

# Principles of Markov Automata

---

A dissertation submitted towards the degree  
Doctor of Engineering  
of the Faculty of Mathematics and Computer Science  
of Saarland University

by  
Christian Georg Eisentraut

Saarbrücken, 2017

Day of Colloquium  
Dean of the Faculty

Chair of the Committee

*Reporters*

First reviewer  
Second reviewer  
Third reviewer

Academic Assistant

17 / 07 / 2017  
Univ.-Prof. Dr. Frank-Olaf Schreyer

Prof. Dr. Christoph Weidenbach

Prof. Dr. Holger Hermanns  
Prof. Rob van Glabbeek, PhD.  
Prof. Dr. Lijun Zhang

Dr. Ernst Moritz Hahn

# Abstract

A substantial amount of today's engineering problems revolve around systems that are concurrent and stochastic by their nature. Solution approaches attacking these problems often rely on the availability of formal mathematical models that reflect such systems as comprehensively as possible. In this thesis, we develop a compositional model, Markov automata, that integrates concurrency, and probabilistic and timed stochastic behaviour. This is achieved by blending two well-studied constituent models, probabilistic automata and interactive Markov chains. A range of strong and weak bisimilarity notions are introduced and evaluated as candidate relations for a natural behavioural equivalence between systems. Among them, weak distribution bisimilarity stands out as a natural notion being more oblivious to the probabilistic branching structure than prior notions. We discuss compositionality, axiomatizations, decision and minimization algorithms, state-based characterizations and normal forms for weak distribution bisimilarity. In addition, we detail how Markov automata and weak distribution bisimilarity can be employed as a semantic basis for generalized stochastic Petri nets, in such a way that known shortcomings of their classical semantics are ironed out in their entirety.



# Zusammenfassung

Ein beträchtlicher Teil gegenwärtiger ingenieurwissenschaftlicher Probleme erstreckt sich auf Systeme, die ihrer Natur nach sowohl stochastisch als auch nebenläufig sind. Lösungsansätze fußen hierbei häufig auf der Verfügbarkeit formaler mathematischer Modelle, die es erlauben, die Spezifika jener Systeme möglichst erschöpfend zu erfassen. In dieser Dissertation entwickeln wir ein kompositionelles Modell namens Markov-Automaten, das Nebenläufigkeit mit probabilistischen und stochastischen Prozessen integriert. Dies wird durch die Verschmelzung der zweier bekannter Modellklassen erreicht, und zwar die der probabilistischen Automaten und die der interaktiven Markovketten. Wir entwickeln dabei ein Spektrum verschiedener, starker und schwacher Bisimulationsrelationen und beurteilen sie im Hinblick auf ihre Eignung als natürliche Verhaltensäquivalenz zwischen Systemen. Die schwache Wahrscheinlichkeitsverteilungsbisimulation sticht dabei als natürliche Wahl hervor, da sie die probabilistische Verzweigungsstruktur treffender abstrahiert als bisher bekannte Bisimulationsrelationen. Wir betrachten des Weiteren Kompositionalitätseigenschaften, Axiomatisierungen, Entscheidungs- und Minimierungsalgorithmen, sowie zustandsbasierte Charakterisierungen und Normalformen für die schwache Wahrscheinlichkeitsverteilungsbisimulation. Abschließend legen wir dar, dass Markov-Automaten und die schwache Wahrscheinlichkeitsverteilungsbisimulation als Grundlage für eine verbesserte Semantik von verallgemeinerten stochastischen Petrinetzen dienen kann, welche bekannte Mängel der klassischen Semantik vollständig behebt.



# Acknowledgement

During my time writing this thesis, I have met many people who inspired me and influenced me for the better. Writing this thesis would not have been possible without them.

First, I would like to thank Holger Hermanns, who supervised my thesis and guided me on my scientific journey from my second year of study onwards. Over the years, he has not only been a teacher of science, methods and scientific passion to me, but also a friend. I especially like to thank him for letting me roam free wherever my passion has led me.

I am grateful to Lijun Zhang for enthusing me with stochastic bisimulations, a topic I originally never wanted to write a thesis about, but then soon loved to be my research topic for the years to come. Additionally, I thank him for reviewing my thesis.

Many ideas of this thesis emerged from discussions and collaborations with Holger Hermanns, Lijun Zhang, Andrea Turrini, Julia Krämer, Pedro D'Argenio, Lei Song and Johann Schuster. Without these people, this thesis would have become a very different one.

I want to thank thank Rob van Glabbeek very much for reviewing my thesis, for his helpful remarks and discussions and the great time we spent, as well as for all his great work that truly inspired and fascinated me since the very beginning of my studies.

I thank all of my group colleagues Sven Johr, Lijun Zhang, Reza Pulungan, Pepijn Crouzen, E. Moritz Hahn, Arnd Hartmanns, Alexander Graf-Brill, Gilles Nies, Hassan Hatefi, Vahid Hashemi, Jan Krčál, Yuliya Butkova, and Hernán Baró Graf for the the great times we spent and the interesting discussions.

My special thank goes to E. Moritz Hahn again, for taking the role of the academic assistant at my thesis' defence, and also to Christoph Weidenbach for being chair of the committee in my colloquium.

Especially, I want to thank Christa Schäfer, our group secretary, for always being most helpful with everything concerning administrative matters, and for always having an open ear for all kind of worries I may have had.

Outside my university life, a few people have supported me writing this thesis in more indirect, but nevertheless crucial ways. I want to thank Konstantin Joseph, for being a friend and teacher of life, Sphen Paul, for providing me with the most valuable insights into leadership and communications, from which I still profit virtually every day of my life.

Finally, I want to whole-heartedly thank my family, who supported me with their love, patience and everything else I needed: my mother Gabriele, my father Hans, and my life partner Julia.





# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Models . . . . .	3
1.2. Observational Equivalence . . . . .	6
1.3. Contributions and Structure of the Thesis . . . . .	11
<b>2. Preliminaries</b>	<b>13</b>
2.1. Sets and Multisets . . . . .	13
2.2. Elementary Topology . . . . .	14
2.3. Graphs . . . . .	15
2.4. Probability Theory . . . . .	17
<b>I. Processes</b>	<b>21</b>
<b>3. Labelled Transition Systems</b>	<b>23</b>
3.1. Model . . . . .	23
3.2. Bisimilarities . . . . .	27
3.3. Decision Algorithm . . . . .	33
3.4. Summary and Discussion . . . . .	35
<b>4. Probabilistic Automata</b>	<b>37</b>
4.1. Model . . . . .	37
4.2. Bisimilarity . . . . .	41
4.3. Decision Algorithm . . . . .	54
4.4. Summary and Discussion . . . . .	56
<b>5. Interactive Markov Chains</b>	<b>61</b>
5.1. Model . . . . .	61
5.2. Bisimilarities . . . . .	66
5.3. Decision Algorithm . . . . .	68
5.4. Summary and Discussion . . . . .	70
<b>6. Comparative Semantics</b>	<b>71</b>
6.1. Languages . . . . .	72
6.2. Axiomatization . . . . .	78
6.3. Probabilistic Forward Simulation . . . . .	84
6.4. Summary and Discussion . . . . .	87

<b>II. Markov Automata Theory</b>	<b>89</b>
<b>7. Markov Automata</b>	<b>91</b>
7.1. Model . . . . .	92
7.2. Parallel Composition and Abstraction . . . . .	99
7.3. Summary and Discussion . . . . .	100
<b>8. Bisimulations</b>	<b>103</b>
8.1. Strong Bisimilarity . . . . .	103
8.2. Weak Bisimilarities . . . . .	104
8.3. Revisiting State-Based (Bi)Simulations . . . . .	127
8.4. A State-Based Characterization . . . . .	131
8.5. Challenge Completed . . . . .	145
8.6. Related Work . . . . .	146
8.7. Summary and Discussion . . . . .	151
<b>9. Axioms</b>	<b>155</b>
9.1. Calculus . . . . .	155
9.2. Axioms . . . . .	158
9.3. Soundness . . . . .	160
9.4. Completeness . . . . .	161
9.5. Open Challenges . . . . .	166
9.6. Summary and Discussion . . . . .	169
<b>III. Markov Automata Algorithmics</b>	<b>171</b>
<b>10. Decision Algorithms</b>	<b>173</b>
10.1. Challenges . . . . .	173
10.2. Algorithm . . . . .	175
10.3. Summary and Discussion . . . . .	180
<b>11. Minimal Normal Forms</b>	<b>183</b>
11.1. Structural Preorders . . . . .	184
11.2. Reductions . . . . .	185
11.3. Normal Forms . . . . .	194
11.4. Summary and Discussion . . . . .	203
<b>12. A Semantics for Every GSPN</b>	<b>205</b>
12.1. Confusion – A Question of Semantic Perspective . . . . .	205
12.2. Generalized Stochastic Petri Nets . . . . .	208
12.3. Markov Automata Semantics for GSPN . . . . .	211
12.4. Bisimulation Semantics . . . . .	213
12.5. Summary and Discussion . . . . .	218

<b>IV. Conclusion</b>	<b>221</b>
<b>13. Conclusion</b>	<b>223</b>
13.1. The Foundations of Markov Automata . . . . .	223
13.2. System Specification and Quantitative Analysis . . . . .	225
13.3. Open Challenges and Further Work . . . . .	227
<b>Appendix</b>	<b>229</b>
<b>A. Alternative Definition of Weak (Hyper)Transitions</b>	<b>231</b>
<b>B. Proofs of Chapter 5</b>	<b>235</b>
B.1. Lemma 5.1 . . . . .	235
<b>C. Proofs of Chapter 7</b>	<b>239</b>
C.1. Lemma 7.5 . . . . .	239
C.2. Lemma 7.6 . . . . .	240
<b>D. Proofs of Chapter 8</b>	<b>243</b>
D.1. Lemma 8.5 . . . . .	243
D.2. Lemma 8.7 . . . . .	248
D.3. Theorem 8.6 . . . . .	250
D.4. Theorem 8.8 . . . . .	256
D.5. Proof of Theorem 8.10 . . . . .	258
D.6. Proof of Theorem 8.13 . . . . .	261
D.7. Proof of Theorem 8.14 . . . . .	264
D.8. Proof of Theorem 8.15 . . . . .	275
<b>E. Proof of Lemma 9.7</b>	<b>285</b>
<b>F. Proofs of Chapter 11</b>	<b>291</b>
F.1. Lemma 11.4 . . . . .	291
F.2. Lemma 11.5 . . . . .	294
F.3. Lemma 11.11 . . . . .	295
<b>Bibliography</b>	<b>297</b>



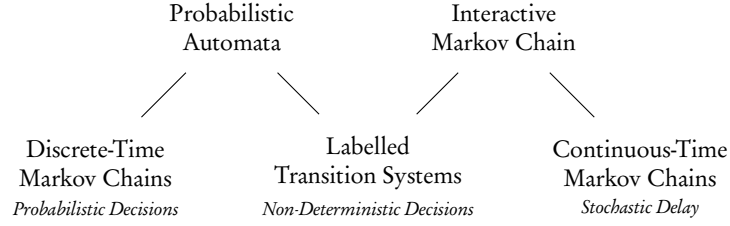
# 1. Introduction

A majority of systems that we study in modern computer science can be classified as concurrent and communicating systems. The primary purpose of these systems is no longer classical computation in the sense of deterministic transformation of input values into output values. Rather, these systems continuously receive information and take it as a stimulus to react upon in different ways. They consist of many components that engage in a steady stream of action and reaction. Hereby, classical computation fades from the spotlight and becomes an abstract activity occurring between the receiving and sending of information. The spectrum of real-world systems that fall into this category is vast. Typical examples range from network protocols to automotive and avionic control systems to complex biological, and nowadays social, networks. Even classical computer programs with user interactions or inter-process communications are typical representatives. So research in computing system has seen a shift from systems that compute a single input/output transformation to complex systems that are composed of a multitude of such simple systems, that moreover continuously receive input and produce output from each other and their broader environment.

The basic intricacies of concurrency have been analysed in formal models for more than thirty years. It has turned out that concurrency is intrinsically linked with a phenomenon called *non-determinism*. It describes the situation that exactly one of several alternative events can occur in the next moment, but it is not specified *by any means* and thus *completely unpredictable*, which of them *will* actually take place.

