and only if for each trace $\alpha \in \mathcal{A}^\star$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{C}(z_s, \alpha)) \le \Pr(\mathcal{C}(z_t, \alpha)).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic trace equivalent*, notation $s \sim_{\mathrm{Tr}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{Tr}} t$ and $t \sqsubseteq_{\mathrm{Tr}} s$.

The following Theorem formalizes the intuition that the proposed trace preorder $\sqsubseteq_{\mathrm{Tr}}$ and equivalence $\sim_{\mathrm{Tr}}$ constitute resp. the kernels of the trace hemimetric and trace metric.

Theorem 4.20. *For all processes $s, t \in \mathcal{S}$, we have:*

1. $\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0$ *if and only if* $s \sqsubseteq_{\mathrm{Tr}} t$*, and*
2. $\mathbf{d}_{\mathrm{Tr},\lambda}(s,t) = 0$ *if and only if* $s \sim_{\mathrm{Tr}} t$*.*

Proof.

1. We have

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

2. We have

$$\begin{aligned} s \sim_{\mathrm{Tr}} t &\iff s \sqsubseteq_{\mathrm{Tr}} t \text{ and } t \sqsubseteq_{\mathrm{Tr}} s \\ &\iff \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0 \text{ and } \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = 0 && \text{(Theorem 4.20.1)} \\ &\iff \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} = 0 \\ &\iff \mathbf{d}_{\mathrm{Tr},\lambda}(s,t) = 0. \end{aligned}$$

■

Definition 4.17 can be seen as a relaxation of the corresponding definition in [29], which requires that $\Pr(\mathcal{C}(z_s,\alpha)) = \Pr(\mathcal{C}(z_t,\alpha))$ instead of $\Pr(\mathcal{C}(z_s,\alpha)) \le \Pr(\mathcal{C}(z_t,\alpha))$. Let $\approx_{\mathrm{Tr}}$ denote the trace equivalence in [29]. Clearly we have $\approx_{\mathrm{Tr}} \subseteq \sim_{\mathrm{Tr}}$. We show now that the inclusion is strict.

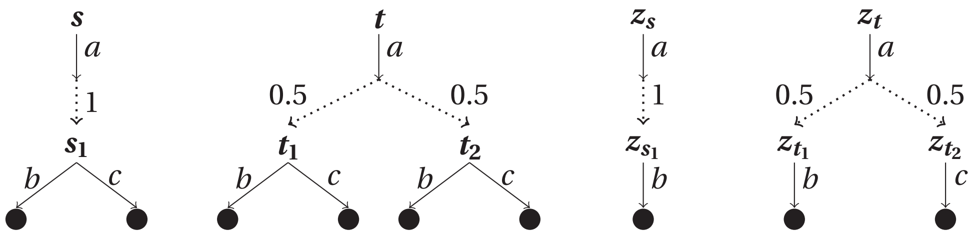


Figure 4.8: *Processes s and t are distinguished by* $\approx_{\mathrm{Tr}}$, *but related by* $\sim_{\mathrm{Tr}}$.

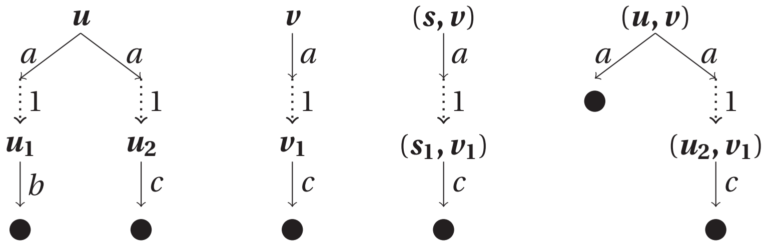


Figure 4.9: *Processes s from Figure 4.8 and u are such that* $s \sim_{\mathrm{Tr}_{\sqcup\sqcap}} u$ *but* $s \parallel v \not\sim_{\mathrm{Tr}_{\sqcup\sqcap}} u \parallel v$.

***Example* 4.10.** Consider processes s, t in Figure 4.8. We have $s \sim_{\mathrm{Tr}} t$ and $s \not\approx_{\mathrm{Tr}} t$. To see that $s \not\approx_{\mathrm{Tr}} t$, we consider the trace $\alpha = ab$ and we note that the resolution $\mathcal{Z}_t \in \mathrm{Res}(t)$ in Figure 4.8 assigns probability 0.5 to α, namely $\Pr(\mathcal{C}(z_t, \alpha)) = 0.5$, whereas the unique resolution for s assigning positive probability to α is $\mathcal{Z}_s$ in Figure 4.8 for which $\Pr(\mathcal{C}(z_s, \alpha)) = 1$. Hence no resolution in $\mathrm{Res}(s)$ matches $\mathcal{Z}_t$ on trace α, thus giving $s \not\approx_{\mathrm{Tr}} t$. Notice that $s \sim_{\mathrm{Tr}} t$ is essential to have $\sim\, \subseteq\, \sim_{\mathrm{Tr}}$. In fact, we have $s \sim t$, which follows from $s_1 \sim t_1$, $s_1 \sim t_2$ and $\delta_{s_1} \sim^{\dagger} (0.5\delta_{t_1} + 0.5\delta_{t_2})$. ◀

In [29] also a notion of trace equivalence on extremal probabilities $\sim_{\mathrm{Tr},\sqcup\sqcap}$ is proposed. Given processes $s, t \in \mathcal{S}$, the idea is to consider, for each trace $\alpha \in \mathcal{A}^{\star}$, the subsets $\mathrm{Res}_{\alpha}(s)$ and $\mathrm{Res}_{\alpha}(t)$ of the resolutions for s and t that do not contain any maximal computation corresponding to a proper prefix of α-compatible computations of the process. Then we compare the extremal execution probabilities of α on this distributions, obtaining that $s \sim_{\mathrm{Tr},\sqcup\sqcap} t$ if and only if for each $\alpha \in \mathcal{A}^{\star}$

$$\sup_{\mathcal{Z}_s \in \mathrm{Res}_{\alpha}(s)} \Pr(\mathcal{C}(z_s, \alpha)) = \sup_{\mathcal{Z}_t \in \mathrm{Res}_{\alpha}(t)} \Pr(\mathcal{C}(z_t, \alpha))$$
$$\inf_{\mathcal{Z}_s \in \mathrm{Res}_{\alpha}(s)} \Pr(\mathcal{C}(z_s, \alpha)) = \inf_{\mathcal{Z}_t \in \mathrm{Res}_{\alpha}(t)} \Pr(\mathcal{C}(z_t, \alpha)).$$

Clearly we have $\sim_{\mathrm{Tr},\sqcup\sqcap}\, \subseteq\, \sim_{\mathrm{Tr}}$. In fact we have that $s \sim_{\mathrm{Tr},\sqcup\sqcap} t$ if and only they assign the same extremal probabilities to all traces, which in particular it means that the suprema probabilities are the same for all traces and thus $s \sim_{\mathrm{Tr}} t$ is guaranteed. We show now that the inclusion is strict.

***Example* 4.11.** Consider process s in Figure 4.8 and process u in Figure 4.9. It holds that $s \sim_{\mathrm{Tr},\sqcup\sqcap} u$. To establish this, it is enough to consider the traces in $\{a, ab, ac\}$ for which the comparison of extremal execution probabilities is immediate. However, if we consider the parallel compositions of s and u with process v as represented in Figure 4.9, we get that $s \parallel v \not\sim_{\mathrm{Tr},\sqcup\sqcap} u \parallel v$. To see this, consider trace ac. We have that $\mathrm{Res}_{ac}(s \parallel v)$ contains only the resolution corresponding to process $s \parallel v$ itself. Conversely, $\mathrm{Res}_{ac}(u \parallel v)$ contains the resolutions corresponding to both a-branches of $u \parallel v$, as the leftmost one has the maximal computation $u \parallel v \overset{a}{\twoheadrightarrow} \mathrm{nil}$, which is not compatible with any proper prefix of the computation corresponding to the rightmost branch. Therefore, we have

$$\begin{aligned} \sup_{\mathcal{Z}_{s,v} \in \mathrm{Res}_{ac}(s,v)} \Pr(\mathcal{C}(z_{s,v}, ac)) &= \sup_{\mathcal{Z}_{u,v} \in \mathrm{Res}_{ac}(u,v)} \Pr(\mathcal{C}(z_{u,v}, ac)) = 1 \\ \inf_{\mathcal{Z}_{s,v} \in \mathrm{Res}_{ac}(s,v)} \Pr(\mathcal{C}(z_{s,v}, ac)) &= 1 \text{ and } \inf_{\mathcal{Z}_{u,v} \in \mathrm{Res}_{ac}(u,v)} \Pr(\mathcal{C}(z_{u,v}, ac)) = 0 \end{aligned}$$

from which we can conclude that $s \parallel v \not\sim_{\mathrm{Tr},\sqcup\sqcap} u \parallel v$. Besides, one can easily check that $s \sim_{\mathrm{Tr}} u$. Then $s \parallel v \sim_{\mathrm{Tr}} u \parallel v$ follows from Theorem 4.22 below, which is essential to guarantee the compositionality of $\sim_{\mathrm{Tr}}$. ◀

Relation $\sim_{\mathrm{Tr}}$ is (like $\approx_{\mathrm{Tr}}$) fully backward compatible with the trace equivalence on fully nondeterministic systems [37], denoted by $\sim_{\mathrm{Tr}}^{\mathbf{N}}$, as well as with the one on fully-probabilistic systems [109], denoted by $\sim_{\mathrm{Tr}}^{\mathbf{P}}$. Moreover, $\sim_{\mathrm{Tr}}$ is preserved by parallel composition.

Proposition 4.21. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ and processes $s, t \in \mathcal{S}$. Then:*

1. *If P is fully-nondeterministic, then $s \sim_{\mathrm{Tr}} t$ if and only if $s \approx_{\mathrm{Tr}} t$ if and only if $s \sim_{\mathrm{Tr}}^{\mathbf{N}} t$.*

2. *If P is fully-probabilistic, then $s \sim_{\mathrm{Tr}} t$ if and only if $s \approx_{\mathrm{Tr}} t$ if and only if $s \sim_{\mathrm{Tr}}^{\mathbf{P}} t$.*

Proof.

1. $s \sim_{\mathrm{Tr}} t \Leftrightarrow s \approx_{\mathrm{Tr}} t$ follows since in the fully nondeterminsitic context, the execution probabilities of the traces are either 0 or 1. Thus, the inequality we use to check the trace equivalence of processes, in this setting becomes an equality and therefore it is no further distinguishable from the approach of [31]. Therefore, $s \sim_{\mathrm{Tr}} t \Leftrightarrow s \sim_{\mathrm{Tr}}^{\mathbf{N}} t$ follows by $s \approx_{\mathrm{Tr}} t \Leftrightarrow s \sim_{\mathrm{Tr}}^{\mathbf{N}} t$ (Theorem 3.4(1) in [31]) and transitivity.

2. $s \sim_{\mathrm{Tr}} t \Leftrightarrow s \approx_{\mathrm{Tr}} t$ follows since in the fully probabilistic context, each process has a single maximal resolution which is the process itself. Thus, the inequality we use to check the trace equivalence of processes, in this setting becomes an equality on the probabilities related to those maximal resolutions for processes and therefore it is no further distinguishable from the approach of [31]. Therefore, $s \sim_{\mathrm{Tr}} t \Leftrightarrow s \sim_{\mathrm{Tr}}^{\mathbf{P}} t$ follows by $s \approx_{\mathrm{Tr}} t \Leftrightarrow s \sim_{\mathrm{Tr}}^{\mathbf{P}} t$ (Theorem 3.4(2) in [31]) and transitivity.

■

Theorem 4.22. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ and processes $s, t \in \mathcal{S}$.*

1. *If $s \sqsubseteq_{\mathrm{Tr}} t$, then we have $s \parallel u \sqsubseteq_{\mathrm{Tr}} t \parallel u$ for all $u \in \mathcal{S}$.*

2. *If $s \sim_{\mathrm{Tr}} t$, then we have $s \parallel u \sim_{\mathrm{Tr}} t \parallel u$ for all $u \in \mathcal{S}$.*

To prove Theorem 4.22 we need first to recall an equivalent definition to trace equivalence inspired by [31]. We start by introducing some auxiliary notation. Let $X, Y \subseteq \mathcal{A}^\star \times \mathbb{R}_{(0,1]}$, $a \in \mathcal{A}$, $\alpha \in \mathcal{A}^\star$:

- ★ $X \Vdash (\alpha, q)$ if and only if either $(\alpha, q) \in X$ or $q = 0$ and $(\alpha, q') \notin X$ for all $q' \in \mathbb{R}_{(0,1]}$;
- ★ $X + Y = \{(\alpha, q_1 + q_2) \mid X \Vdash (\alpha, q_1) \wedge Y \Vdash (\alpha, q_2) \wedge q_1 + q_2 > 0\}$;
- ★ $a.X = \{(a\alpha, q) \mid (\alpha, q) \in X\}$;
- ★ $q \cdot X = \{(\alpha, q \cdot q') \mid (\alpha, q') \in X\}$.

Then we recall the notion of *weighted traces.*

Definition 4.18 (Weighted traces, [31])**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. The set of functions $\mathrm{traces}_i \colon \mathcal{S} \to 2^{\mathcal{A}^\star \times \mathbb{R}_{(0,1]}}$ is defined for each $i \in \mathbb{N}$ inductively as follows:

- ★ $\mathrm{traces}_0(s) = \{(\mathfrak{e}, 1)\}$;
- ★ $\mathrm{traces}_{i+1}(s) = \{(\mathfrak{e}, 1)\} \cup \bigcup\limits_{s \xrightarrow{a} \pi} a.\Big(\sum\limits_{s' \in \mathsf{supp}(\pi)} \pi(s') \cdot \mathrm{traces}_i(s')\Big)$.

We let $\mathrm{traces}(s) = \bigcup\limits_{i \in \mathbb{N}} \mathrm{traces}_i(s)$.

Moreover, the following two Lemmas from [31] are still valid in our context.

Lemma 4.23 ([31, Lemma 3.6])**.** *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$. For all $s \in \mathcal{S}$ and $i \in \mathbb{N}$ it holds that* $\mathrm{traces}_i(s) \subseteq \mathrm{traces}_{i+1}(s)$.

Lemma 4.24 ([31, Lemma 3.7])**.** *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$. For all $s \in \mathcal{S}$, $\alpha \in \mathcal{A}^\star$ and $q \in (0,1]$ it holds that $(\alpha, q) \in \mathrm{traces}_(s)$ if and only if there is a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ with $\mathrm{Pr}(\mathcal{C}(z_s, \alpha)) = q$.*

Given processes $s, t \in \mathcal{S}$, we write that $\mathrm{traces}(s) \leq \mathrm{traces}(t)$ if and only if whenever $(\alpha, q) \in \mathrm{traces}(s)$ then there is a $(\alpha, q') \in \mathrm{traces}(t)$ such that $q \leq q'$.

The following theorem is obtained by adapting Theorem 3.8 of [31] to our definition of trace equivalence.

Theorem 4.25. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$ and consider $s, t \in \mathcal{S}$. Then $s \sqsubseteq_{\mathrm{Tr}} t$ if and only if* $\mathrm{traces}(s) \leq \mathrm{traces}(t)$.

Proof. The proof is analogous to the proof of Theorem 3.8 in [31] by exploiting Lemmas 4.23 and 4.24. ■

We are now ready to prove our Theorem 4.22.

Proof of Theorem 4.22. First of all we notice that the proof of Theorem 4.22.2 is an immediate consequence of Theorem 4.22.1. In fact, we have that

$$\begin{tikzcd}
 & s \sqsubseteq_{\mathrm{Tr}} t \arrow[r, Rightarrow, "\text{Thm.4.22.1}"] & (s,u) \sqsubseteq_{\mathrm{Tr}} (t,u) & \\
s \sim_{\mathrm{Tr}} t \arrow[r, Rightarrow, "\text{Def.4.17}"] & {} & {} \arrow[r, Rightarrow, "\text{Def.4.17}"] & (s,u) \sim_{\mathrm{Tr}} (t,u). \\
 & t \sqsubseteq_{\mathrm{Tr}} s \arrow[r, Rightarrow, "\text{Thm.4.22.1}"] & (t,u) \sqsubseteq_{\mathrm{Tr}} (s,u) &
\end{tikzcd}$$

Then the proof of Theorem 4.22.1 follows by applying the same arguments used in the proof of Theorem 3.9 in [31] by exploiting our Theorem 4.25 in place of Theorem 3.8 in [31]. ■

Probabilistic decorated trace equivalences

We proceed to study the kernels of the decorated trace metrics. As one can expect, their characterization reflects that of kernel of the trace (hemi)metric, accordingly to the chosen decoration.

Definition 4.19 (Probabilistic completed trace equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{TrC}} t$ if and only if $s \sqsubseteq_{\mathrm{Tr}} t$ and for each trace $\alpha \in \mathcal{A}^\star$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{CC}(z_s, \alpha)) \le \Pr(\mathcal{CC}(z_t, \alpha)).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic completed trace equivalent*, notation $s \sim_{\mathrm{TrC}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{TrC}} t$ and $t \sqsubseteq_{\mathrm{TrC}} s$.

Definition 4.20 (Probabilistic failure equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{F}} t$ if and only if for each failure pair $\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{FC}(z_s, \mathfrak{f})) \le \Pr(\mathcal{FC}(z_t, \mathfrak{f})).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic failure equivalent*, notation $s \sim_{\mathrm{F}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{F}} t$ and $t \sqsubseteq_{\mathrm{F}} s$.

Definition 4.21 (Probabilistic failure trace equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{TrF}} t$ if and only if for each failure trace $\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{FC}(z_s, \mathfrak{F})) \le \Pr(\mathcal{FC}(z_t, \mathfrak{F})).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic failure trace equivalent*, notation $s \sim_{\mathrm{TrF}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{TrF}} t$ and $t \sqsubseteq_{\mathrm{TrF}} s$.

Definition 4.22 (Probabilistic ready equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{R}} t$ if and only if for each ready pair $\mathfrak{r} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{RC}(z_s, \mathfrak{r})) \le \Pr(\mathcal{RC}(z_t, \mathfrak{r})).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic ready equivalent*, notation $s \sim_{\mathrm{R}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{R}} t$ and $t \sqsubseteq_{\mathrm{R}} s$.

Definition 4.23 (Probabilistic ready trace equivalence)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{TrR}} t$ if and only if for each ready trace $\mathfrak{R} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ it holds that

$$\text{for each } \mathcal{Z}_s \in \mathrm{Res}(s) \text{ there is a } \mathcal{Z}_t \in \mathrm{Res}(t) \text{ with } \Pr(\mathcal{RC}(z_s, \mathfrak{R})) \le \Pr(\mathcal{RC}(z_t, \mathfrak{R})).$$

We say that $s, t \in \mathcal{S}$ are *probabilistic ready trace equivalent*, notation $s \sim_{\mathrm{TrR}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{TrR}} t$ and $t \sqsubseteq_{\mathrm{TrR}} s$.

The following Theorem formalizes the intuition that the proposed decorated trace preorders and equivalences constitute the kernels of the respective decorated trace hemimetrics and metrics.

Theorem 4.26. *Let* $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$*. For processes* $s, t \in \mathcal{S}$ *we have:*

1. $\mathbf{d}_{\sqsubseteq_x,\lambda}(s,t) = 0$ *if and only if* $s \sqsubseteq_x t$.
2. $\mathbf{d}_{x,\lambda}(s,t) = 0$ *if and only if* $s \sim_x t$.

Proof. The thesis follows by applying the same arguments used in the proof of Theorem 4.20. ■

Also in the case of decorated trace semantics, our definitions can be seen as the relaxation of the corresponding ones in [29], which require the equality of the probabilities on the same decorated trace, instead of our inequalities. For $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$, let $\approx_x$ denote the proper decorated trace equivalence from [29]. Clearly we have $\approx_x \subseteq \sim_x$. In the following example we show that the inclusion $\approx_x \subseteq \sim_x$ is strict, for any $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$.

***Example* 4.12.** Consider processes s, t in Figure 4.8. We have $s \sim_x t$ and $s \not\approx_x t$ for any $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$. To see that $s \not\approx_x t$, we can apply the same arguments used in Example 4.10 to show that $s \not\approx_{\mathrm{Tr}} t$ directly to obtain the case for $x = \mathrm{TrC}$ and by considering in place of the trace $\alpha = ab$, respectively

- the failure pair (resp. the ready pair) $ab\emptyset$ for the case $x = \mathrm{F}$ (resp. $x = \mathrm{R}$);
- the failure trace $a\emptyset b\emptyset$ for the case $x = \mathrm{TrF}$;
- the ready trace $a\{b,c\}b\emptyset$ for the case $x = \mathrm{TrR}$.

Also in this case, we notice that $s \sim_x t$ is essential to have $\sim \subseteq \sim_x$. ◀

Next, we show that each relation $\sim_x$, for $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$ is (like $\approx_x$) fully backward compatible with the corresponding decorated trace equivalence on fully nondeterministic systems [37, 135], denoted by $\sim_x^{\mathbf{N}}$, as well as with the one on fully probabilistic systems [108, 109], denoted by $\sim_x^{\mathbf{P}}$. Then, $\sim_x$ is preserved by parallel composition.

Proposition 4.27. *Assume a PTS* $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ *and processes* $s, t \in \mathcal{S}$*. Let* $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$.

1. *If* P *is fully-nondeterministic, then* $s \sim_x t$ *if and only if* $s \approx_x t$ *if and only if* $s \sim_x^{\mathbf{N}} t$.
2. *If* P *is fully-probabilistic, then* $s \sim_x t$ *if and only if* $s \approx_x t$ *if and only if* $s \sim_x^{\mathbf{P}} t$.

Proof. The thesis follows by applying the same reasoning used in the proof of Proposition 4.21 and the analogous result in [29]. ■

Theorem 4.28. *Assume a PTS* $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ *and processes* $s, t \in \mathcal{S}$*. Let* $x \in \{\mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$.

1. *If* $s \sqsubseteq_x t$*, then we have* $s \parallel x \sqsubseteq_x t \parallel u$ *for all* $u \in \mathcal{S}$.
2. *If* $s \sim_x t$*, then we have* $s \parallel u \sim_x t \parallel u$ *for all* $u \in \mathcal{S}$.

To prove Theorem 4.28 we need to adapt the notion of weighted traces to the different types of decorations. Firstly we deal with the cases of completed traces.

Definition 4.24 (Weighted completed traces). Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS. The set of functions Ctraces: $\mathcal{S} \rightarrow 2^{\mathcal{A}^{\star} \times \mathbb{R}_{(0,1]}}$ is defined inductively as follows:

$$\mathrm{Ctraces}(s) = \begin{cases} (\mathfrak{e}, 1) & \text{if } \mathrm{init}(s) = \emptyset \\ \bigcup\limits_{s \xrightarrow{a} \pi} a.\big(\sum\limits_{s' \in \mathsf{supp}(\pi)} \pi(s') \cdot \mathrm{Ctraces}(s') \big) & \text{otherwise.} \end{cases}$$

We remark that for each process $s \in \mathcal{S}$, we have that $\mathrm{Ctraces}(s) \subset \mathrm{traces}(s)$.

Given processes $s, t \in \mathcal{S}$, we write that $\mathrm{Ctraces}(s) \le \mathrm{Ctraces}(t)$ if and only if whenever $(\alpha, q) \in \mathrm{Ctraces}(s)$ then there is a $(\alpha, q') \in \mathrm{Ctraces}(t)$ such that $q \le q'$.

Lemma 4.29. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$. For all $s \in \mathcal{S}$, $\alpha \in \mathcal{A}^{\star}$ and $q \in (0,1]$ we have that $(\alpha, q) \in \mathrm{Ctraces}(s)$ if and only if there is a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ with $\Pr(\mathcal{CC}(z_s, \alpha)) = q$.*

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.24 (Lemma 3.7 in [31]). ■

Theorem 4.30. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ and consider $s, t \in \mathcal{S}$. Then $s \sqsubseteq_{\mathrm{TrC}} t$ if and only if* $\mathrm{traces}(s) \le \mathrm{traces}(t)$ *and* $\mathrm{Ctraces}(s) \le \mathrm{Ctraces}(t)$.

Proof. The thesis follows as a direct consequence of $\mathrm{Ctraces}(s) \subset \mathrm{traces}(s)$, Lemma 4.29 and Theorem 4.25. ■

Interestingly, the same technique can be applied to obtain the compositionality results for the readies semantics. To this aim, we also need to slightly modify the auxiliary relation $\Vdash$. Let $X, Y \subseteq \mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A}) \times \mathbb{R}_{[0,1]}$, $a \in \mathcal{A}$, $\alpha \in \mathcal{A}^{\star}$ $A \in \mathcal{P}(\mathcal{A})$:

- ★ $X \Vdash (\alpha, A, q)$ if and only if either $(\alpha, A, q) \in X$ or $q = 0$ and $(\alpha, A, q') \not\in X$ for all $q' \in \mathbb{R}_{(0,1]}$;
- ★ $X + Y = \{(\alpha, A, q_1 + q_2) \mid X \Vdash (\alpha, A, q_1) \wedge Y \Vdash (\alpha, A, q_2)\}$;
- ★ $a.X = \{(a\alpha, A, q) \mid (\alpha, A, q) \in X\}$;
- ★ $q \cdot X = \{(\alpha, A, q \cdot q') \mid (\alpha, A, q') \in X\}$.

Notice that in the definition of $X + Y$ we have relaxed the constraint on the summation of the weights q_1, q_2' to be non-zero. This is due to the fact that to capture the failure semantics we need to allow those weights to be 0.

Firstly, we introduce the notion of *weighted ready pairs*.

Definition 4.25 (Weighted ready pairs). Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS. The set of functions $\mathrm{Rpairs}_i \colon \mathcal{S} \rightarrow 2^{\mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A}) \times \mathbb{R}_{(0,1]}}$ is defined for each $i \in \mathbb{N}$ inductively as follows:

- ★ $\mathrm{Rpairs}_0(s) = \{(\mathfrak{e}, \mathrm{init}(s), 1)\}$;
- ★ $\mathrm{Rpairs}_{i+1}(s) = \{(\mathfrak{e}, \mathrm{init}(s), 1)\} \cup \bigcup\limits_{s \xrightarrow{a} \pi} a.\big(\sum\limits_{s' \in \mathsf{supp}(\pi)} \pi(s') \cdot \mathrm{Rpairs}_i(s') \big)$.

We let $\mathrm{Rpairs}(s) = \bigcup\limits_{i \in \mathbb{N}} \mathrm{Rpairs}_i(s)$.

The definition of *weighted failure pairs* is more technical as, intuitively, we need to consider all possible failure sets.

Definition 4.26 (Weighted failure pairs)**.** Let $P = (\mathcal{S}, \mathcal{A}, \to)$ be a PTS. The set of functions $\mathrm{Fpairs}_i \colon \mathcal{S} \times \mathcal{P}(\mathcal{A}) \to 2^{\mathcal{A}^\star \times \mathcal{P}(\mathcal{A}) \times \mathbb{R}_{[0,1]}}$ is defined for each $i \in \mathbb{N}$ inductively as follows:

- ⋆ $\mathrm{Fpairs}_0(s, A) = \begin{cases} \{(\mathfrak{e}, A, 1)\} & \text{if } \mathrm{init}(s) \cap A = \emptyset \\ \{(\mathfrak{e}, A, 0)\} & \text{otherwise;} \end{cases}$

- ⋆ $\mathrm{Fpairs}_{i+1}(s) = \mathrm{Fpairs}_0(s, A) \cup \bigcup_{s \xrightarrow{a} \pi} a.\big(\sum_{s' \in \mathsf{supp}(\pi)} \pi(s') \cdot \mathrm{Fpairs}_i(s', A)\big).$

We let $\mathrm{Fpairs}(s, A) = \bigcup_{i \in \mathbb{N}} \mathrm{Fpairs}_i(s, A)$ and $\mathrm{Fpairs}(s) = \bigcup_{F \in \mathcal{P}(\mathcal{A})} \mathrm{Fpairs}(s, A)$.

Notice that differently from the previous notions of weighted decorated traces, weighted failure pairs also consider the failure sets to which a process assigns probability 0.

Given processes $s, t \in \mathcal{S}$, we write that $\mathrm{Rpairs}(s) \le \mathrm{Rpairs}(t)$ (resp. $\mathrm{Fpairs}(s) \le \mathrm{Fpairs}(t)$) if and only if whenever $(\alpha, A, q) \in \mathrm{Rpairs}(s)$ (resp. $\mathrm{Fpairs}(s)$) then there is a $(\alpha, A, q') \in \mathrm{Rpairs}(t)$ (resp. $\mathrm{Fpairs}(t)$) such that $q \le q'$.

Lemma 4.31. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$. For all $s \in \mathcal{S}, F \in \mathcal{P}(\mathcal{A})$ and $i \in \mathbb{N}$ it holds that $\mathrm{Rpairs}_i(s) \subseteq \mathrm{Rpairs}_{i+1}(s)$ and $\mathrm{Fpairs}_i(s, F) \subseteq \mathrm{Fpairs}_{i+1}(s, F)$.*

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.23 (Lemma 3.6 in [31]). ■

Lemma 4.32. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$. For all $s \in \mathcal{S}$, $\alpha \in \mathcal{A}^\star$, $A \in \mathcal{P}(\mathcal{A})$ and $q \in (0, 1]$ it holds that*

1. $(\alpha, A, q) \in \mathrm{Rpairs}(s)$ *if and only if there is a* $\mathcal{Z}_s \in \mathrm{Res}(s)$ *with* $\Pr(\mathcal{RC}(z_s, \alpha A)) = q$.
2. $(\alpha, A, q) \in \mathrm{Fpairs}(s)$ *if and only if there is a* $\mathcal{Z}_s \in \mathrm{Res}(s)$ *with* $\Pr(\mathcal{FC}(z_s, \alpha A)) = q$.

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.24 (Lemma 3.7 in [31]). ■

Theorem 4.33. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$ and consider $s, t \in \mathcal{S}$. Then*

1. $s \sqsubseteq_{\mathrm{R}} t$ *if and only if* $\mathrm{Rpairs}(s) \le \mathrm{Rpairs}(t)$.
2. $s \sqsubseteq_{\mathrm{F}} t$ *if and only if* $\mathrm{Fpairs}(s) \le \mathrm{Fpairs}(t)$.

Proof. The proof is analogous to the proof of Theorem 3.8 in [31] by exploiting Lemmas 4.31 and 4.32 in place of, resp., Lemmas 4.23 and 4.24. ■

Next, we deal with the case of ready trace equivalence which is obtained by modifying again the auxiliary relation $\Vdash$ in order to capture the ready traces. Let $X, Y \subseteq (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A})) \times \mathbb{R}_{(0,1]}$, $a \in \mathcal{A}$, $\mathfrak{A} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$:

- $X \Vdash (\mathfrak{A}, q)$ if and only if either $(\mathfrak{A}, q) \in X$ or $q = 0$ and $(\mathfrak{A}, q') \not\in X$ for all $q' \in \mathbb{R}_{(0,1]}$;
- $X + Y = \{(\mathfrak{A}, q_1 + q_2) \mid X \Vdash (\mathfrak{A}, q_1) \wedge Y \Vdash (\mathfrak{A}, q_2) \wedge q_1 + q_2 > 0\}$;
- $a.X = \{(a\mathfrak{A}, q) \mid (\mathfrak{e}\mathfrak{A}, q) \in X\}$;
- $[q, A] \cdot X = \{(\mathfrak{e}A\mathfrak{A}, q \cdot q') \mid (\mathfrak{A}, q') \in X\}$.

Definition 4.27 (Weighted ready traces)**.** Let $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ be a PTS. The set of functions $\mathrm{Rtraces}_i \colon \mathcal{S} \rightarrow 2^{(\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A})) \times \mathbb{R}_{(0,1]}}$ is defined for each $i \in \mathbb{N}$ inductively as follows:

- $\mathrm{Rtraces}_0(s) = \{(\mathfrak{e}(\mathrm{init}(s)), 1)\}$;
- $\mathrm{Rtraces}_{i+1}(s) = \{(\mathfrak{e}(\mathrm{init}(s)), 1)\} \cup \bigcup_{s \xrightarrow{a} \pi} a.\Big(\sum_{s' \in \mathsf{supp}(\pi)} [\pi(s'), \mathrm{init}(s')] \cdot \mathrm{Rtraces}_i(s')\Big)$.

We let $\mathrm{Rtraces}(s) = \bigcup_{i \in \mathbb{N}} \mathrm{Rtraces}_i(s)$.

Given processes $s, t \in \mathcal{S}$, we write that $\mathrm{Rtraces}(s) \le \mathrm{Rtraces}(t)$ if and only if whenever $(\mathfrak{A}, q) \in \mathrm{Rtraces}(s)$ then there is a $(\mathfrak{A}, q') \in \mathrm{Rtraces}(t)$ such that $q \le q'$.

Lemma 4.34. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$. For all $s \in \mathcal{S}$ and $i \in \mathbb{N}$ it holds that* $\mathrm{Rtraces}_i(s) \subseteq \mathrm{Rtraces}_{i+1}(s)$.

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.23 (Lemma 3.6 in [31]). ■

Lemma 4.35. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$. For all $s \in \mathcal{S}$, $\mathfrak{A} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ and $q \in (0,1]$ it holds that $(\mathfrak{A}, q) \in \mathrm{Rtraces}(s)$ if and only if there is a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ with* $\Pr(\mathcal{RC}(z_s, \mathfrak{A})) = q$.

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.24 (Lemma 3.7 in [31]). ■

Theorem 4.36. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \rightarrow)$ and consider $s, t \in \mathcal{S}$. Then $s \sqsubseteq_{\mathrm{TrR}} t$ if and only if* $\mathrm{Rtraces}(s) \le \mathrm{Rtraces}(t)$.

Proof. The proof is analogous to the proof of Theorem 3.8 in [31] by exploiting Lemmas 4.34 and 4.35 in place of, resp., Lemmas 4.23 and 4.24. ■

Finally, we consider the case of failure traces. Again, we need to adapt relation $\Vdash$ to the considered semantics.

Let $X, Y \subseteq \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})^\star \times \mathbb{R}_{[0,1]}$, $a \in \mathcal{A}$, $\alpha \in \mathcal{A}^\star$, $F \in \mathcal{P}(\mathcal{A})$, $\mathbb{F} \in \mathcal{P}(\mathcal{A})^\star$:

- $X \Vdash (\alpha, \mathbb{F}, q)$ if and only if either $(\alpha, \mathbb{F}, q) \in X$ or $q = 0$ and $(\alpha, \mathbb{F}, q') \not\in X$ for all $q' \in \mathbb{R}_{(0,1]}$;
- $X + Y = \{(\alpha, \mathbb{F}, q_1 + q_2) \mid X \Vdash (\alpha, \mathbb{F}, q_1) \wedge Y \Vdash (\alpha, \mathbb{F}, q_2)\}$;
- $a.X = \{(a\alpha, \mathbb{F}, q) \mid (\alpha, \mathbb{F}, q) \in X\}$;
- $q \cdot X = \{(\alpha, \mathbb{F}, q \cdot q') \mid (\alpha, \mathbb{F}, q') \in X\}$;

- ★ $(\mathfrak{e}, F, q) \cdot X = \{(\alpha, F\mathbb{F}, q \cdot q') \mid (\alpha, \mathbb{F}, q') \in X\}$.

Definition 4.28 (Weighted failure traces)**.** Let $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ be a PTS. The set of functions $\text{Ftraces}\colon \mathcal{S} \times \mathcal{P}(\mathcal{A})^{\star} \to 2^{\mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A})^{\star} \times \mathbb{R}_{[0,1]}}$ is defined inductively as follows:

- ★ $\text{Ftraces}(s, \emptyset) = \{(\mathfrak{e}, \emptyset, 1)\}$;
- ★ $\text{Ftraces}(s, F_1 \dots F_n) = \bigcup_{s \xrightarrow{a} \pi} a.\Big(\sum_{s' \in \mathsf{supp}(\pi)} \pi(s') \cdot \text{Fpairs}_0(s', F_1) \cdot \text{Ftraces}(s', F_2 \dots F_n)\Big)$.

We let $\text{Ftraces}(s) = \bigcup_{\mathbb{F} \in \mathcal{P}(\mathcal{A})^{\star}} \text{Ftraces}(s, \mathbb{F})$.

Given processes $s, t \in \mathcal{S}$, we write that $\text{Ftraces}(s) \leq \text{Ftraces}(t)$ if and only if whenever $(\alpha, \mathbb{F}, q) \in \text{Ftraces}(s)$ then there is a $(\alpha, \mathbb{F}, q') \in \text{Ftraces}(t)$ such that $q \leq q'$.

Lemma 4.37. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$. For all $s \in \mathcal{S}$ and $F_1, \dots, F_n, F \in \mathcal{P}(\mathcal{A})$ it holds that* $\text{Ftraces}(s, F_1 \dots F_n) \subseteq \text{Ftraces}(s, F_1 \dots F_n F)$.

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.23 (Lemma 3.6 in [31]). ■

Lemma 4.38. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$. For all $s \in \mathcal{S}$, $\alpha = a_1 \dots a_n \in \mathcal{A}^{\star}$, $\mathbb{F} = F_1 \dots F_n \in \mathcal{P}(\mathcal{A})^{\star}$ and $q \in (0,1]$ it holds that $(\alpha, \mathbb{F}, q) \in \text{Ftraces}(s)$ if and only if there is a resolution $\mathcal{Z}_s \in \text{Res}(s)$ with $\Pr(\mathcal{FC}(z_s, a_1 F_1 \dots a_n F_n)) = q$.*

Proof. The thesis follows by applying the same arguments used in the proof of Lemma 4.24 (Lemma 3.7 in [31]). ■

Theorem 4.39. *Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \longrightarrow)$ and consider $s, t \in \mathcal{S}$. Then $s \sqsubseteq_{\text{TrF}} t$ if and only if* $\text{Ftraces}(s) \leq \text{Ftraces}(t)$.

Proof. The proof is analogous to the proof of Theorem 3.8 in [31] by exploiting Lemmas 4.37 and 4.38 in place of, resp., Lemmas 4.23 and 4.24. ■

We are now ready to prove our Theorem 4.28.

Proof of Theorem 4.28. Let $x \in \{\text{TrC}, \text{F}, \text{TrF}, \text{R}, \text{TrR}\}$. First of all we notice that the proof of Theorem 4.28.2 is an immediate consequence of Theorem 4.28.1. In fact, we have that

$$s \sim_x t \overset{\text{Def.}}{\Longrightarrow} \begin{array}{l} s \sqsubseteq_x t \xRightarrow{\text{Thm.4.28.1}} (s,u) \sqsubseteq_x (t,u) \\ t \sqsubseteq_x s \xRightarrow{\text{Thm.4.28.1}} (t,u) \sqsubseteq_x (s,u) \end{array} \overset{\text{Def.}}{\Longrightarrow} (s,u) \sim_x (t,u).$$

Then the proof of Theorem 4.28.1 follows by applying similar arguments to those used in the proof of Theorem 3.9 in [31] by exploiting, in place of Theorem 3.8 in [31], respectively

- ★ Theorem 4.30 for $x = \text{TrC}$;
- ★ Theorem 4.33.1 for $x = \text{R}$;
- ★ Theorem 4.33.2 for $x = \text{F}$;

- ⋆ Theorem 4.36 for $x = \mathrm{TrR}$;
- ⋆ Theorem 4.39 for $x = \mathrm{TrF}$.

■

Probabilistic testing equivalence

Finally, we deal with the kernel of our testing metric. As argued in Section 4.3, we have adopted a *trace-by-trace* view of testing and thus the resulting *trace equivalence* is based on the same approach: for each process we consider the probability of passing a given test with respect to the execution of a single trace and then we compare those probabilities as done in the trace equivalence. Moreover, we will see that the, apparently, forced control on the probabilities of α-compatible maximal computations for an unsuccessful trace α, will actually lead to a testing equivalence which is fully backward compatible with the fully-nondeterministic case.

Definition 4.29 (Probabilistic testing equivalence)**.** Assume a PTS $P = (\mathcal{S}, \mathcal{A}, \to)$ and an NPT $O = (\mathcal{O}, \mathcal{A}, \to_O)$. Given $s, t \in \mathcal{S}$, we write $s \sqsubseteq_{\mathrm{test}} t$ if and only if for each test $o \in \mathcal{O}$ and trace $\alpha \in \mathcal{A}^\star$ it holds that

for each maximal resolution $\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)$ there is a maximal resolution $\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)$ such that $\Pr(\mathcal{SC}(z_{s,o},\alpha)) \le \Pr(\mathcal{SC}(z_{t,o},\alpha))$.

We say that $s,t \in \mathcal{S}$ are *probabilistic testing equivalent*, notation $s \sim_{\mathrm{test}} t$, if and only if it holds that $s \sqsubseteq_{\mathrm{test}} t$ and $t \sqsubseteq_{\mathrm{test}} s$.

The following Theorem formalizes the intuition that the proposed testing preorder and equivalence constitute the kernels of the respective testing premetric and semimetric.

Theorem 4.40. *For processes $s,t \in \mathcal{S}$ we have:*

1. $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0$ *if and only if* $s \sqsubseteq_{\mathrm{test}} t$, *and*
2. $\mathbf{d}_{\mathrm{test},\lambda}(s,t) = 0$ *if and only if* $s \sim_{\mathrm{test}} t$.

Proof.

1. We have

$$\begin{aligned}
& s \sqsubseteq_{\mathrm{test}} t \\
\iff & \forall\, o \in \mathcal{O}\ \forall\, \alpha \in \mathcal{A}^\star\ \forall\ \mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s)\ \exists\, \mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o) \text{ s.t.} \\
& \Pr(\mathcal{SC}(z_{s,o},\alpha)) \le \Pr(\mathcal{SC}(z_{t,o},\alpha)) \\
\iff & \forall\, o \in \mathcal{O}\ \forall\, \alpha \in \mathcal{A}^\star \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) \le \sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) \\
\iff & \forall\, o \in \mathcal{O}\ \forall\, \alpha \in \mathcal{A}^\star\ \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0 \\
\iff & \sup_{o \in \mathcal{O}} \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0 \\
\iff & \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0.
\end{aligned}$$

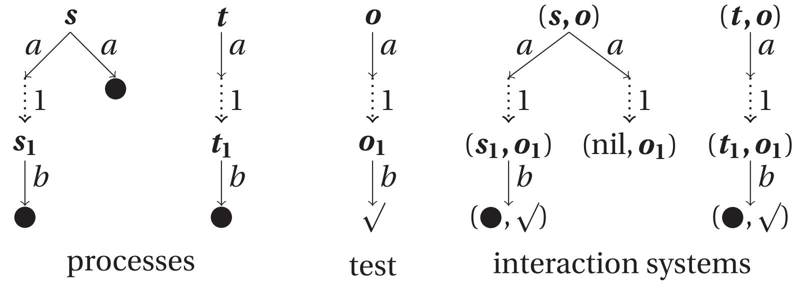


Figure 4.10: *We have that $s \not\sim_{\text{test}} t$ due to the restriction to* $\text{Res}_{\max,\alpha}$.

2. We have

$$\begin{aligned} s \sim_{\text{test}} t &\iff s \sqsubseteq_{\text{test}} t \text{ and } t \sqsubseteq_{\text{test}} s \\ &\iff \mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0 \text{ and } \mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(t,s) = 0 \qquad \text{(by Theorem 4.40.1)} \\ &\iff \max\{\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t), \mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(t,s)\} = 0 \\ &\iff \mathbf{d}_{\text{test},\lambda}(s,t) = 0. \end{aligned}$$

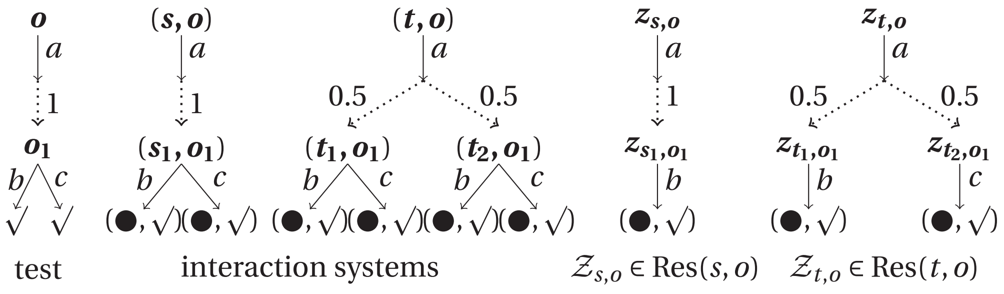
■

In the following Example we give a further intuition on the necessity of the restriction on the class of resolutions used in Definition 4.29 (as those in [31]) to guarantee full backward compatibility with the fully-nondeterministic case.

***Example* 4.13.** Consider processes s,t and the test o represented in Figure 4.10. Intuitively, the test o should discriminate s and t, since t will always pass the test o, whereas the maximal computation $(s,o) \overset{a}{\twoheadrightarrow} (\text{nil},o_1)$ does not reach success. Hence, if in Definition 4.29 we considered resolutions in $\text{Res}_{\max}$ instead of $\text{Res}_{\max,\alpha}$, then the probability $\Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0$ would be matched by the maximal resolution for (t,o), thus giving $s \sqsubseteq_{\text{test}} t$ and $s \sim_{\text{test}} t$. Conversely, the restriction to resolutions in $\text{Res}_{\max,a}$ allows us to distinguish the two processes with respect to their interaction with the test o. Indeed, we have $\text{Res}_{\max,a}(s,o) \neq \emptyset$ and $\text{Res}_{\max,a}(t,o) = \emptyset$, which directly gives $s \not\sqsubseteq_{\text{test}} t$. ◀

Again, Definition 4.29 can be seen as the relaxation of the corresponding definition in [31], which requires that $\Pr(\mathcal{SC}(z_{s,o},\alpha)) = \Pr(\mathcal{SC}(z_{t,o},\alpha))$ instead of $\Pr(\mathcal{SC}(z_{s,o},\alpha)) \le \Pr(\mathcal{SC}(z_{t,o},\alpha))$. Let $\approx_{\text{test}}$ denote the testing equivalence in [31]. Clearly we have $\approx_{\text{test}} \subseteq \sim_{\text{test}}$. The following Example show that the inclusion is strict.

***Example* 4.14.** Consider again processes s, t in Figure 4.8. We argue first that $s \sim_{\text{test}} t$. We start with noticing that, since s_1, t_1 and t_2 execute exactly the same actions, then no test can distinguish them. Then, $t \sqsubseteq_{\text{test}} s$ is due to the fact that for each test o the probability of (t,o) to execute ab or ac will always be matched by a resolution for (s,o). Conversely, $s \sqsubseteq_{\text{test}} t$ follows since whenever o tests trace ab (resp. ac) then there is a resolution for (s,o) assigning probability 1 to such a trace. Then, since both t_1 and t_2 pass the test of trace b (resp. c) we are guaranteed that it will be always possible to choose a resolution for (t,o) assigning



test

interaction systems

$\mathcal{Z}_{s,o} \in \mathrm{Res}(s,o)$

$\mathcal{Z}_{t,o} \in \mathrm{Res}(t,o)$

Figure 4.11: *A test showing that $s \not\approx_{\text{test}} t$ for s,t in Figure 4.8.*

probability 1 to trace ab (resp. ac). However, we get $s \not\approx_{\text{test}} t$ by the NPT test o in Figure 4.11. In fact, if we consider the resolution $\mathcal{Z}_{t,o} \in \mathrm{Res}(t,o)$ in Figure 4.11 for the interaction system (t,o), we have that $\Pr(\mathcal{SC}(z_{t,o},ab)) = 0.5$. The only resolution for (s,o) assigning non-zero success probability to the trace ab is resolution $\mathcal{Z}_{s,o}$ represented in Figure 4.11, for which $\Pr(\mathcal{SC}(z_{s,o},ab)) = 1$. Thus, resolution $\mathcal{Z}_{t,o} \in \mathrm{Res}(t,o)$ cannot be matched by any resolution for (s,o), thus giving $s \not\approx_{\text{test}} t$. Finally, we note that it is essential that $s \sim_{\text{test}} t$ to have $\sim \subseteq \sim_{\text{test}}$, since, as shown in Example 4.10, processes s,t in Figure 4.8 are probabilistic bisimilar. ◄

Relation $\sim_{\text{test}}$ is (like $\approx_{\text{test}}$) fully backward compatible with the testing equivalence on fully nondeterministic systems [62], denoted by $\sim^{\mathbf{N}}_{\text{test}}$, as well as with the one on fully probabilistic systems [51], denoted by $\sim^{\mathbf{P}}_{\text{test}}$. Then, $\sim_{\text{test}}$ is preserved by parallel composition.