In concurrency theory, it is assumed that concurrent components are scheduled non-deterministically. This implies that it is unpredictable, which of them will make progress in any moment of time and how far it will be executed before any of the other components becomes active again. This assumption allows to abstract from details of concrete systems that would actually determine the scheduling, for example, the individual computation speed of the concurrent components or the time messages need to travel in the ether before they arrive at their point of destination. In this way, the models built with this theory are faithful in all possible situations, and the properties derived from them possess general validity.

The generality of non-deterministic abstraction is, however, not always beneficial. Analyses of concurrent models usually have high computational demands, as in general, all possibilities to resolve non-determinism in one or the other way have to be considered. Furthermore, leaving the choice between two possible events completely undetermined may lead to analysis results that are practically irrelevant, as the interdependencies between competing events are known and relevant. This is true especially for real world systems, where informatics meets natural sciences. To model such systems adequately, concurrency theory needs to be extended with quantitative aspects. Quantitative aspects include for example probabilistic behaviour, hard and soft real-time stochastic time as well as cost and reward information. However, extending models of concurrency with quantitative aspects in a semantically sound, and at the same time smooth way has been an ongoing challenge. Still, this endeavour is of key importance for humanity's current and future challenges.



**Figure 1.1.:** Lattice of Models

In this thesis, we focus our attention on two aspects, that play a key role in today’s engineering problems. One of them is the incorporation of events that are determined by probabilistic decisions, for instance, message loss in the modelling of communication protocols. In a classic engineering setting, and without the need to model communication and concurrency, the modelling would rely on *discrete time Markov chains*. However, if concurrency and communication are needed, this elementary mathematical models lacks expressiveness. In fact, establishing a suitable model that allows to incorporate concurrency theory and probabilistic behaviour has proven to be a challenging task, and several competing approaches have been proposed over the last two decades. *Probabilistic automata* as formulated by Roberto Segala and Nancy Lynch [SL94; Seg95] have proven a model with broad range of applications in practical and theoretic research [Bai96; BEM00; BK97; BK00; BS00; JY02; KN98; SV99; SV03; WZH07; HWZ08; TSP11].

The other aspect we are interested in is the incorporation of time. More precisely, the possibility to express that certain processes consume a specific amount of time. We hereby focus on a *stochastic* model of time, where the passage of time is governed by an negative exponential distribution. The classical mathematical models to express this type of behaviour are *continuous-time Markov chains*. They are the model of choice in many engineering contexts. On one hand, they are based on a rather simple mathematical theory, that allows for an efficient computation of many interesting properties, as for example the expected behaviour exhibited by a system in the long run; on the other hand, Markov chains are suited to approximate almost any arbitrary form of stochastic time obtained by extending the Markov chain appropriately. Thus, they allow, at least in principle, to model arbitrary intricate stochastic behaviour in an approximate way, for which often no feasible exact solutions exist.

Unfortunately, as discrete-time Markov chains, continuous-time Markov chains are not directly suited to express concurrency and communication between systems. In the past decades, several models based on continuous-time Markov chains have been proposed to overcome these limitations. Among them, *Interactive Markov Chains* [Her02] stand out. They are extensively applied in both theory and practice-oriented research [Böd+09; Bou+08; BCS07b; BCS07a; BCS10; Boz+09b; Boz+09a; Boz+11; CZM09; Cos+08; Cos+09; HJ08; ZN10; Hav+10; Est+12].

In Figure 1.1, we summarize these models, together with their underlying models, as a lattice, ordered by expressiveness.

For more than one decade, however, it has remained an open challenge to incorporate *both* probabilistic events and stochastic time into one *single* model of concurrency. From an engineering perspective, such a model has been long awaited. Many of today’s engineering problems actually involve both probabilistic events and stochastic timed aspects. Especially in the context of error modelling exists a rising need for formal models with both capabilities [Kat13].

The only existing model that is expressive enough are Generalized Stochastic Petri Nets [Bal07; Mar+91; MCB84]. They incorporate concurrency, probabilistic events and stochastic time. Unfortunately, they are not a fully satisfactory option. Though being broadly applied in industrial strength applications, their semantic foundations are in fact not complete. For certain nets, so called *confused* nets, no semantic interpretation exists. This semantic gap is a sore spot for practical application, as many users of tools based on this formalism are not aware of the problem, and automatic checks are hardly implemented. This may result in the unwitting usage of semantically undefined models.

The aim of this thesis is to provide a model, that blends both probabilistic events and stochastic time with concurrency in a semantically clean and sound way. In addition, we strive to combine and preserve the features of two of the most successful models, probabilistic automata and interactive Markov chains, where it is reasonable, and to extend them where needed.

## 1.1. Models

Labelled Transition Systems (LTS), Probabilistic Automata (PA) [Seg95] and Interactive Markov Chains (IMC) [Her02] are all variations of state-transition systems, and focused on concurrency. Each of them, however, offers a different expressiveness with respect to quantitative aspects. Figure 1.1 depicts the lattice of our models, ordered by expressiveness. LTS is a basic model, which fully captures concurrency, but without means to deal with quantitative aspects. It is the common foundation of all models discussed in this thesis. PA enhance LTS by the probabilistic aspect, drawing inspiration from discrete-timed Markov chains, while IMC enhance them by exponentially distributed stochastic time, drawing their inspiration from continuous-times Markov chains. We will review the differences and commonalities of these models in the following, and discuss reasons of their success.

The foundational idea of all these models is to describe systems in terms of discrete *states* and *state changes*, also called *transitions*. Transitions are labelled by names, so called *actions*, which are abstract representations of the elementary activities a system could ever perform and that are relevant for specific analysis intents. Actions are assumed to be *timeless* and *atomic* entities; they do not bear any structure.