Proposition 4.41. *Assume a PTS $P = (\mathcal{S},\mathcal{A},\longrightarrow)$ and processes $s,t \in \mathcal{S}$.*

1. *If P is fully-nondeterministic, then $s \sim_{\text{test}} t$ if and only if $s \approx_{\text{test}} t$ if and only if $s \sim^{\mathbf{N}}_{\text{test}} t$.*

2. *If P is fully-probabilistic, then $s \sim_{\text{test}} t$ if and only if $s \approx_{\text{test}} t$ if and only if $s \sim^{\mathbf{P}}_{\text{test}} t$.*

Proof.

1. $s \sim_{\text{test}} t \Leftrightarrow s \approx_{\text{test}} t$ follows since in the fully nondeterministic context, the execution probabilities of the traces are either 0 or 1. Thus, the inequality we use to check the testing equivalence of processes, in this setting becomes an equality and therefore it is no further distinguishable from the approach of [31]. Therefore, $s \sim_{\text{test}} t \Leftrightarrow s \sim^{\mathbf{N}}_{\text{test}} t$ follows by $s \approx_{\text{test}} t \Leftrightarrow s \sim^{\mathbf{N}}_{\text{test}} t$ (Theorem 5.4(1) in [31]) and transitivity.

2. $s \sim_{\text{test}} t \Leftrightarrow s \approx_{\text{test}} t$ follows since in the fully probabilistic context, each process has a single maximal resolution which is the process itself. Thus, the inequality we use to check the testing equivalence of processes, in this setting becomes an equality on the probabilities related to those maximal resolutions for processes (and tests) and therefore it is no further distinguishable from the approach of [31]. Therefore, $s \sim_{\text{test}} t \Leftrightarrow s \sim^{\mathbf{P}}_{\text{test}} t$ follows by $s \approx_{\text{test}} t \Leftrightarrow s \sim^{\mathbf{P}}_{\text{test}} t$ (Theorem 5.4(2) in [31]) and transitivity.

■

Theorem 4.42. *Assume a PTS $P = (\mathcal{S},\mathcal{A},\longrightarrow)$ and processes $s,t \in \mathcal{S}$.*

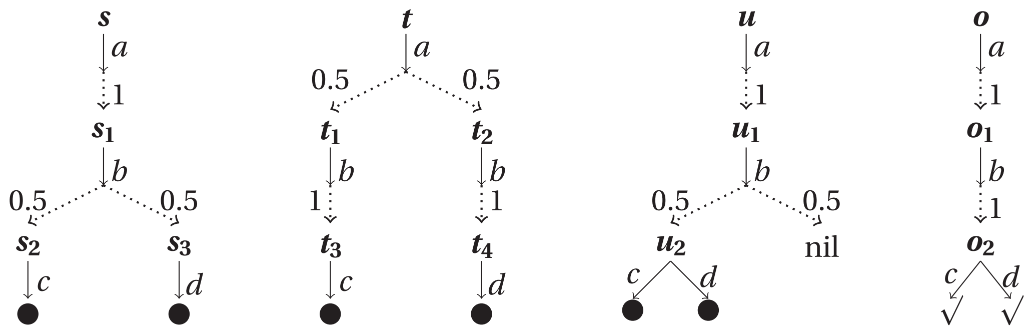


Figure 4.12: *s, t, u cannot simulate each other but $s \sim_{\mathrm{Tr}} t \sim_{\mathrm{Tr}} u$, $s \sim_{\mathrm{test}} t$ and $s \not\sim_{\mathrm{test}} u \not\sim_{\mathrm{test}} t$.*

1. *If $s \sqsubseteq_{\mathrm{test}} t$, then we have $s \parallel u \sqsubseteq_{\mathrm{test}} t \parallel u$ for all $u \in \mathcal{S}$.*

2. *If $s \sim_{\mathrm{test}} t$, then we have $s \parallel u \sim_{\mathrm{test}} t \parallel u$ for all $u \in \mathcal{S}$.*

Proof. The proof of both items follows by applying the same arguments used in the proof of Theorem 5.5 in [31] the idea being that given any test $o \in \mathcal{O}$ the interaction systems $(s \parallel u, o)$ and $(t \parallel u, o)$ can be rewritten respectively as $(s, u \parallel o)$ and $(t, u \parallel o)$, namely as a test for s and t. ■

4.6 A spectrum of probabilistic relations

In this section we show that the behavioral equivalences and preorders discussed so far can be partially ordered in a spectrum by the relation '*makes strictly less identifications than*', as represented in the lower half of Figure 4.1. Notice that part of this spectrum follows from the well-known relations on branching semantics $\sim \subset \sqsubseteq_{\mathrm{r}} \subset \sqsubseteq$.

Theorem 4.43. *1.* $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{TrF}} \subset \sqsubseteq_{\mathrm{F}} \subset \sqsubseteq_{\mathrm{test}} \subset \sqsubseteq_{\mathrm{Tr}}$.

2. $\sqsubseteq_{\mathrm{F}} \subset \sqsubseteq_{\mathrm{TrC}} \subset \sqsubseteq_{\mathrm{Tr}}$.

3. $\sqsubseteq \subset \sqsubseteq_{\mathrm{Tr}}$.

4. $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{TrR}}$ *and* $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{R}}$.

5. $\sim \subset \sim_{\mathrm{TrF}} \subset \sim_{\mathrm{F}} \subset \sim_{\mathrm{test}} \subset \sim_{\mathrm{Tr}}$.

6. $\sim_{\mathrm{F}} \subset \sim_{\mathrm{TrC}} \subset \sim_{\mathrm{Tr}}$.

7. $\sim \subset \sim_{\mathrm{TrR}}$ *and* $\sim \subset \sim_{\mathrm{R}}$.

Given any relation $\mathcal{R} \subset \mathcal{R}'$ in Theorem 4.43, the non strict version $\mathcal{R} \subseteq \mathcal{R}'$ follows by combining the quantitative analogous results in Section 4.4 giving the spectrum in the upper half of Figure 4.1, with the results in Sections 4.2 and 4.5 showing that each red dotted arrow in Figure 4.1 originating from a hemimetric (resp. pseudometric, premetric, semimetric) in the upper half in Figure 4.1 leads to the preorder (resp. equivalence) in the lower half being its kernel. Then, the strict version $\mathcal{R} \subset \mathcal{R}'$ follows from: (i) the non strict version $\mathcal{R} \subseteq \mathcal{R}'$, (ii) the full backward compatibility with the nondeterministic case (Propositions 4.21, 4.27, 4.41), (iii) the same and well-known result on fully-nondeterministic processes. However, by means of the processes in Figure 4.12, in the following examples we show that fully-probabilistic processes suffices to witness the strictness of most of these relations. In detail, Example 4.15 considers the relations in items 1–4 of Theorem 4.43 and Example 4.16 those in items 5–7 of Theorem 4.43. Moreover, Example 4.17 shows the isolation of similarity $\sqsubseteq$ with respect to the testing semantics and the decorated traces semantics. Example 4.18 shows the isolation of the completed trace equivalence $\sim_{\mathrm{TrC}}$ with respect to the testing equivalence $\sim_{\mathrm{test}}$. Finally, Example 4.19 deals with ready trace equivalence and ready equivalence by showing that they are isolated in the spectrum, with the only exceptions of ready similarity and bisimilarity given resp. by Theorem 4.43.4 and Theorem 4.43.7.

Proof of the non strict version of Theorem 4.43.

1.

$$\begin{tikzcd}
s \sqsubseteq_{\mathrm{r}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.3}"] & s \sqsubseteq_{\mathrm{TrF}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sqsubseteq_{\mathrm{F}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sqsubseteq_{\mathrm{test}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.40}"] & s \sqsubseteq_{\mathrm{Tr}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.20}"] \\
\mathbf{d}_{\mathrm{r},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.12}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.13}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.13}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.13}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)=0
\end{tikzcd}$$

2.

$$\begin{tikzcd}
s \sqsubseteq_{\mathrm{F}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sqsubseteq_{\mathrm{TrC}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sqsubseteq_{\mathrm{Tr}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.20}"] \\
\mathbf{d}_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.14}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.14}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)=0
\end{tikzcd}$$

3.

$$\begin{tikzcd}
s \sqsubseteq t \arrow[d, Leftrightarrow, "\text{Thm. 4.5}"] & s \sqsubseteq_{\mathrm{Tr}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.20}"] \\
\mathbf{d}_{\mathrm{s},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.15}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)=0
\end{tikzcd}$$

4.

$$\begin{tikzcd}
s \sqsubseteq_{\mathrm{r}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.3}"] & s \sqsubseteq_{\mathrm{TrR}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sqsubseteq_{\mathrm{r}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.3}"] & s \sqsubseteq_{\mathrm{R}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] \\
\mathbf{d}_{\mathrm{r},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.16}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)=0 & \mathbf{d}_{\mathrm{r},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.16}"'] & \mathbf{d}_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t)=0
\end{tikzcd}$$

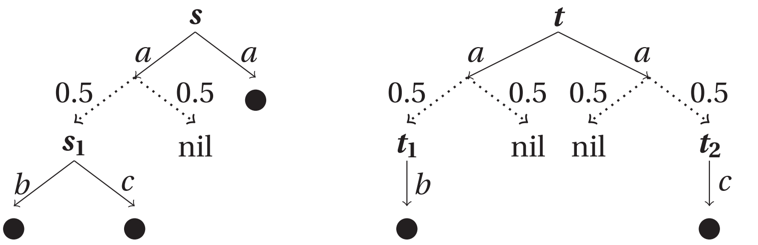


Figure 4.13: *Processes s, t are such that $s \not\sqsubseteq_{\mathrm{F}} t$ and $s \sim_{\mathrm{test}} t$.*

5.

$$\begin{tikzcd}
s \sim t \arrow[d, Leftrightarrow, "\text{Prop. 2.7}"] & s \sim_{\mathrm{TrF}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sim_{\mathrm{F}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sim_{\mathrm{test}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.40}"] & s \sim_{\mathrm{Tr}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.20}"] \\
\mathbf{d}_{\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.17}"'] & \mathbf{d}_{\mathrm{TrF},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.17}"'] & \mathbf{d}_{\mathrm{F},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.17}"'] & \mathbf{d}_{\mathrm{test},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.17}"'] & \mathbf{d}_{\mathrm{Tr},\lambda}(s,t)=0
\end{tikzcd}$$

6.

$$\begin{tikzcd}
s \sim_{\mathrm{F}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sim_{\mathrm{TrC}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sim_{\mathrm{Tr}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.20}"] \\
\mathbf{d}_{\mathrm{F},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.18}"'] & \mathbf{d}_{\mathrm{TrC},\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.18}"'] & \mathbf{d}_{\mathrm{Tr},\lambda}(s,t)=0
\end{tikzcd}$$

7.

$$\begin{tikzcd}
s \sim t \arrow[d, Leftrightarrow, "\text{Prop. 2.7}"] & s \sim_{\mathrm{TrR}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] & s \sim t \arrow[d, Leftrightarrow, "\text{Prop. 2.7}"] & s \sim_{\mathrm{R}} t \arrow[d, Leftrightarrow, "\text{Thm. 4.26}"] \\
\mathbf{d}_{\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.19}"'] & \mathbf{d}_{\mathrm{TrR},\lambda}(s,t)=0 & \mathbf{d}_{\lambda}(s,t)=0 \arrow[r, Rightarrow, "\text{Thm. 4.19}"'] & \mathbf{d}_{\mathrm{R},\lambda}(s,t)=0
\end{tikzcd}$$

■

***Example* 4.15.** Consider Figure 4.12. Firstly, we notice that neither $s \sqsubseteq t$ nor $t \sqsubseteq s$, thus implying $s \not\sqsubseteq_{\mathrm{r}} t$ and $t \not\sqsubseteq_{\mathrm{r}} s$. In fact, process s_1 can be (ready) simulated by neither t_1 nor t_2 and, analogously, s_1 cannot (ready) simulate any of those two processes. Conversely, we have $s \sim_x t$, for all $x \in \{\mathrm{Tr}, \mathrm{test}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$. To see $s \sim_{\mathrm{Tr}} t$ it is enough to notice that both s and t assign probability 1 to the (subtraces of) trace ab, and 0.5 to traces abc and abd, thus being indistinguishable by the trace semantics. Moreover, abc and abd are the only completed traces of s and t, thus giving $s \sim_{\mathrm{TrC}} t$. To derive that the equivalence holds also with respect to the testing semantics and the remaining decorated traces semantics, we can observe that on one hand no test can distinguish the b-action performed by s_1 from the b-actions performed by t_1 and t_2, and, moreover, the maximal probability of executing trace abc or trace abd in the two processes is always the same. On the other hand, we have that whenever processes s_2 and s_3 are distinguished by a decoration on the trace ab then t_3 and t_4 would be distinguished as well and consequently the probability assigned to the decorated trace by s and t will always be 0.5. Summarizing, we have obtained that

- ★ $s \not\sqsubseteq_{\mathrm{r}} t$ and $s \sim_{\mathrm{TrF}} t$, namely $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{TrF}}$;
- ★ $s \not\sqsubseteq_{\mathrm{r}} t$ and $s \sim_{\mathrm{R}} t$, namely $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{R}}$;
- ★ $s \not\sqsubseteq_{\mathrm{r}} t$ and $s \sim_{\mathrm{TrR}} t$, namely $\sqsubseteq_{\mathrm{r}} \subset \sqsubseteq_{\mathrm{TrR}}$;
- ★ $s \not\sqsubseteq t$ and $s \sim_{\mathrm{Tr}} t$, namely $\sqsubseteq \subset \sqsubseteq_{\mathrm{Tr}}$;

and we also remark that $s \not\sqsubseteq t$ but $s \sim_{\mathrm{TrC}} t$ and $s \sim_{\mathrm{test}} t$.

Next, let us study the relations between process s and process u. Since s and u assign probability 1 to the (subtraces of) trace ab and 0.5 to traces abc and abd, we conclude that $u \sim_{\mathrm{Tr}} s$. However, we have that $u \not\sqsubseteq_{x} s$ for any $x \in \{\mathrm{test}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$. To see $u \not\sqsubseteq_{\mathrm{test}} s$ we consider the interaction systems of the two processes with the test o in Figure 4.12. We have that $\mathrm{Res}_{\max,ab}(u,o) \neq \emptyset$, whereas $\mathrm{Res}_{\max,ab}(s,o) = \emptyset$. For $u \not\sqsubseteq_{\mathrm{TrC}} s$ it is enough to notice that ab is a completed trace for u but not for s. To obtain $u \not\sqsubseteq_{\mathrm{F}} s$ (resp: $u \not\sqsubseteq_{\mathrm{TrF}} s$; $u \not\sqsubseteq_{\mathrm{R}} s$; $u \not\sqsubseteq_{\mathrm{TrR}} s$) we consider the failure pair $ab\{cd\}$ (resp: the failure trace $a\emptyset b\{cd\}$; the ready pair $ab\{cd\}$; the ready trace $a\{b\}b\{cd\}$) to which u assigns probability 0.5, whereas s assigns to it probability 0 (we recall that although the probabilities of the traces are evaluated on the resolutions of nondeterminism for the two processes, the decorations are tested directly on the processes). In particular, we have obtained that

- ★ $u \not\sqsubseteq_{\mathrm{test}} s$ and $u \sim_{\mathrm{Tr}} s$, namely $\sim_{\mathrm{test}} \subset \sim_{\mathrm{Tr}}$;
- ★ $u \not\sim_{\mathrm{TrC}} s$ and $u \sim_{\mathrm{Tr}} s$, namely $\sim_{\mathrm{TrC}} \subset \sim_{\mathrm{Tr}}$.

Consider now processes s,t in Figure 4.13. We have $s \not\sqsubseteq_{\mathrm{F}} t$. In fact, if we consider the failure pair $\mathfrak{f} = a\{b,c\}$ and the resolution $\mathcal{Z}_s$ for s corresponding to its rightmost a-branch, we get $\Pr(\mathcal{FC}(z_s,\mathfrak{f})) = 1$. However, whichever resolution of nondeterminism $\mathcal{Z}_t$ we select for t we get that $\Pr(\mathcal{FC}(z_t,\mathfrak{f})) = 0.5$. Next we notice that $s \sim_{\mathrm{test}} t$. The only meaningful tests for s and t are those testing the traces a, ab and ac. Moreover, the structure of the two processes guarantees that for each test o we have $\mathrm{Res}_{\max,\alpha}(s,o) \neq \emptyset$ if and only if $\mathrm{Res}_{\max,\alpha}(t,o) \neq \emptyset$ for any $\alpha \in \{a, ab, ac\}$. It is then clear that the success probabilities assigned to this traces by the interaction systems of the two processes with the tests are always the same. Thus, we have obtained that

- ★ $s \not\sqsubseteq_{\mathrm{F}} t$ and $s \sim_{\mathrm{test}} t$, namely $\sqsubseteq_{\mathrm{F}} \subset \sqsubseteq_{\mathrm{test}}$.

The remaining strict inclusions $\sim_{\mathrm{TrF}} \subset \sim_{\mathrm{F}}$ and $\sim_{\mathrm{F}} \subset \sim_{\mathrm{TrC}}$ follow from the analogous relations in the fully nondeterministic spectrum [159]. ◀

***Example* 4.16.** We show only the strictness of the relations involving bisimilarity. The others trivially follow by the same arguments used in Example 4.15 on the corresponding preorders. To this aim, consider processes s,t in Figure 4.12. We have already argued in Example 4.15 that $s \sim_{\mathrm{TrF}} t$, $s \sim_{\mathrm{R}} t$ and $s \sim_{\mathrm{TrR}} t$. Moreover, we showed that s and t cannot simulate each other, thus implying that $s \not\sim t$. Therefore, we can immediately conclude that:

- ★ $s \not\sim t$ and $s \sim_{\mathrm{TrF}} t$, namely $\sim \subset \sim_{\mathrm{TrF}}$;
- ★ $s \not\sim t$ and $s \sim_{\mathrm{R}} t$, namely $\sim \subset \sim_{\mathrm{R}}$;

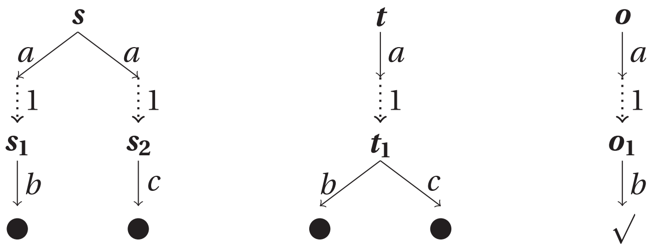


Figure 4.14: *Processes s, t are such that $s \not\sqsubseteq_{\text{test}} t$ and $s \sim_{\text{TrC}} t$.*

- ★ $s \not\sim t$ and $s \sim_{\text{TrR}} t$, namely $\sim \subset \sim_{\text{TrR}}$.

◀

***Example* 4.17.** Consider first processes s, t in Figure 4.10. Clearly, it holds that $s \sqsubseteq t$, but, as shown in Example 4.13, we have that $s \not\sqsubseteq_{\text{test}} t$ and moreover it holds that $s \not\sqsubseteq_{\text{TrC}} t$, $s \not\sqsubseteq_{\text{R}} t$ and $s \not\sqsubseteq_{\text{TrR}} t$. These immediately follow by noticing that the trace $\alpha = a$ is a completed trace for s but not for t and thus also the ready pair (resp. trace) $a\emptyset$ distinguishes them. Consider now processes s, t in Figure 4.12. As shown in Example 4.15, we have that $s \not\sqsubseteq t$ and $t \not\sqsubseteq s$ whereas it holds that $s \sim_{\text{test}} t$ and $s \sim_x t$ for $x \in \{\text{TrC}, \text{TrF}, \text{R}, \text{TrR}\}$. These, together with the already established relations in the spectrum, allow us to conclude that

- ★ $\sqsubseteq$ and $\sqsubseteq_{\text{test}}$ are incomparable;
- ★ $\sqsubseteq$ and $\sqsubseteq_x$ are incomparable for $x \in \{\text{TrC}, \text{F}, \text{TrF}, \text{R}, \text{TrR}\}$.

◀

***Example* 4.18.** Consider processes s, t in Figure 4.13. In Example 4.15 we showed that $s \sim_{\text{test}} t$. However we have that $s \not\sim_{\text{TrC}} t$. In fact, if we consider the completed trace $\alpha = a$, the resolution corresponding to the rightmost a-branch of s assigns probability 1 to it, whereas all the resolutions for t assign at most probability 0.5 to α. Finally, consider processes s, t in Figure 4.14. Clearly, we have that $s \sim_{\text{TrC}} t$. However, we also have that $s \not\sim_{\text{test}} t$. If we consider the interaction systems of processes s, t with the test o in the same Figure, we get that $\text{Res}_{\max,a}(s, o) \neq \emptyset$ and $\text{Res}_{\max,a}(t, o) = \emptyset$, from which we can directly conclude that $s \not\sim_{\text{test}} t$. Therefore we can conclude that

- ★ $\sim_{\text{TrC}}$ and $\sim_{\text{test}}$, are incomparable.

◀

The major difference between our spectrum and the linear time - branching time spectrum of [159] is in the isolation of the two readiness semantics, as already pointed out in [29]. This distinction is mainly due to the close interaction of nondeterminism and probability typical of the PTS model. In the fully-nondeterministic case it holds that $\sqsubseteq^{\mathbf{N}}_{\text{TrR}} \subset \sqsubseteq^{\mathbf{N}}_{\text{R}}$ [159], the intuition being that whenever there is a sequence of processes reachable from a process s such that s admits the ready trace $a_1 R_1 \dots a_n R_n$, then we are automatically guaranteed that s

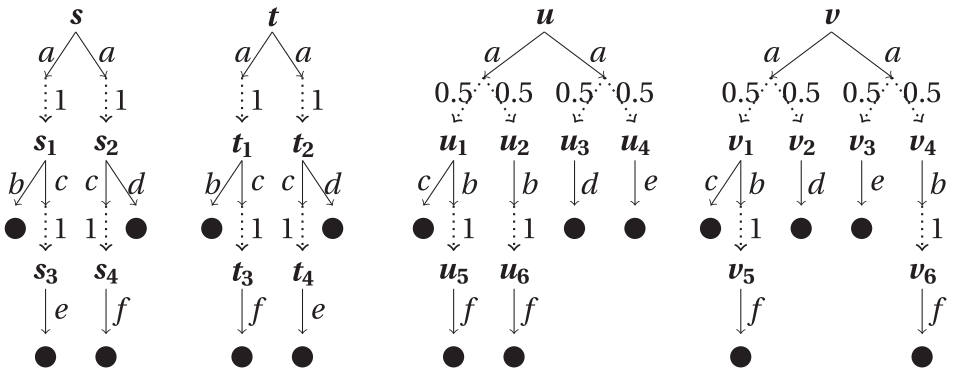


Figure 4.15: *Processes s, t are such that $s \not\sim_{\mathrm{TrR}} t$ and $s \sim_{\mathrm{R}} t$, whereas u, v are such that $u \not\sim_{\mathrm{R}} v$ and $u \sim_{\mathrm{TrR}} v$.*

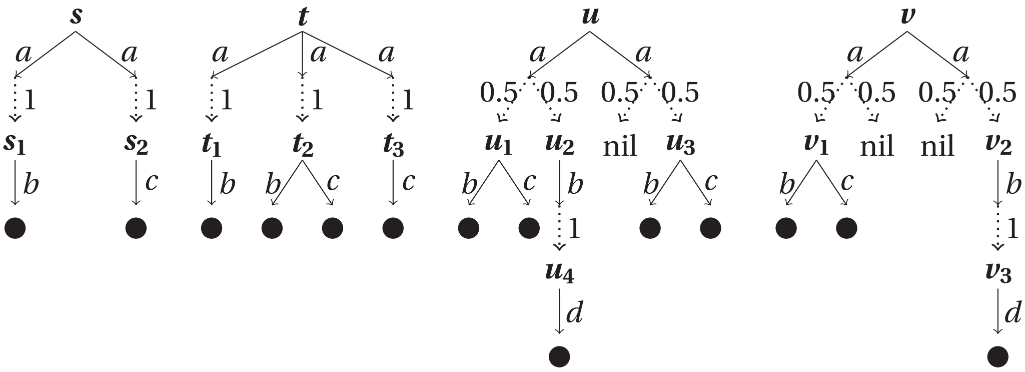


Figure 4.16: *The four processes show the isolation of the readies semantics with respect to the other decorated traces semantics and testing semantics.*

admits the ready pair $a_1 \dots a_n R_n$ as well as any process t such that $s \sqsubseteq^{\mathbf{N}}_{\mathrm{TrR}} t$ will do. When also probability is taken into account this relation between the two readiness semantics does not hold anymore, since as shown in the following Example 4.19, from the comparison of the probabilities of all the ready traces related to the trace $a_1 \dots a_n$ and a ready set R_n, we cannot derive any information on the result of the comparison of the probabilities assigned to the ready pair $a_1 \dots a_n R_n$ in which the probabilities of the aforementioned ready traces maybe summed up or not depending on nondeterminism. Analogously, the coexistence of nondeterminism and probability isolates the two readiness semantics for the other decorated trace semantics in the spectrum.

***Example* 4.19.** Consider first processes s, t in Figure 4.15. We have that $s \not\sim_{\mathrm{TrR}} t$. If we consider the ready trace $\mathfrak{R} = a\{b,c\}c\{e\}$ we have that s has a resolution executing it with probability 1, whereas t cannot execute it. However, we have that $s \sim_{\mathrm{R}} t$ as for the ready pair semantics a trace is executed and the ready set is checked only at the end of it. More precisely, we have that both processes have resolutions assigning probability 1 to the ready pairs $\mathfrak{r}_1 = a\{b,c\}$, $\mathfrak{r}_2 = a\{c,d\}$, $\mathfrak{r}_3 = ac\{e\}$ and $\mathfrak{r}_4 = ac\{f\}$. Consider now processes u, v in the same Figure. We have that $u \not\sim_{\mathrm{R}} v$. If we consider the ready pair $\mathfrak{r} = ab\{f\}$ we have that u has a resolution assigning probability 1 to it, whereas the resolutions for v assign at most probability 0.5 to it. However, we have that $u \sim_{\mathrm{TrR}} v$ since the presence of action c allows us to distinguish the b-move performed by u_1 from the b-move performed by u_2. In particular we have that the probability assigned to the ready trace $\mathfrak{R}_1 = a\{b,c\}b\{f\}$ is 0.5 in both processes and the probability assigned to the ready trace $\mathfrak{R}_2 = a\{b\}b\{f\}$ is 0.5 in both processes. Summarizing, we have obtained that

- ⋆ $\sim_{\mathrm{TrR}}$ and $\sim_{\mathrm{R}}$ are incomparable.

Next, consider processes s, t in Figure 4.16. We have that neither $s \sim_{\mathrm{TrR}} t$ nor $s \sim_{\mathrm{R}} t$. In fact if we consider the ready trace $\mathfrak{R} = a\{b,c\}$ (resp. the ready pair $\mathfrak{r} = a\{b,c\}$) then there is a resolution for t assigning probability 1 to it, whereas all the resolutions for s assign probability 0 to it. However, it is immediate to verify that $s \sim_{\mathrm{TrF}} t$. Finally, consider processes u, v in the same figure. We notice that $u \not\sim_{\mathrm{Tr}} v$. If we consider the trace $\alpha = ab$ we have that there is a resolution for u assigning probability 1 to α, whereas the resolutions for v assign at most probability 0.5 to it. However we have $u \sim_{\mathrm{TrR}} v$ and $u \sim_{\mathrm{R}} v$. This is due to the presence of action d, which allows us to distinguish the b-move performed by u_1 from the one performed by u_2. In conclusion, from these considerations and the already established relations in the spectrum, we get

- ⋆ $\sim_{\mathrm{TrR}}$ and $\sim_x$ are incomparable for $x \in \{\mathrm{Tr}, \mathrm{test}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}\}$;
- ⋆ $\sim_{\mathrm{R}}$ and $\sim_x$ are incomparable for $x \in \{\mathrm{Tr}, \mathrm{test}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}\}$.

◀

4.7 Concluding remarks

We have proposed a spectrum of behavioral distances on the PTS model (upper part of Figure 4.1), considering the bisimilarity metric [64, 72, 157] and novel notions of (hemi)metrics capturing the ready similarity, similarity, testing and (decorated) trace semantics.

We remark that ours is the first proposal of a quantitative analogue to the ready simulation, testing and decorated traces semantics. Moreover, our trace metric is novel with respect to existing ones [14, 43, 53, 59, 148] and its kernel, i.e. our probabilistic trace equivalence, satisfies several desirable properties as compositionality, full backward compatibility with the fully nondeterministic and the fully probabilistic cases and the compatibility with the (bi)simulation semantics. The same good properties hold also for our decorated trace and testing equivalence and, in fact, we have also shown that when we consider the kernels of the discussed (hemi)metrics, we obtain a probabilistic analogous of the linear time-branching time spectrum of [159] (lower part of Figure 4.1).

Ours is the first proposal of a spectrum for distances on PTSs.

In [77] a spectrum of distances defined as the generalization of a chosen trace distance is proposed. It is obtained by applying to LTSs the theory of quantitative Ehrenfeucht-Fraïssé games. However we remark that our results on PTSs cannot be obtained from the ones in [77] since the metric semantics considered are quite different and, moreover, the *well-behavedness* property assumed for the metrics in [77] does not hold for distances on PTSs.

In [59] distances for trace, simulation and bisimulation semantics on Metric Transition Systems (MTSs) have been proposed and ordered in a spectrum. Differently from PTSs, MTSs consist of a process-to-process transition relation and a set of atomic propositions which are evaluated on processes. Hence, all metrics in [59] are defined on a ground distance, called *propositional distance*, which quantifies the maximal distance on processes with respect to the evaluation of atomic propositions at them. Thus, this propositional distance depends on the metric that has been designed to measure the differences in the evaluation of atomic propositions. We also notice that, although our trace metric is technically different from the linear distances of [59], defined as the asymmetric Hausdorff distance on the propositional distance on traces, the idea behind their definition is quite similar. Moreover, it is worth noticing that by exploiting some properties of the (asymmetric) Hausdorff metric we can prove that it coincides with the (asymmetric) total variation distance (see [43] for a proof of this fact) by means of which we have defined our trace (hemi)metric.

In [14, 53] trace metrics on Markov Chains (MCs) are defined as total variation distances on the cones generated by traces. Our definition can be seen as a generalization of theirs: in MCs the transition probability function depends only on the current process and not on its nondeterministic properties. Thus, our quantification over resolutions would become trivial on MCs, resulting in a total variation distance over traces.

For what concerns the probabilistic relations, we have already compared them with the ones in [29–31] and the related spectra.

We recall that in [69] probabilistic similarity is shown to be equivalent to *probabilistic may testing*, in which the supremum success probabilities with respect to all resolutions of nondeterminism and to all tests are compared. Here, we have obtained an opposite result as we have proved that the two semantics are incomparable. This discrepancy is mainly due to the disparity between our testing semantics and the may testing of [69]. Although in both cases the suprema success probabilities are compared, in our semantics these are related to the execution of a single trace per time. Moreover, we have to deal with the additional requirements guaranteeing the full backward compatibility with the fully-nondeterministic case, which are too demanding for the simulation semantics.

The most exciting phrase to hear in science, the one that heralds the most discoveries, is not "Eureka! I found it!", but "That's funny..."

Isaac Asimov

CHAPTER 5

Logical Characterization of Branching Metrics

The aim of this Chapter is to propose a novel approach for the logical characterization of both behavioral distances and behavioral relations. In particular, we will focus on the branching part of the spectrum presented in Chapter 4, that is on (bi)similarity metrics and relations. Our targets are the following: 1. We aim to provide a characterization technique that can be easily generalized to other behavioral metrics and relations proposed in Chapter 4 and in the literature. This will be obtained by identifying for each process a special formula, called *mimicking formula*, expressing the relevant properties of it with respect to the considered semantics. Then, we transform the modal logic into a metric space by defining a *syntactical distance* on formulae. Finally, we define a *logical distance* on processes as the distance between their mimicking formulae: this logical distance characterizes the considered metric semantics. 2. We aim to use a simple boolean-valued modal logic to characterize both the behavioral relations and the (hemi)metrics related to them. Boolean-logics are preferable to the real-valued ones as they will allow for an easy generalization of our characterization technique. 3. We aim to establish whether two processes are related by a given behavioral relation by inspecting only a finite number of formulae. To this purpose, we will show how by comparing the mimicking formulae of two processes s and t for a particular semantics, one can infer whether s and t are related by the behavioral relation for that semantics.

Technically, in order to deal with (possibly) infinite execution sequences of processes, we follow the approach of [4, 120, 143], known as *equational* μ*-calculus*. Informally, a chosen class of formulae is enriched with *recursion*, namely a set of variables which allow for a recursive specification of modal properties. Then, an appropriate interpretation to each variable (called *model* in [120]) is provided as the solution of a system of equations defined using *endodeclarations*, namely functions mapping each variable into an arbitrary formula of the logic. More specifically, we will use endodeclarations to implicitly define a system of equations of the from

$$\gamma(X) = [\![\mathcal{E}(X)]\!]\gamma \tag{5.1}$$

whose solution will correspond to the proper variable interpretation for the formula: the interpretation γ is a solution for the system (5.1) if the semantics of X under γ corresponds to the interpretation of the formula assigned to X by the endodeclaration $\mathcal{E}$.

For our purposes, we will consider the boolean-valued modal logic $\mathcal{L}_{\mathcal{S}}$ obtained by extending the logic $\mathcal{L}$, introduced in Chapter 2.4, with a family of variables, one for each process in the set $\mathcal{S}$.

In [4] the equational μ-calculus is used as a general framework for the construction of characteristic formulae for behavioral equivalences and preorders in the non probabilistic setting. More precisely, it is proved that whenever a behavioral relation is obtained as the fixed point (equally greatest or least) of a monotone endofunction over the complete lattice of binary relations over processes, then the greatest interpretation of an endodeclaration expressing that endofunction can be viewed as the characteristic formula for its fixed-point (Theorem 2.16 [4]). In [143] this technique is applied in the probabilistic setting to obtain characteristic formulae for some behavioral relations on probabilistic automata. Here we use the equational μ-calculus to define a proper semantics for formulae in $\mathcal{L}_{\mathcal{S}}$, but we propose a different approach to obtain the characterization results.

In the case of the bisimilarity metric, the idea is the following: 1. For a process s we consider its mimicking formula, which captures the ability and the inability of s to execute any action and describes also its probabilistic behavior. Mimicking formulae will be defined as the images of the variables corresponding to processes through a particular endodeclaration $\mathcal{M}$ on $\mathcal{L}_{\mathcal{S}}$, called *mimicking endodeclaration*. 2. Then, we transform the modal logic into a metric space by introducing a notion of *syntactical distance* on formulae. This is a 1-bounded pseudometric assigning to each pair of formulae a suitable quantitative analogue of their syntactic disparities. In particular, we define the distance between distribution formulae as the Kantorovich lifting of the distance on the state formulae in their supports and the distance between conjunctions as the Hausdorff lifting of the distance on the state formulae in the two conjunctions. 3. We conclude by defining a *logical bisimulation distance* on processes corresponding to the distance between their mimicking formulae and proving that this logical distance characterizes the considered metric semantics.

Up to our knowledge, this is the first characterization of bisimilarity metric given by means of a boolean-valued logic and of a distance on the logic. Moreover, notice that the distance between two processes can be obtained by simply looking at their mimicking formulae, without analyzing any other formula in the logic.

Our results go even further. Along with mimicking formulae we introduce the *simulation characteristic formulae* of processes , namely the negation-free version of their mimicking formulae, defined through a proper endodeclaration $\mathcal{C}$ on $\mathcal{L}_{\mathcal{S}}$, called *simulation endodeclaration*. Then we show that by slightly modifying the notion of distance on $\mathcal{L}_{\mathcal{S}}$, we can characterize also the branching hemimetrics introduced in Chapter 4, namely the ready simulation metric and the simulation metric. More precisely, firstly we relax the distance on conjunctions to the asymmetric version of the Hausdorff lifting. Secondly, we define the *logical ready simulation distance* on processes as the modified distance between their mimicking formulae and we define the *logical simulation distance* on processes as the modified distance between their simulation characteristic formulae. Through these, we characterize the branching hemimetrics: the logical (ready) simulation distance coincides with the (ready) similarity metric. Being ready similarity metric new, ours is the first characterization for it. Moreover, to the best of our knowledge, this is the first characterization of similarity metric given by means of a boolean-valued logic and of a distance on the logic. Our approach can be easily extended to other notions of behavioral (hemi)metrics by tuning the notion of distance on formulae on a proper class of formulae. This will be further investigated in Chapter 6.

To strengthen our results, we show that the logic $\mathcal{L}_{\mathcal{S}}$ is *weak expressive* for probabilistic bisimilarity, meaning that two processes are probabilistic bisimilar if and only if their mimicking formulae are equivalent under a proper definition of structural equivalence over $\mathcal{L}_{\mathcal{S}}$. We cannot refer to this characterization of probabilistic bisimilarity as to an expressive one since the single mimicking formula of a process is not powerful enough to capture the whole equivalence class of the process, namely it is not its characteristic formula for probabilistic bisimilarity. We remark that the fully expressive characterization of bisimulation of [68] requires a logic much richer than $\mathcal{L}_{\mathcal{S}}$ (see Section 5.7 for a comparison). Finally, we show that the logic $\mathcal{L}_{\mathcal{S}}$ is expressive for probabilistic ready similarity, meaning that the mimicking formula of a process is its characteristic formula for this preorder, and that $\mathcal{L}_{\mathcal{S}}$ is expressive also for probabilistic similarity, meaning that the simulation characteristic formula of a process is its characteristic formula for probabilistic similarity. Up to our knowledge, this is the first paper in which a single logic is used to characterize both (bi)simulation metrics and classic notions of equivalence and preorders. As we will discuss in detail in Section 5.6, this is a great advantage since a modification in the expression of the considered behavioral metric would not affect the characterization of its kernel. Roughly speaking, there is no need of modifying the characterizing class of formulae, as conversely it would happen in the case of real-valued logics, but only the metric defined on it.

Summarizing, by means of our mimicking formulae for processes and logical distances on them we get:

1. Characterization of bisimilarity metric: we define a distance on formulae and we prove that the bisimulation distance between two processes equals the distance between their mimicking formulae (Theorem 5.24).

2. Characterization of (ready) similarity metric: by slightly relaxing the distance on formulae, we prove that the ready simulation distance between two processes equals the modified distance between their mimicking formulae (Theorem 5.33) and that the simulation distance between two processes equals the distance between their simulation characteristic formulae (Theorem 5.37).

3. Weak expressive characterization of probabilistic bisimilarity: we establish whether two processes are probabilistic bisimilar by simply comparing their mimicking formulae (Theorem 5.13).

4. Expressive characterization of ready probabilistic similarity by means of mimicking formulae (Theorem 5.14) and probabilistic similarity by means of simulation characteristic formulae (Theorem 5.17).

Organization of contents

In Section 5.1 we introduce the class of $\mathfrak{I}$-indexed logics $\mathcal{L}_{\mathfrak{I}}$, which includes $\mathcal{L}_{\mathcal{S}}$, and we proceed with some preliminary results on endodeclarations and on the equivalence of formulae (Section 5.2). In Section 5.3 we present mimicking formulae of processes and in Section 5.4 we use them for the characterization of probabilistic bisimilarity and (ready) similarity. In Section 5.5 we introduce the distance on $\mathcal{L}_{\mathcal{S}}$ through which we obtain the characterization of the bisimilarity metric which is then adapted in Section 5.6 to derive the characterizations of the ready similarity and similarity metrics. We end with some conclusions in Section 5.7.

A preliminary version of this chapter dealing with these characterizations of bisimulation metrics, probabilistic bisimilarity and (ready) similarity on *finite processes*, i.e. image finite process with finite execution sequences, can be found as [41].

5.1 The modal logic

In this Section we present the modal logic that we will use to characterize the (bi)simulation metrics and their kernels on image finite processes, for which we allow for infinite execution sequences, in a PTS $(\mathcal{S},\mathcal{A},\to)$.

We extend the modal logic $\mathcal{L}$ introduced in Chapter 2, which allows one to characterize probabilistic bisimilarity [66] and bisimilarity metric for finite processes [41], to the *modal $\mathfrak{I}$-indexed logic $\mathcal{L}_{\mathfrak{I}}$*, which considers also an $\mathfrak{I}$-indexed family of variables $\{X_{\mathfrak{i}} \mid \mathfrak{i} \in \mathfrak{I}\}$ allowing for a recursive specification of formulae.

Definition 5.1 (Modal $\mathfrak{I}$-indexed logic $\mathcal{L}_{\mathfrak{I}}$)**.** Let $\mathfrak{I}$ be a set of at most countable many identifiers. The classes of *modal $\mathfrak{I}$-indexed state formulae* $\mathcal{L}_{\mathfrak{I}}^{\mathrm{s}}$ and of *modal $\mathfrak{I}$-indexed distribution formulae* $\mathcal{L}_{\mathfrak{I}}^{\mathrm{d}}$ over $\mathcal{A}$ are defined by the following BNF-like grammar:

$$\begin{array}{llll}\mathcal{L}_{\mathfrak{I}}^{\mathrm{s}}: & \varphi & ::= & \top \mid X_{\mathfrak{i}} \mid \bar{a} \mid \bigwedge_{j\in\mathcal{J}} \varphi_j \mid \langle a\rangle\psi \\ \mathcal{L}_{\mathfrak{I}}^{\mathrm{d}}: & \psi & ::= & \bigoplus_{i\in I} r_i\varphi_i\end{array}$$

where: 1. $\mathfrak{i} \in \mathfrak{I}$; 2. $a \in \mathcal{A}$; 3. $\mathcal{J}$ is an at most countable set of indexes; 4. for each $j \in \mathcal{J}$ it holds $\varphi_j \neq \bigwedge_{i\in\mathcal{I}} \varphi_i$ for any set of indexes $\mathcal{I}$ with $|\mathcal{I}| > 1$ and $\varphi_j \neq X_{\mathfrak{i}}$ for any $\mathfrak{i} \in \mathfrak{I}$; 5. I is a finite set of indexes, $r_i \in (0,1]$ for all $i \in I$ and $\sum_{i\in I} r_i = 1$.

We note that the constraints in item (4) may be avoided by using a more complicated grammar. For sake of simplicity and readability, we opted for this formulation.

We shall write $\varphi_1 \wedge \varphi_2$ for $\bigwedge_{j\in\mathcal{J}} \varphi_j$ with $\mathcal{J} = \{1,2\}$, $r_1\varphi_1 \oplus r_2\varphi_2$ for $\bigoplus_{i\in I} r_i\varphi_i$ with $I = \{1,2\}$, and $\langle a\rangle\varphi$ for $\langle a\rangle \bigoplus_{i\in I} r_i\varphi_i$ with $I = \{i\}$, $r_i = 1$ and $\varphi_i = \varphi$. Notice that instead of using $\top$ we could use $\bigwedge_{\emptyset}$. We decided to use $\top$ to improve readability.

As usual, the semantics of a state (resp. distribution) formula is embodied in the set of processes (resp. distributions) which satisfy it. However, formulae in $\mathcal{L}_{\mathfrak{I}}$ may contain variables which could be satisfied potentially by any process. Thus the meaning of formulae in $\mathcal{L}_{\mathfrak{I}}$ is subject to *variable interpretations*, namely functions of the form $\gamma : \mathfrak{I} \to \mathcal{P}(\mathcal{S})$ mapping each identifier $\mathrm{i} \in \mathfrak{I}$ into the set of processes which are presumed to satisfy the variable X_{i}. Since each variable is univocally determined by an identifier in $\mathfrak{I}$, to improve readability we abuse of notation and write $\gamma(X_{\mathrm{i}})$ in place of $\gamma(\mathrm{i})$. In what follows, we let $\Gamma_{\mathfrak{I}}$ be the set of variable interpretations defined from $\mathfrak{I}$ to $\mathcal{P}(\mathcal{S})$. The elements of $\Gamma_{\mathfrak{I}}$ can be ordered by means of the ordering $\preceq$ induced by set inclusion, namely $\gamma_1 \preceq \gamma_2$ if and only if $\gamma_1(X_{\mathrm{i}}) \subseteq \gamma_2(X_{\mathrm{i}})$ for all $\mathrm{i} \in \mathfrak{I}$.

Lemma 5.1. *The set* $(\Gamma_{\mathfrak{I}}, \preceq)$ *of variable interpretations is a complete lattice.*

Proof. The proof follows by applying the same arguments used in the proof of Lemma 2.9(2) in [4]. ■

Since the ordering over $\Gamma_{\mathfrak{I}}$ is defined by means of set inclusion over $\mathcal{P}(\mathcal{S})$, we have that the join is defined by means of set union, namely $\left(\bigsqcup_{h \in H} \gamma_h\right)(X_{\mathrm{i}}) = \bigcup_{h \in H} \gamma_h(X_{\mathrm{i}})$, for all $\mathrm{i} \in \mathfrak{I}$. Similarly, the meet is given by set intersection, that is $\left(\bigsqcap_{h \in H} \gamma_h\right)(X_{\mathrm{i}}) = \bigcap_{h \in H} \gamma_h(X_{\mathrm{i}})$, for all $\mathrm{i} \in \mathfrak{I}$.

Next, we introduce the relation $\models_\gamma$ which asserts when a process s (resp. distribution π) satisfies the state formula φ (resp. distribution formula ψ) under a given variable interpretation γ.

Definition 5.2 (Satisfiability)**.** Assume a PTS $(\mathcal{S}, \mathcal{A}, \to)$. For each variable interpretation $\gamma \in \Gamma_{\mathfrak{I}}$, the satisfaction relation $\models_\gamma \subseteq (\mathcal{S} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}) \cup (\Delta(\mathcal{S}) \times \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}})$ is defined recursively as:

- $s \models_\gamma \top$ always;
- $s \models_\gamma X_{\mathrm{i}}$ iff $s \in \gamma(X_{\mathrm{i}})$;
- $s \models_\gamma \bar{a}$ iff $s \overset{a}{\not\to}$;
- $s \models_\gamma \bigwedge_{j \in \mathcal{J}} \varphi_j$ iff $s \models_\gamma \varphi_j$ for all $j \in \mathcal{J}$;
- $s \models_\gamma \langle a \rangle \psi$ iff $s \xrightarrow{a} \pi$ for a distribution $\pi \in \Delta(\mathcal{S})$ such that $\pi \models_\gamma \psi$;
- $\pi \models_\gamma \bigoplus_{i \in I} r_i \varphi_i$ iff $\pi = \sum_{i \in I} r_i \pi_i$ for a family $\{\pi_i\}_{i \in I} \subseteq \Delta(\mathcal{S})$ of distribution such that, for all $i \in I$, whenever $s \in \mathsf{supp}(\pi_i)$ then $s \models_\gamma \varphi_i$.

***Example* 5.1.** Assume $\gamma \in \Gamma_{\mathfrak{I}}$. Consider a process $s \in \mathcal{S}$ with $\mathrm{der}(s,a) = \{\pi\}$, where $\pi = \frac{1}{2}\delta_{s_1} + \frac{1}{2}\delta_{s_2}$. Assume that, for the chosen variable interpretation, it holds that

$$s_1 \models_\gamma \langle b \rangle \top \qquad s_1 \models_\gamma \bar{a} \qquad s_2 \models_\gamma \langle b \rangle \top \qquad s_2 \models_\gamma \langle a \rangle \top.$$

Consider the formula $\varphi \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ defined, for a given index $\mathrm{i} \in \mathfrak{I}$, as

$$\varphi = \langle a \rangle \Big(\frac{2}{3} X_{\mathrm{i}} \oplus \frac{1}{3} \big(\langle a \rangle \top \wedge \langle b \rangle \top \big) \Big).$$

We have that $s \models_\gamma \varphi$ if and only if $\{s_1, s_2\} \subseteq \gamma(X_\mathfrak{i})$. In fact we have that s_2 is the only process in the support of π that satisfies the formula $\langle a \rangle \top$, thus implying that to have $s \models_\gamma \varphi$ we need the convex combination

$$\pi = \frac{2}{3}\Big(\frac{3}{4}\delta_{s_1} + \frac{1}{4}\delta_{s_2}\Big) + \frac{1}{3}\delta_{s_2}$$

such that

- ⋆ for each process s' in the support of $\frac{3}{4}\delta_{s_1} + \frac{1}{4}\delta_{s_2}$ it holds that $s' \models_\gamma X_\mathfrak{i}$;
- ⋆ for each process s' in the support of δ_{s_2} it holds that $s' \models_\gamma \langle a \rangle \top \wedge \langle b \rangle \top$.

Therefore, to guarantee that $s \models_\gamma \varphi$ we should guarantee that for each process s' in the union of the supports of δ_{s_1} and δ_{s_2} it holds that $s' \models_\gamma X_\mathfrak{i}$, that is it must be the case that both $s_1 \models_\gamma X_\mathfrak{i}$ and $s_2 \models_\gamma X_\mathfrak{i}$, namely $\{s_1, s_2\} \subseteq \gamma(X_\mathfrak{i})$. ◀

The meaning of a state (resp. distribution) formula is represented by the set of processes (resp. probability distributions) satisfying it.