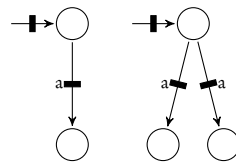
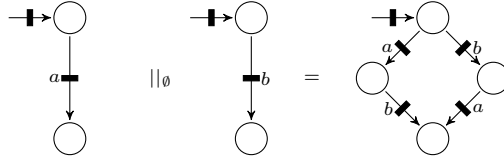
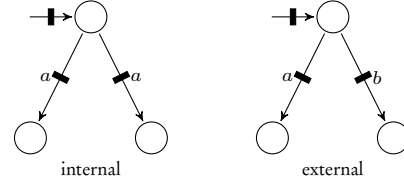


Figure 1.2.: Labelled Transition System

**Example 1.1.** In Figure 1.2 we see two simple labelled transition systems with two, and three states respectively. We represent transitions by arrows with a black bar in the middle:  $\text{---}\!\!\!\!\!\rightarrow$ . Transitions are labelled by action, here  $a$ . In the system on the right we see that from one state more than one transition with the same action label may originate, as long as they end in different states. States remain unlabelled.  $\triangleleft$



**Figure 1.3.:** Interleaving Semantics of Two Parallel Systems



**Figure 1.4.:** Internal and External Non-Determinism

While actions are the elementary building blocks of behaviour, the complete behaviour of a system is the result of the entire graph structure of the state-transition system.

### 1.1.1. Concurrency

Concurrent systems are systems, which consist of several components, that can act independently of each other at the same time, and that can as well interact via communication activities. More generally, we also call a system concurrent when it consists of a single component, but is intended to run concurrently with other components. Thus, every component of a concurrent system is by definition a concurrent system itself. Sometimes, concurrent systems are also called *reactive systems*, as they are susceptible to stimuli from the environment. This term is often used to differentiate these kind of systems from classical computation devices, where the only input is received initially and an output is generated upon termination. Concurrent, or reactive, systems unusually are not designed to terminate.

The main assumption of concurrency theory as we use it in this thesis is that two *independent* and *concurrent* actions can never occur exactly at the same time. This is reasonable since *actions* are assumed to be *timeless* and *atomic* entities. However, concurrent actions are assumed to happen exactly at the same time when they are part of a communication between two concurrent components. This act is called synchronization. It is in fact the only way concurrent components may interact.

As a matter of abstraction, it is further assumed that the order in which independent actions occur cannot be determined — it is *non-deterministic*. As a consequence, the concurrent execution of two components can be understood as the set of all possible *interleavings* of their individual actions, while the relative execution order defined by the components is maintained. Due to this typical pattern models of concurrency that follow this abstraction are called *interleaving (concurrency)* models. Pioneering work in this field has been done by Milner [Mil82; Mil89b] and Hoare [Hoa85].

There is an ongoing debate whether interleaving semantics are really expressive enough, especially as the model does not provide any built-in primitives to distinguish between non-determinism arising from parallel execution of processes, and non-determinism arising from a modeller's choice. For the analysis and verification of concurrent systems though, these differences have proven practically negligible. For further discussions we refer the reader to [Bae93; Bae05; Abr06; ML12].



### 1.1.2. Non-Determinism

In the models we study, non-determinism is closely linked to concurrency. However, non-determinism can occur in a model for many other reasons [Her02; Seg95]:

**Implementation Freedom** When models serve as system specifications, certain behavioural aspects are left deliberately underspecified and to be closed by concrete implementations in one or another way. In the model, this is expressed as a non-deterministic choice between the implementation alternatives.

**Abstraction** When real world systems are modelled, the conditions that lead to decisions are often too complex to be encompassed by the model. A non-deterministic choice between all possible decisions abstract from the details by encompassing all possible outcomes.

**Concurrency** Non-determinism occurs as a consequence of concurrent execution. As we have already discussed, in interleaving semantics it is at the core of any concurrent execution.

**Receptiveness in External Communication** Communication between concurrent components of a system is modelled by synchronized execution of transitions with the same label. A non-deterministic choice between differently labelled transitions is thus necessary in the case that a component needs to be receptive for more than one communication action in a single moment.

One furthermore distinguishes between *external* and *internal* non-determinism. External non-determinism is present when the transition involved in the choice are labelled by pairwise different actions. Two or more transitions are in an internal non-deterministic choice if they have the same label.

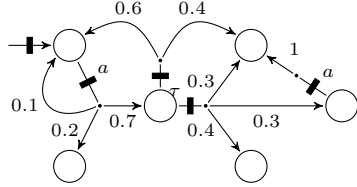
External non-determinism is needed to guarantee receptiveness of external communication. In combination with other communicating processes its behaviour is fully determined, depending on the communication actions offered. Therefore, systems that are only externally non-deterministic are often considered *deterministic* in their behaviour.

Non-determinism is an ingredient of virtually every model of concurrent systems. In isolation, it has been intensively studied over the past decades and is well understood. In combination with quantitative aspects, however, non-determinism and thus also concurrency remains a challenge.

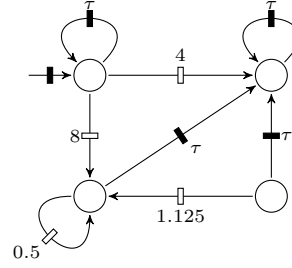
### 1.1.3. Probabilities

In many situations, non-determinism is too abstract to model the choice between alternatives faithfully. For example, when a fair coin is flipped, there are two possible alternative outcomes: head and tail. Clearly, the choice between the two is probabilistically determined. We can predict that each of them will occur with a chance of one to one. If our system under consideration relies on this aspect, for example, because we want to compute winning probability in a game of dice, modelling the coin throw with a non-deterministic choice is inappropriate.

Automata with both non-deterministic and probabilism can be defined in several ways. For a good overview we refer the reader to [SV04], or to Chapter 4 for a brief summary. In our thesis, we focus on Segala's simple probabilistic automata [SL94; SL95; Seg97]. We will refer to them as *probabilistic automata* (PA) in the following. In this model, non-deterministic choice between several labelled transitions is allowed, exactly as for labelled transition systems. As probabilistic



**Figure 1.5.:** Probabilistic Automaton



**Figure 1.6.:** Interactive Markov Chain

extension, each transition leads to a distribution over states instead of a single state. Communication between systems and concurrent execution are modelled completely analogously to labelled transitions systems via synchronization.

### 1.1.4. Stochastic Time

Mathematically, time can be modelled by stochastic processes. The model we consider here is that of continuous-time Markov chains (*CTMC*). It assumes that the time when an activity occurs is distributed probabilistically by an exponential distribution. In principle, a *CTMC* is a state-transition system where transitions are labelled by real numbers, called *rates*. The rates denote the parameters of the exponential distribution. Intuitively, they correspond to the expected time units it takes on average until this transition is taken. Typical applications are failure models of critical systems. From empirically known data rates are derived that represent the meantime to failure of the single components of a system. With the complete *CTMC* model, more general information about the failure behaviour of the system can then be computed, as for example the probability and the average time to a fatal system crash.

*Interactive Markov Chains* (IMC) [Her02] combined the purely stochastic expressiveness of *CTMC* with interleaving concurrency. The genuine idea of IMC is that interactive transitions and time transitions are modelled as different transitions. In this way, they are a conservative extension of LTS and *CTMC*. While many other models have emerged that aim at the combination of stochastic time and concurrency, IMC excels with several superior properties. We will discuss them in a broader context in Chapter 5.

## 1.2. Observational Equivalence

Computing systems are always built with a specific purpose in mind, so that when executed, they pursue a certain goal. For the systems we consider, this goal finds expression in the actions and communications these systems issue and their interdependencies during execution. We call this aspect the *observable behaviour* of a system. We distinguish this notion from the system's *internal structure*, that well induces its observable behaviour, but that is never directly exposed to an observer. For example, when we describe a certain system as one that receives a message, for example a number, and then sends a response, telling, for instance, whether the received

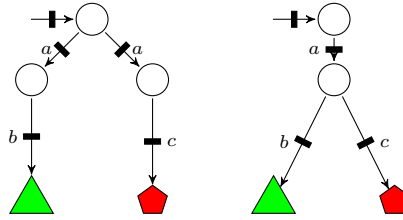
number was a prime number or not, then we describe the system at the level of its observable behaviour. Internally, the system might have a more detailed structure. For example, between receiving the number and sending the result internal computations may happen. Maybe this computation is even transferred to other subcomponents, such that many internal messages are sent and received on the way. So, the internal structure of the system can become fairly complex, while its observable behaviour remains simple. We also see that the same observable behaviour can be implemented in many different ways internally, leading to more or less complex structures. However, how simple or complex these internal structures may be, the resulting observable behaviour must always stay the same.

From a more abstract perspective, one calls the observable behaviour of a system the *process* it implements. Formally, a process is defined as an equivalence class of systems according to some notion of observational equivalence. As the theory of concurrency has been closely linked to the study of observational equivalences, one often refers to a system as a process, especially when one considers it a representative of the behaviour it implements.

In concurrency theory, there is nothing like a canonical observational equivalence. This is rooted in different opinions about which aspects of a system's execution should actually be considered observable, and also about which properties observationally equivalent systems must share. In general, an ideal observational equivalence should be a relation that is the coarsest possible, satisfying a few basic requirements.

*Remark 1.1* (Requirements of Observational equivalences). For this thesis, we stipulate the following requirements.

**Requirement 1: Congruence** When two systems are considered to show the same behaviour, they should do so in every context of other systems. On a formal level, this is called the *congruence* property of a relation. This property enables a powerful proof technique for very large and complex systems: instead of considering the systems as monolithic blocks, which can only be analysed as a whole, substructures can be considered separately, which are smaller and thus simpler to analyse, and then recombined to arrive at a result for the complete systems. The congruence property ensures that such an approach is actually correct. This methodology is often referred to as compositional reasoning.



**Figure 1.7.:** Two Systems with Different Branching Structure, but Identical Sequences of Actions

**Requirement 2: Preservation of the Non-Deterministic Branching Structure** The non-deterministic branching structure of a systems [Gla93; Van94; GW96] refers to the points during an execution of a system, *where* decisions are made what future *observable* behaviour will be possible or impossible. For example in Figure 1.7, in the system on the right, the decision between

action  $b$  and  $c$  occurs at the latest moment, while in the system on the left, the decision occurs already when action  $a$  is executed. As a contrast to the last example, the two systems in Figure 1.2 should not be distinguished, even though they also exhibit a similar difference in their internal structure. The reason is that the system on the right exhibits the same followup behaviour after action  $a$  has occurred independent of which branch has been chosen. In this simple case, the system simply terminates. Thus, their non-deterministic branching structures agree, as no decisions can be taken at any point during the execution that would *influence the future observable behaviour*.

Preservation of the non-deterministic branching structure guarantees the preservation of virtually all kinds of essential observable properties of systems, for instance, preservation of deadlock/livelock behaviour<sup>1</sup>. Deadlock refers to the inability of a system to execute any further actions, even though it has not yet reached an intentional final state. Deadlock occurs if a system waits for communication offers from another system, but no system in its environment offers such communication. Livelock refers to a behaviour, where a system continuously performs some activities without effectively making any progress.

The behavioural properties that are linked to the non-deterministic branching structure can also be characterized as formulas of various temporal logics, such as  $CTL^*$  and its sublogics. In fact, the demand for preservation of the non-deterministic branching structure coincides with the demand to preserve satisfaction of formulas in the common temporal logics.

**Requirement 3: Abstraction from Internal Details** Internal activities of a system should be completely ignored, as long as they do not have indirect consequences towards the non-deterministic branching of the process.

Due to their vague nature, the last two requirements outline concepts that can never be formalized conclusively for all kinds of models at once. They are thus not meant to be properties a notion of equivalence will be formally evaluated against. We rather consider them a guiding principle.

### 1.2.1. Bisimulation

Bisimulation is a co-inductive methodology of defining equivalences between concurrent systems in the form of labelled transition systems [Mil89b; Par81]. Equivalences based on bisimulations are called bisimilarities. The prominent feature of bisimilarities is the preservation of the non-deterministic branching structure of a system. We will discuss this more thoroughly in Chapter 3. We therefore consider bisimilarities as candidates for observational equivalences.

### Important Properties

Bisimulations have proven a valuable tool for the verification and analysis of concurrent systems. We summarize some of their benefits.

---

<sup>1</sup>In a setting where one abstracts from internal transitions, equivalences that preserve the branching structure do still not allow to distinguish between deadlock and livelock, so that any such equivalence will distinguish when livelock or deadlock is present in one system, while not in the other, but it will not distinguish the situation where a deadlock has been replaced by a livelock or vice versa.

**Locality** Even though bisimulation considers the dynamics of system in its completeness, including its evolution over time, it is formulated as relation over states. This allows to proof two systems bisimilar based on arguments that only pertain on the behaviour of the individual states. In this way, global conformance in behaviour is reduced to state-wise local conformance. This makes bisimulation not only a mathematically elegant, but also powerful proof technique.