Definition 5.3 (Semantics of $\mathcal{L}_\mathfrak{I}$)**.** For each state formula φ in $\mathcal{L}_\mathfrak{I}^{\mathrm{s}}$ we define $[\![\varphi]\!] \colon \Gamma_\mathfrak{I} \to \mathcal{P}(\mathcal{S})$ by

$$[\![\varphi]\!]\gamma = \{s \in \mathcal{S} \mid s \models_\gamma \varphi\}$$

and for each distribution formula ψ in $\mathcal{L}_\mathfrak{I}^{\mathrm{d}}$ we define $[\![\psi]\!] \colon \Gamma_\mathfrak{I} \to \mathcal{P}(\Delta(\mathcal{S}))$ by

$$[\![\psi]\!]\gamma = \{\pi \in \Delta(\mathcal{S}) \mid \pi \models_\gamma \psi\}.$$

Notice that, in particular, for all $\gamma \in \Gamma_\mathfrak{I}$ we have

- ⋆ $[\![\top]\!]\gamma = \mathcal{S}$;
- ⋆ $[\![X_\mathfrak{i}]\!]\gamma = \gamma(X_\mathfrak{i})$ for each $\mathfrak{i} \in \mathfrak{I}$;
- ⋆ $[\![\bigwedge_{j \in \mathcal{J}} \varphi_j]\!]\gamma = \bigcap_{j \in \mathcal{J}} [\![\varphi_j]\!]\gamma$;
- ⋆ $[\![\langle a \rangle \psi]\!]\gamma = \bigcup_{\pi \in [\![\psi]\!]\gamma} \{s \in \mathcal{S} \mid s \xrightarrow{a} \pi\}$.

The following result, which proves the monotonicity of the mapping $[\![\,]\!]$ w.r.t. variable interpretations in $\Gamma_\mathfrak{I}$, has been adapted from [4] to fit the probabilistic framework.

Proposition 5.2. *For any $\varphi \in \mathcal{L}_\mathfrak{I}^{\mathrm{s}}$, the mapping $[\![\varphi]\!] \colon \Gamma_\mathfrak{I} \to \mathcal{P}(\mathcal{S})$ is monotone. Analogously, for any $\psi \in \mathcal{L}_\mathfrak{I}^{\mathrm{d}}$, the mapping $[\![\psi]\!] \colon \Gamma_\mathfrak{I} \to \mathcal{P}(\Delta(\mathcal{S}))$ is monotone.*

Proof. Let $\gamma_1, \gamma_2 \in \Gamma_\mathfrak{I}$ be two variable interpretations such that $\gamma_1 \preceq \gamma_2$. We prove that $[\![\phi]\!]\gamma_1 \subseteq [\![\phi]\!]\gamma_2$ by structural induction over $\phi \in \mathcal{L}_\mathfrak{I}^{\mathrm{s}} \cup \mathcal{L}_\mathfrak{I}^{\mathrm{d}}$. We show only the inductive step of the diamond modality, which is the only one that differs from the proof of Lemma 2.9(1) in [4]. Consider $\phi = \langle a \rangle \psi$, with $\psi = \bigoplus_{i \in I} r_i \varphi_i$. We have

$$\begin{aligned}
& [\![\langle a \rangle \psi]\!]\gamma_1 \\
&= \{s \in \mathcal{S} \mid s \models_{\gamma_1} \langle a \rangle \psi\} \\
&= \{s \in \mathcal{S} \mid s \xrightarrow{a} \pi \text{ and } \pi \models_{\gamma_1} \psi\}
\end{aligned}$$

$$
\begin{aligned}
&= \{s \in \mathcal{S} \mid s \xrightarrow{a} \pi, \pi = \sum_{i \in I} r_i \pi_i \text{ and for each } s' \in \mathsf{supp}(\pi_i),\ s' \models_{\gamma_1} \varphi_i\} \\
&\subseteq \{s \in \mathcal{S} \mid s \xrightarrow{a} \pi, \pi = \sum_{i \in I} r_i \pi_i \text{ and for each } s' \in \mathsf{supp}(\pi_i),\ s' \models_{\gamma_2} \varphi_i\} \\
&= \{s \in \mathcal{S} \mid s \xrightarrow{a} \pi \text{ and } \pi \models_{\gamma_2} \psi\} \\
&= \{s \in \mathcal{S} \mid s \models_{\gamma_2} \langle a \rangle \psi\} \\
&= [\![\langle a \rangle \psi]\!] \gamma_2
\end{aligned}
$$

where the inclusion holds by structural induction. Notice that this case includes also the proof for distribution formulae. ■

As a final remark to this Section, we notice that monotonicity of $[\![\,]\!]$ is guaranteed since no variable can occur in the scope of negation (see Section 5.7 for a further discussion on this issue and our choice on negation).

Interpretation through declarations

Definition 5.3 states that formulae in $\mathcal{L}_{\mathfrak{J}}$ can be interpreted only with respect to a given variable interpretation. Thus, our next task is to establish a criterion to obtains suitable interpretations for variables. To this aim, we follow the *equational μ-calculus* approach [4, 120, 143], in which the desired interpretation is provided as the solution of a system of equations defined using *endodeclarations*. Informally speaking, an endodeclaration is a function ascribing a state formula to each identifier, and consequently to each variable.

Definition 5.4 (Endodeclaration)**.** An *endodeclaration* on $\mathcal{L}_{\mathfrak{J}}$ is a mapping $\mathcal{E} \colon \mathfrak{J} \to \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$.

As for variable interpretations, for sake of readability we abuse of notation and write $\mathcal{E}(X_{\mathrm{i}})$ in place of $\mathcal{E}(\mathrm{i})$. Moreover, we remark that since variables belong to the syntactic class of state formulae, then also the formulae assigned to them by endodeclarations are required to be state formulae.

The meaning of an endodeclaration is a mapping from variable interpretations to variable interpretations.

Definition 5.5 (Semantics of endodeclarations)**.** For an endodeclaration $\mathcal{E} \colon \mathfrak{J} \to \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$, we define $\langle\!\langle \mathcal{E} \rangle\!\rangle \colon \Gamma_{\mathfrak{J}} \to \Gamma_{\mathfrak{J}}$ as the mapping such that, for all $\mathrm{i} \in \mathfrak{J}$ and $\gamma \in \Gamma_{\mathfrak{J}}$, we have

$$(\langle\!\langle \mathcal{E} \rangle\!\rangle \gamma)(X_{\mathrm{i}}) := [\![\mathcal{E}(X_{\mathrm{i}})]\!] \gamma.$$

In order to duly interpreting variables, we consider the interpretations of the formulae assigned to them by the endodeclaration. More specifically, we will use endodeclarations to implicitly define a system of equations

$$\gamma(X_{\mathrm{i}}) = [\![\mathcal{E}(X_{\mathrm{i}})]\!] \gamma \text{ for } \mathrm{i} \in \mathfrak{J} \tag{5.2}$$

whose solution will correspond to the proper variable interpretation for the formula: $\gamma \in \Gamma_{\mathfrak{J}}$ is a solution for the system (5.2) if the semantics of X_{i} under γ (namely $[\![X_{\mathrm{i}}]\!] \gamma$, which, by definition, is $\gamma(X_{\mathrm{i}})$) corresponds to the interpretation of the formula $\mathcal{E}(X_{\mathrm{i}})$.

Notice that a variable interpretation $\gamma \in \Gamma_{\mathfrak{J}}$ being fixed point of $\langle\langle \mathcal{E} \rangle\rangle$ guarantees that

$$[\![X_{\mathfrak{i}}]\!] \gamma = [\![\mathcal{E}(X_{\mathfrak{i}})]\!] \gamma$$

for all $\mathfrak{i} \in \mathfrak{J}$, and is therefore a good candidate to be used as variable interpretation.

It is now necessary that the set of fixed points of $\langle\langle \mathcal{E} \rangle\rangle$ is not empty. We show that $\langle\langle \mathcal{E} \rangle\rangle$ is monotone, then, since $(\Gamma_{\mathfrak{J}}, \preceq)$ is a complete lattice (Lemma 5.1), we will conclude that $\langle\langle \mathcal{E} \rangle\rangle$ has the least and greatest fixed points.

Proposition 5.3. *The function $\langle\langle \mathcal{E} \rangle\rangle$ is monotone and thus it has the least and greatest fixed point.*

Proof. The proof follows by applying the same arguments used in the proof of Lemma 2.12 in [4]. ■

The variable interpretation that we will use to interpret formulae is the greatest fixed point of $\langle\langle \mathcal{E} \rangle\rangle$, denoted by $\nu_{\mathcal{E}}$.

Next, we propose an alternative inductive characterization of $\nu_{\mathcal{E}}$, which will be useful to prove some of the results in the following sections.

As first step, we show that the function $\langle\langle \mathcal{E} \rangle\rangle$ is Scott-co-continuous (see Remark 2.1).

Proposition 5.4. *The function $\langle\langle \mathcal{E} \rangle\rangle$ is Scott-co-continuous.*

Proof. Let H be an arbitrary set of, at most, countably many indexes. Let $\{\gamma_h\}_{h \in H}$ be a descending chain on $\Gamma_{\mathfrak{J}}$. We aim to show that

$$\langle\langle \mathcal{E} \rangle\rangle \bigsqcap_{h \in H} \gamma_h = \bigsqcap_{h \in H} \langle\langle \mathcal{E} \rangle\rangle \gamma_h$$

which by the definitions of $\langle\langle \mathcal{E} \rangle\rangle$ (Definition 5.5) and $\bigsqcap$ over $\Gamma_{\mathfrak{J}}$ is equivalent to

$$\text{for each } \mathfrak{i} \in \mathfrak{J}: \quad [\![\mathcal{E}(X_{\mathfrak{i}})]\!] \bigsqcap_{h \in H} \gamma_h = \bigcap_{h \in H} [\![\mathcal{E}(X_{\mathfrak{i}})]\!] \gamma_h. \tag{5.3}$$

To prove Equation 5.3 we proceed by structural induction over $\mathcal{E}(X_{\mathfrak{i}})$. Here we present only the case for the diamond operator, since the proofs for the other cases are immediate from the definition of $[\![\,]\!]$.

Let $\mathcal{E}(X_{\mathfrak{i}}) = \langle a \rangle \bigoplus_{i \in I} r_i \varphi_i$. We proceed by showing the two inclusions separately. By the monotonicity of $[\![\,]\!]$ (Proposition 5.2) we can immediately infer that

$$[\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \bigsqcap_{h \in H} \gamma_h \subseteq \bigcap_{h \in H} [\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \gamma_h \tag{5.4}$$

Let us show now that

$$\bigcap_{h \in H} [\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \gamma_h \subseteq [\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \bigsqcap_{h \in H} \gamma_h. \tag{5.5}$$

Consider any $s \in \bigcap_{h \in H} [\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \gamma_h$, namely $s \in [\![\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i]\!] \gamma_h$ for each $h \in H$. This implies that for each $h \in H$ there exists a probability distribution π_h such that $s \xrightarrow{a} \pi_h$ and

$\pi_h = \sum_{i\in I} r_i \pi_i^h$ for some $\pi_i^h \in \Delta(\mathcal{S})$ such that for each $t \in \mathsf{supp}(\pi_i^h)$ it holds that $t \in [\![\varphi_i]\!]\gamma_h$. Since H is a countable set and $\mathrm{der}(s,a)$ is finite, we can infer that there is a probability distribution $\pi \in \mathrm{der}(s,a)$ which satisfies the formula $\bigoplus_{i\in I} r_i\varphi_i$ w.r.t. γ_h for countably many indexes $h \in H$. Moreover, considering that π has finite support, we have that for each $i \in I$ we can define the set $S_i = \{t \in \mathsf{supp}(\pi) \mid t \in [\![\varphi_i]\!]\gamma_h$ for countably many $h \in H\}$. As $\{\gamma_h\}_{h\in H}$ is a descending chain on $\Gamma_{\mathfrak{J}}$ we obtain that $t \in [\![\varphi_i]\!]\gamma_h$ for countably many $h \in H$ implies $t \in [\![\varphi_i]\!]\gamma_h$ for all $h \in H$, that is $t \in \bigcap_{h\in H}[\![\varphi_i]\!]\gamma_h$ for each $t \in S_i$. By structural induction over the φ_i, this implies that each $t \in S_i$ is such that $t \in [\![\varphi_i]\!]\bigsqcap_{h\in H}\gamma_h$. Hence, to conclude we need to show that we can rewrite π as $\sum_{i\in I} r_i\pi_i$ where the support of each distribution π_i is given by the related set S_i. This can be easily derived by defining the family of probability distributions $\{\pi_i\}_{i\in I}$ satisfying $\pi = \sum_{i\in I} r_i\pi_i$ and $\pi_i(t) > 0$ implies $t \models \varphi_i$ as the solution of the following linear system of equations, for each $i \in I$:

$$\begin{cases} \pi_i(t) = 0 & \text{if } t \notin S_i \\ \pi_i(t) = \dfrac{\pi(t)}{r_i} & \text{if } t \in S_i \text{ and } t \notin S_j \text{ for each } j \in I, j \neq i \\ \pi_i(t) = \dfrac{1}{r_i}\Big(\pi(t) - \sum\limits_{j\in I, j\neq i} r_j\pi_j(t)\Big) & \text{otherwise.} \end{cases}$$

By construction I is finite, so let C_1 be the cardinality of I, analogously $\mathsf{supp}(\pi)$ is finite and let C_2 be its cardinality. Notice that the system above has C_2 equations for each $i \in I$, and therefore we have a total of $C_1 \cdot C_2$ equations. We have that the number of unknowns in our system is also $C_1 \cdot C_2$. In fact for each $i \in I$ we need to establish the value of $\pi_i(t)$ for each $t \in \mathsf{supp}(\pi)$. Finally, notice that we can rewrite all equations of the system in the general form $\pi_i(t) = \dfrac{1}{r_i}\Big(\pi(t) - \sum\limits_{j\in I, j\neq i} r_j\pi_j(t)\Big)$. Then our system is correct since for each $t \in \mathsf{supp}(\pi)$ we have

$$\begin{aligned} \sum_{i\in I} r_i\pi_i(t) &= \sum_{i\in I} r_i \cdot \frac{1}{r_i}\Big(\pi(t) - \sum_{j\in I, j\neq i} r_j\pi_j(t)\Big) \\ &= \sum_{i\in I}\Big(\pi(t) - \sum_{j\in I, j\neq i} r_j\pi_j(t)\Big) \\ &= C_1\cdot\pi(t) - (C_1 - 1)\cdot\sum_{i\in I} r_i\pi_i(t) \end{aligned}$$

from which we gather $C_1 \cdot \sum\limits_{i\in I} r_i\pi_i(t) = C_1\cdot\pi(t)$, thus giving

$$\sum_{i\in I} r_i\pi_i(t) = \pi(t).$$

Moreover, for each $i \in I$ we have $\mathsf{supp}(\pi_i) = S_i$ and thus whenever $t \in \mathsf{supp}(\pi_i)$ then $t \in [\![\varphi_i]\!]\bigsqcap_{h\in H}\gamma_h$. We can therefore conclude that $\pi \in [\![\bigoplus_{i\in I} r_i\varphi_i]\!]\bigsqcap_{h\in H}\gamma_h$ and thus that $s \in [\![\langle a\rangle \bigoplus_{i\in I} r_i\varphi_i]\!]\bigsqcap_{h\in H}\gamma_h$. Since s is arbitrary, Equation (5.5) follows.

Equation (5.4) and Equation (5.5) taken together give the thesis. ■

Consider now $\tilde{\gamma}$ as the top element of the complete lattice $\Gamma_{\mathfrak{I}}$, namely $\tilde{\gamma}$ is the variable interpretation assigning the whole set of processes $\mathcal{S}$ as interpretation to each variable. For $n \in \mathbb{N}$, we define the variable interpretation γ_n as follows:

$$\gamma_n := \begin{cases} \langle\langle \mathcal{E} \rangle\rangle^0 \tilde{\gamma} = \tilde{\gamma} & \text{if } n = 0 \\ \langle\langle \mathcal{E} \rangle\rangle^n \tilde{\gamma} = \langle\langle \mathcal{E} \rangle\rangle \gamma_{n-1} & \text{if } n > 0. \end{cases}$$

Moreover, we define $\gamma_\omega := \bigsqcap_{n\in\mathbb{N}} \langle\langle \mathcal{E} \rangle\rangle^n \tilde{\gamma} = \bigsqcap_{n\in\mathbb{N}} \gamma_n$.

Notice that $\{\gamma_n\}_{n\in\mathbb{N}}$ is the descending Kleene chain of $\langle\langle \mathcal{E} \rangle\rangle$. It follows that γ_ω is the greatest fixed point of $\langle\langle \mathcal{E} \rangle\rangle$, namely $\gamma_\omega = \nu_{\mathcal{E}}$.

Proposition 5.5. *γ_ω is the greatest fixed point of $\langle\langle \mathcal{E} \rangle\rangle$, namely $\nu_{\mathcal{E}} = \gamma_\omega$.*

Proof. Since $\langle\langle \mathcal{E} \rangle\rangle$ is Scott-co-continuous (Proposition 5.4), we can apply the Kleene fixed-point Theorem to the descending Kleene chain $\{\gamma_n\}_{n\in\mathbb{N}}$. ■

5.2 An equivalence relation on formulae

In this Section we introduce a *structural equivalence* on formulae in $\mathcal{L}_{\mathfrak{I}}$ defined as the greatest fixed point of a monotone function over relations on formulae. Equivalence on formulae can be defined either in terms of their semantics or in terms of their syntax. As we will see, structural equivalence of formulae will imply their semantic equivalence (Theorem 5.11), which is defined in the classic way.

Definition 5.6 (Semantic equivalence)**.** Given an endodeclaration $\mathcal{E}$, we say that the state formulae $\varphi, \varphi' \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ are *semantically equivalent* if $[\![\varphi]\!]\nu_{\mathcal{E}} = [\![\varphi']\!]\nu_{\mathcal{E}}$, and the distribution formulae $\psi, \psi' \in \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}}$ are *semantically equivalent* if $[\![\psi]\!]\nu_{\mathcal{E}} = [\![\psi']\!]\nu_{\mathcal{E}}$.

Given some index $\mathrm{i} \in \mathfrak{I}$, we say that the variable X_{i} is *guarded* in a formula $\varphi \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ (resp. $\psi \in \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}}$) if all occurrences of X_{i} appear in φ (resp. ψ) in the scope of the diamond modality. An endodeclaration $\mathcal{E}$ is then said to be *guarded* if it maps variables into formulae in which the occurring variables are all guarded. From now on, all the considered endodeclarations $\mathcal{E}$ are guarded, if not differently specified.

Moreover, we recall that we are considering only processes that are image finite. For this reason the $\mathfrak{I}$-indexed logic $\mathcal{L}_{\mathfrak{I}}$ is too rich to characterize them. In particular, we can simply consider an *image finite* version of the conjunction operator.

Definition 5.7 (Image finiteness)**.** We say that a state formula of the form $\varphi = \bigwedge_{j\in\mathcal{J}} \varphi_j$ is *image finite* if the number of the state formulae φ_j of the form $\varphi_j = \langle a \rangle \psi_j$, for some $\psi_j \in \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}}$, occurring in φ is finite for each $a \in \mathcal{A}$.

Then, we say that an endodeclaration $\mathcal{E}$ is *image finite* if for each $\mathrm{i} \in \mathfrak{I}$ the variable X_{i} is mapped by $\mathcal{E}$ into an image finite formula.

Henceforth we consider only endodeclarations that are image finite, if not differently specified.

Definition 5.8. Let $\mathcal{E}\colon \mathfrak{I} \to \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ be a guarded endodeclaration on the logic $\mathcal{L}_{\mathfrak{I}}$. We define $\mathcal{F}_{\mathcal{E}}\colon \mathcal{P}(\mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}) \to \mathcal{P}(\mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}})$ as the function such that for all relations $\mathcal{R} \in \mathcal{P}(\mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}})$ we have that $\mathcal{F}_{\mathcal{E}}(\mathcal{R})$ is the greatest relation satisfying:

1. $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi$;
2. $\varphi_2\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi_1$ iff $\varphi_1\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi_2$;
3. $X_{\mathfrak{i}}\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi$ iff $\mathcal{E}(X_{\mathfrak{i}})\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi$, for any $\mathfrak{i} \in \mathfrak{I}$;
4. $\bigwedge_{j\in\mathcal{J}} \varphi_j \quad \mathcal{F}_{\mathcal{E}}(\mathcal{R}) \quad \bigwedge_{j\in(\mathcal{J}\setminus\mathcal{I})} \varphi'_j$ for some $\mathcal{I} \neq \emptyset, \mathcal{I} \subset \mathcal{J}$ iff
   - ⋆ for each $j \in \mathcal{J}\setminus\mathcal{I}$ we have $\varphi_j\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi'_j$,
   - ⋆ for each $i \in \mathcal{I}$ we have $\varphi_i\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi'_{j_i}$ for some $j_i \in \mathcal{J}\setminus\mathcal{I}$;
5. $\bigwedge_{j\in\mathcal{J}} \varphi_j\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \bigwedge_{i\in\mathcal{I}} \varphi_i$ iff there is a bijection $f\colon \mathcal{J} \to \mathcal{I}$ with $\varphi_j\, \mathcal{F}_{\mathcal{E}}(\mathcal{R})\, \varphi_{f(j)}$ for all $j \in \mathcal{J}$;
6. $\langle a\rangle\psi\ \mathcal{F}_{\mathcal{E}}(\mathcal{R})\ \langle a\rangle\psi'$ iff $\psi\, \mathcal{R}^{\dagger}\, \psi'$.

***Example* 5.2.** Assume the relation $\mathcal{R} \subseteq \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ given by $\mathcal{R} = \{(\varphi_1,\varphi_2), (\varphi_1,\varphi_3), (\varphi_2,\varphi_4)\}$. Then the distribution formulae $\psi_1 = \frac{1}{2}\varphi_1 \oplus \frac{1}{2}\varphi_2$ and $\psi_2 = \frac{1}{4}\varphi_2 \oplus \frac{1}{4}\varphi_3 \oplus \frac{1}{2}\varphi_4$ are such that $\psi_1\, \mathcal{R}^{\dagger}\, \psi_2$. Therefore we can infer that

$$\langle a\rangle\left(\frac{1}{2}\varphi_1 \oplus \frac{1}{2}\varphi_2\right) \quad \mathcal{F}_{\mathcal{E}}(\mathcal{R}) \quad \langle a\rangle\left(\frac{1}{4}\varphi_2 \oplus \frac{1}{4}\varphi_3 \oplus \frac{1}{2}\varphi_4\right).$$

◀

Lemma 5.6. *The function $\mathcal{F}_{\mathcal{E}}$ is monotone with respect to inclusion.*

Proof. We need to show that $\mathcal{R}_1 \subseteq \mathcal{R}_2$ implies $\mathcal{F}_{\mathcal{E}}(\mathcal{R}_1) \subseteq \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)$ for arbitrary relations $\mathcal{R}_1, \mathcal{R}_2 \in \mathcal{P}(\mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}})$. Hence, we assume $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)\, \varphi'$ for arbitrary $\varphi, \varphi' \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}}$ and prove $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi'$. To infer $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)\, \varphi'$, we apply the rules (1)–(6) in Definition 5.8 n times, for some $n \geq 1$. We prove $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi'$ by induction over n.

The base case $n = 1$ has two subcases, namely either we infer $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)\, \varphi'$ by applying rule (1) or by applying rule (6). In the former case we have $\varphi' = \varphi$, which immediately gives $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi'$ by the same rule (1). In the latter case we have $\varphi = \langle a\rangle\psi$ and $\varphi' = \langle a\rangle\psi'$ with $\psi\, {\mathcal{R}_1}^{\dagger}\psi'$. By the monotonicity of operator $\_^{\dagger}$ we get $\psi\, {\mathcal{R}_2}^{\dagger}\psi'$. Then, by applying rule (6) we get $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi'$.

Consider now the inductive step $n > 1$. The last rule in Definition 5.8 applied to infer $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)\, \varphi'$ is one of the rules in the set (2)–(5). Let (r) be that rule. To apply such a rule (r) it is necessary to have a set of pairs of formulae that are already in $\mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)$, namely we need a set $R \subseteq \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)$ such that each pair of formulae $(\varphi_1, \varphi_2) \in R$ is such that $(\varphi_1\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_1)\, \varphi_2)$ is derived by applying the rules (1)–(6) in Definition 5.8 at most $n-1$ times. By the inductive hypothesis we infer that $(\varphi_1\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi_2)$ for all $(\varphi_1, \varphi_2) \in R$, namely $R \subseteq \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)$, which then gives $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_2)\, \varphi'$ by rule (r). ■

Since $\mathcal{F}_{\mathcal{E}}$ is monotone over the complete lattice $(\mathcal{P}(\mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}), \subseteq)$, by the Knaster-Tarski fixed point Theorem it admits the least and the greatest fixed point.

Definition 5.9 ($\mathcal{L}_{\mathfrak{J}}$-equivalence under $\mathcal{E}$)**.** The $\mathcal{L}_{\mathfrak{J}}$*-equivalence under* an endodeclaration $\mathcal{E}$ is the greatest fixed point of $\mathcal{F}_{\mathcal{E}}$ and will be denoted with $\equiv_{\mathcal{E}}$.

We devote the remaining of this section to an alternative inductive definition of $\equiv_{\mathcal{E}}$. First we show that if the endodeclaration $\mathcal{E}$ is guarded and image finite, then the function $\mathcal{F}_{\mathcal{E}}$ is Scott-co-continuous.

Proposition 5.7. *For a guarded and image finite endodeclaration $\mathcal{E}$ the function $\mathcal{F}_{\mathcal{E}}$ is Scott-co-continuous.*

Proof. We have to show that for each descending chain of subsets $\mathcal{R}_0, \mathcal{R}_1, \ldots$ of $\mathcal{P}(\mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}})$ it holds that $\mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n) = \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)$. More precisely, we need to show that for arbitrary φ, φ' we have

$$\varphi\, \mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n)\, \varphi' \text{ if and only if } \varphi \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$$

Assume first $\varphi\ \mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n)\ \varphi'$. By the monotonicity of $\mathcal{F}_{\mathcal{E}}$ (Lemma 5.6) we get $\varphi\ \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\ \varphi'$ for all $n \in \mathbb{N}$, which gives $\varphi\ \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\ \varphi'$.

Assume now $\varphi \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$, namely $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$ for all $n \in \mathbb{N}$. We aim to show

$$\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi' \text{ for all } n \in \mathbb{N} \text{ implies } \varphi\, \mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n)\, \varphi'.$$

This is proved by structural induction over φ. The interesting cases are the base case for the diamond operator and the non-trivial inductive step related to conjunction.

- ★ Base case $\varphi = \langle a \rangle \bigoplus_{i\in I} r_i \varphi_i$. By Definition 5.8, $\varphi\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$ for each $n \in \mathbb{N}$ requires that $\varphi' = \bigwedge_{h\in\mathcal{H}} \langle a \rangle \psi_h$ with $\psi_h (\mathcal{R}_n)^{\dagger} \bigoplus_{i\in I} r_i \varphi_i$ for all $h \in \mathcal{H}$. We consider only the case of $|\mathcal{H}| = 1$, since the general case for $|\mathcal{H}| > 1$ directly follows from it. Hence, $\varphi' = \langle a \rangle \bigoplus_{j\in J} r_j \varphi_j$. Then from $\langle a \rangle \bigoplus_{i\in I} r_i \varphi_i\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$ for each $n \in \mathbb{N}$, we obtain by Proposition 2.5 that $\varphi' = \langle a \rangle \bigoplus_{\substack{i\in I \\ h_i\in H_i^n}} r_{h_i}^n \varphi_{h_i}^n$ with $\sum_{h_i\in H_i^n} r_{h_i}^n = r_i$ and $\varphi_i \mathcal{R}_n \varphi_{h_i}^n$ for all $h_i \in H_i^n$. As J is finite, for each $i \in I$ there is a set of indexes $J_i \subseteq J$ s.t. $\varphi_i \mathcal{R}_n \varphi_{j_i}$ for countably many $n \in \mathbb{N}$, for each $j_i \in J_i$. In particular, for each $i \in I$ there is an $N_i \in \mathbb{N}$ s.t. for all $n \geq N_i$ it holds that $\varphi_i \mathcal{R}_n \varphi_{j_i}$ for each $j_i \in J_i$. Let $N = \max_{i\in I} N_i$. Then, $\varphi' = \langle a \rangle \bigoplus_{i\in I, j_i\in J_i} L_{j_i} \varphi_{j_i}$ where $L_{j_i} = r_{h_i}^N$ for $h_i \in H_i^N$ s.t. $h_i = j_i$. We remark that by the choice of N, we have that $H_i^N = J_i$. Therefore, from $\langle a \rangle \bigoplus_{i\in I} r_i \varphi_i\, \mathcal{F}_{\mathcal{E}}(\mathcal{R}_N)\, \varphi'$ we obtain that $\sum_{j_i\in J_i} L_{j_i} = r_i$ and moreover, by construction of J_i it holds that $\varphi_i \mathcal{R}_n \varphi_{j_i}$ for all $j_i \in J_i$, for each $n \geq \mathbb{N}$. Furthermore, $\{\mathcal{R}_n\}_{n\in\mathbb{N}}$ is a descending chain on $\mathcal{P}(\mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}})$ and thus $\varphi_i \mathcal{R}_n \varphi_{j_i}$ holds also for all $n \leq N$. Hence we can conclude that $\varphi_i \mathcal{R}_n \varphi_{j_i}$ for all $j_i \in J_i$, for all $n \in \mathbb{N}$ thus implying $\varphi_i \bigcap_{n\in\mathbb{N}} \mathcal{R}_n\, \varphi_{j_i}$ from which we gather that $\langle a \rangle \bigoplus_{i\in I} r_i \varphi_i\, \mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n)\, \varphi'$. We have therefore obtained that

$$\langle a \rangle \bigoplus_{i\in I} r_i \varphi_i \bigcap_{n\in\mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi' \Rightarrow \langle a \rangle \bigoplus_{i\in I} r_i \varphi_i\, \mathcal{F}_{\mathcal{E}}(\bigcap_{n\in\mathbb{N}} \mathcal{R}_n)\, \langle a \rangle \bigoplus_{\substack{i\in I \\ j_i\in J_i}} r_{j_i} \varphi_{j_i}.$$

- ⋆ Inductive step $\varphi = \bigwedge_{j \in \mathcal{J}} \varphi_j$, with $\varphi \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$ for all $n \in \mathbb{N}$ given by Definition 5.8.(5). Notice that $\varphi' \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$ is fixed and it is of the form $\varphi' = \bigwedge_{i \in \mathcal{I}} \varphi_i$ for some set of indexes $\mathcal{I}$. Then, $\bigwedge_{j \in \mathcal{J}} \varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi'$ for all $n \in \mathbb{N}$ implies that for each $n \in \mathbb{N}$ there is a bijection $f^n \colon \mathcal{J} \to \mathcal{I}$ with $\varphi' = \bigwedge_{j \in \mathcal{J}} \varphi_{f^n(j)}$ and $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f^n(j)}$ for all $j \in \mathcal{J}$. Intuitively, each bijection f^n can be seen as a simple reordering of the state formulae in φ' in order to match the corresponding formulae in $\bigwedge_{j \in \mathcal{J}} \varphi_j$. By definition of the functional $\mathcal{F}_{\mathcal{E}}$ (Def. 5.8) whenever φ_j is of the form $\varphi_j = \top$ or $\varphi_j = \bar{a}$, for some $a \in \mathcal{A}$, then $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f^n(j)}$ iff $\varphi_{f^n(j)} = \bigwedge_{h \in \mathcal{H}} \varphi_j$ for some set of indexes $\mathcal{H}$. Moreover whenever $\varphi_j = \langle a \rangle \psi_j$ then $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f^n(j)}$ iff $\varphi_{f^n(j)} = \langle a \rangle \psi_{f^n(j)}$ with $\psi_j \,\mathcal{R}_n^{\dagger}\, \psi_{f^n(j)}$. Since we are considering image finite formulae only, we can infer that for each φ_j, for $j \in \mathcal{J}$, there is a formula φ_i, for some $i \in \mathcal{I}$, s.t. $\varphi_i = \varphi_{f^n(j)}$ for countably many $n \in \mathbb{N}$. In fact, by definition of image finiteness (Def. 5.7), there is only a finite number of formulae in φ' that can be related to each formula φ_j. In particular, for each $j \in \mathcal{J}$ there is an $N_j \in \mathbb{N}$ s.t. $f^n(j) = f^{N_j}(j)$ for all $n \geq N_j$. Let $N = \sup_{j \in \mathcal{J}} N_j$. Then we have that $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f^N(j)}$ for all $n \geq N$. Furthermore, $\{\mathcal{R}_n\}_{n \in \mathbb{N}}$ is a descending chain on $\mathcal{P}(\mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}})$ and $\mathcal{F}_{\mathcal{E}}$ is monotone (Lemma 5.6), and thus f^N is a bijection such that $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f^N(j)}$ also for all $n \leq N$. We have therefore obtained that there exist a bijection $f \colon \mathcal{J} \to \mathcal{I}$ s.t. $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f(j)}$ for all $n \in \mathbb{N}$, $j \in \mathcal{J}$, that is $\varphi_j \bigcap_{n \in \mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n)\, \varphi_{f(j)}$ for all $j \in \mathcal{J}$. By induction, this implies that $\varphi_j \,\mathcal{F}_{\mathcal{E}}(\bigcap_{n \in \mathbb{N}} \mathcal{R}_n)\, \varphi_{f(j)}$ for all $j \in \mathcal{J}$, thus giving $\bigwedge_{j \in \mathcal{J}} \varphi_j \,\mathcal{F}_{\mathcal{E}}(\bigcap_{n \in \mathbb{N}} \mathcal{R}_n)\, \bigwedge_{i \in \mathcal{I}} \varphi_i$. Therefore, we can conclude that

$$\bigwedge_{j \in \mathcal{J}} \varphi_j \bigcap_{n \in \mathbb{N}} \mathcal{F}_{\mathcal{E}}(\mathcal{R}_n) \bigwedge_{i \in \mathcal{I}} \varphi_i \Rightarrow \bigwedge_{j \in \mathcal{J}} \varphi_j \,\mathcal{F}_{\mathcal{E}}\Big(\bigcap_{n \in \mathbb{N}} \mathcal{R}_n\Big) \bigwedge_{i \in \mathcal{I}} \varphi_i.$$

■

Definition 5.10 (Approximated $\mathcal{L}_{\mathfrak{J}}$-equivalence)**.** Let $\mathcal{E}$ be an endodeclaration on the logic $\mathcal{L}_{\mathfrak{J}}$. We define the family of relations $\equiv_{\mathcal{E}}^{n} \subseteq \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$, for $n \in \mathbb{N}$, as follows:

$$\equiv_{\mathcal{E}}^{n} = \begin{cases} \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} & \text{for } n = 0 \\ \mathcal{F}_{\mathcal{E}}(\equiv_{\mathcal{E}}^{n-1}) & \text{for } n > 0. \end{cases}$$

Notice that $\{\equiv_{\mathcal{E}}^{n}\}_{n \in \mathbb{N}}$ is defined as the descending Kleene chain of $\mathcal{F}_{\mathcal{E}}$. By the Scott-co-continuity property of the function $\mathcal{F}_{\mathcal{E}}$ it follows that $\equiv_{\mathcal{E}}^{\omega}$ coincides with the greatest fixed point of $\mathcal{F}_{\mathcal{E}}$, namely $\equiv_{\mathcal{E}}$.

Proposition 5.8. *For any guarded and image finite endodeclaration $\mathcal{E}$, $\equiv_{\mathcal{E}}^{\omega}$ is the greatest fixed point of $\mathcal{F}_{\mathcal{E}}$, namely $\equiv_{\mathcal{E}}^{\omega} = \equiv_{\mathcal{E}}$.*

Proof. Since function $\mathcal{F}_{\mathcal{E}}$ is Scott-co-continuous (Proposition 5.7), we can apply the Kleene fixed-point Theorem to the descending Kleene chain $\{\equiv_{\mathcal{E}}^{n}\}_{n \in \mathbb{N}}$. ■

Now we show that all $\equiv_{\mathcal{E}}^{n}$ and $\equiv_{\mathcal{E}}$ are equivalence relations.

Proposition 5.9. *All $\equiv_{\mathcal{E}}^{n}$ with $n \in \mathbb{N}$ and $\equiv_{\mathcal{E}}^{\omega}$ are equivalence relations.*

Proof. By Definition 5.8.(1) and Definition 5.8.(2) we can immediately infer that $\equiv_{\mathcal{E}}^{n}$ is reflexive and symmetric for each $n \in \mathbb{N}$. Hence, we only have to prove the transitivity property, namely that whenever $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_2$ and $\varphi_2 \equiv_{\mathcal{E}}^n \varphi_3$ we have $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_3$, to conclude that all $\equiv_{\mathcal{E}}^n$ are equivalences.

To this aim, we proceed by induction over $n \in \mathbb{N}$.

(I) The base case $n = 0$ is immediate since by definition $\equiv_{\mathcal{E}}^0 = \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$.

(II) Consider now the inductive step $n > 0$. We proceed by structural induction over φ_1. The interesting cases are those of conjunction and of the diamond modality.

- ⋆ Inductive step $\varphi_1 = \bigwedge_{j \in \mathcal{J}} \varphi_j$. We distinguish two cases.
  - a. $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_2$ for φ_2 of the form $\bigwedge_{j \in (\mathcal{J} \setminus \mathcal{I})} \varphi'_j$ with $\varphi_i \equiv_{\mathcal{E}}^n \varphi'_{j_i}$ for some $j_i \in \mathcal{J} \setminus \mathcal{I}$ for all $i \in \mathcal{I}$ and $\varphi_j \equiv_{\mathcal{E}}^n \varphi'_j$ for all $j \in \mathcal{J} \setminus \mathcal{I}$. We distinguish two cases.
    - i. $\varphi_2 \equiv_{\mathcal{E}}^n \varphi_3$ for φ_3 of the form $\bigwedge_{j \in (\mathcal{J} \setminus \mathcal{I}) \setminus \mathcal{H}} \varphi''_j$ with $\varphi'_h \equiv_{\mathcal{E}}^n \varphi''_{j_h}$ for some $j_h \in (\mathcal{J} \setminus \mathcal{I}) \setminus \mathcal{H}$ for all $h \in \mathcal{H}$ and $\varphi'_j \equiv_{\mathcal{E}}^n \varphi''_j$ for all $j \in (\mathcal{J} \setminus \mathcal{I}) \setminus \mathcal{H}$. Then we have
$$\varphi_3 = \bigwedge_{j \in (\mathcal{J} \setminus (\mathcal{I} \cup \mathcal{H})} \varphi''_j$$
and by structural induction we obtain that for each $k \in \mathcal{I} \cup \mathcal{H}$ $\varphi_k \equiv_{\mathcal{E}}^n \varphi''_{j_k}$ for some $j_k \in \mathcal{J} \setminus (\mathcal{I} \cup \mathcal{H})$. Therefore we can conclude that $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_3$ as required.
    - ii. $\varphi_2 \equiv_{\mathcal{E}}^n \varphi_3$ for φ_3 of the form $\bigwedge_{k \in \mathcal{K}} \varphi''_k$ and there exists a bijection $f : (\mathcal{J} \setminus \mathcal{I}) \to \mathcal{K}$ with $\varphi'_j \equiv_{\mathcal{E}}^n \varphi''_{f(j)}$ for all $j \in (\mathcal{J} \setminus \mathcal{I})$. Therefore we can rewrite
$$\varphi_3 = \bigwedge_{j \in (\mathcal{J} \setminus \mathcal{I})} \varphi''_{f(j)}.$$
Moreover, by structural induction we obtain that for each $i \in \mathcal{I}$ we have that $\varphi_i \equiv_{\mathcal{E}}^n \varphi''_{f(j_i)}$ and for each $j \in \mathcal{J} \setminus \mathcal{I}$ we have that $\varphi_j \equiv_{\mathcal{E}}^n \varphi''_{f(j)}$. Therefore, we can conclude that $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_3$ as required.
  - b. $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_2$ for φ_2 of the form $\bigwedge_{i \in \mathcal{I}} \varphi_i$ and there exists a bijection $f : \mathcal{J} \to \mathcal{I}$ with $\varphi_j \equiv_{\mathcal{E}}^n \varphi_{f(j)}$ for all $j \in \mathcal{J}$. We distinguish three cases.
    - i. $\varphi_2 \equiv_{\mathcal{E}}^n \varphi_3$ for φ_3 of the form $\bigwedge_{i \in (\mathcal{I} \setminus \mathcal{H})} \varphi'_i$ with $\varphi_h \equiv_{\mathcal{E}}^n \varphi'_{k_h}$ for some $k_h \in \mathcal{I} \setminus \mathcal{H}$ for all $h \in \mathcal{H}$ and $\varphi_i \equiv_{\mathcal{E}}^n \varphi'_i$ for all $i \in \mathcal{I} \setminus \mathcal{H}$. Since f is a bijection we obtain
$$\varphi_3 = \bigwedge_{f^{-1}(i) \in (\mathcal{J} \setminus f^{-1}(\mathcal{H}))} \varphi_{f^{-1}(i)}$$
and by structural induction we obtain that $\varphi_{f^{-1}(h)} \equiv_{\mathcal{E}}^n \varphi'_{k_h}$ for all $h \in \mathcal{H}$ and $\varphi_{f^{-1}(i)} \equiv_{\mathcal{E}}^n \varphi'_i$ for all $i \in \mathcal{I} \setminus \mathcal{H}$ thus giving $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_3$ as required.
    - ii. $\varphi_2 \equiv_{\mathcal{E}}^n \varphi_3$ for φ_3 of the form $\bigwedge_{k \in \mathcal{K}} \varphi_k$ and there exists a bijection $g : \mathcal{I} \to \mathcal{K}$ with $\varphi_i \equiv_{\mathcal{E}}^n \varphi_{g(i)}$ for all $i \in \mathcal{I}$. Since f is a bijection, the last relation can be rewritten as $\varphi_{f(j)} \equiv_{\mathcal{E}}^n \varphi_{g(f(j))}$ for all $j \in \mathcal{J}$. Thus, by structural induction over each triple $\varphi_j, \varphi_{f(j)}, \varphi_{g(f(j))}$, we obtain that $\varphi_j \equiv_{\mathcal{E}}^n \varphi_{g(f(j))}$ for each $j \in \mathcal{J}$. Moreover, $h : \mathcal{J} \to \mathcal{K}$ defined by $h(j) = g(f(j))$, for $j \in \mathcal{J}$, is a bijection as composition of bijections. Therefore, we can conclude that $\bigwedge_{j \in \mathcal{J}} \varphi_j \equiv_{\mathcal{E}}^n \bigwedge_{k \in \mathcal{K}} \varphi_k$, namely $\varphi_1 \equiv_{\mathcal{E}}^n \varphi_3$.

- ⋆ Inductive step $\varphi_1 = \langle a \rangle \bigoplus_{i \in I} r_i \varphi_i$. By definition $\varphi_1 \equiv_{\mathcal{E}}^{n} \varphi_2$ if it is of the form $\varphi_2 = \bigwedge_{j \in \mathcal{J}} \langle a \rangle \psi_j$ with $\psi_j (\equiv_{\mathcal{E}}^{n-1})^{\dagger} \bigoplus_{i \in I} r_i \varphi_i$ for all $j \in \mathcal{J}$. However we consider only the case of $|\mathcal{J}| = 1$, since the general case for $|\mathcal{J}| > 1$ directly follows from it. Hence let us consider $\varphi_2 = \langle a \rangle \bigoplus_{\substack{i \in I \\ j_i \in J_i}} r_{j_i} \varphi_{j_i}$ where for each $i \in I$, $\sum_{j_i \in J_i} r_{j_i} = r_i$ and $\varphi_{j_i} \equiv_{\mathcal{E}}^{n-1} \varphi_i$ for all $j_i \in J_i$. Moreover, $\varphi_2 \equiv_{\mathcal{E}}^{n} \varphi_3$ if φ_3 if $\varphi_3 = \langle a \rangle \bigoplus_{\substack{i \in I \\ j_i \in J_i \\ h_{j_i} \in H_{j_i}}} r_{h_{j_i}} \varphi_{h_{j_i}}$ where $\sum_{h_{j_i} \in H_{j_i}} r_{h_{j_i}} = r_{j_i}$ and $\varphi_{h_{j_i}} \equiv_{\mathcal{E}}^{n-1} \varphi_{j_i}$ for all $h_{j_i} \in H_{j_i}$. Therefore, we gather
  - a. $r_i = \sum_{j_i \in J_i} r_{j_i} = \sum_{\substack{j_i \in J_i \\ h_{j_i} \in H_{j_i}}} r_{h_{j_i}}$;
  - b. from $\varphi_i \equiv_{\mathcal{E}}^{n-1} \varphi_{j_i}$ for all $j_i \in J_i$ and $\varphi_{j_i} \equiv_{\mathcal{E}}^{n-1} \varphi_{h_{j_i}}$ for all $h_{j_i} \in H_{j_i}$, the inductive hypothesis gives $\varphi_i \equiv_{\mathcal{E}}^{n-1} \varphi_{h_{j_i}}$ for all $h_{j_i} \in H_{j_i}, j_i \in J_i$.

  From items (i) and (ii) we obtain $\varphi_1 \equiv_{\mathcal{E}}^{n} \varphi_3$ as required.

Finally, since the intersection of equivalence relations is indeed an equivalence relation we conclude that also $\equiv_{\mathcal{E}}^{\omega}$ is an equivalence. ■

It follows that also $\equiv_{\mathcal{E}}$ is an equivalence relation.

Corollary 5.10. *For any guarded and image finite endodeclaration $\mathcal{E}$ the $\mathcal{L}_{\mathfrak{J}}$-equivalence under $\mathcal{E}$ $\equiv_{\mathcal{E}}$ is an equivalence relation.*

Proof. By Proposition 5.8 and Proposition 5.9. ■

Structural equivalence of formulae implies their semantic equivalence, as formalized in the following Theorem. We also notice that the converse implication does not hold in general. In fact, as a trivial counterexample, we can consider the formulae $\varphi = \langle a \rangle \top \wedge \bar{a}$ and $\varphi' = \langle b \rangle \top \wedge \bar{b}$. Clearly, for any endodeclaration $\mathcal{E}$ we have $[\![\varphi]\!] \nu_{\mathcal{E}} = [\![\varphi']\!] \nu_{\mathcal{E}} = \emptyset$ but $\varphi \not\equiv_{\mathcal{E}} \varphi'$.

Theorem 5.11. *$[\![\varphi]\!] \nu_{\mathcal{E}} = [\![\varphi']\!] \nu_{\mathcal{E}}$ for all formulae $\varphi, \varphi' \in \mathcal{L}_{\mathfrak{J}}$ such that $\varphi \equiv_{\mathcal{E}} \varphi'$.*

Proof. We proceed by structural induction over φ.

- ⋆ Base case $\varphi = \top$. By definition $\varphi \equiv_{\mathcal{E}} \varphi'$ if φ' is of the form $\bigwedge_{j \in \mathcal{J}} \top$, namely φ' is a conjunction of an arbitrary number of formulae $\top$. Then, by definition we have $[\![\top]\!] \nu_{\mathcal{E}} = \mathcal{S}$ and $[\![\bigwedge_{j \in \mathcal{J}} \top]\!] \nu_{\mathcal{E}} = \bigcap_{j \in \mathcal{J}} [\![\top]\!] \nu_{\mathcal{E}} = \bigcap_{j \in \mathcal{J}} \mathcal{S} = \mathcal{S}$, thus giving the thesis.

- ⋆ Base case $\varphi = \bar{a}$ for some $a \in \mathcal{A}$. By definition $\varphi \equiv_{\mathcal{E}} \varphi'$ if φ' is of the form $\bigwedge_{j \in \mathcal{J}} \bar{a}$, namely φ' is a conjunction of an arbitrary number of formulae $\bar{a}$. Let $S_{\bar{a}} \subseteq \mathcal{S}$ be the subset of processes that do not perform an a move. Then, by definition we have $[\![\bar{a}]\!] \nu_{\mathcal{E}} = S_{\bar{a}}$ and $[\![\bigwedge_{j \in \mathcal{J}} \bar{a}]\!] \nu_{\mathcal{E}} = \bigcap_{j \in \mathcal{J}} [\![\bar{a}]\!] \nu_{\mathcal{E}} = \bigcap_{j \in \mathcal{J}} S_{\bar{a}} = S_{\bar{a}}$, thus giving the thesis.