**Preservation of Logical Properties** Bisimulation characterizes important modal and temporal logics such as  $\mu$ -calculus, CTL, and CTL\* [BK08], in the sense that bisimilar states or systems are precisely those that cannot be distinguished by any formula expressed in these logics. For the same reason, bisimulation preserves linear time temporal logic. In general, the practical usefulness of bisimulation is rooted in its ability to abstract from state identities and from concrete implementations of behaviour as much as possible, but rather to focus on the observable behaviour itself.

**State-Space Minimization** A favourable strategy often used in practice is to minimize the system – or components thereof – to the quotient under bisimilarity, before further analysis are performed. This can speed up the overall model analysis or turn a too large problem into a tractable one [Che+96; HK00; Kat+07].

**Decidability** Due to its locality property, bisimulations are usually efficiently decidable for finite systems.

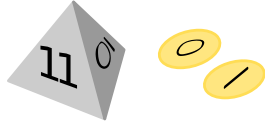
We want to emphasize that the preservation of the non-deterministic branching structure is the property of bisimulation that distinguishes it from all other notions of equivalence for concurrent systems, such as trace equivalence, failure equivalence and so on [Gla90; Gla93]. Even though some of these notions preserve the non-deterministic branching structure partially, bisimulation is the only relation to fully preserves it. Its other prominent features, such as compositionality and preservation of logical properties, are mainly a consequence thereof. For completeness, we want to remark that while bisimulation is in general widely accepted as an essential tool for a complete semantic characterization of concurrent processes based on its branching behaviour, there exist arguments that a linear time perspective on process behaviour and, consequently, trace equivalences, completely suffice [Var01; NV07; NV09].

## Strong vs. Weak Bisimulations

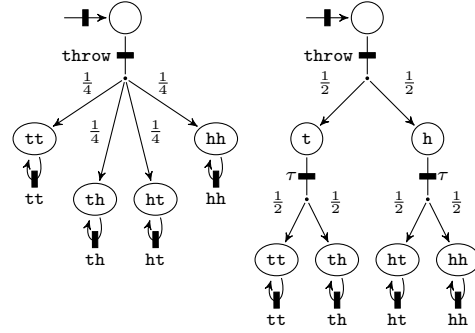
Bisimulation comes in two general variants, usually referred to as *strong* and *weak* bisimulation. The former makes no distinction between internal and observable activities. Weak variants, instead, try to abstract from internal behaviours as much as possible. Weak bisimulation therefore are suitable candidates when it comes to the last of our three properties of observational equivalences.

### 1.2.2. Challenges

Bisimilarities has proven a successful basis for observational equivalences on labelled transition systems. Naturally, the enhancement of labelled transition systems with quantitative aspects has gone along with extensions of the theory of bisimulation. This, however, has turned out to be a challenging task. This is especially true for weak notions of bisimilarity.



**Figure 1.8.:** Rolling a Four-Sided Die vs. Throwing Two Coins



**Figure 1.9.:** Observationally Identical Probabilistic Automata?

For probabilistic automata, weak (and strong) probabilistic bisimilarity [Seg95] has been widely accepted as a reasonable observational equivalence. However, its known weak point is that it distinguishes processes while its observable behaviour would suggest differently.

**Example 1.2.** Assume we want to conduct a probability experiment with four equally distributed outcomes. A natural way to implement this experiment is to throw a four-sided die (see Figure 1.8). If no such die is at hand, we may as well replace the die by two successive coin throws. To an observer who is only interested in the final outcome, these two variants should be completely transparent. Phrased differently, an observer should not be able to reckon how a system produces probabilistic outcomes, as long as the resulting probability distributions are identical.

The two automata in Figure 1.9 represent the two experiments in the form of two Markov automata. To avoid confusion of actions representing outcomes and probabilities, we assume the sides of the die are labelled by *ht*, *th*, *tt* and *hh*, corresponding to *head* and *tail* of the coins.

We can see that the only difference between the two automata is the way the events are determined. In the automaton on the left, each of the four events is fixed by a single probabilistic experiment (one single throw of a four-side *coin*). In contrast, in the automaton on the right the outcome is actually determined by two coin throws (of two-sided coins) in sequence. Notably, the actual procedure of how the final outcomes are determined, is hidden to the observer (by using internal transitions). From the information that is available to an external observer, we only see that each of the four possible outcomes appears with probability one over four after the action *throw* has occurred. Intuitively, it is clear that we do not want to distinguish between the two automata. Yet, probabilistic weak bisimilarity (and, to the best of our knowledge), any other known notion of weak bisimilarity on probabilistic automata, fails to equate the two automata. ◁

For interactive Markov chains the situation seems less debated. Weak bisimilarity, as defined in [Her02], is generally accepted as the coarsest reasonable observational congruence on IMC, except for some for some proposed orthogonal optimizations [Bra02]. Still, no definite answer has been given on the question, if this notion of bisimilarity is in deed the *coarsest* possible equivalence relations satisfying the requirements of an observational equivalence.

In summary, for probabilistic and stochastic systems a famous old question still remains open:

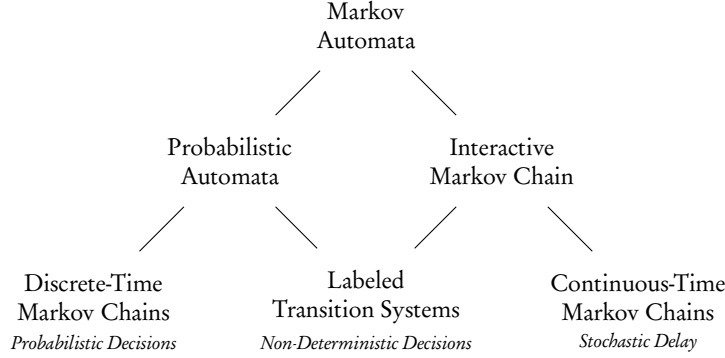


Figure 1.10.: Completing the Lattice

what is the coarsest reasonable notion of observational equivalence on this kind of systems? In addition, the precise role the existing notions of bisimilarity play in this regard needs to be explored. Finally, the combination of both probabilistic immediate transitions *and* stochastic timed behaviour in one model has been investigated never before.

## 1.3. Contributions and Structure of the Thesis

In this thesis, we will develop a model of concurrent and reactive systems that additionally exhibits both probabilistic and stochastic timed behaviour: *Markov automata*. As illustrated by Figure 1.10, it will be a conservative extension of PA and IMC. The thesis summarizes the work that has been done in the context of Markov automata and their semantics since the introduction of Markov automata at LICS 2010 [EHZ10a]. Our presentation follows a historical approach in that it initially takes the perspective from the time before this publication, and develops the theory of Markov automata from ground up.

**Part I – Processes** In the first part of the thesis, we will review the base models of Markov automata, i.e. labelled transition systems, probabilistic automata and interactive Markov chains. Chapters 3 to 5 are respectively dedicated to the three models. For each of them, we introduce and formally define the models and their most common notions of bisimilarity. We end each chapter with a decision algorithm for the respective notion of weak bisimilarity. For the sake of a comparative analysis of the algorithms, especially in foresight of the corresponding algorithm for Markov automata bisimilarity, we present them in a uniform way. This may result in algorithms that are not optimal in practice. However, with respect to their theoretical complexity, the presented algorithms are state-of-the-art.

Chapter 6 provides the axiomatic perspective on the different models and their notions of bisimilarity. It still allows us to identify the similarities and differences between the bisimilarities in a very succinct way. With one exception, the presented axiomatizations are a summary of existing work.

**Part II Markov Automata Theory** The theory of Markov automata is summarized in the second part of the thesis. In Chapter 7, the model of Markov automata is defined formally, and equipped with operators for parallel composition and abstraction, i.e. hiding of internal details. Finally, the model is compared to several other quantitative models with respect to their expressiveness.

Chapter 9 is the climax of the second part. It develops the theory of bisimilarity for Markov automata by discussing several variations of weak bisimilarity with their semantic properties. Among them, we will identify one, weak distribution bisimilarity, that provably is the coarsest notion of weak bisimilarity on Markov automata that satisfies the requirements we demanded for a reasonable observational equivalence in the preceding sections. This notion of bisimilarity is defined directly on distributions, and not on states as it is usually the case. At least for *weak* bisimilarity, this is a novelty, *and* the roots of its coarseness. For a structurally restricted subclass of Markov automata, we present a state-based characterization of this relation. Finally, we compare weak distribution bisimilarity to the bisimilarities from Part I, which will also answer the question in how far they have been the coarsest relation possible for their respective models.

We complement our discussion of comparative semantics for the different models with a sound and complete axiomatization of the coarsest congruence contained in weak distribution bisimilarity.

**Part III** Part III is the final part of the thesis, in which we discuss practically relevant aspects and consequences of our findings of the preceding parts. In Chapter 11, we develop a decision algorithm for weak distribution bisimilarity. Chapter 12 offers a complete treatment of the algorithmic question how to arrive at a minimal quotient of a given automaton with respect to bisimilarity. This problem is of particular practical relevance, as the minimization of an automaton with respect to a given notion of behavioural equivalence prior to further processing is a crucial step in many automated analysis trajectories. We treat this problem not only for the bisimilarity developed in this thesis, but also for all bisimilarities we have discussed in Part I. Our minimization algorithms aim at minimizing several dimensions of the *size* of an automaton at the same time: the number of states and the number of transitions, as well as the size of the support of the individual transitions. We will provide formal results under which conditions minimality with respect to all criteria can be reached. While the basic quotient construction itself is folklore, to the best of knowledge, a complete treatment of the necessary steps to arrive at a truly minimal automaton is novel.

In Chapter 13, we use Markov automata and weak distribution bisimilarity to provide a novel semantics for Generalized Stochastic Petri Nets that is complete in the sense that it provides semantics to every *GSPN*, including confused nets. The semantics is furthermore conservative in the sense that on all other *GSPN*, it agrees with the established semantics.



## 2. Preliminaries

In this chapter, we summarize the basic mathematical backgrounds of process algebra and probabilistic and stochastic models. This summary comprises the fundamentals of sets, probability theory and graphs.

### 2.1. Sets and Multisets

We use the notation  $\{\dots\}$  to denote sets. As we only rely on finite or countably infinite sets of numbers or vertices of graphs, it suffices to use the naïve definition of sets and not to rely on Zermelo-Fraenkel set theory. For arbitrary sets  $X$ , we use  $x \in X$  to denote that  $x$  is contained in  $X$ , e.g. to denote that  $x$  is an element of  $X$ . We write  $X \subseteq Y$  to denote that  $X$  is a subset of  $Y$ , e.g. every element of  $X$  is also an element of  $Y$ . We call two sets disjoint if they do not have an element in common. Moreover, we use the standard definitions of the intersection  $X \cap Y$ , the union  $X \cup Y$  and the difference  $X \setminus Y$  for two sets  $X$  and  $Y$ . Furthermore, we use  $X \uplus Y$  to denote the union of two disjoint sets  $X$  and  $Y$ .

In the following, we use  $\mathbb{N}$  to abbreviate the set of all natural numbers starting at 0, e.g.  $\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$ . Moreover, we use  $\mathbb{R}$  to abbreviate the set of all real numbers. We use square brackets to denote a closed interval, e.g.  $[0, 1]$  denotes the interval of all real numbers between 0 and 1 also including 0 and 1.

**Definition 2.1 (Multiset).** A *multiset* is an unordered collection of well-defined objects.  $\triangleleft$

The difference between sets and multisets is that the later can contain the same object more than once. We use the notation  $\llbracket \dots \rrbracket$  to denote multisets.

**Definition 2.2 (Cartesian Product).** The *Cartesian product*  $A \times B$  of two sets  $A$  and  $B$  is defined as  $A \times B := \{(a, b) \mid a \in A \wedge b \in B\}$ .  $\triangleleft$

Based on the Cartesian product of sets, relations and functions are defined. A *binary relation* over two sets  $A$  and  $B$  is a subset of  $A \times B$ . A *function*  $f$  from a set  $A$  to a set  $B$  is a binary relation over  $A$  and  $B$  such that for every element  $a$  in  $A$ , there exists at most one element  $b$  in  $B$ , such that  $(a, b) \in f$ . For simplicity, we use the standard notion  $f(a) = b$  for functions instead of  $(a, b) \in f$ . A function  $f$  is called *total* if for every  $a$  in  $A$ , there exists  $b$  in  $B$  such that  $f(a) = b$  holds.