- ⋆ $\varphi = \bigwedge_{j \in \mathcal{J}} \varphi_j$. We can distinguish two cases.
  1. $\varphi \equiv_{\mathcal{E}} \varphi'$ with $\varphi' = \bigwedge_{j \in (\mathcal{J} \setminus \mathcal{I})} \varphi'_j$ with $\varphi_i \equiv_{\mathcal{E}} \varphi'_{j_i}$ for some $j_i \in \mathcal{J} \setminus \mathcal{I}$ for each $i \in \mathcal{I}$ and $\varphi_j \equiv_{\mathcal{E}} \varphi'_j$ for each $j \in \mathcal{J} \setminus \mathcal{I}$. We have that

$$[\![\bigwedge_{j \in \mathcal{J}} \varphi_j]\!] \nu_{\mathcal{E}} = \bigcap_{j \in \mathcal{J}} [\![\varphi_j]\!] \nu_{\mathcal{E}}$$

$$
\begin{aligned}
&= \bigcap_{j\in(\mathcal{J}\setminus\mathcal{I})} \llbracket\varphi_j\rrbracket\nu_{\mathcal{E}} \cap \bigcap_{i\in\mathcal{I}} \llbracket\varphi_i\rrbracket\nu_{\mathcal{E}} \\
&= \bigcap_{j\in(\mathcal{J}\setminus\mathcal{I})} \llbracket\varphi'_j\rrbracket\nu_{\mathcal{E}} \cap \bigcap_{i\in\mathcal{I}} \llbracket\varphi_i\rrbracket\nu_{\mathcal{E}} && \text{(by induction over } \varphi_j\text{)} \\
&= \bigcap_{j\in(\mathcal{J}\setminus\mathcal{I})} \llbracket\varphi'_j\rrbracket\nu_{\mathcal{E}} \cap \bigcap_{i\in\mathcal{I}} \llbracket\varphi'_{j_i}\rrbracket\nu_{\mathcal{E}} && \text{(by induction on } \varphi_i\text{)} \\
&= \bigcap_{j\in(\mathcal{J}\setminus\mathcal{I})} \llbracket\varphi'_j\rrbracket\nu_{\mathcal{E}} \cap \bigcap_{j_i\in\mathcal{J}\setminus\mathcal{I}} \llbracket\varphi'_{j_i}\rrbracket\nu_{\mathcal{E}} \\
&= \bigcap_{j\in(\mathcal{J}\setminus\mathcal{I})} \llbracket\varphi'_j\rrbracket\nu_{\mathcal{E}} \\
&= \llbracket \bigwedge_{j\in(\mathcal{J}\setminus\mathcal{I})} \varphi'_j\rrbracket\nu_{\mathcal{E}}
\end{aligned}
$$

2. $\varphi \equiv_{\mathcal{E}} \varphi'$ with $\varphi' = \bigwedge_{i\in\mathcal{I}}\varphi_i$ and there is a bijection $f\colon \mathcal{J}\to\mathcal{I}$ with $\varphi_j \equiv_{\mathcal{E}} \varphi_{f(j)}$, for all $j\in\mathcal{J}$. We have that

$$
\begin{aligned}
\llbracket \bigwedge_{j\in\mathcal{J}} \varphi_j\rrbracket\nu_{\mathcal{E}} &= \bigcap_{j\in\mathcal{J}} \llbracket\varphi_j\rrbracket\nu_{\mathcal{E}} \\
&= \bigcap_{j\in\mathcal{J}} \llbracket\varphi_{f(j)}\rrbracket\nu_{\mathcal{E}} && \text{(by structural induction)} \\
&= \llbracket \bigwedge_{j\in\mathcal{J}} \varphi_{f(j)}\rrbracket\nu_{\mathcal{E}}
\end{aligned}
$$

- ⋆ Inductive step $\varphi = \langle a\rangle \bigoplus_{i\in I} r_i\varphi_i$. By definition $\varphi_1 \equiv_{\mathcal{E}} \varphi_2$ if it is of the form $\varphi_2 = \bigwedge_{j\in\mathcal{J}}\langle a\rangle\psi_j$ with $\psi_j (\equiv_{\mathcal{E}})^{\dagger} \bigoplus_{i\in I} r_i\varphi_i$ for all $j\in\mathcal{J}$. However we consider only the case of $|\mathcal{J}| = 1$, since the general case for $|\mathcal{J}| > 1$ directly follows from it. Hence let us consider $\varphi' = \langle a\rangle \bigoplus_{i\in I,\, j_i\in J_i} r_{j_i}\varphi_{j_i}$ with $\sum_{j_i\in J_i} r_{j_i} = r_i$ and $\varphi_{j_i} \equiv_{\mathcal{E}} \varphi_i$ for all $j_i\in J_i$.

  From $s \models_{\nu_{\mathcal{E}}} \varphi$, we infer that $s \xrightarrow{a} \pi$ for a probability distribution $\pi\in\Delta(\mathcal{S})$ with $\pi \models_{\nu_{\mathcal{E}}} \bigoplus_{i\in I} r_i\varphi_i$, namely there are some distributions $\pi_i\in\Delta(\mathcal{S})$ such that $\pi = \sum_{i\in I} r_i\pi_i$ and, for all $i\in I$, $s' \models_{\nu_{\mathcal{E}}} \varphi_i$ for all states s' with $\pi_i(s') > 0$.

  By structural induction, $s' \models_{\nu_{\mathcal{E}}} \varphi_i$ implies that $s' \models_{\nu_{\mathcal{E}}} \tilde{\varphi}$ for each $\tilde{\varphi} \equiv_{\mathcal{E}} \varphi_i$. In particular, this implies that $s' \models_{\nu_{\mathcal{E}}} \varphi_{j_i}$ for each $j_i\in J_i$. Moreover,

$$
\pi = \sum_{i\in I} r_i\pi_i = \sum_{i\in I}\Big(\sum_{j_i\in J_i} r_{j_i}\Big)\pi_i = \sum_{i\in I}\Big(\sum_{j_i\in J_i} r_{j_i}\pi_i\Big) = \sum_{i\in I}\Big(\sum_{j_i\in J_i} r_{j_i}\pi_{j_i}\Big)
$$

  where $\pi_{j_i} = \pi_i$ for all $j_i\in J_i$.

  Hence, we have obtained that there is a probability distribution $\pi\in\operatorname{der}(s,a)$ such that $\pi = \sum_{i\in I}\sum_{j_i\in J_i} r_{j_i}\pi_{j_i}$, for some distributions π_{j_i}, and moreover for each $s'\in\mathcal{S}$ such that $\pi_{j_i}(s') > 0$ we have $s' \models_{\nu_{\mathcal{E}}} \varphi_{j_i}$. We can then conclude that $\pi \models_{\nu_{\mathcal{E}}} \bigoplus_{\substack{i\in I\\ j_i\in J_i}} r_{j_i}\varphi_{j_i}$, namely $s \models_{\nu_{\mathcal{E}}} \varphi'$.

- ⋆ Inductive step $\varphi = X_{\mathrm{i}}$, for some $\mathrm{i}\in\mathfrak{I}$. By definition, $X_{\mathrm{i}} \equiv_{\mathcal{E}} \mathcal{E}(X_{\mathrm{i}})$. Moreover, $s\in\llbracket X_{\mathrm{i}}\rrbracket\nu_{\mathcal{E}}$ if and only if $s\in\nu_{\mathcal{E}}(X_{\mathrm{i}})$ and, by definition of $\nu_{\mathcal{E}}$, $\nu_{\mathcal{E}}(X_{\mathrm{i}}) = \llbracket\mathcal{E}(X_{\mathrm{i}})\rrbracket\nu_{\mathcal{E}}$. Therefore, the thesis follows from the previous cases.

■

5.3 Mimicking formulae of processes

In this Section we introduce the notion of *mimicking formula* of a process which will allow us to obtain the characterizations of probabilistic bisimilarity, ready similarity and similarity (Section 5.4) and those of their quantitative versions (Sections 5.5 and 5.6).

We consider the logic $\mathcal{L}_{\mathcal{S}}$, which allows us to associate a variable X_s to each process $s \in \mathcal{S}$. Then we introduce the notion of *mimicking formula* of a process $s \in \mathcal{S}$ as a formula capturing the branching and probabilistic features of s, and we define the *mimicking endodeclaration* $\mathcal{M}$ on $\mathcal{L}_{\mathcal{S}}$ such that $\mathcal{M}(X_s)$ is the mimicking formula of s.

Definition 5.11 (Mimicking formula)**.** For a process $s \in \mathcal{S}$, the *mimicking formula of* s is denoted with φ_s and is defined by

$$\varphi_s = \bigwedge_{(s,a,\pi)\in\rightarrow} \langle a \rangle \bigoplus_{s' \in \mathsf{supp}(\pi)} \pi(s') X_{s'} \wedge \bigwedge_{b \notin \mathsf{init}(s)} \bar{b}.$$

Then the *mimicking endodeclaration* $\mathcal{M}\colon \mathcal{S} \to \mathcal{L}^{\mathrm{s}}_{\mathcal{S}}$ is defined for all processes $s \in \mathcal{S}$ by

$$\mathcal{M}(X_s) = \varphi_s.$$

Intuitively, φ_s characterizes the branching structure of s by specifying which transitions are enabled for s as well as all the actions that it cannot perform. Moreover, as states evolve to distributions, the mimicking formula of s captures the probabilistic behavior of s by associating to each process s' in the support of π, for each $\pi \in \mathrm{der}(s,a)$, its own mimicking formula $\varphi_{s'}$ weighted by $\pi(s')$.

Remark 5.1. Notice that all the variables occurring in the mimicking formulae are guarded, so that the mimicking endodeclaration $\mathcal{M}$ is guarded. Moreover, since the processes in $\mathcal{S}$ are image finite, we infer that all mimicking formulae are image finite and so is the mimicking endodeclaration $\mathcal{M}$. Furthermore, we remark that to simplify reading and presentation we have written the mimicking formulae as a nested conjunction of state formulae, although non allowed in $\mathcal{L}_{\mathcal{S}}$ (see Definition 5.1). However, mimicking formulae can always be expressed without nested conjunctions, as showed in Example 5.3.

***Example* 5.3.** Assume $\mathcal{A} = \{a, b, c\}$. Let us consider the process $s \in \mathcal{S}$ represented in Figure 5.1. Then the mimicking formula of s is obtained by the following assignments:

$$\begin{aligned}
\mathcal{M}(X_s) &= \langle a \rangle \Big(\frac{1}{4} X_{s_1} \oplus \frac{1}{4} X_{s_2} \oplus \frac{1}{2} X_{s_3} \Big) \wedge \langle a \rangle 1 X_{s_4} \wedge \bar{b} \wedge \bar{c} \\
\mathcal{M}(X_{s_1}) &= \langle a \rangle 1 X_{s_1} \wedge \bar{b} \wedge \bar{c} \\
\mathcal{M}(X_{s_2}) &= \langle a \rangle 1 X_{\mathrm{nil}} \wedge \langle c \rangle 1 X_{\mathrm{nil}} \wedge \bar{b} \\
\mathcal{M}(X_{s_3}) &= \langle b \rangle 1 X_{\mathrm{nil}} \wedge \bar{a} \wedge \bar{c} \\
\mathcal{M}(X_{s_4}) &= \langle c \rangle 1 X_{s_4} \wedge \bar{a} \wedge \bar{b} \\
\mathcal{M}(X_{\mathrm{nil}}) &= \bar{a} \wedge \bar{b} \wedge \bar{c}.
\end{aligned}$$

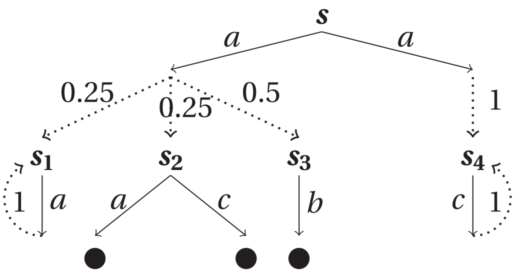


Figure 5.1: *An arbitrary image finite process.*

◀

As expected, each process s satisfies its own mimicking formula φ_s.

Theorem 5.12. *For any process $s \in \mathcal{S}$, $s \in \nu_{\mathcal{M}}(X_s)$.*

Proof. By Proposition 5.5 we have $\nu_{\mathcal{M}} = \gamma_\omega$, where, by definition, $\gamma_\omega = \bigsqcap_{n\in\mathbb{N}} \gamma_n$, for the variable interpretations γ_n defined as

$$\gamma_n := \begin{cases} \tilde{\gamma} & \text{if } n = 0 \\ \langle\langle \mathcal{M} \rangle\rangle \gamma_{n-1} & \text{if } n > 0 \end{cases}$$

where $\tilde{\gamma}$ is the top element of the complete lattice $\Gamma_{\mathcal{S}}$, namely the variable interpretation assigning $\mathcal{S}$ as meaning to each variable. Since $\nu_{\mathcal{M}}(X_s) = \gamma_\omega(X_s) = \bigcap_{n\in N} \gamma_n(X_s)$, to prove the thesis we show that

$$\text{for all } n \in \mathbb{N},\ s \in \gamma_n(X_s). \tag{5.6}$$

We prove Equation (5.6) by induction over n.

Consider the base case $n = 0$. Then $s \in \gamma_n(X_s)$ follows immediately by $\gamma_0(X_s) = \mathcal{S}$.

Consider now the inductive step $n > 0$: we assume that for each $t \in \mathcal{S}$ it holds $t \in \gamma_{n-1}(X_t)$ and we show that under this assumption we have $s \in \gamma_n(X_s)$, for each $s \in \mathcal{S}$. We have

$$\gamma_n(X_s) = \big(\langle\langle \mathcal{M} \rangle\rangle \gamma_{n-1}\big)(X_s) = [\![\mathcal{M}(X_s)]\!] \gamma_{n-1} = [\![\varphi_s]\!] \gamma_{n-1}$$

and by definition of mimicking formula (Definition 5.11) we have

$$\varphi_s = \bigwedge_{(s,a,\pi)\in\to} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi)} \pi(s') X_{s'} \wedge \bigwedge_{b\notin\mathsf{init}(s)} \bar{b}.$$

Hence, we need to prove that $s \in [\![\varphi_s]\!] \gamma_{n-1}$, namely

$$s \models_{\gamma_{n-1}} \bigwedge_{(s,a,\pi)\in\to} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi)} \pi(s') X_{s'} \wedge \bigwedge_{b\notin\mathsf{init}(s)} \bar{b}. \tag{5.7}$$

It is immediate to see that for all action types b such that $s \overset{b}{\not\to}$ we have $s \models_{\gamma_{n-1}} \bar{b}$.

Thus to complete the proof of Equation (5.7), we need to show that for each distribution $\pi \in \Delta(\mathcal{S})$ we have

$$s \xrightarrow{a} \pi \text{ implies } s \models_{\gamma_{n-1}} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi)} \pi(s') X_{s'}. \tag{5.8}$$

To prove Equation (5.8) it is enough to prove that for each distribution $\pi \in \Delta(\mathcal{S})$ we have

$$s \xrightarrow{a} \pi \text{ implies } \pi \models_{\gamma_{n-1}} \bigoplus_{s' \in \mathsf{supp}(\pi)} \pi(s') X_{s'}. \tag{5.9}$$

We have $\pi = \sum_{s' \in \mathsf{supp}(\pi)} \pi(s')\delta_{s'}$ and moreover by inductive hypothesis $s' \in \gamma_{n-1}(X_{s'})$ for each $s' \in \mathsf{supp}(\pi)$, which gives $\pi \models_{\gamma_{n-1}} \bigoplus_{s' \in \mathsf{supp}(\pi)} \pi(s') X_{s'}$ and, then, Equation (5.9). ■

5.4 $\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATION OF PROBABILISTIC RELATIONS

In this Section we present our characterizations of probabilistic bisimilarity, probabilistic ready similarity and probabilistic similarity.

By means of mimicking formulae we are able to characterize probabilistic bisimilarity in a *weak expressive* fashion: two processes are probabilistic bisimilar if and only if their mimicking formulae are $\mathcal{L}_{\mathcal{S}}$-equivalent under $\mathcal{M}$ (Theorem 5.13). Moreover, we will prove that the mimicking formula of a process coincides with the characteristic formula of that process with respect to ready similarity thus allowing for an expressive characterization of that preorder: a process t satisfies the mimicking formula of process s, that is $t \in \nu_{\mathcal{M}}(X_s)$, if and only if t ready simulates s (Theorem 5.14). Finally, we define the *simulation endo-declaration* $\mathcal{C}$ on $\mathcal{L}_{\mathcal{S}}$ from which we obtain the characteristic formulae for simulation from the negation free subformulae of the mimicking formulae, thus obtaining an expressive characterization of similarity: a process t satisfies the *simulation characteristic formula* of process s, that is $t \in \nu_{\mathcal{C}}(X_s)$, if and only if t simulates s (Theorem 5.17).

$\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATION OF PROBABILISTIC BISIMILARITY

We obtain the characterization of probabilistic bisimilarity through the comparison of mimicking formulae of processes. Informally, we exploit $\mathcal{L}_{\mathcal{S}}$-equivalence: two processes are bisimilar if and only if their mimicking formulae are $\mathcal{L}_{\mathcal{S}}$-equivalent under $\mathcal{M}$.

Theorem 5.13. *Given any processes $s, t \in \mathcal{S}$, $X_s \equiv_{\mathcal{M}} X_t$ if and only if $s \sim t$.*

Proof. ($\Leftarrow$). Assume first that $s \sim t$. We have to show that $X_s \equiv_{\mathcal{M}} X_t$. To this aim, we prove a stronger result, namely that

$$\text{for all } n \in \mathbb{N} \quad s \sim_n t \text{ implies } X_s \equiv^n_{\mathcal{M}} X_t. \tag{5.10}$$

Since processes are image finite, the thesis will then follow from Equation (5.10) by $\sim = \bigcap_{n \in \mathbb{N}} \sim_n$ and $\equiv_{\mathcal{M}} = \bigcap_{n \in \mathbb{N}} \equiv^n_{\mathcal{M}}$. We prove Equation (5.10) by induction over $n \in \mathbb{N}$.

The base case $n = 0$ is immediate since by Definition 5.10 we have $\equiv^0_{\mathcal{M}} = \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$.

Consider now the inductive step $n > 0$. Let $s \sim_n t$. By Definition 5.10 we have $X_s \equiv^n_{\mathcal{M}} X_t$ iff $\mathcal{M}(X_s) \equiv^n_{\mathcal{M}} \mathcal{M}(X_t)$, i.e. $\varphi_s \equiv^n_{\mathcal{M}} \varphi_t$, where by Definition 5.11 we have

$$\varphi_s = \bigwedge_{(s,a,\pi_s) \in \to} \langle a \rangle \bigoplus_{s' \in \mathsf{supp}(\pi_s)} \pi_s(s') X_{s'} \wedge \bigwedge_{b \notin \mathsf{init}(s)} \bar{b}$$

$$\varphi_t = \bigwedge_{(t,a,\pi_t)\in\rightarrow} \langle a\rangle \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'} \wedge \bigwedge_{b\not\in\mathsf{init}(t)} \bar{b}.$$

From the assumption $s \sim_n t$ with $n > 0$, it follows $\mathsf{init}(s) = \mathsf{init}(t)$, thus giving that there is a bijection $f\colon \mathsf{init}(s)^c \to \mathsf{init}(t)^c$ (which is actually given by the identity function) such that given any $b \not\in \mathsf{init}(s)$ we have that $\bar{b} \equiv^n_{\mathcal{M}} f(\bar{b})$ is obtained by applying Definition 5.8.(1).

To prove that $\varphi_s \equiv^n_{\mathcal{M}} \varphi_t$ it remains to show that for each $a \in \mathsf{init}(s)$ there is a bijection $g_a\colon \mathsf{der}(s,a) \to \mathsf{der}(t,a)$ such that

$$\langle a\rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'} \equiv^n_{\mathcal{M}} \langle a\rangle \bigoplus_{t'\in\mathsf{supp}(g_a(\pi_s))} g_a(\pi_s)(t')X_{t'}. \tag{5.11}$$

Consider an arbitrary transition $s \xrightarrow{a} \pi_s$. Since $s \sim_n t$, there exists a probability distribution π_t such that $t \xrightarrow{a} \pi_t$ and $\pi_s \sim^\dagger_{n-1} \pi_t$. By Definition 2.17 we have $\pi_s \sim^\dagger_{n-1} \pi_t$ iff whenever $\pi_s = \sum_{i\in I} p_i\delta_{s_i}$ then $\pi_t = \sum_{i\in I} p_i\delta_{t_i}$ and $s_i \sim_{n-1} t_i$ for all $i \in I$. By the inductive hypothesis, $s_i \sim_{n-1} t_i$ implies that $X_{s_i} \equiv^{n-1}_{\mathcal{M}} X_{t_i}$. Therefore,

$$\bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'} = \bigoplus_{i\in I} p_i X_{s_i} (\equiv^{n-1}_{\mathcal{M}})^\dagger \bigoplus_{i\in I} p_i X_{t_i} = \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'}$$

from which we get (by Definition 5.8.(6))

$$\langle a\rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'} \equiv^n_{\mathcal{M}} \langle a\rangle \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'}. \tag{5.12}$$

Analogously, for any $t \xrightarrow{a} \pi_t$ there is a transition $s \xrightarrow{a} \pi_s$ such that Equation (5.12) holds. Hence by combining Definition 5.8.(4) and Definition 5.8.(5) we obtain the bijection g_a we were looking for. Briefly, whenever a distribution $\pi_s \in \mathsf{der}(s,a)$ is related to more than one distribution in $\mathsf{der}(t,a)$, we can use Definition 5.8.(4) to add to φ_s as many occurrences of the formula $\langle a\rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'}$ as the number of distributions in $\mathsf{der}(t,a)$ to which π_s is related. By applying this reasoning to all distributions in $\mathsf{der}(s,a)$ and $\mathsf{der}(t,a)$, we obtain the bijection g_a satisfying Equation (5.11) as $\pi_t = g_a(\pi_s)$ if and only if $\pi_s \sim^\dagger_{n-1} \pi_t$. Then $\varphi_s \equiv^n_{\mathcal{M}} \varphi_t$ is obtained from Definition 5.8.(5) and the transitivity of $\equiv^n_{\mathcal{M}}$.

($\Rightarrow$). Assume now that $X_s \equiv_{\mathcal{M}} X_t$. We aim to show that $s \sim t$. To this aim, it is enough to prove that there is a probabilistic bisimulation relating s and t. To this purpose, we prove that the relation

$$\mathcal{R} := \{(s,t) \mid X_s \equiv_{\mathcal{M}} X_t\}$$

is a probabilistic bisimulation relation. Let $s \mathcal{R} t$. We aim to prove that

$$\text{whenever } s \xrightarrow{a} \pi_s \text{ there is a transition } t \xrightarrow{a} \pi_t \text{ with } \pi_s \mathcal{R}^\dagger \pi_t. \tag{5.13}$$

Consider any transition $s \xrightarrow{a} \pi_s$. By definition, we have $X_s \equiv_{\mathcal{M}} X_t$ if and only if $\mathcal{M}(X_s) \equiv_{\mathcal{M}} \mathcal{M}(X_t)$, namely $\varphi_s \equiv_{\mathcal{M}} \varphi_t$. Moreover, from Theorem 5.12 $s \models_{\nu_{\mathcal{M}}} \varphi_s$ and $t \models_{\nu_{\mathcal{M}}} \varphi_t$, where, by definition of mimicking formula (Definition 5.11), we have

$$\varphi_s = \bigwedge_{(s,a,\pi_s)\in\rightarrow} \langle a\rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'} \wedge \bigwedge_{b\not\in\mathsf{init}(s)} \bar{b}$$

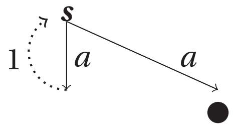


Figure 5.2: *Processes s, t are not probabilistic bisimilar but $s \in \nu_{\mathcal{M}}(X_t)$.*

$$\varphi_t = \bigwedge_{(t,a,\pi_t)\in\to} \langle a\rangle \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'} \wedge \bigwedge_{b\not\in\mathsf{init}(t)} \bar{b}.$$

From $\varphi_s \equiv_{\mathcal{M}} \varphi_t$ and $s \xrightarrow{a} \pi_s$ we get that there is a distribution π_t with $t \xrightarrow{a} \pi_t$ and $\langle a\rangle\psi_{\pi_s} \equiv_{\mathcal{M}} \langle a\rangle\psi_{\pi_t}$, where $\psi_{\pi_s} = \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'}$ and $\psi_{\pi_t} = \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'}$. To derive Equation (5.13) it is enough to prove that $\pi_s \mathcal{R}^\dagger \pi_t$. From $\langle a\rangle\psi_{\pi_s} \equiv_{\mathcal{M}} \langle a\rangle\psi_{\pi_t}$ we derive $\psi_{\pi_s}(\equiv_{\mathcal{M}})^\dagger\psi_{\pi_t}$, which by Definition 2.15 implies that

$$\text{if } \psi_{\pi_s} = \bigoplus_{i\in I} r_i\varphi_i, \text{ then } \psi_{\pi_t} = \bigoplus_{i\in I,\, j_i\in J_i} r_{j_i}\varphi_{j_i} \text{ with } \sum_{j_i\in J_i} r_{j_i} = r_i \text{ and } \varphi_i \equiv_{\mathcal{M}} \varphi_{j_i}. \tag{5.14}$$

Moreover, we notice that by definition of mimicking formula (Definition 5.11), for all $i \in I$ and $j_i \in J_i$ the formulae φ_i and φ_{j_i} in Equation (5.14) have to be the $\mathcal{S}$-indexed variables for some appropriate processes $s_i \in \mathsf{supp}(\pi_s)$ and $t_{j_i} \in \mathsf{supp}(\pi_t)$, namely

$$\varphi_i = X_{s_i} \text{ for all } i \in I \text{ and } \varphi_{j_i} = X_{t_{j_i}} \text{ for all } j_i \in J_i \text{ and } i \in I.$$

Therefore, from $\varphi_i \equiv_{\mathcal{M}} \varphi_{j_i}$ we infer $X_{s_i} \equiv_{\mathcal{M}} X_{t_{j_i}}$. By definition of $\mathcal{R}$, from $X_{s_i} \equiv_{\mathcal{M}} X_{t_{j_i}}$ we get $s_i \mathcal{R}\, t_{j_i}$. Hence we have (i) $\pi_s = \sum_{\substack{i\in I\\ j_i\in J_i}} r_{j_i}\delta_{s_i}$; (ii) $s_i \mathcal{R}\, t_{j_i}$ for all $i \in I$ and $j \in J_i$; (iii) $\pi_t = \sum_{\substack{i\in I\\ j_i\in J_i}} r_{j_i}\delta_{t_{j_i}}$, thus giving $\pi_s \mathcal{R}^\dagger \pi_t$ as required.

Summarizing, we have shown that the transition $s \xrightarrow{a} \pi_s$ is matched by a transition $t \xrightarrow{a} \pi_t$ such that Equation (5.13) holds. With the same argument it can be shown that

$$\text{whenever } t \xrightarrow{a} \pi_t \text{ there is a transition } s \xrightarrow{a} \pi_s \text{ with } \pi_t \mathcal{R}^\dagger \pi_s.$$

Therefore we can conclude that $\mathcal{R}$ is a probabilistic bisimulation as required. ■

A consequence of Theorem 5.13 is that it is enough to inspect two formulae to establish whether two processes are equivalent. This is a *weak* form of expressive characterization since, as discussed in Chapter 2.4, the classic expressive characterization result states that to check whether two processes are equivalent with respect to a given behavioral relation it is enough to verify that one of the two satisfies the characteristic formula of the other process (see for example [68, 133, 149]). This is due to the fact that mimicking formulae are slightly less expressive than characteristic formulae, as shown in the following example.

***Example* 5.4.** Let $\mathcal{A} = \{a, b\}$ and consider processes s and t in Figure 5.2. We have

$$\mathcal{M}(X_s) = \langle a\rangle X_s \wedge \langle a\rangle(\bar{a} \wedge \bar{b}) \wedge \bar{b}$$
$$\mathcal{M}(X_t) = \langle a\rangle X_t \wedge \bar{b}.$$

Clearly $s \in \nu_{\mathcal{M}}(X_t)$, but $s \not\sim t$ since nil, reached by s via the rightmost a transition, cannot simulate t. ◀

$\mathcal{L}_{\mathcal{S}}$-characterization of probabilistic ready simulation

Mimicking formulae capture all possible resolutions of nondeterminism of processes as well as their inability to perform a specific action. Consequently, they give us enough power to expressively characterize ready similarity: the mimicking formula of a process s is the characteristic formula of s with respect to ready similarity.

Theorem 5.14. *Given any processes $s,t \in \mathcal{S}$, $t \in \nu_{\mathcal{M}}(X_s)$ if and only if $s \sqsubseteq_{\mathrm{r}} t$.*

Proof. ($\Leftarrow$) Assume first that $s \sqsubseteq_{\mathrm{r}} t$. We have to show that $t \in \nu_{\mathcal{M}}(X_s)$. We recall that by Proposition 5.5 we have $\nu_{\mathcal{M}} = \gamma_\omega$, where, by definition, $\gamma_\omega = \bigsqcap_{n\in\mathbb{N}} \gamma_n$, for the variable interpretations γ_n defined as

$$\gamma_n := \begin{cases} \tilde{\gamma} & \text{if } n = 0 \\ \langle\!\langle \mathcal{M} \rangle\!\rangle \gamma_{n-1} & \text{if } n > 0 \end{cases}$$

where $\tilde{\gamma}$ is the top element of the complete lattice $\Gamma_{\mathcal{S}}$, namely the variable interpretation assigning $\mathcal{S}$ as meaning to each variable. Since processes are image finite, we have $\nu_{\mathcal{M}}(X_s) = \gamma_\omega(X_s) = \bigcap_{n\in N} \gamma_n(X_s)$ and $\sqsubseteq_{\mathrm{r}} = \bigcap_{n\in\mathbb{N}} \sqsubseteq_n^{\mathrm{r}}$. Therefore, to prove the thesis it is sufficient to show the stronger property that

$$\text{for all } n \in \mathbb{N} \;\; s \sqsubseteq_n^{\mathrm{r}} t \text{ implies } t \in \gamma_n(X_s). \tag{5.15}$$

We prove Equation (5.15) by induction over $n \in \mathbb{N}$.

The base case $n = 0$ is immediate since by definition we have that $\gamma_0(X_s) = \mathcal{S}$.

Consider now the inductive step $n > 0$. Since $s \sqsubseteq_n^{\mathrm{r}} t$, by Definition 2.17 we have that whenever $s \xrightarrow{a} \pi_s$ for some action $a \in \mathcal{A}$ and probability distribution $\pi_s \in \Delta(\mathcal{S})$, then there exists a probability distribution π_t such that $t \xrightarrow{a} \pi_t$ and $\pi_s \sqsubseteq_{n-1}^{\mathrm{r}^\dagger} \pi_t$. This implies (i) $\pi_s = \sum_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')\delta_{s'}$; (ii) for each $s' \in \mathsf{supp}(\pi_s)$ there is a $t' \in \mathsf{supp}(\pi_t)$ such that $s' \sqsubseteq_{n-1}^{\mathrm{r}} t'$; (iii) $\pi_t = \sum_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')\delta_{t'}$.

By the inductive hypothesis, $s' \sqsubseteq_{n-1}^{\mathrm{r}} t'$ implies $t' \in \gamma_{n-1}(X_{s'})$, namely $t' \models_{\gamma_{n-1}} \varphi_{s'}$. Hence, we have that $\pi_t = \sum_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')\delta_{t'}$ for some processes t' with $t' \models_{\gamma_{n-1}} X_{s'}$. By Definition 5.2 this gives that $\pi_t \models_{\gamma_{n-1}} \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'}$ and thus that $t \models_{\gamma_{n-1}} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'}$. Since a and π_s are arbitrary, we can conclude that

$$t \models_{\gamma_{n-1}} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'} \text{ for each } (s,a,\pi_s) \in \to . \tag{5.16}$$

Moreover from $s \sqsubseteq_n^{\mathrm{r}} t$, by Definition 2.17 we have that $\mathrm{init}(s) = \mathrm{init}(t)$. Hence we can immediately infer that

$$t \models_{\gamma_{n-1}} \bar{b} \text{ for each } b \not\in \mathrm{init}(s). \tag{5.17}$$

From Equations (5.16) and (5.17) we obtain that

$$t \models_{\gamma_{n-1}} \bigwedge_{(s,a,\pi)\in\to} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'} \wedge \bigwedge_{b\not\in\mathrm{init}(s)} \bar{b}$$

namely $t \models_{\gamma_{n-1}} \mathcal{M}(X_s)$, that is $t \in [\![\mathcal{M}(X_s)]\!] \gamma_{n-1} = \langle\!\langle \mathcal{M} \rangle\!\rangle \gamma_{n-1}(X_s) = \gamma_n(X_s)$, thus concluding this part of the proof.

($\Rightarrow$) Assume now that $t \in \nu_\mathcal{M}(X_s)$. We have to show that $s \sqsubseteq_r t$. To this aim, it is enough to prove that there is a ready probabilistic simulation relating s and t. Hence, we will prove that the relation

$$\mathcal{R} := \{(s,t) \mid t \in \nu_\mathcal{M}(X_s)\}$$

is a ready probabilistic simulation. Let $s \mathcal{R} t$. We aim to prove that

$$\text{whenever } s \xrightarrow{a} \pi_s \text{ there is a transition } t \xrightarrow{a} \pi_t \text{ with } \pi_s \mathcal{R}^\dagger \pi_t \tag{5.18}$$

$$\text{whenever } s \overset{b}{\not\to} \text{ we have } t \overset{b}{\not\to}. \tag{5.19}$$

Consider any transition $s \xrightarrow{a} \pi_s$. From the hypothesis we have that $t \in \nu_\mathcal{M}(X_s)$, thus implying $t \models_{\nu_\mathcal{M}} \mathcal{M}(X_s)$, namely $t \models_{\nu_\mathcal{M}} \varphi_s$, where, by definition of mimicking formula for s (Definition 5.11), we have

$$\varphi_s = \bigwedge_{(s,a,\pi)\in\to} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'} \wedge \bigwedge_{b \notin \mathsf{init}(s)} \bar{b}.$$

Hence, for any transition $s \xrightarrow{a} \pi_s$ we have that $t \models_{\nu_\mathcal{M}} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'}$. By definition of relation $\models_{\nu_\mathcal{M}}$ (Definition 5.2), this implies that there exists a distribution π_t such that $t \xrightarrow{a} \pi_t$ and $\pi_t \models_{\nu_\mathcal{M}} \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'}$. To derive Equation (5.18) it is now enough to prove that π_t is such that $\pi_s \mathcal{R}^\dagger \pi_t$. From $\pi_t \models_{\nu_\mathcal{M}} \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') \varphi_{s'}$ and the definition of relation $\models_{\nu_\mathcal{M}}$ (Definition 5.2) it follows that for all $s' \in \mathsf{supp}(\pi_s)$ there is some distribution $\pi_{s'}$ such that $\pi_t = \sum_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') \pi_{s'}$ and for each $t' \in \mathsf{supp}(\pi_{s'})$ it holds that $t' \models_{\nu_\mathcal{M}} X_{s'}$, thus giving $s' \mathcal{R} t'$. From Proposition 2.3 it follows that $\delta_{s'} \mathcal{R}^\dagger \delta_{t'}$ and, by the same Proposition 2.3, we gather $\delta_{s'} \mathcal{R}^\dagger \pi_{s'}$ for all $s' \in \mathsf{supp}(\pi_s)$ and thus that $\sum\limits_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')\delta_{s'} \mathcal{R}^\dagger \sum\limits_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')\pi_{s'}$, namely $\pi_s \mathcal{R}^\dagger \pi_t$, which completes the proof of Equation (5.18).

Consider now Equation (5.19). From the hypothesis $t \models_{\nu_\mathcal{M}} X_s$, namely $t \models_{\nu_\mathcal{M}} \varphi_s$ and by definition of mimicking formula for s (Definition 5.11)

$$\varphi_s = \bigwedge_{(s,a,\pi)\in\to} \langle a \rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s') X_{s'} \wedge \bigwedge_{b \notin \mathsf{init}(s)} \neg\bar{b}$$

it follows that whenever $s \overset{b}{\not\to}$ we have $t \models_{\nu_\mathcal{M}} \bar{b}$. By definition of relation $\models_{\nu_\mathcal{M}}$ (Definition 5.2) this implies $t \overset{b}{\not\to}$, which gives Equation (5.19).

Hence both Equations (5.18) and (5.19) have been proved and the proof is complete. ■

$\mathcal{L}_\mathcal{S}$-characterization of probabilistic simulation

We notice that whenever a process t satisfies the mimicking formula φ_s of process s, we are guaranteed that all transitions performed by s are mimicked by transitions by t. Thus, the following soundness results with respect to similarity is natural.

Theorem 5.15. *Given any processes $s, t \in \mathcal{S}$, if $t \in \nu_\mathcal{M}(X_s)$ then $s \sqsubseteq t$.*

Proof. It is enough to prove that there is a probabilistic simulation relating s and t. Hence, we will prove that the relation

$$\mathcal{R} := \{(s,t) \mid t \in \nu_{\mathcal{M}}(X_s)\}$$

is a probabilistic simulation. Let $s \mathcal{R} t$. We need to prove that

$$\text{whenever } s \xrightarrow{a} \pi_s \text{ there is a transition } t \xrightarrow{a} \pi_t \text{ with } \pi_s \sqsubseteq^{\dagger} \pi_t. \tag{5.20}$$

Equation (5.20) can be proved as done for Equation (5.18) in the proof of Theorem 5.14. ■

The distinguishing power of mimicking formulae is too strong to obtain completeness: a process s with $\mathrm{init}(s) = \emptyset$ is simulated by any process t, but the mimicking formula of s, $\varphi_s = \bigwedge_{a \in \mathcal{A}} \bar{a}$, is satisfied only by those t with $\mathrm{init}(t) = \emptyset$. However, if we consider the negation free subformula of a mimicking formula then we obtain the *simulation characteristic formula* of a process (Theorem 5.17).

Definition 5.12 (Simulation characteristic formula)**.** For a process $s \in \mathcal{S}$, the *simulation characteristic formula of s* is denoted with ϑ_s and is defined by

$$\vartheta_s := \bigwedge_{(s,a,\pi) \in \to} \langle a \rangle \bigoplus_{s' \in \mathsf{supp}(\pi)} \pi(s') X_{s'}.$$

Then the *simulation endodeclaration* $\mathcal{C} \colon \mathcal{S} \to \mathcal{L}^{\mathrm{s}}_{\mathcal{S}}$ is defined for all processes $s \in \mathcal{S}$ by

$$\mathcal{C}(X_s) = \vartheta_s.$$

As expected, each process s satisfies its own simulation characteristic formula ϑ_s.

Theorem 5.16. *For any process $s \in \mathcal{S}$, $s \in \nu_{\mathcal{C}}(X_s)$.*

Proof. The same reasoning used in the proof of Theorem 5.12 applies. ■

Theorem 5.17. *Given any processes $s,t \in \mathcal{S}$, $t \in \nu_{\mathcal{C}}(X_s)$ if and only if $s \sqsubseteq t$.*

Proof. ($\Leftarrow$) Assume first that $s \sqsubseteq t$. We have to show that $t \in \nu_{\mathcal{C}}(X_s)$. We recall that by Proposition 5.5 we have $\nu_{\mathcal{C}} = \gamma_\omega$, where, by definition, $\gamma_\omega = \bigsqcap_{n \in \mathbb{N}} \gamma_n$, for the variable interpretations γ_n defined as

$$\gamma_n := \begin{cases} \tilde{\gamma} & \text{if } n = 0 \\ \langle\langle \mathcal{C} \rangle\rangle \gamma_{n-1} & \text{if } n > 0 \end{cases}$$

where $\tilde{\gamma}$ is the top element of the complete lattice $\Gamma_{\mathcal{S}}$, namely the variable interpretation assigning $\mathcal{S}$ as meaning to each variable. Since processes are image finite, $\nu_{\mathcal{C}}(X_s) = \gamma_\omega(X_s) = \bigcap_{n \in N} \gamma_n(X_s)$ and $\sqsubseteq = \bigcap_{n \in \mathbb{N}} \sqsubseteq_n$, to prove the thesis we show the stronger property that

$$\text{for all } n \in \mathbb{N} \;\; s \sqsubseteq_n t \text{ implies } t \in \gamma_n(X_s). \tag{5.21}$$

Equation (5.21) is then proved by induction over $n \in \mathbb{N}$ with the same arguments used to prove Equation (5.15) in the proof of Theorem 5.14.

($\Rightarrow$) Assume now that $t \in \nu_{\mathcal{C}}(X_s)$. We have to prove that $s \sqsubseteq t$. To this aim, it is enough to prove that there is a probabilistic simulation relating s and t. Hence, we will prove that the relation

$$\mathcal{R} \colon = \{(s,t) \mid t \in \nu_{\mathcal{C}}(X_s)\}$$

is a probabilistic simulation. Let $s \mathcal{R} t$. We need to prove that

$$\text{whenever } s \xrightarrow{a} \pi_s \text{ there is a transition } t \xrightarrow{a} \pi_t \text{ with } \pi_s \sqsubseteq^{\dagger} \pi_t. \tag{5.22}$$

Equation (5.22) can be proved as done for Equation (5.18) in the proof of Theorem 5.14. ■

As a final remark to this Section, we notice that our characterizations of probabilistic ready similarity and similarity are not simple consequences of the results in [143]. In fact the examples of characterizations presented in [143] are built on a two-sorted modal logic which is quite different from $\mathcal{L}_{\mathcal{S}}$. Their logic allows for the occurrence of conjunctions of distribution formulae in the scope of the diamond modality and moreover the quantitative properties of processes are not captured through the probabilistic operator $\bigoplus$ but by means of an alternative version of the quantitative diamond modality in [123]: they say that a probability distribution π satisfies the distribution formula $L_p\varphi$ if and only if the total probability weight that π assigns to processes satisfying φ is at least p.

5.5 $\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATION OF BISIMILARITY METRIC

In this section we present the logical characterization of bisimilarity metric. To this aim, we introduce a suitable notion of *distance between formulae* in the $\mathcal{S}$-indexed logic $\mathcal{L}_{\mathcal{S}}$. Our distance is defined inductively over the structure of formulae and will allow us to capture the syntactic and probabilistic disparities between formulae. Then, we characterize bisimilarity metric as the distance between the mimicking formulae of processes (Theorem 5.24).

DISTANCE ON $\mathcal{L}_{\mathfrak{J}}$

We take a generic logic $\mathcal{L}_{\mathfrak{J}}$ and an image finite and guarded endodeclaration $\mathcal{E}$. Then we propose a notion of *up-to-k distance under $\mathcal{E}$* for formulae for each $k \in \mathbb{N}$, and we define the *distance under $\mathcal{E}$* as the limit of the up-to-k-distances.

Definition 5.13 (Up-to-k distance under $\mathcal{E}$)**.** Given an image finite and guarded endodeclaration $\mathcal{E}$ on $\mathcal{L}_{\mathfrak{J}}$, the *up-to-$k$ distance under $\mathcal{E}$* for $k \in \mathbb{N}$ is defined over state formulae as the mapping $\mathfrak{d}\lambda^k_{\mathcal{E}} \colon \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}} \to [0,1]$ such that

1. $\mathfrak{d}\lambda^0_{\mathcal{E}}(\varphi_1, \varphi_2) = 0$ for all $\varphi_1, \varphi_2 \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$,

2. $$\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{E}}(\varphi_1,\varphi_2) = \begin{cases} 0 & \text{if } \varphi_1 = \top = \varphi_2 \text{ or } \varphi_1 = \bar{a} = \varphi_2 \\ \lambda \cdot \mathfrak{D}_{\lambda}{}^{k}_{\mathcal{E}}(\psi_1,\psi_2) & \text{if } \varphi_1 = \langle a\rangle\psi_1 \text{ and } \varphi_2 = \langle a\rangle\psi_2 \\ \mathbf{H}(\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{E}})(\{\varphi_j \mid j \in \mathcal{J}\},\{\varphi_i \mid i \in \mathcal{I}\}) & \text{if } \varphi_1 = \bigwedge_{j\in\mathcal{J}} \varphi_j \text{ and } \varphi_2 = \bigwedge_{i\in\mathcal{I}} \varphi_i \\ \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{E}}(\mathcal{E}(X_{\mathfrak{i}}),\varphi_2) & \text{if } \varphi_1 = X_{\mathfrak{i}} \text{ for some } \mathfrak{i} \in \mathfrak{J} \\ \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{E}}(\varphi_1,\mathcal{E}(X_{\mathfrak{i}})) & \text{if } \varphi_2 = X_{\mathfrak{i}} \text{ for some } \mathfrak{i} \in \mathfrak{J} \\ 1 & \text{otherwise} \end{cases}$$

and on distribution formulae as the mapping $\mathfrak{D}_{\lambda}{}^{k}_{\mathcal{E}} \colon \mathcal{L}^{\mathrm{d}}_{\mathfrak{J}} \times \mathcal{L}^{\mathrm{d}}_{\mathfrak{J}} \to [0,1]$ such that

★ $\mathfrak{D}_{\lambda}{}^{k}_{\mathcal{E}}(\psi_1,\psi_2) = \mathbf{K}(\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{E}})(\psi_1,\psi_2)$.

Notice that since distribution formulae are probability distributions over state formulae, the Kantorovich metric is well defined on them. Moreover, we recall that the Hausdorff metric is used in the definition of bisimulation metrics to capture nondeterministic choices (Definition 2.7). Here, we use it to quantify the distance between conjunctions of formulae, which is natural since in mimicking formulae the conjunction is used to capture nondeterminism. The close relation between our distance on $\mathcal{L}_{\mathfrak{J}}$ and the Hausdorff and Kantorovich metrics will be crucial in the characterization of bisimilarity metric (Theorem 5.24).

***Example* 5.5.** Let us consider the state formulae

$$\begin{aligned} &\varphi_1^1 = \langle a\rangle\Big(\frac{1}{2}\langle b\rangle\top \oplus \frac{1}{2}\langle c\rangle\top\Big) && \varphi_1^2 = \langle a\rangle\langle b\rangle\top && \varphi_1^3 = \bar{b} \\ &\varphi_2^1 = \langle a\rangle\Big(\frac{1}{4}X_{\mathfrak{i}} \oplus \frac{3}{4}\langle c\rangle\top\Big) && \varphi_2^2 = \langle a\rangle\Big(\frac{5}{6}\langle b\rangle\top \oplus \frac{1}{6}\langle c\rangle\top\Big) && \varphi_2^3 = \bar{b} \end{aligned}$$

and, for for $\mathcal{J} = \{1,2,3\} = \mathcal{I}$, define

$$\varphi_1 = \bigwedge_{j\in\mathcal{J}} \varphi_1^j \quad \text{and} \quad \varphi_2 = \bigwedge_{i\in\mathcal{I}} \varphi_2^i.$$

We aim to evaluate $\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{E}}(\varphi_1,\varphi_2)$ for all $k \in \mathbb{N}$ and any guarded endodeclaration $\mathcal{E}$ on $\mathcal{L}_{\mathfrak{J}}$.

For $k = 0$, Definition 5.13 directly gives $\mathfrak{d}_{\lambda}{}^{0}_{\mathcal{E}}(\varphi_1,\varphi_2) = 0$.

Let us consider now the case $k = 1$. It is immediate to see that $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^3,\varphi_2^3) = 0$ and that $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^3,\varphi_2^i) = 1$ for $i \neq 3$ and $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^j,\varphi_2^3) = 1$ for $j \neq 3$. Next, we aim to evaluate $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^1,\varphi_2^1)$. By Definition 5.13 we have

$$\begin{aligned} \mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^1,\varphi_2^1) &= \lambda \cdot \mathfrak{D}_{\lambda}{}^{0}_{\mathcal{E}}\Big(\frac{1}{2}\langle b\rangle\top \oplus \frac{1}{2}\langle c\rangle\top, \frac{1}{4}X_{\mathfrak{i}} \oplus \frac{3}{4}\langle c\rangle\top\Big) \\ &= \lambda \cdot \mathbf{K}(\mathfrak{d}_{\lambda}{}^{0}_{\mathcal{E}})\Big(\frac{1}{2}\langle b\rangle\top \oplus \frac{1}{2}\langle c\rangle\top, \frac{1}{4}X_{\mathfrak{i}} \oplus \frac{3}{4}\langle c\rangle\top\Big) \\ &= 0 \end{aligned}$$

where the last equality follows from the definition of $\mathfrak{d}_{\lambda}{}^{0}_{\mathcal{E}}$. Analogously we obtain $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^1,\varphi_2^2) = \mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^2,\varphi_2^1) = \mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^2,\varphi_2^2) = 0$. Then we have

$$\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1,\varphi_2) = \mathbf{H}(\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}})(\{\varphi_1^j \mid j \in \mathcal{J}\},\{\varphi_2^i \mid i \in \mathcal{I}\})$$

$$
\begin{aligned}
&= \max\Big\{\sup_{j\in\mathcal{J}}\inf_{i\in\mathcal{I}} \mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^j,\varphi_2^i), \sup_{i\in\mathcal{I}}\inf_{j\in\mathcal{J}} \mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^j,\varphi_2^i)\Big\}\\
&= \max\{0,0\} = 0
\end{aligned}
$$

where the second equality follows since for each $j\in\mathcal{J}$ (resp. $i\in\mathcal{I}$) there is at least one $i\in\mathcal{I}$ (resp. $j\in\mathcal{J}$) such that $\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\varphi_1^j,\varphi_2^i)=0$.

Let us deal with the case $k=2$. As in case $k=1$ we have

$$
\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^3,\varphi_2^i) = \begin{cases} 0 & \text{if } i=3\\ 1 & \text{otherwise}\end{cases} \quad \text{and, analogously, } \mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^j,\varphi_2^3) = \begin{cases} 0 & \text{if } j=3\\ 1 & \text{otherwise.}\end{cases}
$$

Next, we evaluate $\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^1,\varphi_2^2)$. To this aim, let $\psi_1^1=\frac{1}{2}\langle b\rangle\top\oplus\frac{1}{2}\langle c\rangle\top$ and $\psi_2^2=\frac{5}{6}\langle b\rangle\top\oplus\frac{1}{6}\langle c\rangle\top$. Then we have

$$
\begin{aligned}
\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^1,\varphi_2^2) &= \lambda\cdot\mathbf{K}(\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}})(\psi_1^1,\psi_2^2)\\
&= \lambda\cdot \min_{\mathfrak{w}\in\mathfrak{W}(\psi_1^1,\psi_2^2)} \sum_{\phi\in\mathsf{supp}(\psi_1^1),\phi'\in\mathsf{supp}(\psi_2^2)} \mathfrak{w}(\phi,\phi')\,\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\phi,\phi'). \qquad (5.23)
\end{aligned}
$$

As $\mathsf{supp}(\psi_1^1)=\{\langle b\rangle\top,\langle c\rangle\top\}=\mathsf{supp}(\psi_2^2)$, for all $\phi\in\mathsf{supp}(\psi_1^1)$ and $\phi'\in\mathsf{supp}(\psi_2^2)$ we have

$$
\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\phi,\phi') = \begin{cases} 0 & \text{if } \phi=\phi'\\ 1 & \text{otherwise.}\end{cases}
$$

It is not hard to see that the matching $\tilde{\mathfrak{w}}\in\mathfrak{W}(\psi_1^1,\psi_2^2)$ defined as

$$
\tilde{\mathfrak{w}}(\langle b\rangle\top,\langle b\rangle\top)=\frac{1}{2},\quad \tilde{\mathfrak{w}}(\langle b\rangle\top,\langle c\rangle\top)=0,\quad \tilde{\mathfrak{w}}(\langle c\rangle\top,\langle c\rangle\top)=\frac{1}{6},\quad \tilde{\mathfrak{w}}(\langle c\rangle\top,\langle b\rangle\top)=\frac{1}{3}
$$

realizes the minimum in Equation (5.23). More precisely, we gather

$$
\begin{aligned}
(5.23) &= \lambda\cdot\Big(\frac{1}{2}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle b\rangle\top,\langle b\rangle\top)+\frac{1}{6}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle c\rangle\top,\langle c\rangle\top)+\frac{1}{3}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle c\rangle\top,\langle b\rangle\top)\Big)\\
&= \lambda\cdot\Big(\frac{1}{2}\cdot 0+\frac{1}{6}\cdot 0+\frac{1}{3}\cdot 1\Big)=\frac{\lambda}{3}.
\end{aligned}
$$

We have obtained that $\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^1,\varphi_2^2)=\lambda/3$. In a similar fashion we obtain $\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^2,\varphi_2^2)=\lambda/6$. To conclude we need to evaluate $\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^1,\varphi_2^1)$ and $\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^2,\varphi_2^1)$. These values will depend on the endodeclaration $\mathcal{E}$. In fact, if we denote $\psi_2^1=\frac{1}{4}X_{\mathrm{i}}\oplus\frac{3}{4}\langle c\rangle\top$ we have

$$
\begin{aligned}
\mathfrak{d}_{\lambda}{}^{2}_{\mathcal{E}}(\varphi_1^1,\varphi_2^1) &= \lambda\cdot\mathbf{K}(\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}})(\psi_1^1,\psi_2^1)\\
&= \lambda\cdot \min_{\mathfrak{w}\in\mathfrak{W}(\psi_1^1,\psi_2^1)} \sum_{\phi\in\mathsf{supp}(\psi_1^1),\phi'\in\mathsf{supp}(\psi_2^1)} \mathfrak{w}(\phi,\phi')\,\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\phi,\phi'). \qquad (5.24)
\end{aligned}
$$

We have $\mathsf{supp}(\psi_2^1)=\{X_{\mathrm{i}},\langle c\rangle\top\}$ and each optimal matching $\tilde{\mathfrak{w}}\in\mathfrak{W}(\psi_1^1,\psi_2^1)$ with respect to Equation (5.24) should be such that $\tilde{\mathfrak{w}}(\langle c\rangle\top,\langle c\rangle\top)=1/2$, thus implying $\tilde{\mathfrak{w}}(\langle c\rangle\top,X_{\mathrm{i}})=0$ and therefore $\tilde{\mathfrak{w}}(\langle b\rangle\top,X_{\mathrm{i}})=1/4$. Consequently, we obtain $\tilde{\mathfrak{w}}(\langle b\rangle\top,\langle c\rangle\top)=1/4$. Then we have

$$
(5.24)=\lambda\cdot\Big(\frac{1}{2}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle c\rangle\top,\langle c\rangle\top)+\frac{1}{4}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle b\rangle\top,\langle c\rangle\top)+\frac{1}{4}\mathfrak{d}_{\lambda}{}^{1}_{\mathcal{E}}(\langle b\rangle\top,X_{\mathrm{i}})\Big)
$$

$$
\begin{aligned}
&= \lambda \cdot \Big(\frac{1}{2} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{4} \mathfrak{d}_{\lambda\mathcal{E}}^{1}(\langle b\rangle\top, X_{\mathrm{i}}) \Big) \\
&= \frac{\lambda}{4}\Big(1 + \mathfrak{d}_{\lambda\mathcal{E}}^{1}(\langle b\rangle\top, X_{\mathrm{i}})\Big) \\
&= \frac{\lambda}{4}\Big(1 + \mathfrak{d}_{\lambda\mathcal{E}}^{1}\big(\langle b\rangle\top, \mathcal{E}(X_{\mathrm{i}})\big)\Big).
\end{aligned}
$$

We have therefore obtained that

$$
\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1, \varphi_2^1) = \begin{cases} \frac{\lambda}{4} & \text{if } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h \text{ for any } \psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}} \\ \frac{\lambda}{2} & \text{otherwise.} \end{cases}
$$

In a similar fashion we obtain that

$$
\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^2, \varphi_2^1) = \begin{cases} 3\frac{\lambda}{4} & \text{if } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h \text{ for any } \psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}} \\ \lambda & \text{otherwise.} \end{cases}
$$

To conclude, if $\mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h$ for any $\psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}}$, then we have

$$
\begin{aligned}
&\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1, \varphi_2) \\
&= \mathbf{H}(\mathfrak{d}_{\lambda\mathcal{E}}^{2})(\{\varphi_1^j \mid j \in \mathcal{J}\}, \{\varphi_2^i \mid i \in \mathcal{I}\}) \\
&= \max\Big\{ \sup_{j\in\mathcal{J}} \inf_{i\in\mathcal{I}} \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^j, \varphi_2^i), \sup_{i\in\mathcal{I}} \inf_{j\in\mathcal{J}} \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^j, \varphi_2^i) \Big\} \\
&= \max\{\max\{\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1, \varphi_2^1), \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^2, \varphi_2^2)\}, \max\{\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1, \varphi_2^1), \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^2, \varphi_2^2)\}\} \\
&= \max\Big\{ \max\Big\{\frac{\lambda}{4}, \frac{\lambda}{6}\Big\}, \max\Big\{\frac{\lambda}{4}, \frac{\lambda}{6}\Big\} \Big\} \\
&= \frac{\lambda}{4}.
\end{aligned}
$$

If, conversely, $\mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h$ for any $\psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}}$, then

$$
\begin{aligned}
&\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1, \varphi_2) \\
&= \mathbf{H}(\mathfrak{d}_{\lambda\mathcal{E}}^{2})(\{\varphi_1^j \mid j \in \mathcal{J}\}, \{\varphi_2^i \mid i \in \mathcal{I}\}) \\
&= \max\Big\{ \sup_{j\in\mathcal{J}} \inf_{i\in\mathcal{I}} \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^j, \varphi_2^i), \sup_{i\in\mathcal{I}} \inf_{j\in\mathcal{J}} \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^j, \varphi_2^i) \Big\} \\
&= \max\{\max\{\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1, \varphi_2^2), \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^2, \varphi_2^2)\}, \max\{\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1, \varphi_2^1), \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^2, \varphi_2^2)\}\} \\
&= \max\Big\{ \max\Big\{\frac{\lambda}{3}, \frac{\lambda}{6}\Big\}, \max\Big\{\frac{\lambda}{2}, \frac{\lambda}{6}\Big\} \Big\} \\
&= \frac{\lambda}{2}.
\end{aligned}
$$

Finally, let us deal with the case $k \geq 3$. First of all we notice that if $\mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h$ for any $\psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}}$, then $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1, \varphi_2) = \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1, \varphi_2) = \frac{\lambda}{2}$ for all $k \geq 3$. If $\mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}} \langle b\rangle \psi_h$ for any $\psi_h \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{d}}$, we obtain that $\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1, \varphi_2) \geq \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1, \varphi_2) = \frac{\lambda}{4}$, with $\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1, \varphi_2) = \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1, \varphi_2)$, if $\mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \langle b\rangle\top$, and $\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1,\varphi_2) > \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1,\varphi_2)$, otherwise. Then we have $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1,\varphi_2) = \mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1,\varphi_2)$ for all $k > 3$. For instance, consider the evaluation of $\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1^1,\varphi_2^1)$. From previous considerations we gather

$$\begin{aligned}
\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1^1,\varphi_2^1) &= \lambda\cdot\mathbf{K}(\mathfrak{d}_{\lambda\mathcal{E}}^{1})(\psi_1^1,\psi_2^1)\\
&= \lambda\cdot \min_{\mathfrak{w}\in\mathfrak{W}(\psi_1^1,\psi_2^1)} \sum_{\phi\in\mathsf{supp}(\psi_1^1),\phi'\in\mathsf{supp}(\psi_2^1)} \mathfrak{w}(\phi,\phi')\,\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\phi,\phi')\\
&= \lambda\cdot\Big(\frac{1}{2}\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle c\rangle\top,\langle c\rangle\top) + \frac{1}{4}\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle b\rangle\top,\langle c\rangle\top) + \frac{1}{4}\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle b\rangle\top,X_{\mathrm{i}})\Big)\\
&= \lambda\cdot\Big(\frac{1}{2}\cdot 0 + \frac{1}{4}\cdot 1 + \frac{1}{4}\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle b\rangle\top,X_{\mathrm{i}})\Big)\\
&= \frac{\lambda}{4}\Big(1 + \mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle b\rangle\top,X_{\mathrm{i}})\Big)\\
&= \frac{\lambda}{4}\Big(1 + \mathfrak{d}_{\lambda\mathcal{E}}^{2}\big(\langle b\rangle\top,\mathcal{E}(X_{\mathrm{i}})\big)\Big) \qquad (5.25)
\end{aligned}$$

where

$$\mathfrak{d}_{\lambda\mathcal{E}}^{2}\big(\langle b\rangle\top,\mathcal{E}(X_{\mathrm{i}})\big) = \begin{cases} 0 & \text{if } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \langle b\rangle\top\\ \lambda & \text{if } \mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \langle b\rangle\top \text{ and } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}}\langle b\rangle\psi_h\\ 1 & \text{otherwise.}\end{cases}$$

Therefore we obtain that

$$\mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1^1,\varphi_2^1) = \begin{cases} \frac{\lambda}{4} & \text{if } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \langle b\rangle\top\\ \frac{\lambda}{4}(1+\lambda) & \text{if } \mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \langle b\rangle\top \text{ and } \mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}}\langle b\rangle\psi_h\\ \frac{\lambda}{2} & \text{otherwise.}\end{cases}$$

In particular this implies that for $\mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \langle b\rangle\top$ and $\mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}}\langle b\rangle\psi_h$ we have

$$\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\varphi_1^1,\varphi_2^1) = \frac{\lambda}{4} \le \frac{\lambda}{4}(1+\lambda) = \mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1^1,\varphi_2^1).$$

Finally we notice that if $\mathcal{E}(X_{\mathrm{i}}) \not\equiv_{\mathcal{E}} \langle b\rangle\top$ and $\mathcal{E}(X_{\mathrm{i}}) \equiv_{\mathcal{E}} \bigwedge_{h\in\mathcal{H}}\langle b\rangle\psi_h$ then $\mathfrak{d}_{\lambda\mathcal{E}}^{2}(\langle b\rangle\top,\mathcal{E}(X_{\mathrm{i}})) = \lambda$ independently from the structure of the distribution formula ψ, thus giving $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\langle b\rangle\top,\mathcal{E}(X_{\mathrm{i}})) = \lambda$ for all $k \ge 2$. Therefore we can conclude that $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1^1,\varphi_2^1) = \mathfrak{d}_{\lambda\mathcal{E}}^{3}(\varphi_1^1,\varphi_2^1)$ for all $k > 3$. ◀

We show now that each mapping $\mathfrak{d}_{\lambda\mathcal{E}}^{k}$ is actually a pseudometric bounded by 1.

Proposition 5.18. *All mappings $\mathfrak{d}_{\lambda\mathcal{E}}^{k}$ with $k \in \mathbb{N}$ are 1-bounded pseudometrics.*

Proof. We prove first that each $\mathfrak{d}_{\lambda\mathcal{E}}^{k}$ is a pseudometric, namely that we have:

1. $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi,\varphi) = 0$, for all $\varphi \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$,

2. $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1,\varphi_2) = \mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_2,\varphi_1)$, for all $\varphi_1,\varphi_2 \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$,

3. $\mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1,\varphi_2) \le \mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_1,\varphi_3) + \mathfrak{d}_{\lambda\mathcal{E}}^{k}(\varphi_3,\varphi_2)$ for all $\varphi_1,\varphi_2,\varphi_3 \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$.

The proofs of the three properties are by induction over $k \in \mathbb{N}$ and over the structure of formulae. The interesting case is property 3, where the base case $k = 0$ of the inductive reasoning follows immediately by $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{0}(\varphi_i, \varphi_j) = 0$ for all $i, j \in \{1,2,3\}$, and the proof of the inductive step $k > 0$ is by structural induction over φ_1. In particular we notice that the proof for the case of the distance on conjunctions of formulae follows from its definition through the Hausdorff functional and Proposition 2.2. Similarly, the proof for the case of the distance between diamond operators follows from the definition of the distance on distribution formulae in terms of the Kantorovich functional and Proposition 2.1.

To conclude, we need to show that for all $k \in \mathbb{N}$ the pseudometric $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}$ is 1-bounded, namely $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi') \leq 1$. This can be done through a simple induction over $k \in \mathbb{N}$ and over the structure of formulae. ■

As expected, $(\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi'))_{k \in \mathbb{N}}$ is a non decreasing sequence of distances.

Proposition 5.19. *For all $k \in \mathbb{N}$ and $\varphi, \varphi' \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$, we have $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k+1}(\varphi, \varphi') \geq \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi')$.*

Proof. The proof is by induction over $k \in \mathbb{N}$ and over the structure of formulae. ■

Being all $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}$ 1-bounded, the sequence $(\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi'))_{k \in \mathbb{N}}$ has a limit in $[0,1]$.

Proposition 5.20. *For all $\varphi, \varphi' \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$, $\lim_{k \to \infty} \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi') \in [0,1]$.*

Proof. Since $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}$ is a 1-bounded pseudometric (Proposition 5.18), for all $k \in \mathbb{N}$ we have $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi') \leq 1$. Then, by Proposition 5.19 we have $\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi') \leq \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k+1}(\varphi, \varphi')$. Hence $(\mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi'))_{k \in \mathbb{N}}$ is a 1-bounded non decreasing sequence. This ensures that $\lim_{k \to \infty} \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi')$ exists and $\lim_{k \to \infty} \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi') \in [0,1]$. ■

Hence we can define the distance under $\mathcal{E}$ between formuale as the limit of their up-to-k distances under $\mathcal{E}$.

Definition 5.14 (Distance under $\mathcal{E}$)**.** Given an image finite and guarded endodeclaration $\mathcal{E}$ on $\mathcal{L}_{\mathfrak{J}}$, the *distance under* $\mathcal{E}$ is defined over state formulae as the mapping $\mathfrak{d}_{\lambda_{\mathcal{E}}} \colon \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \times \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}} \to [0,1]$ such that

$$\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') = \lim_{k \to \infty} \mathfrak{d}_{\lambda_{\mathcal{E}}}^{k}(\varphi, \varphi').$$

By Proposition 5.20 we are guaranteed that $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ is well defined. Now we show that $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ is a 1-bounded pseudometric.

Proposition 5.21. *The mapping $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ is a 1-bounded pseudometric.*

Proof. By Proposition 5.20 we already know that $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi) \leq 1$ for all $\varphi \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$. To prove that $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ is a pseudometric we need to show the following properties:

1. $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi) = 0$, for all $\varphi \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$,

2. $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_1, \varphi_2) = \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_2, \varphi_1)$, for all $\varphi_1, \varphi_2 \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$,

3. $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_1, \varphi_2) \leq \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_1, \varphi_3) + \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_3, \varphi_2)$ for all $\varphi_1, \varphi_2, \varphi_3 \in \mathcal{L}_{\mathfrak{J}}^{\mathrm{s}}$.

Since by definition $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') = \lim_{k\to\infty} \mathfrak{d}^k_{\lambda_{\mathcal{E}}}(\varphi, \varphi')$, the three properties follow by Proposition 5.18 and the linearity of the limit. ■

We show now that the kernel of the distance $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ is given by the $\mathcal{L}_{\mathfrak{J}}$-equivalence under $\mathcal{E}$, namely $\equiv_{\mathcal{E}}$.

Proposition 5.22. *Given $\varphi, \varphi' \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$, we have $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') = 0$ if and only if $\varphi \equiv_{\mathcal{E}} \varphi'$.*

Proof. (⇒). Assume first $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') = 0$. To prove $\varphi \equiv_{\mathcal{E}} \varphi'$ we can proceed by structural induction over φ. The interesting cases are the inductive steps of conjunction and diamond operator.

- ★ Inductive step $\varphi = \bigwedge_{j\in\mathcal{J}} \varphi_j$. Consider $\varphi' = \bigwedge_{i\in\mathcal{I}} \varphi_i$. Then

$$\begin{aligned}
0 &= \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') \\
&= \mathfrak{d}_{\lambda_{\mathcal{E}}}(\bigwedge_{j\in\mathcal{J}} \varphi_j, \bigwedge_{i\in\mathcal{I}} \varphi_i) \\
&= \max\{\sup_{j\in\mathcal{J}} \inf_{i\in\mathcal{I}} \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_i), \sup_{i\in\mathcal{I}} \inf_{j\in\mathcal{J}} \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_j)\} \\
\text{iff } & \sup_{j\in\mathcal{J}} \inf_{i\in\mathcal{I}} \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_i) = 0 \text{ and } \sup_{i\in\mathcal{I}} \inf_{j\in\mathcal{J}} \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_i) = 0 \\
\text{iff } & \text{for all } j \in \mathcal{J} \text{ there is an } i_j \in \mathcal{I} \text{ such that } \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_{i_j}) = 0 \\
& \text{and for all } i \in \mathcal{I} \text{ there is a } j_i \in \mathcal{J} \text{ such that } \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_{j_i}, \varphi_i) = 0.
\end{aligned}$$

  By structural induction on φ_j and φ_{j_i}, this implies that $\varphi_j \equiv_{\mathcal{E}} \varphi_{i_j}$ and $\varphi_{j_i} \equiv_{\mathcal{E}} \varphi_i$. Hence, if we define a function $f\colon \mathcal{J} \to \mathcal{I}$ as $f(j) = i_j$ and a function $g\colon \mathcal{I} \to \mathcal{J}$ as $g(i) = j_i$ then we obtain that $\varphi_j \equiv_{\mathcal{E}} \varphi_{f(j)}$, for all $j \in \mathcal{J}$, and $\varphi_i \equiv_{\mathcal{E}} \varphi_{g(i)}$, for all $i \in \mathcal{I}$. Next, we note that if f is not injective then we can reduce $\mathcal{J}$ modulo $\mathcal{L}_{\mathfrak{J}}$-equivalence under $\mathcal{E}$ in order to make it so. In detail, assume that there are two indexes $j_1, j_2 \in \mathcal{J}$ such that $f(j_1) = f(j_2)$, from which we draw $\varphi_{j_1} \equiv_{\mathcal{E}} \varphi_{f(j_1)} = \varphi_{f(j_2)}$ and $\varphi_{j_2} \equiv_{\mathcal{E}} \varphi_{f(j_2)} = \varphi_{f(j_1)}$. Hence, by transitivity of $\equiv_{\mathcal{E}}$, we obtain $\varphi_{j_1} \equiv_{\mathcal{E}} \varphi_{j_2}$. Thus, we can consider the formula $\bigwedge_{j\in(\mathcal{J}\setminus\{j_2\})} \varphi_j \equiv_{\mathcal{E}} \bigwedge_{j\in\mathcal{J}} \varphi_j$ (by Definition 5.9). We can repeat this way of reasoning until we obtain a set of indexes $\mathcal{J}' \subseteq \mathcal{J}$ such that $\varphi \equiv_{\mathcal{E}} \bigwedge_{j\in\mathcal{J}'} \varphi_j$ and $f\colon \mathcal{J}' \to \mathcal{I}$ is injective. With an analogous reasoning, we obtain that there exists a set of indexes $\mathcal{I}' \subseteq \mathcal{I}$ such that $\varphi' \equiv_{\mathcal{E}} \bigwedge_{i\in\mathcal{I}'} \varphi_i$ and $g\colon \mathcal{I}' \to \mathcal{J}$ is injective. Moreover, we note that given $j \in \mathcal{J}'$, by construction, we have $f(j) = i_j$, which implies $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_j, \varphi_{i_j}) = 0$. Thus, $j = j_{i_j}$ that is $j = g(i_j)$. Furthermore, if we restrict the co-domain of f to $\mathcal{I}'$, by construction we have that i_j is unique modulo $\equiv_{\mathcal{E}}$ and for all $i \in \mathcal{I}'$ there is a j in $\mathcal{J}'$ such that $i = f(j)$. Hence we gather that $f\colon \mathcal{J}' \to \mathcal{I}'$ is also surjective. Since a similar argument holds for g, we can conclude that f is a bijective function with g as inverse. Therefore, we have that there is a bijective function $f\colon \mathcal{J}' \to \mathcal{I}'$ such that $\varphi_j \equiv_{\mathcal{E}} \varphi_{f(j)}$ for all $j \in \mathcal{J}'$. Hence we have

$$\begin{aligned}
\bigwedge_{j\in\mathcal{J}} \varphi_j &\equiv_{\mathcal{E}} \bigwedge_{j\in\mathcal{J}'} \varphi_j && \text{(by construction of } \mathcal{J}'\text{)} \\
&\equiv_{\mathcal{E}} \bigwedge_{i\in\mathcal{I}'} \varphi_i && \text{(by the choice of } f \text{ and Definition 5.9)}
\end{aligned}$$

$$\equiv_{\mathcal{E}} \bigwedge_{i \in \mathcal{I}} \varphi_i \qquad\qquad \text{(by construction of } \mathcal{I}'\text{).}$$

★ Inductive step $\varphi = \langle a \rangle \bigoplus_{i \in I} r_i \varphi_i$. By definition, $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') < 1$ only if $\varphi' \equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{j \in J} r_j \varphi_j$. Let $\psi_\varphi = \bigoplus_{i \in I} r_i \varphi_i$ and $\psi_{\varphi'} = \bigoplus_{j \in J} r_j \varphi_j$. We have

$$\begin{aligned} 0 &= \mathfrak{d}_{\lambda_{\mathcal{E}}}(\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i, \langle a \rangle \bigoplus_{j \in J} r_j \varphi_j) \\ &= \lambda \cdot \mathfrak{D}_{\lambda_{\mathcal{E}}}(\bigoplus_{i \in I} r_i \varphi_i, \bigoplus_{j \in J} r_j \varphi_j) \\ &= \lambda \min_{\mathfrak{w} \in \mathfrak{W}(\psi_\varphi, \psi_{\varphi'})} \sum_{i \in I, j \in J} \mathfrak{w}(\varphi_i, \varphi_j) \mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_i, \varphi_j). \end{aligned}$$

For each $i \in I$ and $j \in J$ we can distinguish two cases:

- ∗ either $\mathfrak{w}(\varphi_i, \varphi_j) = 0$,
- ∗ or $\mathfrak{w}(\varphi_i, \varphi_j) > 0$, implying $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi_i, \varphi_j) = 0$. By induction we get $\varphi_i \equiv_{\mathcal{E}} \varphi_j$.

For each i, let J_i be the set of indexes j_i for which $\mathfrak{w}(\varphi_i, \varphi_{j_i}) > 0$ and, symmetrically, for each j let I_j be the set of indexes i_j for which $\mathfrak{w}(\varphi_{i_j}, \varphi_j) > 0$. So we have

$$\begin{aligned} \varphi' &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{j \in J} \Big(\sum_{i \in I} \mathfrak{w}(\varphi_j, \varphi_i) \Big) \varphi_j \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{j \in J} \Big(\sum_{i_j \in I} \mathfrak{w}(\varphi_j, \varphi_{i_j}) \Big) \varphi_j \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{j \in J, i_j \in I} \mathfrak{w}(\varphi_j, \varphi_{i_j}) \varphi_{i_j} \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{j \in J, i_j \in I, j'_{i_j} \in J} \mathfrak{w}(\varphi_{j'_{i_j}}, \varphi_{i_j}) \varphi_{j'_{i_j}} \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{i \in I, j_i \in J} \mathfrak{w}(\varphi_{j_i}, \varphi_i) \varphi_{j_i} \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{i \in I} \Big(\sum_{j_i \in J} \mathfrak{w}(\varphi_{j_i}, \varphi_i) \Big) \varphi_i \\ &\equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{i \in I} \Big(\sum_{j \in J} \mathfrak{w}(\varphi_j, \varphi_i) \Big) \varphi_i \\ &\equiv_{\mathcal{E}} \varphi. \end{aligned}$$

($\Leftarrow$). Assume now that $\varphi \equiv_{\mathcal{E}} \varphi'$. To prove $\mathfrak{d}_{\lambda_{\mathcal{E}}}(\varphi, \varphi') = 0$ we can proceed by structural induction over φ. The interesting case is that of the diamond operator.

Consider the inductive step $\varphi = \langle a \rangle \bigoplus_{i \in I} r_i \varphi_i$. By definition, $\varphi' \equiv_{\mathcal{E}} \langle a \rangle \bigoplus_{i \in I} r_i \varphi_i$ would imply $\varphi' = \bigwedge_{j \in \mathcal{J}} \langle a \rangle \psi_j$ for an arbitrary set of indexes $\mathcal{J} \neq \emptyset$ so that $\psi_j (\equiv_{\mathcal{E}})^{\dagger} \bigoplus_{i \in I} r_i \varphi_i$ for all $j \in \mathcal{J}$. We consider only the case of $|\mathcal{J}| = 1$, since the general case for $|\mathcal{J}| > 1$ directly follows from it. Hence let us consider $\varphi' = \langle a \rangle \bigoplus_{\substack{i \in I \\ j_i \in J_i}} r_{j_i} \varphi_{j_i}$ with $\sum_{j_i \in J_i} r_{j_i} = r_i$ and $\varphi_{j_i} \equiv_{\mathcal{E}} \varphi_i$ for all $j_i \in J_i$. Let $\psi_\varphi = \bigoplus_{i \in I} r_i \varphi_i$ and $\psi_{\varphi'} = \bigoplus_{\substack{i \in I \\ j_i \in J_i}} r_{j_i} \varphi_{j_i}$. Then

$$\mathfrak{d}_{\lambda_{\mathcal{E}}}(\langle a \rangle \bigoplus_{i \in I} r_i \varphi_i, \langle a \rangle \bigoplus_{i \in I, j_i \in J_i} r_{j_i} \varphi_{j_i})$$

$$
\begin{aligned}
&= \lambda\, \mathfrak{D}_{\lambda\mathcal{E}}\big(\bigoplus_{i\in I} r_i\varphi_i, \bigoplus_{i\in I,\, j_i\in J_i} r_{j_i}\varphi_{j_i}\big)\\
&= \lambda \min_{\mathfrak{w}\in\mathfrak{W}(\psi_\varphi,\psi_{\varphi'})} \sum_{i\in I,\, j_h\in J_h} \mathfrak{w}(\varphi_i,\varphi_{j_h})\,\mathfrak{d}_{\lambda\mathcal{E}}(\varphi_i,\varphi_{j_h})\\
&\le \lambda \sum_{i\in I,\, j_h\in J_h} \tilde{\mathfrak{w}}(\varphi_i,\varphi_{j_h})\,\mathfrak{d}_{\lambda\mathcal{E}}(\varphi_i,\varphi_{j_h})\\
&= \lambda \sum_{i\in I,\, j_i\in J_i} r_{j_i}\,\mathfrak{d}_{\lambda\mathcal{E}}(\varphi_i,\varphi_{j_i})\\
&= \lambda \sum_{i\in I,\, j_i\in J_i} r_{j_i}0 && \text{(by structural induction over } \varphi_i\text{)}\\
&= 0
\end{aligned}
$$

where the inequality follows by observing that function $\tilde{\mathfrak{w}}$ defined by

$$\tilde{\mathfrak{w}}(\varphi_i,\varphi_{j_h}) = \begin{cases} r_{j_i} & \text{if } h = i\\ 0 & \text{otherwise}\end{cases}$$

is a matching in $\mathfrak{W}(\psi_\varphi,\psi_{\varphi'})$. ■

Logical distance between processes

Let us focus now on $\mathcal{L}_{\mathcal{S}}$ and on the mimicking endodeclaration $\mathcal{M}$. By exploiting the distance between formulae under $\mathcal{M}$ and the close relation between processes and their own mimicking formulae, we define a distance between processes. All distances between probabilistic processes proposed so far take into account the disparities in their branching structures as well as the differences between the probabilistic choices, in order to conciliate behavioral equivalence with quantitative properties. By construction, each mimicking formula is univocally determined by the process and in turn the branching and probabilistic structure of the process are univocally captured by the mimicking formula. Hence, we define the *logical distance* on processes as the distance between their mimicking formulae.

Definition 5.15 (Logical bisimulation distance)**.** Let $\lambda \in (0,1]$. For any $k \in \mathbb{N}$, the *up-to-k logical bisimulation distance* over processes $\ell_\lambda^k\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by

$$\ell_\lambda^k(s,t) = \mathfrak{d}_{\lambda\mathcal{M}}^k(X_s,X_t).$$

Then the *logical bisimulation distance* over processes $\ell_\lambda\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by

$$\ell_\lambda(s,t) = \mathfrak{d}_{\lambda\mathcal{M}}(X_s,X_t).$$

Notice that both ℓ_λ^k and ℓ_λ are well-defined since $\mathcal{M}$ is image finite and guarded.

Notice also that $\ell_\lambda(s,t) = \lim_{k\to\infty}\ell_\lambda^k(s,t)$ (see Definition 5.14).

We give now the characterization result for up-to-k-bisimilarity metric.

Theorem 5.23. *For all $k\in\mathbb{N}$ and for all processes $s,t\in\mathcal{S}$ we have*

$$\ell_\lambda^k(s,t) = \mathbf{d}_\lambda^k(s,t).$$

Proof. We proceed by induction over k.

The base case $k=0$ is immediate, since both $\mathfrak{d}_{\lambda}{}^{0}_{\mathcal{M}}$ and $\mathbf{d}_{\lambda}^{0}$ equal the zero function $\mathbf{0}$.

Consider now the inductive step $k>0$. To simplify reasoning we exploit the nesting of conjunctions in the definition of mimicking formulae. However, we remark that the proof can be written without nesting the conjunctions but the presentation of results and general reasoning would turn out to be technically heavier.

For any process s, we define

$$\phi_s = \bigwedge_{(s,a,\pi_s)\in\rightarrow} \langle a\rangle \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'} \qquad \text{and} \qquad \theta_s = \bigwedge_{b\not\in\mathrm{init}(s)} \bar{b}$$

so that the mimicking formula of s can be rewritten as $\varphi_s = \phi_s \wedge \theta_s$. Hence, for any pair of processes $s,t\in\mathcal{S}$ we have

$$\begin{aligned}
\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(X_s,X_t) &= \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\varphi_s,\varphi_t)\\
&= \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s\wedge\theta_s,\phi_t\wedge\theta_t)\\
&= \max\left\{\begin{array}{l}
\max\left\{\begin{array}{l}\min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\theta_t)\}\\ \min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\phi_t),\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}\end{array}\right\}\\
\max\left\{\begin{array}{l}\min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\phi_t)\}\\ \min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\theta_t),\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}\end{array}\right\}
\end{array}\right\}\\
&= \max\left\{\begin{array}{l}
\max\left\{\begin{array}{l}\min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\, 1\}\\ \min\{1,\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}\end{array}\right\}\\
\max\left\{\begin{array}{l}\min\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\, 1\}\\ \min\{1,\, \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}\end{array}\right\}
\end{array}\right\}\\
&= \max\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}
\end{aligned}$$

where the second last equality follows by

$$\begin{aligned}
\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\theta_t) &= \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}\left(\bigwedge_{(s,a,\pi_s)\in\rightarrow}\langle a\rangle\bigoplus_{s'\in\mathsf{supp}(\pi_s)}\pi_s(s')X_{s'},\ \bigwedge_{b'\not\in\mathrm{init}(t)}\bar{b}'\right)\\
&= \max\left\{\begin{array}{l}
\sup\limits_{(s,a,\pi_s)\in\rightarrow}\ \inf\limits_{b'\not\in\mathrm{init}(t)}\ \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}\left(\langle a\rangle\bigoplus\limits_{s'\in\mathsf{supp}(\pi_s)}\pi_s(s')X_{s'},\, \bar{b}'\right)\\
\sup\limits_{b'\not\in\mathrm{init}(t)}\ \inf\limits_{(s,a,\pi_s)\in\rightarrow}\ \mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}\left(\langle a\rangle\bigoplus\limits_{s'\in\mathsf{supp}(\pi_s)}\pi(s')X_{s'},\, \bar{b}'\right)
\end{array}\right\}\\
&= \max\left\{\sup_{(s,a,\pi_s)\in\rightarrow}\ \inf_{b'\not\in\mathrm{init}(t)}\{1\},\ \sup_{b'\not\in\mathrm{init}(t)}\ \inf_{(s,a,\pi_s)\in\rightarrow}\{1\}\right\} = 1
\end{aligned}$$

and, analogously, $\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\phi_t)=1$.

Summarizing, we need to show that

$$\mathbf{d}_{\lambda}^{k+1}(s,t) = \max\{\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\phi_s,\phi_t),\mathfrak{d}_{\lambda}{}^{k+1}_{\mathcal{M}}(\theta_s,\theta_t)\}. \tag{5.26}$$

To prove Equation (5.26) we distinguish two cases: either $\mathrm{init}(s) \neq \mathrm{init}(t)$ or $\mathrm{init}(s) = \mathrm{init}(t)$. Consider the case $\mathrm{init}(s) \neq \mathrm{init}(t)$. Assume wlog. that $\hat{b} \in \mathrm{init}(s) \setminus \mathrm{init}(t)$. The case $\hat{b} \in \mathrm{init}(t) \setminus \mathrm{init}(s)$ is analogous. Under this assumption, we have

$$\begin{aligned} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\theta_s, \theta_t) &= \max\{ \sup_{b \notin \mathrm{init}(s)} \inf_{b' \notin \mathrm{init}(t)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}'), \sup_{b' \notin \mathrm{init}(t)} \inf_{b \notin \mathrm{init}(s)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}')\} \\ &\geq \max\{ \sup_{b \notin \mathrm{init}(s)} \inf_{b' \notin \mathrm{init}(t)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}'), \inf_{b \notin \mathrm{init}(s)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{\hat{b}})\} \\ &= \max\{ \sup_{b \notin \mathrm{init}(s)} \inf_{b' \notin \mathrm{init}(t)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}'), 1\} = 1 \end{aligned}$$

where the second last equality follows from $\hat{b} \in \mathrm{init}(s)$. Hence Equation (5.26) instantiates as

$$\mathbf{d}_{\lambda}^{k+1}(s,t) = \max\{\mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\phi_s, \phi_t), \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\theta_s, \theta_t)\} = \max\{\mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\phi_s, \phi_t), 1\} = 1$$

which holds by $\mathrm{init}(s) \neq \mathrm{init}(t)$ and the following result from [92] (Proposition 2.16): *Let $s, t \in \mathcal{S}$ be two processes with* $\mathrm{init}(s) \neq \mathrm{init}(t)$. *Then, for all $k > 0$ it holds that* $\mathbf{d}_{\lambda}^{k}(s,t) = 1$.

The second case is $\mathrm{init}(s) = \mathrm{init}(t)$. We prove first that $\mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\theta_s, \theta_t) = 0$. From $\mathrm{init}(s) = \mathrm{init}(t)$ it follows that for each $b \in \mathcal{A}$ we have $s \overset{b}{\not\to}$ if and only if $t \overset{b}{\not\to}$. Hence we have

$$\begin{aligned} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\theta_s, \theta_t) &= \max\{ \sup_{b \notin \mathrm{init}(s)} \inf_{b' \notin \mathrm{init}(t)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}'), \sup_{b' \notin \mathrm{init}(t)} \inf_{b \notin \mathrm{init}(s)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}')\} \\ &\leq \max\{ \sup_{b \notin \mathrm{init}(s)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}, \bar{b}), \sup_{b' \notin \mathrm{init}(t)} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bar{b}', \bar{b}')\} \\ &= \max\{ \sup_{b \notin \mathrm{init}(s)} 0, \sup_{b' \notin \mathrm{init}(t)} 0\} = 0. \end{aligned}$$

Therefore, Equation (5.26) becomes

$$\mathbf{d}_{\lambda}^{k+1}(s,t) = \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\phi_s, \phi_t) \tag{5.27}$$

which follows by

$$\begin{aligned} &\mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\phi_s, \phi_t) \\ &= \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}(\bigwedge_{(s,a,\pi_s) \in \to} \langle a \rangle \bigoplus_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') X_{s'}, \bigwedge_{(t,a,\pi_t) \in \to} \langle a \rangle \bigoplus_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') X_{t'}) \\ &= \max \left\{ \begin{array}{l} \sup\limits_{(s,a,\pi_s) \in \to} \inf\limits_{(t,a,\pi_t) \in \to} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}\left(\langle a \rangle \bigoplus\limits_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') X_{s'}, \langle a \rangle \bigoplus\limits_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') X_{t'} \right) \\ \sup\limits_{(t,a,\pi_t) \in \to} \inf\limits_{(s,a,\pi_s) \in \to} \mathfrak{d}_{\lambda\mathcal{M}}^{k+1}\left(\langle a \rangle \bigoplus\limits_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') X_{s'}, \langle a \rangle \bigoplus\limits_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') X_{t'} \right) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} \sup\limits_{a \in \mathcal{A}} \sup\limits_{(s,a,\pi_s) \in \to} \inf\limits_{(t,a,\pi_t) \in \to} \lambda \mathfrak{D}_{\lambda\mathcal{M}}^{k}\left(\bigoplus\limits_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') X_{s'}, \bigoplus\limits_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') X_{t'} \right) \\ \sup\limits_{a \in \mathcal{A}} \sup\limits_{(t,a,\pi_t) \in \to} \inf\limits_{(s,a,\pi_s) \in \to} \lambda \mathfrak{D}_{\lambda\mathcal{M}}^{k}\left(\bigoplus\limits_{s' \in \mathrm{supp}(\pi_s)} \pi_s(s') X_{s'}, \bigoplus\limits_{t' \in \mathrm{supp}(\pi_t)} \pi_t(t') X_{t'} \right) \end{array} \right\} \\ &= \max \left\{ \begin{array}{l} \sup\limits_{a \in \mathcal{A}} \sup\limits_{(s,a,\pi_s) \in \to} \inf\limits_{(t,a,\pi_t) \in \to} \lambda \min\limits_{\mathfrak{w} \in \mathfrak{W}(\psi_{\pi_s}, \psi_{\pi_t})} \sum\limits_{s', t' \in \mathcal{S}} \mathfrak{w}(X_{s'}, X_{t'}) \mathfrak{d}_{\lambda\mathcal{M}}^{k}(X_{s'}, X_{t'}) \\ \sup\limits_{a \in \mathcal{A}} \sup\limits_{(t,a,\pi_t) \in \to} \inf\limits_{(s,a,\pi_s) \in \to} \lambda \min\limits_{\mathfrak{w} \in \mathfrak{W}(\psi_{\pi_s}, \psi_{\pi_t})} \sum\limits_{s', t' \in \mathcal{S}} \mathfrak{w}(X_{s'}, X_{t'}) \mathfrak{d}_{\lambda\mathcal{M}}^{k}(X_{s'}, X_{t'}) \end{array} \right\} \end{aligned}$$

$$= \max\left\{\begin{array}{l} \sup\limits_{a\in\mathcal{A}} \sup\limits_{(s,a,\pi_s)\in\rightarrow} \inf\limits_{(t,a,\pi_t)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\,\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{M}}(\mathcal{M}(s'),\mathcal{M}(t')) \\ \sup\limits_{a\in\mathcal{A}} \sup\limits_{(t,a,\pi_t)\in\rightarrow} \inf\limits_{(s,a,\pi_s)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\,\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{M}}(\mathcal{M}(s'),\mathcal{M}(t')) \end{array}\right\}$$

$$= \max\left\{\begin{array}{l} \sup\limits_{a\in\mathcal{A}} \sup\limits_{(s,a,\pi_s)\in\rightarrow} \inf\limits_{(t,a,\pi_t)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\,\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{M}}(\varphi_{s'},\varphi_{t'}) \\ \sup\limits_{a\in\mathcal{A}} \sup\limits_{(t,a,\pi_t)\in\rightarrow} \inf\limits_{(s,a,\pi_s)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\,\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{M}}(\varphi_{s'},\varphi_{t'}) \end{array}\right\}$$

$$= \max\left\{\begin{array}{l} \sup\limits_{a\in\mathcal{A}} \sup\limits_{(s,a,\pi_s)\in\rightarrow} \inf\limits_{(t,a,\pi_t)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\mathbf{d}_{\lambda}^{k}(s',t') \\ \sup\limits_{a\in\mathcal{A}} \sup\limits_{(t,a,\pi_t)\in\rightarrow} \inf\limits_{(s,a,\pi_s)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(X_{s'},X_{t'})\mathbf{d}_{\lambda}^{k}(s',t') \end{array}\right\}$$

$$= \max\left\{\begin{array}{l} \sup\limits_{a\in\mathcal{A}} \sup\limits_{(s,a,\pi_s)\in\rightarrow} \inf\limits_{(t,a,\pi_t)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\pi_s,\pi_t)} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(s',t')\mathbf{d}_{\lambda}^{k}(s',t') \\ \sup\limits_{a\in\mathcal{A}} \sup\limits_{(t,a,\pi_t)\in\rightarrow} \inf\limits_{(s,a,\pi_s)\in\rightarrow} \lambda \min\limits_{\mathfrak{w}\in\mathfrak{W}(\pi_s,\pi_t)} \sum\limits_{s',t'\in\mathcal{S}} \mathfrak{w}(s',t')\mathbf{d}_{\lambda}^{k}(s',t') \end{array}\right\}$$

$$= \max\left\{\begin{array}{l} \sup\limits_{a\in\mathcal{A}} \sup\limits_{(s,a,\pi_s)\in\rightarrow} \inf\limits_{(t,a,\pi_t)\in\rightarrow} \lambda\mathbf{K}(\mathbf{d}_{\lambda}^{k})(\pi_s,\pi_t) \\ \sup\limits_{a\in\mathcal{A}} \sup\limits_{(t,a,\pi_t)\in\rightarrow} \inf\limits_{(s,a,\pi_s)\in\rightarrow} \lambda\mathbf{K}(\mathbf{d}_{\lambda}^{k})(\pi_s,\pi_t) \end{array}\right\}$$

$$= \mathbf{d}_{\lambda}^{k+1}(s,t)$$

where

- $\psi_{\pi_s} = \bigoplus_{s'\in\mathsf{supp}(\pi_s)} \pi_s(s')X_{s'}$;
- $\psi_{\pi_t} = \bigoplus_{t'\in\mathsf{supp}(\pi_t)} \pi_t(t')X_{t'}$;
- the fifth equality follows by Definition 5.13;
- the sixth equality follows by definition of the mimicking endodeclaration $\mathcal{M}$;
- the seventh equality follows by the inductive hypothesis;
- the eight equality holds since each matching $\mathfrak{w}$ for mimicking distribution formulae $\psi_{\pi_s},\psi_{\pi_t}$ is indeed a matching in $\mathfrak{W}(\pi_s,\pi_t)$. In fact the functions $\mathfrak{w}\times\mathcal{L}^{\mathrm{s}}_{\mathcal{S}}\times\mathcal{L}^{\mathrm{s}}_{\mathcal{S}}\to[0,1]$ and $\mathfrak{w}'\times\mathcal{S}\times\mathcal{S}\to[0,1]$ with $\mathfrak{w}(\varphi_{s'},\varphi_{t'})=\mathfrak{w}'(s',t')$ for all $s',t'\in\mathcal{S}$, are such that $\mathfrak{w}\in\mathfrak{W}(\psi_{\pi_s},\psi_{\pi_t})$ if and only if $\mathfrak{w}'\in\mathfrak{W}(\pi_s,\pi_t)$, which is equivalent to have both

$$\sum_{\varphi\in\mathcal{L}^{\mathrm{s}}_{\mathcal{S}}} \mathfrak{w}(\varphi_{s'},\varphi) = \pi_s(s') \text{ if and only if } \sum_{t'\in\mathcal{S}} \mathfrak{w}'(s',t') = \pi_s(s')$$
$$\sum_{\varphi\in\mathcal{L}^{\mathrm{s}}_{\mathcal{S}}} \mathfrak{w}(\varphi,\varphi_{t'}) = \pi_t(t') \text{ if and only if } \sum_{s'\in\mathcal{S}} \mathfrak{w}'(s',t') = \pi_t(t')$$

which follow immediately from $\mathfrak{w}(\varphi_{s'}, \varphi_{t'}) = \mathfrak{w}'(s', t')$ for all $s', t' \in \mathcal{S}$.

■

From the characterization result for the up-to-k-bisimilarity metrics we derive that for the bisimilarity metric.

Theorem 5.24. *For all processes $s, t \in \mathcal{S}$ we have*

$$\ell_\lambda(s,t) = \mathbf{d}_\lambda(s,t).$$

Proof. By definition $\ell_\lambda(s,t) = \mathfrak{d}_{\lambda_{\mathcal{M}}}(X_s, X_t) = \mathfrak{d}_{\lambda_{\mathcal{M}}}(\varphi_s, \varphi_t) = \lim_{k\to\infty} \mathfrak{d}_{\lambda_{\mathcal{M}}}^{k}(\varphi_s, \varphi_t)$ and $\mathbf{d}_\lambda(s,t) = \lim_{k\to\infty} \mathbf{d}_\lambda^k(s,t)$. Moreover, by Theorem 5.23 for each $k \in \mathbb{N}$ it holds that $\mathfrak{d}_{\lambda_{\mathcal{M}}}^{k}(\varphi_s, \varphi_t) = \mathbf{d}_\lambda^k(s,t)$. Then the thesis follows by the uniqueness of the limit. ■

We infer that two processes are bisimilar if and only if they are at logical bisimulation distance 0.

Corollary 5.25. *For all processes $s, t \in \mathcal{S}$ we have*

$$s \sim t \text{ if and only if } \ell_\lambda(s,t) = 0.$$

Proof.

$$\begin{aligned} s \sim t \text{ iff } \mathbf{d}_\lambda(s,t) &= 0 && \text{(by Proposition 2.7)} \\ \text{iff } \ell_\lambda(s,t) &= 0 && \text{(by Theorem 5.24).} \end{aligned}$$

■

Moreover, the mimicking formulae of two processes are $\mathcal{L}_{\mathcal{S}}$-equivalent under $\mathcal{M}$ if and only if those processes are at logical bisimulation distance 0.

Corollary 5.26. *For all processes $s, t \in \mathcal{S}$, we have*

$$X_s \equiv_{\mathcal{M}} X_t \text{ if and only if } \ell_\lambda(s,t) = 0.$$

Proof.

$$\begin{aligned} X_s \equiv_{\mathcal{M}} X_t &\text{ iff } s \sim t && \text{(by Theorem 5.13)} \\ &\text{iff } \ell(s,t) = 0 && \text{(by Corollary 5.25).} \end{aligned}$$

■

5.6 $\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATIONS OF BRANCHING HEMIMETRICS

We have defined the metrics $\mathfrak{d}_{\lambda_{\mathcal{E}}}$ and $\mathfrak{D}_{\lambda_{\mathcal{E}}}$ as the exact transposition of the Hausdorff and Kantorovich lifting functionals over the elements of $\mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$ and $\mathcal{L}^{\mathrm{d}}_{\mathfrak{J}}$, respectively. This is one of the key features that allowed us to obtain the characterization of bisimulation metric. However, we need to answer to the natural question that may arise: what happens if we change one or both lifting functionals in the definition of bisimilarity metric (Definition 2.18)? In this case, the metric of Definition 5.13 would not be useful for the characterization result. However the ideas exceeding the technical definition would be still valid, thus confirming the robustness of our approach: to obtain the logical characterization, we have to define the logical distance between processes as a suitable distance between the mimicking formulae. Hence, if in the definition of bisimulation metric the Kantorovich lifting functional $\mathbf{K}$ is substituted by another lifting functional P, then we should modify the distance between distribution formulae as $\mathfrak{D}_{\lambda}{}^{k}_{\mathcal{E}}(\psi_1,\psi_2) = P(\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{E}})(\psi_1,\psi_2)$. If conversely the Hausdorff lifting functional $\mathbf{H}$ is changed with another lifting functional, then we would have to modify the definition of $\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{E}}$ on the boolean operator $\bigwedge$ accordingly.

Notice that no change would need to be done on the class of formulae characterizing the kernel of the metric. This means that the logical characterization obtained for the kernel of the metric would be still valid. In this case of real-valued logics, a variation in the definition of the metric would imply syntactic and semantic modifications to the class of formulae and thus the characterization of the kernel would have to be proven again.

As an example, we consider the two branching distances for probabilistic ready similarity and probabilistic similarity introduced in Chapter 4, and we characterize them using the same approach of previous Section 5.5.

$\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATION OF READY SIMILARITY METRIC

We exploit the logical characterization of ready similarity obtained in Section 5.4 and we slightly modify the distance on formulae defined in Section 5.5 to obtain a logical distance on processes characterizing the ready similarity metric. More precisely, given a generic logic $\mathcal{L}_{\mathfrak{I}}$ and an endodeclaration $\mathcal{E}$ on $\mathcal{L}_{\mathfrak{I}}$, for each $k \in \mathbb{N}$, the *up-to-k ready simulation distance under $\mathcal{E}$* is the mapping $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}} \colon \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{I}} \to [0,1]$ obtained from $\mathfrak{d}_{\lambda}{}^{k}_{\mathcal{E}}$ by modifying the distance on conjunction as follows

$$\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}\Big(\bigwedge_{j\in\mathcal{J}} \varphi_j, \bigwedge_{i\in\mathcal{I}} \varphi_i\Big) = \sup_{j\in\mathcal{J}} \inf_{i\in\mathcal{I}} \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}(\varphi_j,\varphi_i).$$

The *up-to-k ready simulation distance under $\mathcal{E}$* over $\mathcal{L}^{\mathrm{d}}_{\mathfrak{I}}$ is then the mapping $\mathfrak{D}^{\mathrm{r},k}_{\lambda,\mathcal{E}} \colon \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}} \times \mathcal{L}^{\mathrm{d}}_{\mathfrak{I}} \to [0,1]$ defined by

$$\mathfrak{D}^{\mathrm{r},k}_{\lambda,\mathcal{E}}(\psi_1,\psi_2) = \mathbf{K}(\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}})(\psi_1,\psi_2).$$

Proposition 5.27. *All mappings $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}$ and $\mathfrak{D}^{\mathrm{r},k}_{\lambda,\mathcal{E}}$ with $k \in \mathbb{N}$, are 1-bounded hemimetrics.*

Proof. The proof follows by applying the same arguments used in the proof of Proposition 5.18 and noticing that since $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}$ is asymmetric on conjunction, then neither $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}$ nor $\mathfrak{D}^{\mathrm{r},k}_{\lambda,\mathcal{E}}$ can be symmetric. ■

We define the *ready simulation distance under $\mathcal{E}$* over $\mathcal{L}_{\mathfrak{I}}$, notation $\mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{E}}$, as the limit of the up-to-k ready simulation distances under $\mathcal{E}$. As in previous Section 5.5, the existence of such a limit is guaranteed by the monotonicity and 1-boundedness of the distances.

Proposition 5.28. *For $k \in \mathbb{N}$ and $\varphi, \varphi' \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$, we have $\mathfrak{d}^{\mathrm{r},k+1}_{\lambda,\mathcal{E}}(\varphi,\varphi') \geq \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}(\varphi,\varphi')$.*

Proof. The proof follows by applying the same arguments used in the proof of Proposition 5.19. ■

Proposition 5.29. *For all $\varphi, \varphi' \in \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$, $\lim_{k\to\infty} \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}(\varphi,\varphi') \in [0,1]$.*

Proof. The proof follows by applying the same arguments used in the proof of Proposition 5.20. ■

Hence we can define the ready simulation distance under $\mathcal{E}$ between formulae as the limit of their up-to-k counterparts.

Definition 5.16 (Ready simulation distance under $\mathcal{E}$)**.** Given an image finite and guarded endodeclaration $\mathcal{E}$ on $\mathcal{L}_{\mathfrak{J}}$, the *ready simulation distance under* $\mathcal{E}$ is defined over state formulae as the mapping $\mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{E}} \colon \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}} \times \mathcal{L}^{\mathrm{s}}_{\mathfrak{J}} \to [0,1]$ such that

$$\mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{E}}(\varphi,\varphi') = \lim_{k\to\infty} \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{E}}(\varphi,\varphi').$$

By Proposition 5.29 we are guaranteed that $\mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{E}}$ is well defined. Now we show it is a 1-bounded hemimetric on $\mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$.

Proposition 5.30. *The mapping $\mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{E}}$ is a 1-bounded hemimetric on $\mathcal{L}^{\mathrm{s}}_{\mathfrak{J}}$.*

Proof. The proof follows by applying the same arguments used in the proof of Proposition 5.18. ■

We can now lift the ready simulation distance on formulae to a *logical ready simulation distance* over processes.

Definition 5.17 (Logical ready simulation distance)**.** Let $\lambda \in (0,1]$. For any $k \in \mathbb{N}$, the *up-to-k logical ready simulation distance* over processes $\ell^k_{\mathrm{r},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined for all $s,t \in \mathcal{S}$ by

$$\ell^k_{\mathrm{r},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{M}}(X_s, X_t).$$

Then, the *logical ready simulation distance* over processes $\ell_{\mathrm{r},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined for all $s,t \in \mathcal{S}$ as

$$\ell_{\mathrm{r},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{M}}(X_s, X_t).$$

Notice that both $\ell^k_{\mathrm{r},\lambda}$ and $\ell_{\mathrm{r},\lambda}$ are well defined since $\mathcal{M}$ is image finite and guarded. Notice also that $\ell_{\mathrm{r},\lambda}(s,t) = \lim_{k\to\infty} \ell^k_{\mathrm{r},\lambda}(s,t)$.

Proposition 5.31.

1. *For any $k \in \mathbb{N}$ the mapping $\ell^k_{\mathrm{r},\lambda}$ is a 1-bounded hemimetric.*
2. *The mapping $\ell_{\mathrm{r},\lambda}$ is a 1-bounded hemimetric.*

Proof.

1. Directly by Proposition 5.27.
2. Directly by Proposition 5.30.

■

We can now formalize the characterization result for the up-to-k ready simulation metric.

Theorem 5.32. *For all $k \in \mathbb{N}$ and processes $s, t \in \mathcal{S}$ we have $\ell^k_{\mathrm{r},\lambda}(s,t) = \mathbf{d}^k_{\mathrm{r},\lambda}(s,t)$.*

Proof. The proof follows by applying the same arguments used in the proof of Theorem 5.23. ■

From the characterization result for the up-to-k ready similarity metrics we derive that for ready similarity metric.

Theorem 5.33. *For all processes $s, t \in \mathcal{S}$ we have*

$$\ell_{\mathrm{r},\lambda}(s,t) = \mathbf{d}_{\mathrm{r},\lambda}(s,t).$$

Proof. By definition $\ell_{\mathrm{r},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{M}}(X_s, X_t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{M}}(\varphi_s, \varphi_t) = \lim_{k\to\infty} \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{M}}(\varphi_s, \varphi_t)$ and $\mathbf{d}_{\mathrm{r},\lambda}(s,t) = \lim_{k\to\infty} \mathbf{d}^k_{\mathrm{r},\lambda}(s,t)$ (by Proposition 4.1). Moreover, by Theorem 5.32 for each $k \in \mathbb{N}$ it holds that $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{M}}(\varphi_s, \varphi_t) = \mathbf{d}^k_{\lambda}(s,t)$. Then the thesis follows by the uniqueness of the limit. ■

As an immediate consequence of Theorem 5.33 we obtain that ready similarity is the kernel of the logical ready similarity distance.

Corollary 5.34. *For all $s, t \in \mathcal{S}$ we have that $s \sqsubseteq_{\mathrm{r}} t$ if and only if $\ell_{\mathrm{r},\lambda}(s,t) = 0$.*

Proof.

$$\begin{aligned} s \sqsubseteq_{\mathrm{r}} t \text{ iff } \mathbf{d}_{\mathrm{r},\lambda}(s,t) = 0 & \qquad\qquad \text{(by Theorem 4.3)} \\ \text{iff } \ell_{\mathrm{r},\lambda}(s,t) = 0 & \qquad\qquad \text{(by Theorem 5.33).} \end{aligned}$$

■

$\mathcal{L}_{\mathcal{S}}$-CHARACTERIZATION OF SIMILARITY METRIC

We exploit the logical characterization of similarity obtained in Section 5.4 and the ready simulation distance on formulae to obtain a logical distance on processes characterizing the similarity metric. More precisely, as distance on formulae we consider exactly the ready simulation distance under an endodeclaration. Then the *logical simulation distance* is obtained by choosing the proper endodeclaration: the simulation endodeclaration $\mathcal{C}$.

Definition 5.18 (Logical simulation distance)**.** Let $\lambda \in (0,1]$. For any $k \in \mathbb{N}$, the *up-to-k logical simulation distance* over processes $\ell^k_{\mathrm{s},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined for all $s, t \in \mathcal{S}$ by

$$\ell^k_{\mathrm{s},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{C}}(X_s, X_t).$$

Then, the *logical simulation distance* over processes $\ell_{\mathrm{s},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined for all $s, t \in \mathcal{S}$ as

$$\ell_{\mathrm{s},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{C}}(X_s, X_t).$$

Notice that $\ell_{\mathrm{s},\lambda}(s,t) = \lim_{k\to\infty} \ell^k_{\mathrm{s},\lambda}(s,t)$.

Proposition 5.35. 1. *For any $k \in \mathbb{N}$ the mapping $\ell^k_{\mathrm{s},\lambda}$ is a 1-bounded hemimetric.*

2. *The mapping $\ell_{\mathrm{s},\lambda}$ is a 1-bounded hemimetric.*

Proof.

1. Directly by Proposition 5.27.

2. Directly by Proposition 5.30.

■

We can now formalize the characterization result for the similarity metric. We start from the characterization of the up-to-k similarity metrics.

Theorem 5.36. *For all $k \in \mathbb{N}$ and processes $s,t \in \mathcal{S}$ we have $\ell^k_{\mathrm{s},\lambda}(s,t) = \mathbf{d}^k_{\mathrm{s},\lambda}(s,t)$.*

Proof. The proof follows by applying the same arguments used in the proof of Theorem 5.23. ■

We are now ready to derive the characterization result for the similarity metric.

Theorem 5.37. *For all processes $s,t \in \mathcal{S}$ we have*

$$\ell_{\mathrm{s},\lambda}(s,t) = \mathbf{d}_{\mathrm{s},\lambda}(s,t).$$

Proof. By definition $\ell_{\mathrm{s},\lambda}(s,t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{C}}(X_s, X_t) = \mathfrak{d}^{\mathrm{r}}_{\lambda,\mathcal{C}}(\vartheta_s, \vartheta_t) = \lim_{k\to\infty} \mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{C}}(\vartheta_s, \vartheta_t)$ and $\mathbf{d}_{\mathrm{s},\lambda}(s,t) = \lim_{k\to\infty} \mathbf{d}^k_{\mathrm{s},\lambda}(s,t)$ (by Proposition 4.1). Moreover, by Theorem 5.36 for each $k \in \mathbb{N}$ it holds that $\mathfrak{d}^{\mathrm{r},k}_{\lambda,\mathcal{C}}(\varphi_s, \varphi_t) = \mathbf{d}^k_{\mathrm{s},\lambda}(s,t)$. Then the thesis follows by the uniqueness of the limit. ■

As an immediate consequence of Theorem 5.37 we obtain that similarity is the kernel of the logical similarity distance.

Corollary 5.38. *For all processes $s,t \in \mathcal{S}$ we have $s \sqsubseteq t$ if and only if $\ell_{\mathrm{s},\lambda}(s,t) = 0$.*

Proof.

$$\begin{aligned} s \sqsubseteq t \text{ iff } \mathbf{d}_{\mathrm{s},\lambda}(s,t) = 0 & \qquad\qquad \text{(by Theorem 4.5)} \\ \text{iff } \ell_{\mathrm{s},\lambda}(s,t) = 0 & \qquad\qquad \text{(by Theorem 5.37).} \end{aligned}$$

■

5.7 Concluding remarks

In this Chapter we have proposed modal characterizations of branching (hemi)metrics, as bisimilarity and (ready) similarity metric, on image finite nondeterministic probabilistic processes. To obtain them, we have introduced the novel notions of *mimicking formula* of a process and *distance on a class of formulae*. More precisely, we have proved that the bisimilarity (resp. (ready) similarity) metric coincides with the *logical bisimulation* (resp. (*ready*) *simulation*) *distance* on processes, that is the bisimilarity (resp. (ready) similarity) distance between two processes is equal to the bisimulation (resp. (ready) simulation) distance between their mimicking formulae defined on the $\mathcal{S}$-indexed logic $\mathcal{L}_{\mathcal{S}}$. This modal logic has been obtained by extending the probabilistic version of HML from [66] with a family of variables, one for each process in $\mathcal{S}$, allowing for the recursive specification of formulae. Following the equational μ-calculus approach of [4, 120, 143] we have provided an appropriate interpretation to each variable as the solution of a system of equations defined using endodeclarations, namely functions $\mathcal{E}$ mapping each variable into an arbitrary formula of the logic. As solution we have considered the variable interpretation corresponding to the greatest fixed point of the system. The mimicking formula of a process s is defined as the image of the variable corresponding to s through a particular endodeclaration $\mathcal{M}$ on $\mathcal{L}_{\mathcal{S}}$, called *mimicking endodeclaration*, and it captures all possible resolutions of nondeterminism for the process by also exactly specifying the reached probability distributions. These properties allowed also for a *weak expressive* characterization of probabilistic bisimilarity: two processes are bisimilar if and only if their mimicking formulae are $\mathcal{L}_{\mathcal{S}}$-equivalent under $\mathcal{M}$. Moreover, we have proved that the mimicking formula of a process s coincides with the characteristic formula of s with respect to probabilistic ready similarity, thus obtaining an expressive characterization of this preorder. Finally, we have showed how to derive the characteristic formula of a process with respect to probabilistic similarity by means of an endodeclaration $\mathcal{C}$ mapping each variable into the negation free subformula of the mimicking formula of the process.

In [4, 143] the equational μ-calculus has been employed to provide a general framework for the definition of characteristic formulae for a wide class of behavioral equivalences and preorders. However the $\mathcal{S}$-indexed logic $\mathcal{L}_{\mathcal{S}}$ is not powerful enough to allow for the construction of characteristic formulae for probabilistic bisimilarity. For instance, in [68] expressive characterizations of strong and weak probabilistic (bi)simulations for image-finite processes are provided by constructing the characteristic formulae of processes by means of the *probabilistic μ-calculus*, which is reacher than $\mathcal{L}_{\mathcal{S}}$ since it allows arbitrary formulae to occur in the scope of the diamond modality. The extension of $\mathcal{L}_{\mathcal{S}}$ with that feature would have certainly allowed for an expressive characterization of probabilistic bisimilarity. Nevertheless, this would have implied a much more technical definition of our distance between formulae, which we recall is defined on the structure of formulae, thus making the characterization of the bisimilarity metric more complex.

Another difference with respect to [68] relies on negation. In [68], to guarantee the monotonicity of the function $[\![\,]\!]$, variables are allowed to occur only in the scope of an even number of negations. We could have applied the same idea to $\mathcal{L}_{\mathcal{S}}$. However, as we have shown, to characterize the chosen probabilistic relations and the bisimilarity metric the negation expressed as formulae $\bar{a}$ is sufficient. Hence, we decided to consider only this form of negation, thus also simplifying to some extent the presentation of some technical results.

The one proposed in this Chapter is not the first logical characterization of a behavioral metric over processes, but it is indeed the first one based on a boolean-valued logic and a proper distance on formulae. Characterizations of bisimilarity metric based on real-valued logics are given in [14, 58, 61, 72, 75, 155, 157]. In detail, in [72] the authors define a metric between labeled Markov processes (LMP) by giving a real-valued semantics to a probabilistic modal logic. Roughly speaking, boolean and modal operators are translated into functional expressions and the satisfaction relation is interpreted as integration. Then, the distance between two processes is defined as the maximal disparity between functionals distinguishing them, obtaining that probabilistic bisimilar processes are the ones at distance 0. Later, in [156, 157] the authors proved that the logic in [72] coincides with the bisimilarity metric based on the Kantorovich lifting and defined co-inductively.

With a similar approach, in [58] the characterization of two classes of behavioral metrics is proposed. They consider Metric Transition Systems (MTS), namely transition systems in which the atomic propositions, at each state, take values in a bounded metric space. They define four metrics characterizing as much system relations: an asymmetric *linear distance* generalizing trace inclusion and its symmetric version for trace equivalence; simulation is characterized by asymmetric *branching distance* whereas bisimulation is the kernel of the symmetric branching distance. Then, they exploit the *Quantitative μ-calculus* of [60] to characterize the branching distances. The same logic is used in [61], for stochastic game structures, to characterize their *a priori* metric, defined as the distance between the expected payoffs of the players.

Finally, in [75] a real-valued logic is proposed for the characterization of a *state-based bisimulation metric* which coincides with the one of [64] and of a *distribution-based bisimulation metric* which is directly defined over distributions without using any lifting functional [67, 79, 103]. Many metrics for distribution-based bisimulations have been recently proposed along with some logical characterizations for them (see for instance [78, 165] and the references therein). However, they all follow the (standard) approach of [75]: the considered logic is real-valued and the metric is characterized as the total-variation distance on the values of formulae. Notice that our approach can be easily modified to capture also the case of distribution-based bisimulations. It would suffices to lift the transition relation and the semantics of formulae on distributions, as in [69, 128, 137]. By applying our characterization method to this Kleisli-like construction we obtain the characterization of the metrics for distribution-based bisimulations.

The originality of our notion of distance on $\mathcal{L}_S$ relies on the fact that it is not defined in terms of any ground distance between processes. As a matter of fact, our distance on formulae is independent from the metric properties of the process space. A first proposal of a distance between formulae can be found in [122], related to the study of approximate reasoning principles for both discrete-time (DMPs) and continuous-time Markov processes (CMPs) with continuous state space. The authors provide their solution to the problem of relating the behavior of approximations to the limit behavior of the system itself. Roughly speaking, given a sequence of processes $\{s_k\}_{k\in\mathbb{N}}$ approximating a given system s, one wishes to know whether it is possible to infer that the limit of such a sequence meets the specification of s and, viceversa, whether one can infer that the specification of s agrees with the limit specification of the approximants. To this aim they introduce the property of *dynamical* *continuity* for a pseudometric: a metric is dynamically continuous if it allows to identify convergent sequences of processes or formulae. Then, they define a metric space for the *Discrete Markovian Logic* (DML) and the *Continuous Markovian Logic* (CML) [129] by considering as distance between formulae the Hausdorff distance on the sets of processes satisfying them. In this way, they are able to topologically characterize the logical properties induced by a dynamically continuous metric for both DMPs and CMPs. Moreover, the identify the requirements necessary to guarantee that parallel sequences of formulae and processes converge to give satisfaction in the limit.

We have already argued that we defined the distance between formulae with the exact purpose of simulating the Hausdorff and Kantorovich lifting functionals on which the bisimilarity metric is defined. Despite this kind of reasoning may seem too restrictive at first glance, we believe that having a distance between formulae instead of a real-valued semantics for the logic turns out to be an advantage in case one wishes to modify the lifting functionals in the definition of (bi)similarity metric (cf. first part of Section 5.6).

Finally, to prove the robustness of our approach, in the next Chapter we apply it to obtain logical characterization of the (decorated) trace metrics, testing metric and their kernels introduced in Chapter 4.

Let's think the unthinkable,
let's do the undoable.
Let us prepare to grapple with the ineffable
itself, and see if we may not eff it after all.

Douglas Adams,
Dirk Gently's Holistic Detective Agency

Chapter 6

Logical Characterization of Linear Metrics

This Chapter generalizes the characterization method proposed in previous Chapter 5 to the linear semantics. In detail, we provide a logical characterization of (decorated) trace and testing metrics as well as of their kernels. These characterizations are defined on modal logics consisting in two classes of formulae, the *linear formulae* expressing traces and the decorations on them, and the *probabilistic formulae* associating a probability to each linear formula. The main difference with respect to Chapter 5 is in the expressive power of mimicking formulae of processes. Probabilistic linear semantics are based on the comparison of the probabilities of semantic-specific events to occur during process execution. Thus, mimicking formulae of processes for these semantics will capture such probabilities rather than the ability or impossibility of a process to perform a particular computation step. By means of these mimicking formulae we obtain weak expressive characterizations of the kernels of the considered linear metrics, in the sense of Chapter 5: to establish whether two processes are related we simply need to compare their mimicking formulae. Due to the simpler structure of the modal operators in the considered classes, in place of the structural equivalence of formulae, we introduce an ordering over probabilistic formulae obtained by comparing the probabilities assigned to the same linear formulae. This ordering will allow us to compare the mimicking formulae of processes.

Our characterization method is mainly based on two ingredients, the mimicking formulae and the metric over a proper class of formulae, by means of which we can define a *logical distance* on processes. The characterizations of behavioral metrics obtained with this approach state that our logical distances are as expressive as the corresponding behavioral distance. Therefore, the logical distances measure the disparities in the behavior of processes with respect to the chosen probabilistic semantics and moreover they are defined solely in terms of a modal logic. Consequently, these logical distances could be used to capture the desired quantitative semantics of processes in place of behavioral metrics. Thus, we order them in a spectrum of logical distances on processes with the purpose of emphasizing the differences in their distinguishing power, so that one can choose the most suitable logical distance with respect to the intended application context. As ordering relation we consider the same used in Chapter 4 to define the spectrum of behavioral metrics, namely the relation '*makes processes further than*'. Interestingly, this spectrum is obtained directly by combining the logical characterizations of branching metrics, investigated in Chapter 5, and the ones on linear metrics, that we will present in this Chapter, with the Theorems and technical results in Chapter 4 that allowed for the construction of the spectrum on behavioral metrics. We remark that ours is the first example of a spectrum of metrics on processes obtained solely from modal logics.

As briefly outlined in Chapter 5, our characterization approach differs from the ones proposed in the literature in that, in general, logics equipped with a real-valued semantics are used for the characterization, which is then expressed as

$$d(s,t) = \sup_{\varphi \in L} |[\varphi](s) - [\varphi](t)| \tag{6.1}$$

where d is the behavioral metric of interest, L is the considered logic and $[\varphi](s)$ denotes the value of the formula φ at process s accordingly to the real-valued semantics [59,61,72,73,75]. However, we notice that with this approach to obtain a spectrum of logical distances similar to ours, one would be forced to prove the semantic inclusion of the considered classes of modal formulae, to guarantee the suprema of the distances on these classes to be ordered. With our approach these inclusions, which furthermore are not true in general, are not needed: each class of formulae expresses the proper semantic properties of processes and the logical distances, capturing the disparities in the satisfaction of such properties, are then ordered in terms of their distinguishing power.

The contributions of this Chapter can be summarized as follows:

1. We provide a logical characterization of linear metrics such as a. trace (hemi)metric; b. completed trace (hemi)metric; c. failure (hemi)metric; d. failure trace (hemi)metric; e. readiness (hemi)metric; f. ready trace (hemi)metric; g. testing (pre)metric.

2. We provide a weak expressive characterization of the kernels of linear metrics such as a. trace preorder and equivalence; b. completed trace preorder and equivalence; c. failure preorder and equivalence; d. failure trace preorder and equivalence; e. readiness preorder and equivalence; f. ready trace preorder ad equivalence; g. testing preorder and equivalence.

3. By combining the characterization results from Chapter 5 with those in this Chapter, we obtain the first example of a spectrum of metrics on processes obtained solely from modal logics that comprehends:

   a. *logical bisimulation distance* (ℓ_λ);

   b. *logical ready simulation distance* ($\ell_{\mathrm{r},\lambda}$);

   c. *logical simulation distance* ($\ell_{\mathrm{s},\lambda}$);

   d. *logical ready trace pre-distance* ($\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}$);

e. *logical ready trace distance* ($\ell_{\mathrm{TrR},\lambda}$);

f. *logical readiness pre-distance* ($\ell_{\sqsubseteq_{\mathrm{R}},\lambda}$);

g. *logical readiness distance* ($\ell_{\mathrm{R},\lambda}$);

h. *logical failure trace pre-distance* ($\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}$);

i. *logical failure trace distance* ($\ell_{\mathrm{TrF},\lambda}$);

j. *logical failure pre-distance* ($\ell_{\sqsubseteq_{\mathrm{F}},\lambda}$);

k. *logical failure distance* ($\ell_{\mathrm{F},\lambda}$);

l. *logical completed trace pre-distance* ($\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}$);

m. *logical completed trace distance* ($\ell_{\mathrm{TrC},\lambda}$);

n. *logical trace pre-distance* ($\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}$);

o. *logical trace distance* ($\ell_{\mathrm{Tr},\lambda}$);

p. *logical testing pre-distance* ($\ell_{\sqsubseteq_{\mathrm{test}}.\lambda}$);

q. *logical testing distance* ($\ell_{\mathrm{test},\lambda}$).

Organization of contents

In Section 6.1 we introduce the modal logic expressing the (decorated) trace semantics whose characterization is then proposed in Section 6.2. In Section 6.4 we present the analogous results for testing semantics based on the modal logic defined in Section 6.3. Section 6.5 contains the spectrum of logical distance obtained by combining the characterizations proposed in this Chapter and those in Chapter 5. We conclude with the discussion of related work in Section 6.6.

6.1 A modal logic for decorated traces

In this Section we present the modal logic $\mathfrak{L}$ that will allow us to characterize the trace metric as well as its decorated versions and their kernels. Interestingly, we can use a single logic to characterize all these semantics because of the expressing power of mimicking formulae. They identify, and isolate, the properties characterizing the particular semantics to which they are related and thus we do not need to distinguish different classes of formulae for different semantics.

The logic $\mathfrak{L}$ can be seen either as a simplified version of the modal logic $\mathcal{L}$ presented in Chapter 2.4, or more naturally as a probabilistic version of the general class of formulae of which the logics characterizing (decorated) trace semantics in the fully nondeterministic case are subclasses (cf. [33]). More precisely, $\mathfrak{L}$ consists of two classes of formulae: the class $\mathfrak{L}^{\mathrm{l}}$ of *linear formulae*, which are constituted by (finite) sequences of diamond operators and that will be used to represent traces and the decorations on them, and the class $\mathfrak{L}^{\mathrm{p}}$ of *probabilistic formulae*, which will be used to capture the quantitative properties of processes.

Definition 6.1 (Modal logic $\mathfrak{L}$)**.** The classes of *linear formulae* $\mathfrak{L}^{\mathrm{l}}$ and of *probabilistic formulae* $\mathfrak{L}^{\mathrm{p}}$ over $\mathcal{A}$ are defined by the following BNF-like grammar:

$$\begin{aligned}\mathfrak{L}^{\mathrm{l}}: \quad & \Phi ::= \quad \top \mid \bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi \mid \bigwedge_{j\in\mathcal{J}} \langle a_j\rangle\top \wedge \Phi \mid \langle a\rangle\Phi \\ \mathfrak{L}^{\mathrm{p}}: \quad & \Psi ::= \quad r\Phi \mid \bigwedge_{i\in\mathcal{I}} \Psi_i\end{aligned}$$

where: (i) Φ, Φ_j range over $\mathfrak{L}^{\mathrm{l}}$, (ii) Ψ, Ψ_i range over $\mathfrak{L}^{\mathrm{p}}$, (iii) $a, a_j \in \mathcal{A}$; (iv) $\mathcal{I}, \mathcal{J}$ are at most countable sets of indexes, (v) in the formula $\bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi$, the a_j are pairwise distinct for all $j \in \mathcal{J}$ and $\Phi \neq \bigwedge_{i\in\mathcal{I}} \bar{a}_i \wedge \Phi'$ for any set of indexes $\mathcal{I}$ with $|\mathcal{I}| \geq 1$ and for all $\Phi' \in \mathfrak{L}^{\mathrm{l}}$, (vi) in the formula $\bigwedge_{j\in\mathcal{J}}\langle a_j\rangle\top \wedge \Phi$, the a_j are pairwise distinct for all $j \in \mathcal{J}$ and $\Phi \neq \bigwedge_{i\in\mathcal{I}}\langle a_i\rangle\top \wedge \Phi'$ for any set of indexes $\mathcal{I}$ with $|\mathcal{I}| \geq 1$ and for all $\Phi' \in \mathfrak{L}^{\mathrm{l}}$, (vii) $r \in [0,1]$, (viii) for each $i \in \mathcal{I}$ it holds $\Psi_i \neq \bigwedge_{j\in\mathcal{J}} \Psi_j$ for any set of indexes $\mathcal{J}$ with $|\mathcal{J}| > 1$.

We shall write $\bigwedge_{j\in J} \bar{a}_j$ and $\bigwedge_{j\in\mathcal{J}} \langle a_j\rangle\top$ as a shorthand for resp. $\bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \top$ and $\bigwedge_{j\in\mathcal{J}} \langle a_j\rangle\top\wedge\top$. Moreover, we recall that $\top$ stands for $\bigwedge_{\emptyset}$.

The notion of *depth* of formulae is standard and expresses the length of the longest sequence of diamond operators in the formulae.

Definition 6.2 (Depth)**.** The *depth of probabilistic formulae* in $\mathfrak{L}^{\mathrm{p}}$ is defined as

- ⋆ $\mathrm{dpt}(r\Phi) = \mathrm{dpt}(\Phi)$ and
- ⋆ $\mathrm{dpt}(\bigwedge_{i\in\mathcal{I}} \Psi_i) = \sup_{i\in\mathcal{I}} \mathrm{dpt}(\Psi_i)$

where the *depth of linear formulae* in $\mathfrak{L}^{\mathrm{l}}$ is defined by induction on their structure as

- ⋆ $\mathrm{dpt}(\top) = 0$;
- ⋆ $\mathrm{dpt}(\bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi) = \max\{1, \mathrm{dpt}(\Phi)\}$;
- ⋆ $\mathrm{dpt}(\bigwedge_{j\in\mathcal{J}} \langle a_j\rangle\top \wedge \Phi) = \max\{1, \mathrm{dpt}(\Phi)\}$;
- ⋆ $\mathrm{dpt}(\langle a\rangle\Phi) = 1 + \mathrm{dpt}(\Phi)$.

Formulae are interpreted over PTSs.

Definition 6.3 (Semantics of $\mathfrak{L}^{\mathrm{l}}$)**.** Given any process $s \in \mathcal{S}$, the *satisfaction relation* $\models \subseteq \mathcal{S} \times \mathfrak{L}^{\mathrm{l}}$ is defined by structural induction over linear formulae in $\mathfrak{L}^{\mathrm{l}}$ by

- ⋆ $s \models \top$ always;
- ⋆ $s \models \bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi$ iff $s \xrightarrow{a_j}\!\!\!\!\!\!/\;\;$ for all $j \in \mathcal{J}$ and $s \models \Phi$;
- ⋆ $s \models \bigwedge_{j\in\mathcal{J}} \langle a_j\rangle\top \wedge \Phi$ iff $s \xrightarrow{a_j}$ for all $j \in \mathcal{J}$ and $s \models \Phi$;

- $s \models \langle a\rangle\Phi$ iff $s \xrightarrow{a} \pi$ for some π such that $s' \models \Phi$ for some $s' \in \mathrm{supp}(\pi)$.

We say that a computation c from process s is compatible with the linear formula Φ, notation $c \in \mathcal{C}^{\mathrm{t}}(s,\Phi)$, if $|c| = \mathrm{dpt}(\Phi)$ and $s \models \Phi$ is obtained by verifying $\models$ on the processes reached by s through c. Moreover, given any resolution $\mathcal{Z}_s$ for process s we say that $z \models \Phi$ if and only if $\mathrm{corr}_{\mathcal{Z}_s}(z) \models \Phi$.

Definition 6.4 (Semantics of $\mathfrak{L}^{\mathrm{p}}$)**.** The *satisfaction relation* $\models \subseteq \mathcal{S} \times \mathfrak{L}^{\mathrm{p}}$ is defined by

- $s \models r\Phi$ if and only if there is a resolution $\mathcal{Z}_s \in \mathrm{Res}(s)$ such that $\Pr(\mathcal{C}^{\mathrm{t}}(z_s,\Phi)) = r$.
- $s \models \bigwedge_{i\in\mathcal{I}} \Psi_i$ if and only if $s \models \Psi_i$ for all $i \in \mathcal{I}$.

A distance on $\mathfrak{L}$

To obtain the logical characterization of metric semantics we need to transform the modal logic into a metric space. For this reason, we propose a syntactical distance on formulae in $\mathfrak{L}$. To obtain it, we introduce an auxiliary distance, denoted by K_δ, that acts like the *Kronecker delta function* on formulae of the form $\bar{a}$ and $\langle a\rangle\top$.

Definition 6.5. Let $\bar{\mathcal{A}}$ denote the set of formulae $\{\bar{a} \mid a \in \mathcal{A}\}$ and $\langle\mathcal{A}\rangle$ denote the set of formulae $\{\langle a\rangle\top \mid a \in \mathcal{A}\}$. The function $K_\delta \colon (\bar{\mathcal{A}} \times \bar{\mathcal{A}}) \cup (\langle\mathcal{A}\rangle \times \langle\mathcal{A}\rangle) \to \{0,1\}$ is defined for all $a,b \in \mathcal{A}$ as follows:

$$K_\delta(\bar{a},\bar{b}) = \begin{cases} 1 & \text{if } a \neq b \\ 0 & \text{otherwise.} \end{cases} \qquad K_\delta(\langle a\rangle\top,\langle b\rangle\top) = \begin{cases} 1 & \text{if } a \neq b \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, to correctly quantify the distances on conjunctions of linear formulae we need to consider them up-to reordering. This means that, in what follows, we subsume that whenever Φ_1 is of the form $\bigwedge_{i\in\mathcal{I}}\langle a_i\rangle\top \wedge \bigwedge_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi$, with Φ not containing any conjunction, then Φ_1 is considered up-to reordering, namely $\Phi_1 = \bigwedge_{j\in\mathcal{J}} \bar{a}_i \wedge \bigwedge_{i\in\mathcal{I}}\langle a_i\rangle\top \wedge \Phi$.

We are now ready to formally introduce a distance on $\mathfrak{L}$.

Definition 6.6 (Distance on $\mathfrak{L}$)**.** Let $\lambda \in (0,1]$. The function $\mathcal{D}^{\mathrm{l}} \colon \mathfrak{L}^{\mathrm{l}} \times \mathfrak{L}^{\mathrm{l}} \to [0,1]$ is defined by structural induction over $\mathfrak{L}^{\mathrm{l}}$ as follows:

$$\mathcal{D}^{\mathrm{l}}(\Phi_1,\Phi_2) = \begin{cases} 0 & \text{if } \Phi_1 = \Phi_2 = \top \\ \max\left\{ \begin{array}{l} \mathbf{H}(K_\delta)(\{\bar{a}_j \mid j \in \mathcal{J}\},\{\bar{a}_i \mid i \in \mathcal{I}\}), \\ \mathcal{D}^{\mathrm{l}}(\Phi_1',\Phi_2') \end{array} \right\} & \begin{array}{l} \text{if } \Phi_1 = \bigwedge\limits_{j\in\mathcal{J}} \bar{a}_j \wedge \Phi_1' \\ \text{and } \Phi_2 = \bigwedge\limits_{i\in\mathcal{I}} \bar{a}_i \wedge \Phi_2' \end{array} \\ \max\left\{ \begin{array}{l} \mathbf{H}(K_\delta)(\{\langle a_j\rangle\top \mid j \in \mathcal{J}\},\{\langle a_i\rangle\top \mid i \in \mathcal{I}\}), \\ \mathcal{D}^{\mathrm{l}}(\Phi_1',\Phi_2') \end{array} \right\} & \begin{array}{l} \text{if } \Phi_1 = \bigwedge\limits_{j\in\mathcal{J}} \langle a_j\rangle\top \wedge \Phi_1' \\ \text{and } \Phi_2 = \bigwedge\limits_{i\in\mathcal{I}} \langle a_i\rangle\top \wedge \Phi_2' \end{array} \\ \mathcal{D}^{\mathrm{l}}(\Phi_1',\Phi_2') & \text{if } \Phi_1 = \langle a\rangle\Phi_1' \text{ and } \Phi_2 = \langle a\rangle\Phi_2' \\ 1 & \text{otherwise.} \end{cases}$$

The function $\mathcal{D}^{\mathrm{p}}_{\lambda} \colon \mathfrak{L}^{\mathrm{p}} \times \mathfrak{L}^{\mathrm{p}} \to [0,1]$ is defined over $\mathfrak{L}^{\mathrm{p}}$ as follows:

$$\mathcal{D}^{\mathrm{p}}_{\lambda}(r_1\Phi_1, r_2\Phi_2) = \begin{cases} \max\{0, \lambda^{\mathrm{dpt}(\Phi_1)-1}(r_1 - r_2)\} & \text{if } \mathcal{D}^{\mathrm{l}}(\Phi_1,\Phi_2) = 0 \\ 1 & \text{otherwise} \end{cases}$$

$$\mathcal{D}^{\mathrm{p}}_{\lambda}(\bigwedge_{i\in\mathcal{I}} \Psi_i, \bigwedge_{j\in\mathcal{J}} \Psi_j) = \sup_{i\in\mathcal{I}} \inf_{j\in\mathcal{J}} \mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi_i,\Psi_j).$$

Notice that since $\mathrm{dpt}(\bigwedge_{j\in\mathcal{J}} \bar{a}_j) = \mathrm{dpt}(\bigwedge_{j\in\mathcal{J}} \langle a_j \rangle \top) = 1$, the choice of the exponent for the discount factor follows from the same arguments used in Chapter 4.3.

Proposition 6.1. *The function $\mathcal{D}^{\mathrm{l}}$ is a 1-bounded metric over $\mathfrak{L}^{\mathrm{l}}$.*

Proof. The proof follows by an easy induction over the structure of linear formulae. ■

Proposition 6.2. *The function $\mathcal{D}^{\mathrm{p}}_{\lambda}$ is a 1-bounded hemimetric over $\mathfrak{L}^{\mathrm{p}}$.*

Proof. The proof follows by an easy induction over the structure of probabilistic formulae. In particular, the base case $\Psi = r\Phi$ follows by applying the same arguments used in the proof of Theorem 4.6. ■

We now present the kernel of $\mathcal{D}^{\mathrm{p}}_{\lambda}$ which can be characterized in terms of an ordering relation $\le$ over formulae in $\mathfrak{L}^{\mathrm{p}}$. For the linear formulae $\Phi_1, \Phi_2 \in \mathfrak{L}^{\mathrm{l}}$ we write $\Phi_1 = \Phi_2$ if they are syntactically indistinguishable, with the equality of conjunctions considered up-to reordering. Then we introduce the relation of *ordering on formulae* $\le$ as follows.

Definition 6.7 (Ordering of formulae in $\mathfrak{L}^{\mathrm{p}}$)**.** The relation $\le \colon \mathfrak{L}^{\mathrm{p}} \times \mathfrak{L}^{\mathrm{p}}$ is defined inductively over the structure of probabilistic formulae as follows

★ $(r_1\Phi_1, r_2\Phi_2) \in \le$ if and only if $\Phi_1 = \Phi_2$ and $r_1 \le r_2$.

★ $(\bigwedge_{i\in\mathcal{I}} \Psi_i, \bigwedge_{j\in\mathcal{J}} \Psi_j) \in \le$ if and only if for each $i \in \mathcal{I}$ there is a $j \in \mathcal{J}$ such that $(\Psi_i, \Psi_j) \in \le$.

To simplify notation, we write $\Psi_1 \le \Psi_2$ in place of $(\Psi_1, \Psi_2) \in \le$. As for linear formulae, the equality of conjunctions is intended up-to reordering. More precisely, we have that $\Psi_1 = \Psi_2$ if and only if $\Psi_1 \le \Psi_2$ and $\Psi_2 \le \Psi_1$. Notice that due to the syntactical simplicity of the class of probabilistic formulae, their equality defined as a symmetric ordering relation coincides with their structural equivalence.

Finally, we prove that the kernel of $\mathcal{D}^{\mathrm{p}}_{\lambda}$ coincides with the relation $\le$ over formulae in $\mathfrak{L}^{\mathrm{p}}$.

Proposition 6.3. *For all $\Psi_1, \Psi_2 \in \mathfrak{L}^{\mathrm{p}}$, we have that $\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi_1, \Psi_2) = 0$ if and only if $\Psi_1 \le \Psi_2$.*

Proof. The proof follows by an easy induction over the structure of $\Psi_1 \in \mathfrak{L}^{\mathrm{p}}$. ■

6.2 Logical characterizations of (decorated) trace semantics

In this section we apply the same technique proposed in previous Chapter 5 for the characterizations of branching relations and metrics to obtain original characterizations of the (decorated) trace relations (Theorem 6.5) and metrics (Theorem 6.8) defined in Chapter 4.

$\mathfrak{L}$-characterizations of (decorated) trace relations

We start by studying the characterizations of (decorated) trace relations. To this aim we need to introduce the mimicking formulae of processes for these semantics. Each mimicking formula has to express the properties of the process that are relevant for the considered semantics, namely the probabilities assigned to traces and their decorated versions. Hence, we need to identify the linear formulae capturing these traces and decorations.

Definition 6.8 (Trace formula)**.** Given any trace $\alpha \in \mathcal{A}^\star$ we define the *trace formula* of α, notation $\Phi_\alpha \in \mathfrak{L}^{\mathrm{l}}$, inductively on the structure of α as follows:

$$\Phi_\alpha = \begin{cases} \top & \text{if } \alpha = \mathfrak{e} \\ \langle a \rangle \Phi_{\alpha'} & \text{if } \alpha = a\alpha', \alpha' \in \mathcal{A}^\star. \end{cases}$$

Definition 6.9 (Completed trace formula)**.** Given any trace $\alpha \in \mathcal{A}^\star$ we define the *completed trace formula* of α, notation $\Phi^C_\alpha \in \mathfrak{L}^{\mathrm{l}}$, inductively on the structure of α as follows:

$$\Phi^C_\alpha = \begin{cases} \bigwedge_{a \in \mathcal{A}} \bar{a} & \text{if } \alpha = \mathfrak{e} \\ \langle a \rangle \Phi^C_{\alpha'} & \text{if } \alpha = a\alpha', \alpha' \in \mathcal{A}^\star. \end{cases}$$

Definition 6.10 (Failure formula)**.** Given any failure pair $\mathfrak{f} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ we define the *failure formula* of $\mathfrak{f}$, notation $\Phi_{\mathfrak{f}} \in \mathfrak{L}^{\mathrm{l}}$, inductively on the structure of $\mathfrak{f}$ as follows:

$$\Phi_{\mathfrak{f}} = \begin{cases} \bigwedge_{b \in F} \bar{b} & \text{if } \mathfrak{f} = \mathfrak{e}F \\ \langle a \rangle \Phi_{\mathfrak{f}'} & \text{if } \mathfrak{f} = a\mathfrak{f}', \mathfrak{f}' \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A}). \end{cases}$$

Definition 6.11 (Failure trace formula)**.** Given any failure trace $\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ we define the *failure trace formula* of $\mathfrak{F}$, notation $\Phi_{\mathfrak{F}} \in \mathfrak{L}^{\mathrm{l}}$, inductively on the structure of $\mathfrak{F}$ as follows:

$$\Phi_{\mathfrak{F}} = \begin{cases} \bigwedge_{b \in F} \bar{b} & \text{if } \mathfrak{F} = \mathfrak{e}F \\ \langle a \rangle \bigwedge_{b \in F} \bar{b} & \text{if } \mathfrak{F} = aF \\ \langle a \rangle (\bigwedge_{b \in F} \bar{b} \wedge \Phi_{\mathfrak{F}'}) & \text{if } \mathfrak{F} = aF\mathfrak{F}', \mathfrak{F}' \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^\star \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A})). \end{cases}$$

Definition 6.12 (Ready formula)**.** Given any ready pair $\mathfrak{r} \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A})$ we define the *ready formula* of $\mathfrak{r}$, notation $\Phi_{\mathfrak{r}} \in \mathfrak{L}^{\mathrm{l}}$, inductively on the structure of $\mathfrak{r}$ as follows:

$$\Phi_{\mathfrak{r}} = \begin{cases} \bigwedge_{b \in R} \langle a \rangle \top & \text{if } \mathfrak{r} = \mathfrak{e}R \\ \langle a \rangle \Phi_{\mathfrak{r}'} & \text{if } \mathfrak{r} = a\mathfrak{r}', \mathfrak{r}' \in \mathcal{A}^\star \times \mathcal{P}(\mathcal{A}). \end{cases}$$

Definition 6.13 (Ready trace formula)**.** Given any ready trace $\mathfrak{R} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$ we define the *ready trace formula* of $\mathfrak{R}$, notation $\Phi_{\mathfrak{R}} \in \mathfrak{L}^{l}$, inductively on the structure of $\mathfrak{R}$ as follows:

$$\Phi_{\mathfrak{R}} = \begin{cases} \bigwedge_{b \in R} \langle b \rangle \top & \text{if } \mathfrak{R} = \mathfrak{e}R \\ \langle a \rangle \bigwedge_{b \in R} \langle b \rangle \top & \text{if } \mathfrak{R} = aR \\ \langle a \rangle (\bigwedge_{b \in R} \langle b \rangle \top \wedge \Phi_{\mathfrak{R}'}) & \text{if } \mathfrak{R} = aR\mathfrak{R}', \mathfrak{R}' \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A})). \end{cases}$$

We are now ready to extend our concept of *mimicking formula* of a process to capture the (decorated) trace semantics.

Definition 6.14 (Mimicking formulae for decorated trace semantics)**.** Let $s \in \mathcal{S}$.

- The *mimicking formula for trace equivalence of* s is denoted by Ψ_s^{Tr} and defined by

$$\Psi_s^{\mathrm{Tr}} = \bigwedge_{\alpha \in \mathcal{A}^{\star}} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \right) \Phi_{\alpha}$$

  where for each trace $\alpha \in \mathcal{A}^{\star}$, Φ_{α} is the trace formula of α.

- The *mimicking formula for completed trace equivalence of* s is denoted by Ψ_s^{TrC} and defined by

$$\Psi_s^{\mathrm{TrC}} = \Psi_s^{\mathrm{Tr}} \wedge \bigwedge_{\alpha \in \mathcal{A}^{\star}} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{CC}(z_s, \alpha)) \right) \Phi_{\alpha}^{C}$$

  where Ψ_s^{Tr} is the mimicking formula for trace equivalence for s and for each trace $\alpha \in \mathcal{A}^{\star}$, Φ_{α}^{C} is the completed trace formula of α.

- The *mimicking formula for failure equivalence of* s is denoted by Ψ_s^{F} and defined by

$$\Psi_s^{\mathrm{F}} = \bigwedge_{\mathfrak{f} \in \mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A})} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{f})) \right) \Phi_{\mathfrak{f}}$$

  where for each failure pair $\mathfrak{f} \in \mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A})$, $\Phi_{\mathfrak{f}}$ is the failure formula of $\mathfrak{f}$.

- The *mimicking formula for failure trace equivalence of* s is denoted by Ψ_s^{TrF} and defined by

$$\Psi_s^{\mathrm{TrF}} = \bigwedge_{\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{FC}(z_s, \mathfrak{F})) \right) \Phi_{\mathfrak{F}}$$

  where for each failure trace $\mathfrak{F} \in (\mathcal{A} \times \mathcal{P}(\mathcal{A}))^{\star} \cup (\mathfrak{e} \times \mathcal{P}(\mathcal{A}))$, $\Phi_{\mathfrak{F}}$ is the failure trace formula of $\mathfrak{F}$.

- The *mimicking formula for readiness equivalence of* s is denoted by Ψ_s^{R} and defined by

$$\Psi_s^{\mathrm{R}} = \bigwedge_{\mathfrak{r} \in \mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A})} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s, \mathfrak{r})) \right) \Phi_{\mathfrak{r}}$$

  where for each ready pair $\mathfrak{r} \in \mathcal{A}^{\star} \times \mathcal{P}(\mathcal{A})$, $\Phi_{\mathfrak{r}}$ is the ready formula of $\mathfrak{r}$.

- ⋆ The *mimicking formula for ready trace equivalence of* s is denoted by Ψ_s^{TrR} and defined by
$$\Psi_s^{\mathrm{TrR}} = \bigwedge_{\mathfrak{R}\in(\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star\cup(\mathfrak{e}\times\mathcal{P}(\mathcal{A}))} \left(\sup_{\mathcal{Z}_s\in\mathrm{Res}(s)} \Pr(\mathcal{RC}(z_s,\mathfrak{R})) \right) \Phi_{\mathfrak{R}}$$
where for each ready trace $\mathfrak{R}\in(\mathcal{A}\times\mathcal{P}(\mathcal{A}))^\star\cup(\mathfrak{e}\times\mathcal{P}(\mathcal{A}))$, $\Phi_{\mathfrak{R}}$ is the ready trace formula of $\mathfrak{R}$.

Clearly, each process satisfies its own mimicking formula for the desired semantics.

Theorem 6.4. *Let $x\in\{\mathrm{Tr},\mathrm{TrC},\mathrm{F},\mathrm{TrF},\mathrm{R},\mathrm{TrR}\}$. For each $s\in\mathcal{S}$, it holds that $s\models\Psi_s^x$.*

Proof. The proof is immediate by Definition 6.14. ■

By means of mimicking formulae we obtain the $\mathfrak{L}$-characterizations of (decorated) trace semantics for image finite processes. Given $x\in\{\mathrm{Tr},\mathrm{TrC},\mathrm{F},\mathrm{TrF},\mathrm{R},\mathrm{TrR}\}$, we have that $s\sqsubseteq_x t$ if and only if the mimicking formulae of s and t for the semantics x are related by the relation of inequality of formulae. When equivalences are considered, the characterizations are derived from the equality of mimicking formulae.

Theorem 6.5 ($\mathfrak{L}$-characterizations of (decorated) trace relations)**.** *Let $x\in\{\mathrm{Tr},\mathrm{TrC},\mathrm{F},\mathrm{TrF},\mathrm{R},\mathrm{TrR}\}$.*

1. *For all $s,t\in\mathcal{S}$ we have that $s\sqsubseteq_x t$ if and only if $\Psi_s^x\le\Psi_t^x$.*

2. *For all $s,t\in\mathcal{S}$ we have that $s\sim_x t$ if and only if $\Psi_s^x=\Psi_t^x$.*

Proof. By Theorems 4.20 and 4.26 we have that $s\sqsubseteq_x t$ if and only if $\boldsymbol{d}_{\sqsubseteq_{x,\lambda}}(s,t)=0$ and $s\sim_x t$ if and only if $\boldsymbol{d}_{x,\lambda}(s,t)=0$. Therefore, the thesis follows by Definition 6.15, Theorem 6.8 and Proposition 6.3. ■

$\mathfrak{L}$-characterizations of (decorated) trace metrics

In this section we present the logical characterization of (decorated) trace metrics (Theorem 6.8). We obtain it by lifting the distance on $\mathfrak{L}$ defined in Section 6.1 to a distance on processes by exploiting the mimicking formulae introduced in Section 6.2: the *logical distance* on processes is defined as the syntactical distance on their mimicking formulae.

Definition 6.15 (Logical decorated trace distances)**.** Let $\lambda\in(0,1]$.

- ⋆ The *logical trace pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}\colon\mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)=\mathcal{D}_\lambda^{\mathrm{p}}(\Psi_s^{\mathrm{Tr}},\Psi_t^{\mathrm{Tr}}).$$
The *logical trace distance* on processes $\ell_{\mathrm{Tr},\lambda}\colon\mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{Tr},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\}.$$

- ⋆ The *logical completed trace pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t)=\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{TrC}}_{s},\Psi^{\mathrm{TrC}}_{t}).$$
The *logical completed trace distance* on processes $\ell_{\mathrm{TrC},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{TrC},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}(t,s)\}.$$

- ⋆ The *logical failure pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{F}},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t)=\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{F}}_{s},\Psi^{\mathrm{F}}_{t}).$$
The *logical failure distance* on processes $\ell_{\mathrm{F},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{F},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{F}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{F}},\lambda}(t,s)\}.$$

- ⋆ The *logical failure trace pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t)=\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{TrF}}_{s},\Psi^{\mathrm{TrF}}_{t}).$$
The *logical failure trace distance* on processes $\ell_{\mathrm{TrF},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{TrF},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}(t,s)\}.$$

- ⋆ The *logical readiness pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{R}},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t)=\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{R}}_{s},\Psi^{\mathrm{R}}_{t}).$$
The *logical readiness distance* on processes $\ell_{\mathrm{R},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{R},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{R}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{R}},\lambda}(t,s)\}.$$

- ⋆ The *logical ready trace pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t)=\mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{TrR}}_{s},\Psi^{\mathrm{TrR}}_{t}).$$
The *logical ready trace distance* on processes $\ell_{\mathrm{TrR},\lambda}\colon \mathcal{S}\times\mathcal{S}\to[0,1]$ is defined, for all $s,t\in\mathcal{S}$, by
$$\ell_{\mathrm{TrR},\lambda}(s,t)=\max\{\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}(s,t),\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}(t,s)\}.$$

Our logical (pre) distances are well-defined (hemimetrics) pseudometrics on $\mathcal{S}$, as formalized in the following Proposition.

Proposition 6.6. *Let $x\in\{\mathrm{Tr},\mathrm{TrC},\mathrm{F},\mathrm{TrF},\mathrm{R},\mathrm{TrR}\}$ and $\lambda\in(0,1]$.*

1. *The mapping $\ell_{\sqsubseteq_x,\lambda}$ is a 1-bounded hemimetric on $\mathcal{S}$.*
2. *The mapping $\ell_{x,\lambda}$ is a 1-bounded pseudometric on $\mathcal{S}$.*

Proof.

1. The proof follows immediately by Proposition 6.2.

2. The proof follows immediately by Proposition 6.6.1.

■

From our $\mathfrak{L}$-characterization of decorated trace relations (Theorem 6.5) we obtain the following characterization of the kernels of our logical (pre) distances.

Theorem 6.7. *Let $x \in \{\mathrm{Tr}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$ and $\lambda \in (0,1]$. For all processes $s, t \in \mathcal{S}$ we have*

1. $\ell_{\sqsubseteq_x,\lambda}(s,t) = 0$ *if and only if* $s \sqsubseteq_x t$.

2. $\ell_{x,\lambda}(s,t) = 0$ *if and only if* $s \sim_x t$.

Proof. We present only the proof for trace preorder and trace equivalence, namely the case of $x = \mathrm{Tr}$. All remaining cases follow by the same arguments.

1.
$$\begin{aligned} s \sqsubseteq_{\mathrm{Tr}} t \text{ iff } & \Psi_s^{\mathrm{Tr}} \leq \Psi_t^{\mathrm{Tr}} && \text{(by Theorem 6.5.1)} \\ \text{iff } & \mathcal{D}_\lambda^{\mathrm{p}}(\Psi_s^{\mathrm{Tr}}, \Psi_t^{\mathrm{Tr}}) = 0 && \text{(by Proposition 6.3)} \\ \text{iff } & \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0. \end{aligned}$$

2.
$$\begin{aligned} s \sim_{\mathrm{Tr}} t \text{ iff } & s \sqsubseteq_{\mathrm{Tr}} t \text{ and } t \sqsubseteq_{\mathrm{Tr}} s \\ \text{iff } & \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) = 0 \text{ and } \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s) = 0 && \text{(by Theorem 6.7.1)} \\ \text{iff } & \max\{\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\ \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} = 0 \\ \text{iff } & \ell_{\mathrm{Tr},\lambda}(s,t) = 0. \end{aligned}$$

■

Finally, we obtain the logical characterization of the decorated trace (hemi)metrics.

Theorem 6.8 ($\mathfrak{L}$-characterization of (decorated) trace (hemi)metrics)**.** *Let $x \in \{\mathrm{Tr}, \mathrm{TrC}, \mathrm{F}, \mathrm{TrF}, \mathrm{R}, \mathrm{TrR}\}$ and $\lambda \in (0,1]$. For all $s, t \in \mathcal{S}$ we have*

1. $\mathbf{d}_{\sqsubseteq_x,\lambda}(s,t) = \ell_{\sqsubseteq_x,\lambda}(s,t)$.

2. $\mathbf{d}_{x,\lambda}(s,t) = \ell_{x,\lambda}(s,t)$.

Proof. We present only the proof for trace preorder and trace equivalence, namely the case of $x = \mathrm{Tr}$. All remaining cases follow by the same arguments.

1. First of all, we notice that whenever $\mathcal{D}^{\mathrm{l}}(\Phi_1, \Phi_2) = 0$, then

$$\max\{0, \lambda^{\mathrm{dpt}(\Phi_1)}(r_1 - r_2)\} \le 1 \text{ for all } r_1, r_2 \in [0,1]. \tag{6.2}$$

This is due to the fact that $\lambda \le 1$ and $r_1, r_2 \in [0,1]$. Therefore, we have

$$\begin{aligned}
\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) &= \mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi^{\mathrm{Tr}}_s, \Psi^{\mathrm{Tr}}_t) \\
&= \mathcal{D}^{\mathrm{p}}_{\lambda}\left(\bigwedge_{\alpha \in \mathcal{A}^\star} \left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \right) \Phi_\alpha, \bigwedge_{\beta \in \mathcal{A}^\star} \left(\sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t, \beta)) \right) \Phi_\beta \right) \\
&= \sup_{\alpha \in \mathcal{A}^\star} \inf_{\beta \in \mathcal{A}^\star} \mathcal{D}^{\mathrm{p}}_{\lambda}\left(\left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) \right) \Phi_\alpha, \left(\sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t, \beta)) \right) \Phi_\beta \right) \\
&= \sup_{\alpha \in \mathcal{A}^\star} \max\left\{ 0, \lambda^{\mathrm{dpt}(\Phi_\alpha) - 1}\left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t, \alpha)) \right) \right\} \\
&= \sup_{\alpha \in \mathcal{A}^\star} \max\left\{ 0, \lambda^{|\alpha| - 1}\left(\sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}(z_s, \alpha)) - \sup_{\mathcal{Z}_t \in \mathrm{Res}(t)} \Pr(\mathcal{C}(z_t, \alpha)) \right) \right\} \\
&= \sup_{\alpha \in \mathcal{A}^\star} \mathbf{d}^{\alpha}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t) \\
&= \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t)
\end{aligned}$$

where the fourth step follows by Equation (6.2).

2.

$$\begin{aligned}
\mathbf{d}_{\mathrm{Tr},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\ \mathbf{d}_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} \\
&= \max\{\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(s,t),\ \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}(t,s)\} && \text{(by Theorem 6.8.1)} \\
&= \ell_{\mathrm{Tr},\lambda}(s,t).
\end{aligned}$$

■

Notice that differently from the total-variation approach used in the literature, which requires to evaluate the disparities of processes in the evaluation of every single formula in the considered logic, our approach considers only mimicking formulae that capture all the relevant semantic properties of processes and thus the ones necessary to measure their behavioral distances.

6.3 A modal logic for testing

We introduce the modal logic $\mathfrak{L}_{\surd}$ which refines $\mathfrak{L}$ to deal with tests. More precisely, we refine the class $\mathfrak{L}^{\mathrm{l}}$ to the class of linear formulae $\mathfrak{L}^{\mathrm{l}}_{\surd}$ by adding two particular modalities capturing the successful process and termination without success.

Definition 6.16 (Modal logic $\mathfrak{L}_{\surd}$)**.** The classes of *linear formulae* $\mathfrak{L}^{\mathrm{l}}_{\surd}$ and *probabilistic formulae* $\mathfrak{L}^{\mathrm{p}}_{\surd}$ over $\mathcal{A}$ are defined by the following BNF-like grammar:

$$\begin{aligned} \mathfrak{L}^{\mathrm{l}}_{\surd}: &\quad \Phi ::= \surd \mid \perp \mid \langle a \rangle \Phi \\ \mathfrak{L}^{\mathrm{p}}_{\surd}: &\quad \Psi ::= r\Phi \mid \bigwedge_{i \in \mathcal{I}} \Psi_i \end{aligned}$$

where: (i) Φ ranges over $\mathfrak{L}^{\mathrm{l}}_{\surd}$, (ii) Ψ ranges over $\mathfrak{L}^{\mathrm{p}}_{\surd}$, (iii) $a \in \mathcal{A}$, (iv) $r \in (0,1]$, (v) $\mathcal{I}$ is an at most countable set of indexes, (vi) for each $i \in \mathcal{I}$ it holds $\Psi_i \neq \bigwedge_{j \in \mathcal{J}} \Psi_j$ for any set of indexes $\mathcal{J}$ with $|\mathcal{J}| > 1$.

The notion of *depth* of formulae can be naturally derived from Definition 6.2 by adding the two following cases:

⋆ $\mathrm{dpt}(\surd) = 0$ ⋆ $\mathrm{dpt}(\perp) = 0$.

Since we work with finite tests, it is enough to consider linear formulae of finite depth. Given $\Phi \in \mathfrak{L}^{\mathrm{l}}_{\surd}$ we say that $\mathrm{last}(\Phi) = \surd$ (resp. $\mathrm{last}(\Phi) = \perp$) if $\Phi = \langle a_1 \rangle \dots \langle a_n \rangle \surd$ (resp. $\Phi = \langle a_1 \rangle \dots \langle a_n \rangle \perp$) for some $n \in \mathbb{N}$.

Formulae in $\mathfrak{L}_{\surd}$ are interpreted over PTSs and the semantics of $\mathfrak{L}^{\mathrm{p}}_{\surd}$ and $\mathfrak{L}^{\mathrm{l}}_{\surd}$ naturally follow from, respectively, Definition 6.4 and Definition 6.3 in which we add the following constraints for the satisfaction of the novel diamond modalities:

- ⋆ $s \models \surd$ iff $s = \surd$.
- ⋆ $s \models \perp$ iff $s \neq \surd$ and $\mathrm{init}(s) = \emptyset$.

A distance on $\mathfrak{L}_{\surd}$

We introduce a syntactical distance on $\mathfrak{L}_{\surd}$ that we will use to characterize the testing (hemi)metric. The distance on linear formulae simply measure the disparities in the labels of diamond operators.

The distance on probabilistic formulae is technically more complicated. As discussed in Chapter 4.3, when considering testing semantics, processes can be distinguished by both their probability to reach success and the possibility of failure. This discriminating power has to be transferred to the formulae expressing this semantics. Therefore, accordingly to the trace-by-trace approach we have applied to testing, the distance between probabilistic formulae $r_1\Phi_1$ and $r_2\Phi_2$ should depend on: 1. the distance between Φ_1 and Φ_2, 2. the last modality occurring in the two linear formulae, 3. the comparison of the weights r_1 and r_2. Firstly, we need to compare the behavior of processes on the same trace and thus if the distance between Φ_1 and Φ_2 is not 0 then the distance between $r_1\Phi_1$ and $r_2\Phi_2$ can be directly set to 1. Conversely, if Φ_1 and Φ_2 express the same trace, and thus their distance is 0, then we need to investigate whether this trace leads to success or not. If the answer is positive, namely we have that $\mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \surd$, then the distance between $r_1\Phi_1$ and $r_2\Phi_2$ should follow from the comparison of r_1 and r_2, that is by setting it to $r_1 - r_2$ if this difference is positive and to 0 otherwise. In the opposite case having $\mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \perp$ the value of $r_1 - r_2 > 0$

is significant only if $r_2 = 0$. This is due to the fact that to guarantee the compatibility with testing semantics in the fully-nondeterministic case, a positive probability of failure has to be matched by a positive probability of failure (cf. Chapter 4.3). Let us consider now the mixed cases. If $\mathrm{last}(\Phi_1) = \bot$ and $\mathrm{last}(\Phi_2) = \surd$, then there is no need to compare r_1 and r_2 and the distance is set to 0 (Lemma 4.8 in Chapter 4.3 and the upcoming definition of mimicking formula for testing Definition 6.21 ensure that a comparison of two such formulae may occur only if $r_1 = 0$). Symmetrically, if $\mathrm{last}(\Phi_1) = \surd$ and $\mathrm{last}(\Phi_2) = \bot$ then the distance is set to r_1 since the "success probability" of Φ_2 is 0.

The following Definition formalize the intuitions discussed above.

Definition 6.17 (Distance on $\mathfrak{L}_{\surd}$)**.** Let $\lambda \in (0,1]$. The function $\mathbb{D}^{\mathrm{l}} \colon \mathfrak{L}^{\mathrm{l}}_{\surd} \times \mathfrak{L}^{\mathrm{l}}_{\surd} \to [0,1]$ is defined by structural induction over $\mathfrak{L}^{\mathrm{l}}_{\surd}$ as follows:

$$\mathbb{D}^{\mathrm{l}}(\Phi_1, \Phi_2) = \begin{cases} 0 & \text{if } \mathrm{dpt}(\Phi_1) = \mathrm{dpt}(\Phi_2) = 0 \\ \mathbb{D}^{\mathrm{l}}(\Phi_1', \Phi_2') & \text{if } \Phi_1 = \langle a \rangle \Phi_1' \text{ and } \Phi_2 = \langle a \rangle \Phi_2' \\ 1 & \text{otherwise.} \end{cases}$$

The function $\mathbb{D}^{\mathrm{p}}_{\lambda} \colon \mathfrak{L}^{\mathrm{p}}_{\surd} \times \mathfrak{L}^{\mathrm{p}}_{\surd} \to [0,1]$ is defined over $\mathfrak{L}^{\mathrm{p}}_{\surd}$ as follows:

$$\mathbb{D}^{\mathrm{p}}_{\lambda}(r_1\Phi_1, r_2\Phi_2) = \begin{cases} 0 & \text{if } \mathbb{D}^{\mathrm{l}}(\Phi_1,\Phi_2) = 0 \text{ and } \mathrm{last}(\Phi_1) = \bot, \mathrm{last}(\Phi_2) = \surd \\ & \text{or } \mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \bot \text{ and } r_2 > 0 \\ \max\{0, \lambda^{\mathrm{dpt}(\Phi_1)-1}(r_1 - r_2)\} & \text{if } \mathbb{D}^{\mathrm{l}}(\Phi_1,\Phi_2) = 0 \text{ and } \mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \surd \\ \lambda^{\mathrm{dpt}(\Phi_1)} r_1 & \text{if } \mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \bot \text{ and } r_2 = 0 \\ \lambda^{\mathrm{dpt}(\Phi_1)-1} r_1 & \text{if } \mathbb{D}^{\mathrm{l}}(\Phi_1,\Phi_2) = 0 \text{ and } \mathrm{last}(\Phi_1) = \surd, \mathrm{last}(\Phi_2) = \bot \\ 1 & \text{otherwise} \end{cases}$$

$$\mathbb{D}^{\mathrm{p}}_{\lambda}\Big(\bigwedge_{i \in \mathcal{I}} \Psi_i, \bigwedge_{j \in \mathcal{J}} \Psi_j\Big) = \sup_{i \in \mathcal{I}} \inf_{j \in \mathcal{J}} \mathcal{D}^{\mathrm{p}}_{\lambda}(\Psi_i, \Psi_j).$$

Proposition 6.9. *The function $\mathbb{D}^{\mathrm{p}}_{\lambda}$ is a 1-bounded premetric over $\mathfrak{L}^{\mathrm{p}}_{\surd}$.*

Proof. The proof follows by an easy induction over the structure of probabilistic formulae. In particular, the base case $\Psi = r\Phi$ follows by applying the same arguments used in the proof of Theorem 4.9. ■

Also in this case, we can characterize the kernel of the distance on formulae in terms of an ordering relation $\leq$ over formulae in $\mathfrak{L}_{\surd}$.

Definition 6.18 (Ordering of formulae in $\mathfrak{L}^{\mathrm{l}}_{\surd}$)**.** The relation $\leq \colon \mathfrak{L}^{\mathrm{l}}_{\surd} \times \mathfrak{L}^{\mathrm{l}}_{\surd}$ is defined inductively over the structure of probabilistic formulae as follows

- ★ $(\bot, \surd) \in \leq$;
- ★ $(\langle a \rangle \Phi_1, \langle a \rangle \Phi_2) \in \leq$ if and only if $(\Phi_1, \Phi_2) \in \leq$;
- ★ $(\Phi, \Phi) \in \leq$ for all $\Phi \in \mathfrak{L}^{\mathrm{l}}_{\surd}$.

To simplify notation, we write $\Phi_1 \leq \Phi_2$ in place of $(\Phi_1, \Phi_2) \in \leq$. Relation $\leq$ coincides with the kernel of $\mathbb{D}^{\mathrm{l}}$.

Proposition 6.10. *For all $\Phi_1, \Phi_2 \in \mathfrak{L}^{\mathrm{l}}_{\surd}$, we have that $\mathbb{D}^{\mathrm{l}}(\Phi_1, \Phi_2) = 0$ if and only if $\Phi_1 \leq \Phi_2$.*

Proof. The proof follows a simple induction analysis over the structure of $\Phi_1 \in \mathfrak{L}^{\mathrm{l}}_{\surd}$. ■

The ordering relation on linear formulae can be exploited to define that on probabilistic formulae.

Definition 6.19 (Ordering of formulae in $\mathfrak{L}^{\mathrm{p}}_{\surd}$)**.** The relation $\leq\colon \mathfrak{L}^{\mathrm{p}}_{\surd} \times \mathfrak{L}^{\mathrm{p}}_{\surd}$ is defined inductively over the structure of probabilistic formulae as follows

- ★ $(r_1\Phi_1, r_2\Phi_2) \in \leq$ if and only if $\Phi_1 \leq \Phi_2$ and
    - ∗ either $\mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \surd$ and $r_1 \leq r_2$;
    - ∗ or $\mathrm{last}(\Phi_1) = \mathrm{last}(\Phi_2) = \bot$ and $r_2 > 0$;
    - ∗ or $\mathrm{last}(\Phi_1) = \bot$ and $\mathrm{last}(\Phi_2) = \surd$.
- ★ $(\bigwedge_{i \in \mathcal{I}} \Psi_i, \bigwedge_{j \in \mathcal{J}} \Psi_j) \in \leq$ if and only if for each $i \in \mathcal{I}$ there is a $j \in \mathcal{J}$ such that $(\Psi_i, \Psi_j) \in \leq$.

To simplify notation, we write $\Psi_1 \leq \Psi_2$ in place of $(\Psi_1, \Psi_2) \in \leq$. As for linear formulae, the equality of conjunctions is intended up-to reordering. More precisely, we have that $\Psi_1 = \Psi_2$ if and only if $\Psi_1 \leq \Psi_2$ and $\Psi_2 \leq \Psi_1$.

Finally, we prove that the kernel of $\mathbb{D}^{\mathrm{p}}_{\lambda}$ coincides with the relation $\leq$ over formulae in $\mathfrak{L}^{\mathrm{p}}_{\surd}$.

Proposition 6.11. *For all $\Psi_1, \Psi_2 \in \mathfrak{L}^{\mathrm{p}}_{\surd}$, we have that $\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi_1, \Psi_2) = 0$ if and only if $\Psi_1 \leq \Psi_2$.*

Proof. The proof follows a simple induction analysis over the structure of $\Psi_1 \in \mathfrak{L}^{\mathrm{p}}_{\surd}$. ■

6.4 Logical characterization of testing semantics

In this Section we propose our characterizations of testing equivalence and metric obtained on the logic $\mathfrak{L}_{\surd}$.

$\mathfrak{L}_{\surd}$-characterization of testing equivalence

The characterization of probabilistic testing preorder and equivalence is obtained from the mimicking formulae for testing. First of all we introduce the notions of *(un)successful trace formulae* expressing trace formulae ending with the (un)success formula.

Definition 6.20 (Trace formulae in $\mathfrak{L}^{\mathrm{l}}_{\surd}$)**.** Consider any trace $\alpha \in \mathcal{A}^\star$ and let $x \in \{\surd, \bot\}$. We define the formula $\Phi_{\alpha,x} \in \mathfrak{L}^{\mathrm{l}}_{\surd}$, inductively on the structure of α as follows:

$$\Phi_{\alpha,x} = \begin{cases} x & \text{if } \alpha = \mathfrak{e} \\ \langle a \rangle x & \text{if } \alpha = a \\ \langle a \rangle \Phi_{\alpha',x} & \text{if } \alpha = a\alpha'. \end{cases}$$

Then the formula $\Phi_{\alpha,\surd}$ is called the *successful trace formula of* α. Analogously, the formula $\Phi_{\alpha,\perp}$ is called the *unsuccessful trace formula of* α.

In the testing equivalence and metric defined in Chapter 4, the success probabilities are evaluated on the interaction systems of processes with tests. Thus, in place of the mimicking formula of a process for testing semantics, we introduce *mimicking formulae of interaction systems*. They are defined by following the trace-by-trace approach and moreover they exploit the normalization introduced to define the testing metric. The reason why we need normalization is the same discussed in Chapter 4.3: ensuring the comparability of the upcoming *logical testing distance* with the logical distances proposed do far. Besides normalization, the definition of the mimicking formulae for testing is built on the same intuitions that led us to the definition of testing metric. Briefly, consider a process s, a test o and a trace α. Assume first that α is successful for (s,o). Then the *mimicking formula on* α *for* (s,o) will be defined as the probabilistic formula assigning to the successful trace formula of α the (normalized) supremal probability of (s,o) to reach success by executing α, with respect to all resolutions of nondeterminism. Notice that α is successful for (s,o) if and only if it is successful for o and there is a computation from s that is compatible with it. Conversely, if α is not successful for (s,o) then the mimicking formula will be defined as the probabilistic formula assigning to the unsuccessful trace formula of α the (normalized) supremal probability of (s,o) of executing a maximal computation compatible with α. This case corresponds to the second case of Definition 4.16 and it captures the maximal probability of the interaction system to fail by executing the trace α.

Definition 6.21 (Mimicking formula for testing)**.** Assume a PTS $P = (\mathcal{S},\mathcal{A},\longrightarrow)$ and an NPT $O = (\mathcal{O},\mathcal{A},\longrightarrow_O)$. Let $s \in \mathcal{S}$ and $o \in \mathcal{O}$. For each $\alpha \in \mathcal{A}^\star$ we define the *mimicking formula on* α for (s,o) as

$$\Psi^{\alpha}_{s,o} = \begin{cases} \dfrac{\sup_{\mathcal{Z}_{s,o}\in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha))}{\sup_{\mathcal{Z}_{o}\in \mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha))} \Phi_{\alpha,\surd} & \text{if } \sup_{\mathcal{Z}_{s,o}\in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) > 0 \\[2ex] \dfrac{\sup_{\mathcal{Z}_{s,o}\in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup_{\mathcal{Z}_{o}\in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))} \Phi_{\alpha,\perp} & \text{otherwise.} \end{cases}$$

Then we define the *mimicking formula for testing of the interaction system* (s,o), notation $\Psi_{s,o} \in \mathfrak{L}^{\mathrm{p}}_{\surd}$, as the probabilistic formula

$$\Psi_{s,o} = \bigwedge_{\alpha\in\mathcal{A}^\star} \Psi^{\alpha}_{s,o}.$$

We remark that the construction of the mimicking formula on a trace, guarantees that for each trace $\alpha \in \mathcal{A}^\star$ in the mimicking formula for the interaction system $s \parallel o$ occurs either the successful trace formula of α or its unsuccessful version. This property will be fundamental for the characterization result. Moreover, mimicking formulae satisfy the following feature.

Lemma 6.12. *For any* $s \in \mathcal{S}, o \in \mathcal{O}, \alpha \in \mathcal{A}^\star$ *we have* $\Psi^{\alpha}_{s,o} = 0\Phi_{\alpha,\perp}$ *iff* $\mathrm{Res}_{\max,\alpha}(s,o) = \emptyset$.

Proof. ($\Rightarrow$) From the assumption $\Psi^{\alpha}_{s,o} = 0\Phi_{\alpha,\perp}$ and Definition 6.21 we can infer that

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0 \qquad \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha)) = 0.$$

Therefore, we can conclude that there is no maximal computation from (s,o) which is compatible with the trace α, that is $\mathrm{Res}_{\max,\alpha}(s,o) = \emptyset$.

($\Leftarrow$) As $\mathrm{Res}_{\max,\alpha}(s,o) = \emptyset$ implies that no maximal computation from (s,o) is compatible with α, we can immediately infer that

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = 0 \qquad \sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha)) = 0$$

thus giving (by Definition 6.21) that $\Psi^{\alpha}_{s,o} = 0\Phi_{\alpha,\perp}$. ■

By means of mimicking formulae we obtain the $\mathfrak{L}_{\surd}$-characterizations of testing semantics. We obtain that $s \sqsubseteq_{\mathrm{test}} t$ if and only if for all tests o the mimicking formulae of the interaction systems of s and t with o, $\Psi_{s,o}$ and $\Psi_{t,o}$, are related by the relation of ordering of formulae. When testing equivalence is considered, the characterization is derived from pointwise equality of mimicking formulae with respect to all tests.

Theorem 6.13. *Assume a PTS $P = (\mathcal{S},\mathcal{A},\rightarrow)$ and an NPT $O = (\mathcal{O},\mathcal{A},\rightarrow_{O})$.*

1. *For all $s,t \in \mathcal{S}$ we have that $s \sqsubseteq_{\mathrm{test}} t$ if and only if for all $o \in \mathcal{O}$ we have $\Psi_{s,o} \le \Psi_{t,o}$.*
2. *For all $s,t \in \mathcal{S}$ we have that $s \sim_{\mathrm{test}} t$ if and only if for all $o \in \mathcal{O}$ we have $\Psi_{s,o} = \Psi_{t,o}$.*

Proof. By Theorem 4.40 we have that $s \sqsubseteq_{\mathrm{test}} t$ if and only if $\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = 0$ and $s \sim_{\mathrm{test}} t$ if and only if $\mathbf{d}_{\mathrm{test},\lambda}(s,t) = 0$. Therefore, the thesis follows by Definition 6.22, Theorem 6.16 and Proposition 6.11. ■

$\mathfrak{L}_{\surd}$-CHARACTERIZATION OF TESTING METRIC

To obtain the characterization of testing metric, we lift the distance on formulae to a distance on processes.

Definition 6.22. Let $\lambda \in (0,1]$. Assume a PTS $P = (\mathcal{S},\mathcal{A},\rightarrow)$ and an NPT $O = (\mathcal{O},\mathcal{A},\rightarrow_{O})$. The *logical testing pre-distance* on processes $\ell_{\sqsubseteq_{\mathrm{test}},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined, for all $s,t \in \mathcal{S}$, by

$$\ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \sup_{o \in \mathcal{O}} \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi_{s,o},\Psi_{t,o}).$$

The *logical testing distance* on processes $\ell_{\mathrm{test},\lambda} \colon \mathcal{S} \times \mathcal{S} \to [0,1]$ is defined, for all $s,t \in \mathcal{S}$, by

$$\ell_{\mathrm{test},\lambda}(s,t) = \max\{\ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t), \ell_{\sqsubseteq_{\mathrm{test}},\lambda}(t,s)\}.$$

Our logical (pre) distance is a well-defined (premetric) semimetric on $\mathcal{S}$, as formalized in the following Proposition.

Proposition 6.14. *Let $\lambda \in (0,1]$.*

1. *The mapping $\ell_{\sqsubseteq_{\text{test}},\lambda}$ is a 1-bounded premetric on $\mathcal{S}$.*
2. *The mapping $\ell_{\text{test},\lambda}$ is a 1-bounded semimetric on $\mathcal{S}$.*

Proof.

1. The proof follows immediately by Proposition 6.9.
2. The proof follows immediately by Proposition 6.14.1.

■

From our $\mathfrak{L}_{\surd}$-characterization of testing equivalence (Theorem 6.13) we obtain the following result.

Theorem 6.15. *Let $\lambda \in (0,1]$. For all processes $s,t \in \mathcal{S}$ we have*

1. *$\ell_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0$ if and only if $s \sqsubseteq_{\text{test}} t$.*
2. *$\ell_{\text{test},\lambda}(s,t) = 0$ if and only if $s \sim_{\text{test}} t$.*

Proof.

1.

$$\begin{aligned}
s \sqsubseteq_{\text{test}} t &\text{ iff } \forall\, o \in \mathcal{O}\ \Psi_{s,o} \le \Psi_{t,o} && \text{(by Theorem 6.13.1)}\\
&\text{ iff } \forall\, o \in \mathcal{O}\ \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi_{s,o},\Psi_{t,o}) = 0 && \text{(by Proposition 6.11)}\\
&\text{ iff } \sup_{o\in\mathcal{O}} \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi_{s,o},\Psi_{t,o}) = 0\\
&\text{ iff } \ell_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0
\end{aligned}$$

2.

$$\begin{aligned}
s \sim_{\text{test}} t &\text{ iff } s \sqsubseteq_{\text{test}} t \text{ and } t \sqsubseteq_{\text{test}} s\\
&\text{ iff } \ell_{\sqsubseteq_{\text{test}},\lambda}(s,t) = 0 \text{ and } \ell_{\sqsubseteq_{\text{test}},\lambda}(t,s) = 0 && \text{(by Theorem 6.15.1)}\\
&\text{ iff } \max\{\ell_{\sqsubseteq_{\text{test}},\lambda}(s,t),\ \ell_{\sqsubseteq_{\text{test}},\lambda}(t,s)\} = 0\\
&\text{ iff } \ell_{\text{test},\lambda}(s,t) = 0.
\end{aligned}$$

■

Finally, we obtain the characterization of the testing metric.

Theorem 6.16 ($\mathfrak{L}_{\surd}$-characterization of testing (pre)metric)**.** *Let $\lambda \in (0,1]$. For all $s,t \in \mathcal{S}$ we have*

1. $\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t) = \ell_{\sqsubseteq_{\text{test}},\lambda}(s,t)$.
2. $\mathbf{d}_{\text{test},\lambda}(s,t) = \ell_{\text{test},\lambda}(s,t)$.

Proof.

1. First of all we notice that for all $\Phi \in \mathfrak{L}^{\mathrm{l}}_{\surd}$

$$\begin{aligned} \max\{0, \lambda^{\mathrm{dpt}(\Phi)-1}(r_1 - r_2)\} &\le 1 \text{ for any } r_1, r_2 \in [0,1] \\ \lambda^{\mathrm{dpt}(\Phi)-1} r_1 &\le 1 \text{ for any } r_1 \in [0,1] \\ \lambda^{\mathrm{dpt}(\Phi)} r_1 &\le 1 \text{ for any } r_1 \in [0,1]. \end{aligned} \tag{6.3}$$

This is due to the fact that $\lambda \le 1$ and $r_1, r_2 \in [0,1]$. Therefore, we have

$$\begin{aligned} \ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) &= \sup_{o\in\mathcal{O}} \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi_{s,o}, \Psi_{t,o}) \\ &= \sup_{o\in\mathcal{O}} \mathbb{D}^{\mathrm{p}}_{\lambda}\big(\bigwedge_{\alpha\in\mathcal{A}^\star} \Psi^{\alpha}_{s,o}, \bigwedge_{\beta\in\mathcal{A}^\star} \Psi^{\beta}_{t,o}\big) \\ &= \sup_{o\in\mathcal{O}} \sup_{\alpha\in\mathcal{A}^\star} \inf_{\beta\in\mathcal{A}^\star} \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\beta}_{t,o}) \\ &= \sup_{o\in\mathcal{O}} \sup_{\alpha\in\mathcal{A}^\star} \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}) \end{aligned}$$

where the last step follows by Equation (6.3). To prove the thesis we will prove that for all $s, t \in \mathcal{S}$ it holds that

$$\ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \le \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \tag{6.4}$$

$$\mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) \le \ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t). \tag{6.5}$$

Proof of Equation (6.4).

By definition of supremum we have that for each $\varepsilon > 0$ there are a test o_ε and a trace α_ε such that $\ell_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) < \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha_\varepsilon}_{s,o_\varepsilon}, \Psi^{\alpha_\varepsilon}_{t,o_\varepsilon}) + \varepsilon$. Hence, to prove Equation (6.4) we need to show that for all $\varepsilon > 0$ we have

$$\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha_\varepsilon}_{s,o_\varepsilon}, \Psi^{\alpha_\varepsilon}_{t,o_\varepsilon}) \le \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t). \tag{6.6}$$

Let $\varepsilon > 0$ and, for simplicity of notation, let $o = o_\varepsilon$ and $\alpha = \alpha_\varepsilon$. We proceed by a case analysis to prove Equation (6.6).

a. Assume first that $\mathrm{Res}_{\max,\alpha}(s,o) = \emptyset$. Then we have $0 = \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}) \le \mathbf{d}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t)$. In fact by Lemma 6.12, $\mathrm{Res}_{\max,\alpha}(s,o) = \emptyset$ implies that $\Psi^{\alpha}_{s,o} = 0\Phi_{\alpha,\perp}$, and thus $\Psi^{\alpha}_{s,o} \le \Psi^{\alpha}_{t,o}$ whichever the structure of $\Psi^{\alpha}_{t,o}$ is. Hence, by Proposition 6.11 we can conclude that $\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}) = 0$.

b. Assume now that $\mathrm{Res}_{\max,\alpha}(s,o) \neq \emptyset$. Accordingly to Definition 6.21, we can distinguish two cases:

   i. $\Psi^{\alpha}_{s,o} = \Big(\sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) / \sup_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha))\Big) \Phi_{\alpha,\surd}$. Then we can distinguish two cases:

★ $\mathrm{Res}_{\max,\alpha}(t,o)=\emptyset$. In this case, we have that

$$\sup_{z_{t,o}\in\mathrm{Res}_{\max,\alpha}(t,o)}\Pr(\mathcal{SC}(z_{t,o},\alpha))=0$$

$$\Psi^{\alpha}_{t,o}=0\Phi_{\alpha,\perp}.$$

Therefore we have

$$\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o},\Psi^{\alpha}_{t,o})=\mathbb{D}^{\mathrm{p}}_{\lambda}\left(\frac{\sup\limits_{z_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)}\Pr(\mathcal{SC}(z_{s,o},\alpha))}{\sup\limits_{z_{o}\in\mathrm{Res}_{\max,\alpha}(o)}\Pr(\mathcal{SC}(z_{o},\alpha))}\Phi_{\alpha,\surd},0\Phi_{\alpha,\perp}\right)$$

$$=\lambda^{\mathrm{dpt}(\Phi_{\alpha,\surd})-1}\frac{\sup\limits_{z_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)}\Pr(\mathcal{SC}(z_{s,o},\alpha))}{\sup\limits_{z_{o}\in\mathrm{Res}_{\max,\alpha}(o)}\Pr(\mathcal{SC}(z_{o},\alpha))}.$$

Notice that $\mathrm{dpt}(\Phi_{\alpha,\surd})=|\alpha|$. If $|\alpha|=\mathrm{dpt}(o)$ then we have

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t)$$

$$=\lambda^{\mathrm{dpt}(o)-1}\frac{\sup\limits_{z_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)}\Pr(\mathcal{SC}(z_{s,o},\alpha))-\sup\limits_{z_{t,o}\in\mathrm{Res}_{\max,\alpha}(t,o)}\Pr(\mathcal{SC}(z_{t,o},\alpha))}{\sup\limits_{z_{o}\in\mathrm{Res}_{\max,\alpha}(o)}\Pr(\mathcal{SC}(z_{o},\alpha))}$$

$$=\lambda^{\mathrm{dpt}(o)-1}\frac{\sup\limits_{z_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)}\Pr(\mathcal{SC}(z_{s,o},\alpha))}{\sup\limits_{z_{o}\in\mathrm{Res}_{\max,\alpha}(o)}\Pr(\mathcal{SC}(z_{o},\alpha))}$$

by which Equation (6.6) directly follows. Conversely, if $|\alpha|<\mathrm{dpt}(o)$ then let o' be the test obtained from o by eliminating all transitions that make $\mathrm{dpt}(o)>|\alpha|$. Then we have

$$\sup_{z_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)}\Pr(\mathcal{SC}(z_{s,o},\alpha))=\sup_{z_{s,o'}\in\mathrm{Res}_{\max,\alpha}(s,o')}\Pr(\mathcal{SC}(z_{s,o'},\alpha))$$

$$\sup_{z_{o}\in\mathrm{Res}_{\max,\alpha}(o)}\Pr(\mathcal{SC}(z_{o},\alpha))=\sup_{z_{o'}\in\mathrm{Res}_{\max,\alpha}(o')}\Pr(\mathcal{SC}(z_{o'},\alpha))$$

and thus

$$\mathbf{d}^{o',\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t)$$

$$=\lambda^{\mathrm{dpt}(o')-1}\frac{\sup\limits_{z_{s,o'}\in\mathrm{Res}_{\max,\alpha}(s,o')}\Pr(\mathcal{SC}(z_{s,o'},\alpha))-\sup\limits_{z_{t,o'}\in\mathrm{Res}_{\max,\alpha}(t,o')}\Pr(\mathcal{SC}(z_{t,o'},\alpha))}{\sup\limits_{z_{o'}\in\mathrm{Res}_{\max,\alpha}(o')}\Pr(\mathcal{SC}(z_{o'},\alpha))}$$

$$=\lambda^{\mathrm{dpt}(o')-1}\frac{\sup\limits_{z_{s,o'}\in\mathrm{Res}_{\max,\alpha}(s,o')}\Pr(\mathcal{SC}(z_{s,o'},\alpha))}{\sup\limits_{z_{o'}\in\mathrm{Res}_{\max,\alpha}(o')}\Pr(\mathcal{SC}(z_{o'},\alpha))}$$

$$= \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o})$$

by which Equation (6.6) follows as well.

- ⋆ $\mathrm{Res}_{\max,\alpha}(t,o) \neq \emptyset$. By Lemma 4.8, accordingly to the structure of $\Psi^{\alpha}_{s,o}$, this implies that $\sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) > 0$ and thus

$$\Psi^{\alpha}_{t,o} = \frac{\sup\limits_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha))}{\sup\limits_{\mathcal{Z}_{o} \in \mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha))} \Phi_{\alpha,\surd}.$$

Then, we have

$$\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}) = \max\left\{0, \lambda^{\mathrm{dpt}(\Phi_{\alpha,\surd})-1} \frac{d(s,t,o,\alpha)}{\sup\limits_{\mathcal{Z}_{o} \in \mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha))}\right\}$$

and, by reasoning as in the previous item, if $|\alpha| = \mathrm{dpt}(o)$ then

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \max\left\{0, \lambda^{\mathrm{dpt}(o)-1} \frac{d(s,t,o,\alpha)}{\sup\limits_{\mathcal{Z}_{o} \in \mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha))}\right\}$$

by which Equation (6.6) directly follows, whereas if $|\alpha| < \mathrm{dpt}(o)$ then we choose o' as above for which

$$\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{SC}(z_{s,o},\alpha)) = \sup_{\mathcal{Z}_{s,o'} \in \mathrm{Res}_{\max,\alpha}(s,o')} \Pr(\mathcal{SC}(z_{s,o'},\alpha))$$
$$\sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) = \sup_{\mathcal{Z}_{t,o'} \in \mathrm{Res}_{\max,\alpha}(t,o')} \Pr(\mathcal{SC}(z_{t,o'},\alpha))$$
$$\sup_{\mathcal{Z}_{o} \in \mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{SC}(z_{o},\alpha)) = \sup_{\mathcal{Z}_{o'} \in \mathrm{Res}_{\max,\alpha}(o')} \Pr(\mathcal{SC}(z_{o'},\alpha))$$

and thus

$$\mathbf{d}^{o',\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \max\left\{0, \lambda^{\mathrm{dpt}(o')-1} \frac{d(s,t,o',\alpha)}{\sup\limits_{\mathcal{Z}_{o'} \in \mathrm{Res}_{\max,\alpha}(o')} \Pr(\mathcal{SC}(z'_{o},\alpha))}\right\}$$
$$= \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}).$$

Thus, Equation (6.6) follows as well.

ii. $\Psi^{\alpha}_{s,o} = \left(\sup_{\mathcal{Z}_{s,o} \in \mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha)) / \sup_{\mathcal{Z}_{o} \in \mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))\right) \Phi_{\alpha,\perp}$. Then we can distinguish two cases:

- ⋆ $\mathrm{Res}_{\max,\alpha}(t,o) = \emptyset$. In this case, we have that

$$\sup_{\mathcal{Z}_{t,o} \in \mathrm{Res}_{\max,\alpha}} \Pr(\mathcal{C}_{\max}(z_{t,o},\alpha)) = 0$$

$$\Psi^{\alpha}_{t,o} = 0\Phi_{\alpha,\perp}.$$

Therefore we have

$$\begin{aligned}\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o}, \Psi^{\alpha}_{t,o}) &= \mathbb{D}^{\mathrm{p}}_{\lambda}\left(\frac{\sup\limits_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup\limits_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))}\Phi_{\alpha,\perp}, 0\Phi_{\alpha,\perp}\right) \\ &= \lambda^{\mathrm{dpt}(\Phi_{\alpha,\perp})} \frac{\sup\limits_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup\limits_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))}.\end{aligned}$$

Notice that $\mathrm{dpt}(\Phi_{\alpha,\perp}) = |\alpha|$. Moreover, since the testing distance is determined by the compatibility with the fully-nondeterministic case, we have that $\mathrm{dpt}(o) \geq |\alpha| + 1$. Hence, if the equality holds, namely $\mathrm{dpt}(o) = |\alpha| + 1$ then it is enough to consider

$$\mathbf{d}^{o,\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \lambda^{\mathrm{dpt}(o)-1} \frac{\sup\limits_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha))}{\sup\limits_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))}$$

for which Equation (6.6) holds. Otherwise, let o' be the test obtained from o by deleting all the transitions that make $\mathrm{dpt}(o) > |\alpha| + 1$. Then we have

$$\begin{aligned}\sup_{\mathcal{Z}_{s,o}\in\mathrm{Res}_{\max,\alpha}(s,o)} \Pr(\mathcal{C}_{\max}(z_{s,o},\alpha)) &= \sup_{\mathcal{Z}_{s,o'}\in\mathrm{Res}_{\max,\alpha}(s,o')} \Pr(\mathcal{C}_{\max}(z_{s,o'},\alpha)) \\ \sup_{\mathcal{Z}_{t,o}\in\mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{C}_{\max}(z_{t,o},\alpha)) &= \sup_{\mathcal{Z}_{t,o'}\in\mathrm{Res}_{\max,\alpha}(t,o')} \Pr(\mathcal{C}_{\max}(z_{t,o'},\alpha)) \\ \sup_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max,\alpha}(o)} \Pr(\mathcal{C}(z_{o},\alpha)) &= \sup_{\mathcal{Z}_{o'}\in\mathrm{Res}_{\max,\alpha}(o')} \Pr(\mathcal{C}(z_{o'},\alpha))\end{aligned}$$

and thus

$$\mathbf{d}^{o',\alpha}_{\sqsubseteq_{\mathrm{test}},\lambda}(s,t) = \lambda^{\mathrm{dpt}(o')-1} \frac{\sup\limits_{\mathcal{Z}_{s,o'}\in\mathrm{Res}_{\max,\alpha}(s,o')} \Pr(\mathcal{C}_{\max}(z_{s,o'},\alpha))}{\sup\limits_{\mathcal{Z}_{o'}\in\mathrm{Res}_{\max}(o')} \Pr(\mathcal{C}(z_{o'},\alpha))}$$

for which Equation (6.6) holds.

★ $\mathrm{Res}_{\max,\alpha}(t,o) \neq \emptyset$. By Lemma 4.8, accordingly to the structure of $\Psi^{\alpha}_{s,o}$, this implies that $\sup_{\mathcal{Z}_{t,o}\in\mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{SC}(z_{t,o},\alpha)) = 0$ and thus

$$\Psi^{\alpha}_{t,o} = \frac{\sup\limits_{\mathcal{Z}_{t,o}\in\mathrm{Res}_{\max,\alpha}(t,o)} \Pr(\mathcal{C}_{\max}(z_{t,o},\alpha))}{\sup\limits_{\mathcal{Z}_{o}\in\mathrm{Res}_{\max}(o)} \Pr(\mathcal{C}(z_{o},\alpha))}\Phi_{\alpha,\perp}.$$

Then, by Definition 6.17 we have

$$\mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha}_{s,o},\Psi^{\alpha}_{t,o}) = 0$$

for which Equation (6.6) holds.

Proof of Equation (6.5).

By definition of supremum we have that for each $\varepsilon > 0$ there are a test o_ε and a trace α_ε such that $\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t) < \mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\text{test}},\lambda} + \varepsilon$. Moreover we notice that in this case, for each $\varepsilon > 0$, o_ε is guaranteed to give the supremal testing distance with respect to α_ε. This implies that either $\mathrm{dpt}(o_\varepsilon) = |\alpha_\varepsilon|$, for $\mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$ defined as in the first case of Definition 4.16, or $\mathrm{dpt}(o_\varepsilon) = |\alpha_\varepsilon| + 1$, for $\mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\text{test}},\lambda}(s,t)$ defined as in the second case of Definition 4.16. Hence, to prove to prove Equation (6.5) we need to show that for all $\varepsilon > 0$ we have

$$\mathbf{d}^{o_\varepsilon,\alpha_\varepsilon}_{\sqsubseteq_{\text{test}},\lambda}(s,t) \le \mathbb{D}^{\mathrm{p}}_{\lambda}(\Psi^{\alpha_\varepsilon}_{s,o_\varepsilon},\Psi^{\alpha_\varepsilon}_{t,o_\varepsilon}). \tag{6.7}$$

Equation (6.6) follows by the same case analysis used to prove Equation (6.6).

2.

$$\begin{aligned}
\mathbf{d}_{\text{test},\lambda}(s,t) &= \max\{\mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(s,t),\ \mathbf{d}_{\sqsubseteq_{\text{test}},\lambda}(t,s)\} \\
&= \max\{\ell_{\sqsubseteq_{\text{test}},\lambda}(s,t),\ \ell_{\sqsubseteq_{\text{test}},\lambda}(t,s)\} && \text{(by Theorem 6.16.1)} \\
&= \ell_{\text{test},\lambda}(s,t).
\end{aligned}$$

■

6.5 A spectrum of logical distances

In this Chapter and in previous Chapter 5 we have proposed a novel method for the logical characterization of behavioral metrics. Briefly, we have considered boolean logics and for each process we have identified a special formula, called mimicking, expressing all the properties of the given process that are relevant with respect to the considered semantics. Then we have transformed the logic into a metric space by defining a distance on formulae measuring their syntactical disparities. Finally, we have proved that the distance between the mimicking formulae of two processes equals the distance of the two processes with respect to the considered semantics.

Interestingly, this kind of characterization gives that the distance on formulae is as expressive as the behavioral metric. As a further evidence of this fact, in this Section we combine the logical characterizations of branching metrics presented in Chapter 5 with those of this Chapter to obtain the first example of a spectrum of behavioral distances on processes obtained solely by means of modal logics. More precisely, we order our *logical distances*, obtained as the distance on the mimicking formulae of processes, by the relation '*make processes farther than*'.

This is formalized in the following Theorem.

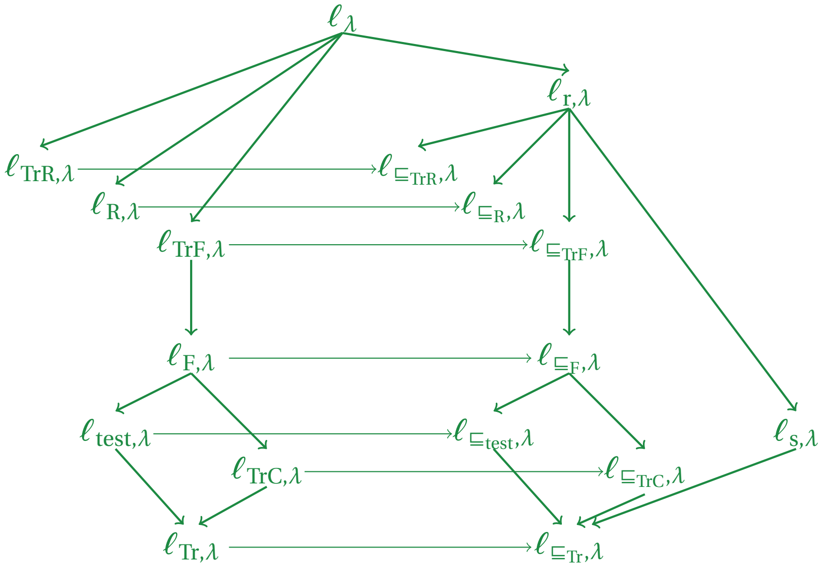

ℓ_λ
$\ell_{\mathrm{r},\lambda}$
$\ell_{\mathrm{TrR},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}$
$\ell_{\mathrm{R},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{R}},\lambda}$
$\ell_{\mathrm{TrF},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{TrF}},\lambda}$
$\ell_{\mathrm{F},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{F}},\lambda}$
$\ell_{\mathrm{test},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{test}},\lambda}$
$\ell_{\mathrm{s},\lambda}$
$\ell_{\mathrm{TrC},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{TrC}},\lambda}$
$\ell_{\mathrm{Tr},\lambda}$
$\ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}$

Figure 6.1: *The spectrum of logical distances. An arrow $\ell \to \ell'$ between two logical distances stands for $\ell(s,t) \geq \ell'(s,t)$ for all processes s,t, and $\ell(s,t) > \ell'(s,t)$ for some processes s,t.*

Theorem 6.17. *Let $\lambda \in (0,1]$. For each $s,t \in \mathcal{S}$ it holds that:*

1. $\ell_\lambda > \ell_{\mathrm{r},\lambda} > \ell_{\mathrm{s},\lambda}$.
2. $\ell_{\mathrm{r},\lambda} > \ell_{\sqsubseteq_{\mathrm{TrF}},\lambda} > \ell_{\sqsubseteq_{\mathrm{F}},\lambda} > \ell_{\sqsubseteq_{\mathrm{test}},\lambda} > \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
3. $\ell_{\sqsubseteq_{\mathrm{F}},\lambda} > \ell_{\sqsubseteq_{\mathrm{TrC}},\lambda} > \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
4. $\ell_{\mathrm{s},\lambda} > \ell_{\sqsubseteq_{\mathrm{Tr}},\lambda}$.
5. $\ell_{\mathrm{r},\lambda} > \ell_{\sqsubseteq_{\mathrm{TrR}},\lambda}$ *and* $\ell_{\mathrm{r},\lambda} > \ell_{\sqsubseteq_{\mathrm{R}},\lambda}$.
6. $\ell_\lambda > \ell_{\mathrm{TrF},\lambda} > \ell_{\mathrm{F},\lambda} > \ell_{\mathrm{test},\lambda} > \ell_{\mathrm{Tr},\lambda}$.
7. $\ell_{\mathrm{F},\lambda} > \ell_{\mathrm{TrC},\lambda} > \ell_{\mathrm{Tr},\lambda}$.
8. $\ell_\lambda > \ell_{\mathrm{TrR},\lambda}$ *and* $\ell_\lambda > \ell_{\mathrm{R},\lambda}$.

Proof. Each relation follows from the corresponding one in Theorem 4.10 and from the proper logical characterization $\ell_{x,\lambda} = \mathbf{d}_{x,\lambda}$ given in Chapters 5 and 6. ■

6.6 Concluding remarks

In this Chapter we have extended the logical characterization method of branching metrics proposed in Chapter 5 to the linear metrics defined in Chapter 4. In detail, we have considered the probabilistic version of the class of formulae characterizing the (decorated) trace relations in the fully-nondeterministic case [33] and a minimal boolean logic, obtained by extending the class of formulae characterizing trace equivalence in the fully-nondeterministic case with special modalities capturing the successful process and failure, to characterize testing metrics and relations. The presence of the modality capturing failure is fundamental to preserve the full backward compatibility of our testing relations with the fully-nondeterministic case. The formulae in this classes express the semantic-specific events and thus the mimicking formulae of processes for the considered semantics specify the supremal probabilities of semantic-specific events to be performed by the related process. We remark that there was no need to introduce recursion in this logic since (decorated) trace semantics are defined in terms of (decorated) traces of finite length. Thus the logical characterization of linear metrics results technically simpler in this Chapter but equally effective to that in Chapter 5.

We have already noticed that our characterization method differs from the total-variation distance approach considered in the literature [59,61,72,73,75]. However, we notice that the disparity in the two characterization techniques diminishes if we consider trace metric only. In fact, due to the simple syntactic structure of formulae in $\mathfrak{L}$, our logical trace distance can be easily translated into a total-variation distance on a real-valued version of the subclass of $\mathfrak{L}^{\mathrm{l}}$

$$\mathfrak{L}^{\mathrm{l}}{}_t :: \Phi := \top \mid \langle a \rangle \Phi.$$

We can define the real-valued semantics of a formula Φ in $\mathfrak{L}^{\mathrm{l}}{}_t$ in process s as

$$[\Phi](s) ::= \sup_{\mathcal{Z}_s \in \mathrm{Res}(s)} \Pr(\mathcal{C}^{\mathrm{t}}(z_s, \Phi)).$$

Then our logical trace distance would become

$$\ell_{\mathrm{Tr},\lambda}(s,t) := \sup_{\Phi \in \mathfrak{L}^{\mathrm{l}}{}_t} |\, [\Phi](s) - [\Phi](t) \,| \,.$$

A similar approach would hold for the other decorated trace metrics. Nevertheless, notice that our approach is more general and it allowed for the construction of the spectrum of logical distances on processes. Moreover, we remark that while the total-variation approach requires to evaluate the disparities of processes in the evaluation of every single formula in the considered logic, our approach considers only mimicking formulae that capture all the relevant semantic properties of processes and thus the ones necessary to measure their behavioral distances. Admittedly, this does not hold for the testing semantics for which we have a mimicking formula for each interaction of the process with a test.

In [59] the authors consider Metric Transition Systems (MTS), namely transition systems in which the atomic propositions, at each state, take values in a bounded metric space. They define an asymmetric *linear distance* generalizing trace inclusion and its symmetric version for trace equivalence and they show that by means of the *Quantitative Linear-Time Temporal Logic* (QLTL), it is possible to characterize those distances. Linear-time properties are also studied in [13], in order to capture approximate reasoning on Stochastic Markov Models (SMMs). SMMs are a generalization of CTMCs in the sense that exit-time probabilities follow generic distributions on the positive real line. Then, for the specification of SMMs properties, they proposed the *Metric Temporal Logic* (MTL), built on implication and the temporal operators *next* and *until*. The authors defined several equivalent distances on SMMs, one of which was the MTL-*variation pseudometric*, defined as the total variation distance on the probability measure on the measurable space of timed paths. As this variation pseudometric is neither computable nor can be approximated (see also [53] for a discussion on the approximation of total variation distances), an over approximation is proposed which is obtained as a convex combination of the total variation distance on the exit-time probabilities and the Kantorovich distance on transition probability functions. In [14] it is proved that the trace metric on Markov Chains (MCs) can be characterized in terms of the probabilistic LTL-model checking problem. Roughly speaking, a characterization as in (6.1) is obtained from the boolean logic LTL by assigning a real-valued semantics to it, defined by exploiting the probabilistic properties of the MC: the value of a formula $\varphi \in$ LTL at state s is given by the probability of s to execute a run satisfying φ.

God's final message to its creation:
'We apologize for the inconvenience.'

Douglas Adams,
So Long and Thank for All the Fish

Chapter 7

Conclusions

In this thesis we have discussed new techniques to study and compare the semantics of processes with nondeterminism and probability, as in Segala's PTS model [145], in terms of both classic behavioral relations and the more recent behavioral metrics.

In detail we have provided a *method for decomposing modal formulae* equipped with a probabilistic choice operator that allowed us to derive the compositional properties of probabilistic strong branching relations. To obtain the decomposition method we have introduced an *SOS-like machinery* specifying the behavior of *distribution terms* as probability distributions over process terms.

Then we focused on behavioral metrics: we have proposed original notions of metrics measuring the disparities in the behavior of processes with respect to (*decorated*) *trace and testing semantics*. To capture the differences in the expressive power of the novel metrics and the ones for probabilistic (bi)simulations we have ordered them by the relation '*makes processes further than*'. Thus we have obtained the first *spectrum of behavioral metrics* on processes in the PTS model. Interestingly, from this spectrum we derived an analogous one for the kernels of the considered metrics, which have been ordered by the relation '*make strictly less identification than*'. Our *spectrum of probabilistic relations* is the probabilistic generalization of the linear time - branching time spectrum of [159].

Finally we have introduced a novel technique for the *logical characterization* of both behavioral metrics and their kernels, based on the notions of *mimicking formula* and *distance on formulae*. In the case of behavioral metrics, the idea is the following: **1.** Once we have chosen a class L of modal boolean-valued formulae suitable for the considered semantics, for a process s we identify a special formula expressing the relevant properties of s with respect to the considered semantics, called *mimicking formula* of s. This is a formula in L that captures the nondeterministic and probabilistic behavior of the process that is relevant for the considered semantics. **2.** Then, we transform the modal logic L into a metric space by introducing a notion of *syntactical distance* on formulae. This is a 1-bounded pseudometric assigning to each pair of formulae a suitable quantitative analogue of their syntactic disparities. 3. We conclude by defining a *logical distance* on processes corresponding to the distance between their mimicking formulae and proving that this logical distance characterizes the considered metric semantics. This kind of characterization allowed us to obtain the first example of a spectrum of behavioral distances on processes obtain directly from modal logics. Moreover, we have showed that the kernels of the considered metrics can be characterized by simply comparing the mimicking formulae of processes.

We conclude by proposing some directions for future work related to the topics addressed in this dissertation.

Modal decomposition We will investigate the application of our decomposition method to the classes of modal formulae characterizing different behavioral semantics for nondeterministic probabilistic processes, as *convex (bi)similarity semantics* [146]. We will also consider the weak semantics [10, 126, 127], and we will derive robust (pre)congruence formats for them from their modal characterizations, as done in the non probabilistic setting. Our claim is that in the case of weak (bi)simulations the congruence results can be obtained by simply combining the formats in [33, 80, 82, 84, 85] and our method of decomposing the probabilistic modalities. A sketch on how we can achieve this purpose has been presented in Chapter 3.5.

Moreover, it would be interesting to combine our decomposition method with the *general rules* and *ruloids* recently introduced in [82] and the characterizations of behavioral metrics that we have provided in Chapters 5 and 6. Inspired by those result, we aim to start a new research line, that is deriving the compositional properties of behavioral metrics from the modal decomposition of formulae characterizing them. As the metric semantics provide notions of *distance* on processes, the formats for them guarantee that a small variance in the behavior of the subprocesses leads to a bounded small variance in the behavior of the composed processes (*uniform continuity* [91,92,94]). Then, we aim to use the decomposition method to re-obtain the formats for bisimilarity metric proposed in [94] and to automatically derive original formats for the (ready) simulation, trace and testing semantics presented in Chapter 4, as well as for weak metric semantics [73, 115] and metric variants of branching bisimulation equivalence [10, 126, 127].

Logical characterizations We have already argued in Chapter 5 that we defined the distance between formulae with the exact purpose of simulating the Hausdorff and Kantorovich lifting functionals on which the bisimilarity metric is defined. Despite this kind of reasoning may seem too restrictive at first glance, we believe that having a distance between formulae instead of a real-valued semantics for the logic turns out to be an advantage in case one wishes to modify the lifting functionals in the definition of (bi)similarity metric (cf. first part of Chapter 5.6). Thus, we aim to extend our results to other lifting functionals, like the generalized Kantorovich lifting functional $\mathbf{K}_V$ [44].

Moreover, we will apply our characterization approach to various behavioral metrics as *convex (bi)simulation metrics* [146] and *weak (bi)simulation metrics* [73]. We aim also to apply the approach to the notion of ϵ*-bisimulation* [7, 8, 74], for which a modal decomposition of formulae characterizing the compositional results in [152, 153] can be given.

Further, it would be interesting to extend our results to the decorated trace and testing *distribution* equivalences defined in [29, 144, 145] and to the metrics capturing them. A first attempt in this direction has been done in [43] where we have proposed a logical characterization of a revised version of the trace metric from [148] by means of the notions of mimicking formula and logical distance.

Behavioral metrics We aim to investigate the behavior of the metrics defined in Chapter 4 and related kernels in the interaction with other types of schedulers, like the randomized ones from [144], the probabilistic ones from [51] and also their balanced versions, in terms of distinguishing power, from [95, 164]. Furthermore, we aim to extend the spectrum of metrics to distribution-based distances, like the trace metric in [148], and to compare the derived spectrum of kernels with its analogous from [30].

Following [92, 94], we are also interested in studying the compositional properties of the behavioral metrics introduced in Chapter 4, like non-extensiveness, non-expansiveness and (Lipschitz) continuity. Furthermore, we aim to define uniform continuity specification formats for them.

Finally, we would like to investigate novel applications for the metric semantics and their logical characterizations. For instance, we can study a quantitative analogue to the *probabilistic barbed bisimulation* from [65]. *Barbed bisimulation* was introduced in [132] to uniformly describe the observable behavior of process calculi equipped with a reduction semantics. Thus, *barbed bisimulation metric* could be exploited to equip process calculi for wireless sensor networks, mobile ad hoc networks and cyber-physical systems with a robust semantics.

Another possible application of behavioral metrics is related to biological systems modeling. Many process algebras have been introduced to model these systems (see for instance [19, 20, 22, 23, 38, 48, 49]) and recently also some probabilistic algebras have been proposed [15, 18, 21, 24]. Metric semantics could be used in this setting to strengthen the expressive power of these models, thus obtaining more effective representations.

We would also study efficient ways to express and verify privacy properties in concurrent and interactive settings. To this purpose we will consider the bisimulation distance for *differential-privacy* [76]. We will study efficient ways to compute such distance by considering the formulation of the Kantorovich lifting in terms of the transportation problem, and we will exploit the fact that differential-privacy has strong separation properties to reduce the space of possibility in the minimization process. We will also explore a logic for characterizing the bisimulation distance for differential-privacy, in the style of the probabilistic Hennessy-Milner logic $\mathcal{L}$ [66].

Bibliography

[1] Alessandro Abate, Alessandro D'Innocenzo, and Maria Domenica Di Benedetto. Approximate abstractions of stochastic hybrid systems. *IEEE Trans. Automat. Contr.*, 56(11):2688–2694, 2011.

[2] Alessandro Abate, Joost-Pieter Katoen, and Alexandru Mereacre. Quantitative automata model checking of autonomous stochastic hybrid systems. In *Proceedings of HSCC 2011*, pages 83–92, 2011.

[3] Luca Aceto, Wan J. Fokkink, and Chris Verhoef. Structural operational semantics. In *Handbook of Process Algebra*, pages 197–292. Elsevier, 2001.

[4] Luca Aceto, Anna Ingólfsdóttir, Paul Blain Levy, and Joshua Sack. Characteristic formulae for fixed-point semantics: a general framework. *Mathematical Structures in Computer Science*, 22(2):125–173, 2012.

[5] Alessandro Aldini and Marco Bernardo. A general framework for nondeterministic, probabilistic, and stochastic noninterference. In *Proceedings of ARSPA-WITS 2009*, volume 5511 of *Lecture Notes in Computer Science*, pages 18–33, 2009.

[6] Alessandro Aldini, Marco Bernardo, and Jeremy Sproston. Performability measure specification: Combining CSRL and MSL. In *Proceedings of FMICS 2011*, volume 6959 of *Lecture Notes in Computer Science*, pages 165–179, 2011.

[7] Alessandro Aldini, Mario Bravetti, Alessandra di Pierro, Roberto Gorrieri, Chris Hankin, and Herbert Wiklicky. Two formal approaches for approximating noninterference properties. In *Proceedings of FOSAD 2001/2002*, volume 2946 of *Lecture Notes in Computer Science*, pages 1–43, 2002.

[8] Alessandro Aldini, Mario Bravetti, and Roberto Gorrieri. A process-algebraic approach for the analysis of probabilistic noninterference. *Journal of Computer Security*, 12(2):191–245, 2004.

[9] Alessandro Aldini and Alessandra Di Pierro. A quantitative approach to noninterference for probabilistic systems. *Electr. Notes Theor. Comput. Sci.*, 99:155–182, 2004.

[10] Suzana Andova and Tim A.C. Willemse. Branching bisimulation for probabilistic systems: characteristics and decidability. *Theoret. Comput. Sci.*, 356(3):325–355, 2006.

[11] James Aspnes and Maurice Herlihy. Fast randomized consensus using shared memory. *J. Algorithms*, 11(3):441–461, 1990.

[12] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-fly exact computation of bisimilarity distances. In *Proceedings of TACAS 2013*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15, 2013.

[13] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Topologies of stochastic markov models: Computational aspects. *CoRR*, abs/1403.6032, 2014.

[14] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Converging from branching to linear metrics on markov chains. In *Proceedings of ICTAC 2015*, volume 9399 of *Lecture Notes in Computer Science*, pages 349–367, 2015.

[15] Giorgio Bacci and Marino Miculan. Measurable stochastics for brane calculus. *Theor. Comput. Sci.*, 431:117–136, 2012.

[16] Jos C. M. Baeten, Jan A. Bergstra, and Scott A. Smolka. Axiomatizing probabilistic processes: ACP with generative probabilities. *Inf. Comput.*, 121(2):234–255, 1995.

[17] Christel Baier. *On Algorithmic Verification Methods for Probabilistic Systems*. PhD thesis, Fakultät für Mathematik and Informatik Uviversität Mannheim, 1998.

[18] Roberto Barbuti, Giulio Caravagna, Andrea Maggiolo-Schettini, and Paolo Milazzo. An intermediate language for the stochastic simulation of biological systems. *Theor. Comput. Sci.*, 410(33-34):3085–3109, 2009.

[19] Roberto Barbuti, Giulio Caravagna, Andrea Maggiolo-Schettini, Paolo Milazzo, and Simone Tini. Foundational aspects of multiscale modeling of biological systems with process algebras. *Theor. Comput. Sci.*, 431:96–116, 2012.

[20] Roberto Barbuti, Giulio Caravagna, Paolo Milazzo, Andrea Maggiolo-Schettini, and Simone Tini. Aspects of multiscale modelling in a process algebra for biological systems. In *Proceedings of MeCBIC 2010*, volume 40 of *EPTCS*, pages 54–69, 2010.

[21] Roberto Barbuti, Stefano Cataudella, Andrea Maggiolo-Schettini, Paolo Milazzo, and Angelo Troina. A probabilistic model for molecular systems. *Fundam. Inform.*, 67(1-3):13–27, 2005.

[22] Roberto Barbuti, Andrea Maggiolo-Schettini, Paolo Milazzo, Giovanni Pardini, and Aureliano Rama. A process calculus for molecular interaction maps. In *Proceedings of MeCBIC 2009*, volume 11 of *EPTCS*, page 35, 2009.

[23] Roberto Barbuti, Andrea Maggiolo-Schettini, Paolo Milazzo, and Simone Tini. Membrane systems working in generating and accepting modes: Expressiveness and encodings. In *Proceedings of CMC 2010*, volume 6501 of *Lecture Notes in Computer Science*, pages 103–118, 2010.

[24] Roberto Barbuti, Andrea Maggiolo-Schettini, Paolo Milazzo, and Angelo Troina. Bisimulations in calculi modelling membranes. *Formal Asp. Comput.*, 20(4-5):351–377, 2008.

[25] Falk Bartels. GSOS for probabilistic transition systems. *Electr. Notes Theor. Comput. Sci.*, 65(1):29–53, 2002.

[26] Falk Bartels. *On Generalised Coinduction and Probabilistic Specification Formats: Distributive laws in coalgebraic modelling.* PhD thesis, VU University Amsterdam, 2004.

[27] Ezio Bartocci, Luca Bortolussi, and Scott A. Smolka. Hybrid systems and biology. *Inf. Comput.*, 236:1–2, 2014.

[28] Marco Bernardo, Rocco De Nicola, and Michele Loreti. Uniform labeled transition systems for nondeterministic, probabilistic, and stochastic processes. In *Proceedings of TGC 2010*, volume 6084 of *Lecture Notes in Computer Science*, pages 35–56, 2010.

[29] Marco Bernardo, Rocco De Nicola, and Michele Loreti. The spectrum of strong behavioral equivalences for nondeterministic and probabilistic processes. In *Proceedings of QAPL 2013*, volume 117 of *EPTCS*, pages 81–96, 2013.

[30] Marco Bernardo, Rocco De Nicola, and Michele Loreti. Relating strong behavioral equivalences for processes with nondeterminism and probabilities. *Theor. Comput. Sci.*, 546:63–92, 2014.

[31] Marco Bernardo, Rocco De Nicola, and Michele Loreti. Revisiting trace and testing equivalences for nondeterministic and probabilistic processes. *Logical Methods in Computer Science*, 10(1), 2014.

[32] Marco Bernardo, Rocco De Nicola, and Michele Loreti. Revisiting bisimilarity and its modal logic for nondeterministic and probabilistic processes. *Acta Inf.*, 52(1):61–106, 2015.

[33] Bard Bloom, Wan J. Fokkink, and Rob J. van Glabbeek. Precongruence formats for decorated trace semantics. *ACM Trans. Comput. Log.*, 5(1):26–78, 2004.

[34] Bard Bloom, Sorin Istrail, and Albert R. Meyer. Bisimulation can't be traced. *J. ACM*, 42(1):232–268, 1995.

[35] Luca Bortolussi, Vashti Galpin, Jane Hillston, and Mirco Tribastone. Hybrid semantics for PEPA. In *Proceedings of QEST 2010*, pages 181–190, 2010.

[36] Luca Bortolussi, Jane Hillston, and Mirco Tribastone. Fluid performability analysis of nested automata models. *Electr. Notes Theor. Comput. Sci.*, 310:27–47, 2015.

[37] Stephen D. Brookes, C. A. R. Hoare, and A. W. Roscoe. A theory of communicating sequential processes. *J. ACM*, 31(3):560–599, 1984.

[38] Muffy Calder and Jane Hillston. Process algebra modelling styles for biomolecular processes. *Trans. Computational Systems Biology*, 11:1–25, 2009.

[39] Luca Cardelli, Mirco Tribastone, Max Tschaikowski, and Andrea Vandin. Forward and backward bisimulations for chemical reaction networks. In *Proceedings of CONCUR 2015*, volume 42 of *LIPIcs*, pages 226–239, 2015.

[40] Luca Cardelli, Mirco Tribastone, Max Tschaikowski, and Andrea Vandin. Syntactic markovian bisimulation for chemical reaction networks. In *Proceedings of KiMfest 2017*, volume 10460 of *Lecture Notes in Computer Science*, pages 466–483, 2017.

[41] Valentina Castiglioni, Daniel Gebler, and Simone Tini. Logical characterization of bisimulation metrics. In *Proceedings of QAPL'16*, volume 227 of *EPTCS*, pages 44–62, 2016.

[42] Valentina Castiglioni, Daniel Gebler, and Simone Tini. Modal decomposition on nondeterministic probabilistic processes. In *Proceedings of CONCUR 2016*, volume 59 of *LIPIcs*, pages 36:1–36:15, 2016.

[43] Valentina Castiglioni and Simone Tini. Logical characterization of trace metrics. In *Proceedings of QAPL@ETAPS 2017*, volume 250 of *EPTCS*, pages 39–74, 2017.

[44] Konstantinos Chatzikokolakis, Daniel Gebler, Catuscia Palamidessi, and Lili Xu. Generalized bisimulation metrics. In *Proceedings of CONCUR 2014*, volume 8704 of *Lecture Notes in Computer Science*, pages 32–46, 2014.

[45] Konstantinos Chatzikokolakis, Sebastian A. Mödersheim, Catuscia Palamidessi, and Jun Pang. Foundational aspects of security. *Journal of Computer Security*, 22(2):201–202, 2014.

[46] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Christelle Braun. Compositional methods for information-hiding. *Mathematical Structures in Computer Science*, 26(6):908–932, 2016.

[47] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Marco Stronati. Location privacy via geo-indistinguishability. *SIGLOG News*, 2(3):46–69, 2015.

[48] Federica Ciocchetta and Jane Hillston. Process algebras in systems biology. In *Proceedings of SFM 2008*, volume 5016 of *Lecture Notes in Computer Science*, pages 265–312, 2008.

[49] Federica Ciocchetta and Jane Hillston. Bio-pepa: A framework for the modelling and analysis of biological systems. *Theor. Comput. Sci.*, 410(33-34):3065–3084, 2009.

[50] Keith L. Clark. Negation as failure. In *Logic and Data Bases*, pages 293–322, 1977.

[51] Rance Cleaveland, Zeynep Dayar, Scott A. Smolka, and Shoji Yuen. Testing preorders for probabilistic processes. *Inf. Comput.*, 154(2):93–148, 1999.

[52] Alessio Coletta, Roberta Gori, and Francesca Levi. Approximating probabilistic behaviors of biological systems using abstract interpretation. *Electr. Notes Theor. Comput. Sci.*, 229(1):165–182, 2009.

[53] Przemyslaw Daca, Thomas A. Henzinger, Jan Křetínský, and Tatjana Petrov. Linear distances between markov chains. In *Proceedings of CONCUR 2016*, volume 59 of *LIPIcs*, pages 20:1–20:15, 2016.

[54] Pedro R. D'Argenio, Daniel Gebler, and Matias D. Lee. Axiomatizing bisimulation equivalences and metrics from probabilistic SOS rules. In *Proceedings of FoSSaCS 2014*, volume 8412 of *Lecture Notes in Computer Science*, pages 289–303. Springer, 2014.

[55] Pedro R. D'Argenio and Matias David Lee. Probabilistic transition system specification: Congruence and full abstraction of bisimulation. In *Proceedings of FoSSaCS 2012*, volume 7213 of *Lecture Notes in Computer Science*, pages 452–466, 2012.

[56] Pedro R. D'Argenio, Pedro Sánchez Terraf, and Nicolás Wolovick. Bisimulations for non-deterministic labelled markov processes. *Mathematical Structures in Computer Science*, 22(1):43–68, 2012.

[57] Pedro R. D'Argenio, Nicolás Wolovick, Pedro Sánchez Terraf, and Pablo Celayes. Non-deterministic labeled markov processes: Bisimulations and logical characterization. In *Proceedings of QEST 2009*, pages 11–20, 2009.

[58] Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching metrics for quantitative transition systems. In *Proceedings of ICALP 2004*, volume 3142 of *Lecture Notes in Computer Science*, pages 97–109, 2004.

[59] Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching system metrics. *IEEE Trans. Software Eng.*, 35(2):258–273, 2009.

[60] Luca de Alfaro, Thomas A. Henzinger, and Rupak Majumdar. Discounting the Future in Systems Theory. In *Proceedings of ICALP'03*, volume 2719 of *Lecture Notes in Computer Science*, pages 1022–1037. 2003.

[61] Luca de Alfaro, Rupak Majumdar, Vishwanath Raman, and Mariëlle Stoelinga. Game refinement relations and metrics. *Logical Methods in Computer Science*, 4(3), 2008.

[62] Rocco De Nicola and Matthew Hennessy. Testing equivalences for processes. *Theor. Comput. Sci.*, 34:83–133, 1984.

[63] Robert de Simone. Higher-level synchronising devices in MEIJE-SCCS. *Theor. Comput. Sci.*, 37:245–267, 1985.

[64] Yuxin Deng, Tom Chothia, Catuscia Palamidessi, and Jun Pang. Metrics for action-labelled quantitative transition systems. *Electr. Notes Theor. Comput. Sci.*, 153(2):79–96, 2006.

[65] Yuxin Deng and Wenjie Du. Probabilistic barbed congruence. *Electronic Notes in Theoretical Computer Science*, 190(3):185–203, 2007.

[66] Yuxin Deng and Wenjie Du. Logical, metric, and algorithmic characterisations of probabilistic bisimulation. *CoRR*, abs/1103.4577, 2011.

[67] Yuxin Deng, Yuan Feng, and Ugo Dal Lago. On coinduction and quantum lambda calculi. In *Proceedings of CONCUR 2015*, volume 42 of *LIPIcs*, pages 427–440, 2015.

[68] Yuxin Deng and Rob J. van Glabbeek. Characterising probabilistic processes logically - (extended abstract). In *Proceedings of LPAR-17*, pages 278–293, 2010.

[69] Yuxin Deng, Rob J. van Glabbeek, Matthew Hennessy, and Carroll Morgan. Characterising testing preorders for finite probabilistic processes. *Logical Methods in Computer Science*, 4(4), 2008.

[70] Josee Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. *Inf. Comput.*, 179(2):163–193, 2002.

[71] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Approximating labeled markov processes. In *Proceedings of IEEE Symposium on Logic in Computer Science*, pages 95–106, 2000.

[72] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled markov processes. *Theor. Comput. Sci.*, 318(3):323–354, 2004.

[73] Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proceedings of LICS 2002*, pages 413–422, 2002.

[74] Alessandra di Pierro, Chris Hankin, and Herbert Wiklicky. Quantitative relations and approximate process equivalences. In *Proceedings of CONCUR 2003*, volume 2761 of *Lecture Notes in Computer Science*, pages 498–512, 2003.

[75] Wenjie Du, Yuxin Deng, and Daniel Gebler. Behavioural pseudometrics for nondeterministic probabilistic systems. In *Proceedings of SETTA 2016*, volume 9984 of *Lecture Notes in Computer Science*, pages 67–84, 2016.

[76] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284, 2006.

[77] Uli Fahrenberg and Axel Legay. The quantitative linear-time-branching-time spectrum. *Theor. Comput. Sci.*, 538:54–69, 2014.

[78] Yuan Feng, Lei Song, and Lijun Zhang. Distribution-based bisimulation and bisimulation metric in probabilistic automata. *CoRR*, abs/1512.05027, 2015.

[79] Yuan Feng and Lijun Zhang. When equivalence and bisimulation join forces in probabilistic automata. In *Proceedings of FM 2014*, pages 247–262, 2014.

[80] Wan Fokkink and Rob J. van Glabbeek. Divide and congruence II: delay and weak bisimilarity. In *Proceedings of LICS '16*, pages 778–787, 2016.

[81] Wan J. Fokkink. Rooted branching bisimulation as a congruence. *J. Comput. Syst. Sci.*, 60(1):13–37, 2000.

[82] Wan J. Fokkink and Rob J. van Glabbeek. Precongruence formats with lookahead through modal decomposition. In *Proceedings of CSL 2017*, volume 82 of *LIPIcs*, pages 25:1–25:20, 2017.

[83] Wan J. Fokkink, Rob J. van Glabbeek, and Paulien de Wind. Compositionality of hennessy-milner logic by structural operational semantics. *Theor. Comput. Sci.*, 354(3):421–440, 2006.

[84] Wan J. Fokkink, Rob J. van Glabbeek, and Paulien de Wind. Divide and congruence: From decomposition of modal formulas to preservation of branching and η-bisimilarity. *Inf. Comput.*, 214:59–85, 2012.

[85] Wan J. Fokkink, Rob J. van Glabbeek, and Bas Luttik. Divide and congruence III: stability & divergence. In *Proceedings of CONCUR 2017*, volume 85 of *LIPIcs*, pages 15:1–15:16, 2017.

[86] Lucia Gallina, Sardaouna Hamadou, Andrea Marin, and Sabina Rossi. A probabilistic energy-aware model for mobile ad-hoc networks. In *Proceedings of ASMTA 2011*, volume 6751 of *Lecture Notes in Computer Science*, pages 316–330, 2011.

[87] Lucia Gallina, Andrea Marin, Sabina Rossi, Tingting Han, and Marta Z. Kwiatkowska. A process algebraic framework for estimating the energy consumption in ad-hoc wireless sensor networks. In *Proceedings of MSWiM '13*, pages 255–262, 2013.

[88] Vashti Galpin. Hybrid semantics for bio-pepa. *Inf. Comput.*, 236:122–145, 2014.

[89] Daniel Gebler. *Robust SOS specifications for probabilistic processes.* PhD thesis, VU University Amsterdam, 2015.

[90] Daniel Gebler and Wan J. Fokkink. Compositionality of probabilistic hennessy-milner logic through structural operational semantics. In *Proceedings of CONCUR 2012*, volume 7454 of *Lecture Notes in Computer Science*, pages 395–409, 2012.

[91] Daniel Gebler, Kim G. Larsen, and Simone Tini. Compositional metric reasoning with probabilistic process calculi. In *Proceedings of FoSSaCS 2015*, volume 9034 of *Lecture Notes in Computer Science*, pages 230–245, 2015.

[92] Daniel Gebler, Kim G. Larsen, and Simone Tini. Compositional bisimulation metric reasoning with probabilistic process calculi. *Logical Methods in Computer Science*, 12(4), 2016.

[93] Daniel Gebler and Simone Tini. Fixed-point characterization of compositionality properties of probabilistic processes combinators. In *Proceedings of EXPRESS/SOS 2014*, volume 160 of *EPTCS*, pages 63–78, 2014.

[94] Daniel Gebler and Simone Tini. SOS specifications of probabilistic systems by uniformly continuous operators. In *Proceedings CONCUR 2015*, volume 42 of *LIPIcs*, pages 155–168. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015.

[95] Sonja Georgievska and Suzana Andova. Probabilistic may/must testing: retaining probabilities by restricted schedulers. *Formal Asp. Comput.*, 24(4-6):727–748, 2012.

[96] Alessandro Giacalone, Chi-Chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of IFIP Work, Conf. on Programming, Concepts and Methods*, pages 443–458, 1990.

[97] Susanne Graf and Joseph Sifakis. A modal characterization of observational congruence on finite terms of CCS. *Information and Control*, 68(1-3):125–145, 1986.

[98] Jan Friso Groote. Transition system specifications with negative premises. *Theoret. Comput. Sci.*, 118(2):263–299, 1993.

[99] Jan Friso Groote and Fritz W. Vaandrager. Structured operational semantics and bisimulation as a congruence. *Information and Computation*, 100(2):202–260, 1992.

[100] Hans Hansson and Bengt Jonsson. A logic for reasoning about time and reliability. *FAC*, 6(5):512–535, 1994.

[101] Matthew Hennessy. Exploring probabilistic bisimulations, part I. *Formal Asp. Comput.*, 24(4-6):749–768, 2012.

[102] Matthew Hennessy and Robin Milner. Algebraic laws for nondeterminism and concurrency. *J. Assoc. Comput. Mach.*, 32:137–161, 1985.

[103] Holger Hermanns, Jan Krcál, and Jan Křetínský. Probabilistic bisimulation: Naturally on distributions. In *Proceedings of CONCUR 2014*, volume 8704 of *Lecture Notes in Computer Science*, pages 249–265, 2014.

[104] Holger Hermanns, Augusto Parma, Roberto Segala, Björn Wachter, and Lijun Zhang. Probabilistic logical characterization. *Inf. Comput.*, 209(2):154–172, 2011.

[105] Jane Hillston. *A compositional approach to performance modelling*. PhD thesis, University of Edinburgh, UK, 1994.

[106] C. A. R. Hoare. *Communicating Sequential Processes*. Prentice-Hall, 1985.

[107] Ronald A. Howard. *Dynamic Probabilistic Systems, Volume I: Markov Models (Dynamic Probabilistic Systems)*. Dover Publications, 1971.

[108] Dung T. Huynh and Lu Tian. On some equivalence relations for probabilistic processes. *Fundam. Inform.*, 17(3):211–234, 1992.

[109] Chi-Chang Jou and Scott A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In *Proceedings of CONCUR '90*, volume 458 of *Lecture Notes in Computer Science*, pages 367–383, 1990.

[110] Leonid V. Kantorovich. On the transfer of masses. *Doklady Akademii Nauk*, 37(2):227–229, 1942.

[111] Robert M. Keller. Formal verification of parallel programs. *Commun. ACM*, 19(7):371–384, 1976.

[112] Dexter Kozen. Results on the propositional mu-calculus. *Theor. Comput. Sci.*, 27:333–354, 1983.

[113] Marta Kwiatkowska. Cognitive reasoning and trust in human-robot interactions. In *Proceedings of TAMC 2017*, volume 10185 of *Lecture Notes in Computer Science*, pages 3–11, 2017.

[114] Marta Z. Kwiatkowska and Gethin Norman. Probabilistic metric semantics for a simple language with recursion. In *Proceedings of MFCS'96*, volume 1113 of *Lecture Notes in Computer Science*, pages 419–430, 1996.

[115] Ruggero Lanotte, Massimo Merro, and Simone Tini. Weak simulation quasimetric in a gossip scenario. In *Proceedings of FORTE 2017*, volume 10321 of *Lecture Notes in Computer Science*, pages 139–155, 2017.

[116] Ruggero Lanotte and Simone Tini. Probabilistic congruence for semistochastic generative processes. In *Proceedings of FoSSaCS'05*, volume 3441 of *Lecture Notes in Computer Science*, pages 63–78. 2005.

[117] Ruggero Lanotte and Simone Tini. Probabilistic bisimulation as a congruence. *ACM Trans. Comput. Log.*, 10:1–48, 2009.

[118] Kim G. Larsen. *Context-dependent bisimulation between processes.* PhD thesis, University of Edinburgh, 1986.

[119] Kim G. Larsen. Proof system for hennessy-milner logic with recursion. In *Proceedings of CAAP '88*, volume 299 of *Lecture Notes in Computer Science*, pages 215–230, 1988.

[120] Kim G. Larsen. Proof systems for satisfiability in Hennessy-Milner logic with recursion. *Theor. Comput. Sci.*, 72(2&3):265–288, 1990.

[121] Kim G. Larsen, Axel Legay, Louis-Marie Traonouez, and Andrzej Wasowski. Robust specification of real time components. In *Proceedings of FORMATS 2011*, volume 6919 of *Lecture Notes in Computer Science*, pages 129–144, 2011.

[122] Kim G. Larsen, Radu Mardare, and Prakash Panangaden. Taking it to the limit: Approximate reasoning for markov processes. In *Proceedings of MFCS 2012*, volume 7464 of *Lecture Notes in Computer Science*, pages 681–692, 2012.

[123] Kim G. Larsen and Arne Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94(1):1–28, 1991.

[124] Kim G. Larsen, Bernhard Steffen, and Carsten Weise. Continuous modeling of real-time and hybrid systems: From concepts to tools. *STTT*, 1(1-2):64–85, 1997.

[125] Kim G. Larsen and Liu Xinxin. Compositionality through an operational semantics of contexts. *J. Log. Comput.*, 1(6):761–795, 1991.

[126] Matias D. Lee and Erik P. de Vink. Rooted branching bisimulation as a congruence for probabilistic transition systems. In *Proceedings of QAPL 2015*, volume 194 of *EPTCS*, pages 79–94, 2015.

[127] Matias D. Lee and Erik P. de Vink. Logical characterization of bisimulation for transition relations over probability distributions with internal actions. In *Proceedings of MFCS 2016*, volume 58 of *LIPIcs*, pages 29:1–29:14, 2016.

[128] Nancy A. Lynch, Roberto Segala, and Frits W. Vaandrager. Observing branching structure through probabilistic contexts. *SIAM J. Comput.*, 37(4):977–1013, 2007.

[129] Radu Mardare, Luca Cardelli, and Kim G. Larsen. Continuous markovian logics - axiomatization and quantified metatheory. *Logical Methods in Computer Science*, 8(4), 2012.

[130] George Markowsky. Chain-complete posets and directed sets with applications. *Algebra Univ.*, 6:53–68, 1976.

[131] Robin Milner. Calculi for synchrony and asynchrony. *Theor. Comput. Sci.*, 25:267–310, 1983.

[132] Robin Milner and Davide Sangiorgi. Barbed bisimulation. In *Proceedings of ICALP‘92*, volume 623 of *Lecture Notes in Computer Science*, pages 685–695, 1992.

[133] Markus Müller-Olm. Derivation of characteristic formulae. *Electr. Notes Theor. Comput. Sci.*, 18:159–170, 1998.

[134] Laura Nenzi, Luca Bortolussi, Vincenzo Ciancia, Michele Loreti, and Mieke Massink. Qualitative and quantitative monitoring of spatio-temporal properties. In *Proceedings of RV 2015*, volume 9333 of *Lecture Notes in Computer Science*, pages 21–37, 2015.

[135] Ernst-Rüdiger Olderog and C. A. R. Hoare. Specification-oriented semantics for communicating processes. *Acta Inf.*, 23(1):9–66, 1986.

[136] James B. Orlin. A faster strongly polynominal minimum cost flow algorithm. In *Proceedings of ACM Symposium on Theory of Computing*, pages 377–387, 1988.

[137] Augusto Parma and Roberto Segala. Logical characterizations of bisimulations for discrete probabilistic systems. In *Proceedings of FoSSaCS 2007*, volume 4423 of *Lecture Notes in Computer Science*, pages 287–301, 2007.

[138] Gordon D. Plotkin. An operational semantics for CSO. In *Proceedings of Logics of Programs and Their Applications*, pages 250–252, 1980.

[139] Gordon D. Plotkin. A structural approach to operational semantics. Report DAIMI FN-19, Aarhus University, 1981.

[140] Amir Pnueli. The temporal logic of programs. In *Symposium on Foundations of Computer Science 1977*, pages 46–57, 1977.

[141] Amir Pnueli. Linear and branching structures in the semantics and logics of reactive systems. In *Proceedings of ICALP 1985*, volume 194 of *Lecture Notes in Computer Science*, pages 15–32, 1985.

[142] Svetlozar T. Rachev, Lev B. Klebanov, Stoyan V. Stoyanov, and Frank J. Fabozzi. *The Methods of Distances in the Theory of Probability and Statistics*. Springer, 2013.

[143] Joshua Sack and Lijun Zhang. A general framework for probabilistic characterizing formulae. In *Proceedings of VMCAI 2012*, pages 396–411, 2012.

[144] Roberto Segala. A compositional trace-based semantics for probabilistic automata. In *Proceedings of CONCUR '95*, volume 962 of *Lecture Notes in Computer Science*, pages 234–248, 1995.

[145] Roberto Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995.

[146] Roberto Segala and Nancy A. Lynch. Probabilistic simulations for probabilistic processes. *Nord. J. Comput.*, 2(2):250–273, 1995.

[147] Roberto Segala and Andrea Turrini. Comparative analysis of bisimulation relations on alternating and non-alternating probabilistic models. In *Proceedings of QEST 2005*, pages 44–53, 2005.

[148] Lin Song, Yuxin Deng, and Xiaojuan Cai. Towards automatic measurement of probabilistic processes. In *Proceedings of QSIC 2007*, pages 50–59, 2007.

[149] Bernhard Steffen and Anna Ingólfsdóttir. Characteristic formulae for processes with divergence. *Inf. Comput.*, 110(1):149–163, 1994.

[150] William J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.

[151] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[152] Simone Tini. Non expansive epsilon-bisimulations. In *Proceedings of AMAST 2008*, volume 5140 of *Lecture Notes in Computer Science*, pages 362–376, 2008.

[153] Simone Tini. Non-expansive epsilon-bisimulations for probabilistic processes. *Theor. Comput. Sci.*, 411(22-24):2202–2222, 2010.

[154] Franck van Breugel. On behavioural pseudometrics and closure ordinals. *Inf. Process. Lett.*, 112(19):715–718, 2012.

[155] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic transition systems. In *Proceedings of CONCUR 2001*, volume 2154 of *Lecture Notes in Computer Science*, pages 336–350, 2001.

[156] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic transition systems. In *Proceedings of ICALP 2001*, volume 2076 of *Lecture Notes in Computer Science*, pages 421–432, 2001.

[157] Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331(1):115–142, 2005.

[158] Franck van Breugel and James Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theor. Comput. Sci.*, 360(1-3):373–385, 2006.

[159] Rob J. van Glabbeek. The linear time - branching time spectrum I - the semantics of concrete, sequential processes. In *Proc. CONCUR '90*, volume 458 of *Lecture Notes in Computer Science*, pages 278–297. Springer, 1990.

[160] Rob J. van Glabbeek. The meaning of negative premises in transition system specifications II. In *Proceedings of ICALP'96*, Lecture Notes in Computer Science, pages 502–513. Springer, 1996.

[161] Rob J. van Glabbeek, Scott A. Smolka, and Bernhard Steffen. Reactive, generative and stratified models of probabilistic processes. *Inf. Comput.*, 121(1):59–80, 1995.

[162] Chris Verhoef. A congruence theorem for structured operational semantics with predicates and negative premises. *Nord. J. Comput.*, 2(2):274–302, 1995.

[163] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

[164] Nicolás Wolovick and Sven Johr. A characterization of meaningful schedulers for continuous-time markov decision processes. In *Proceedings of FORMATS 2006*, volume 4202 of *Lecture Notes in Computer Science*, pages 352–367, 2006.

[165] Pengfei Yang, David N. Jansen, and Lijun Zhang. Distribution-based bisimulation for labelled markov processes. *CoRR*, abs/1706.10049v1, 2017.

[166] Wang Yi and Kim G. Larsen. Testing probabilistic and nondeterministic processes. In *Proceedings of PSTV'92*, volume C-8 of *IFIP Transactions*, pages 47–61, 1992.