*Notation 2.1 (Domain of a Function).* Let  $f : A \rightarrow B$  be a function. We write  $\text{dom}(f)$  to denote the subset of  $A$  for which  $f$  is defined.

## 2.2. Elementary Topology

We here provide basic definitions and notations of topology.

**Definition 2.3 (Metric Space).** A *metric space*  $(X, d)$  is a set  $X$  and a total function  $d: X \times X \rightarrow \mathbb{R}$  such that

- $d(x, y) \geq 0$  for all  $x, y \in X$ ,
- $d(x, y) = 0$  if and only if  $x = y$ ,
- $d(x, y) = d(y, x)$
- and  $d(x, z) \leq d(x, y) + d(y, z)$  for every  $x, y, z \in X$

◁

For notion of convergence of sequences, the following set theoretical terminologies and properties are needed.

**Definition 2.4 (Open and Closed Sets).** An arbitrary subset  $Y$  of a metric space  $(X, d)$  is called *open* if for any point  $x \in Y$  there exists a real number  $\epsilon > 0$  such that for any point  $y \in X$  satisfying  $d(x, y) < \epsilon$  holds:  $y \in Y$ .

A subset is called *closed* if its complement is open.

◁

For every point of an open set there exists a radius  $\epsilon$  such that a spherical with diameter  $\epsilon$  (or smaller) is still a subset of  $X$ . Here is a small example: The set  $(0, 1)$  (the open interval from 0 to 1) is open. For example, if we define  $\epsilon := \frac{\min(x, 1-x)}{2}$  for every point  $x$ , all points with a shorter distance from  $x$  than  $\epsilon$  must also be in the interval. On the other side, the closed interval  $[0, 1]$  is closed because its complement can be written as  $(-\infty, 0) \cup (1, \infty)$  and this set can easily (and similarly to the construction above) be shown to be open.

**Definition 2.5 (Bounded set).** A subset  $Y$  of a metric space  $(X, d)$  is called *bounded* if there exists an  $y \in Y$  and an  $\epsilon > 0$  such that for all  $x \in Y$  holds:  $d(x, y) < \epsilon$ .

◁

If we consider  $\mathbb{R}$  with the function  $d(x, y) := |x - y|$  as a metric space and again look at the interval  $[0, 1]$ , we will see, that it is bounded. If we set  $\epsilon$  to 0.5 and  $y$  to 0.5 no point in the interval will be more than 0.5 away.

Now, we can define compact sets. Compact sets offer convenient properties with respect to convergence of sequences.

**Definition 2.6 (Compactness).** A subset  $Y$  of a metric space  $(X, d)$  is called *compact*, if for every superimposition  $Y = \bigcup_i M_i$  where for every  $i$  the set  $M_i$  is open there exists a finite superimposition  $M_{i_1} \dots M_{i_n}$  such that  $Y = \bigcup_j M_{i_j}$ .

◁

Compactness is a very strong criterion, which allows for many applications especially in the theory of converging sequences.

**Definition 2.7 (Sequence).** A *sequence of elements* of a set  $X$  is a mapping  $m: N \rightarrow X$ , where  $N$  is either the set  $\{1, \dots, n\}$  for some  $n \in \mathbb{N}$  or  $N = \mathbb{N}$ . In the first case, the sequence is called *finite* and  $n$  is called the length of the sequence. In the second case, the sequence is called *infinite*.

◁

We usually denote sequences in sloppy notation. For example by  $a_1, a_2, \dots, a_n$  we denote a finite sequence  $m$  of length  $n$  with  $m(i) = a_i$  for all  $i \in \{1, \dots, n\}$ . Accordingly, we denote infinite sequences as  $a_1, a_2, \dots$ . We also write  $(a_i)_{i=1, \dots, n}$  for  $a_1, a_2, \dots, a_n$ .

For infinite sequences, it is also important if they converge. Intuitively, we ask whether there exists a number which we approximate better and better until we reach it at infinity.

**Definition 2.8 (Convergence).** A infinite sequence  $(a_n)_{n \in \mathbb{N}}$  in a metric space  $(X, d)$  is called *convergent* with limes  $a$  if for every  $\epsilon > 0$  there exists an index  $N \in \mathbb{N}$  such that for every  $m \geq N$  holds:  $d(a_m, a) < \epsilon$ .  $\triangleleft$

With the help of convergent sequences, we can define another version of compactness:

**Definition 2.9 (Sequentially Compactness).** A subset  $Y$  of a metric space  $(X, d)$  is called *sequentially compact* if for every sequence  $(a_n)_{n \in \mathbb{N}}$  there exists a convergent subsequence with limes in  $Y$ .  $\triangleleft$

In the following chapters, we will frequently use the following theorem:

**Theorem 2.1.** A subset of a metric space is compact if and only if it is sequentially compact.

If we only talk about subsets of  $\mathbb{R}^n$  (for  $n \in \mathbb{N}$ ), we will also rely on the theorem of Heine-Borel:

**Theorem 2.2.** A subset of  $\mathbb{R}^n$  is compact if and only if it is bounded and closed.

Finally, we will also use special kinds of sequences, namely Cauchy sequences:

**Definition 2.10 (Cauchy Sequence).** We call an infinite sequence  $(a_n)_{n \in \mathbb{N}}$  *Cauchy* if for every  $\epsilon > 0$  there exists an index  $N \in \mathbb{N}$  such that for every  $m, l \geq N$  holds:  $d(a_m, a_l) < \epsilon$ .  $\triangleleft$

Intuitively, a Cauchy sequence is a sequence where the distance between to members of the sequence becomes arbitrarily small.

**Definition 2.11 (Completeness).** A subset of a metric space is called *complete* if every Cauchy sequence converges.  $\triangleleft$

In the following chapters, we will mainly use the complete sets  $[0, 1]$  and  $\mathbb{R}$ .

## 2.3. Graphs

Graphs are one of the most foundational models in computer science. In our setting, they serve as the fundamental underlying concept of all our semantic models of process behaviour.

**Definition 2.12 (Labelled Directed Graph).** A *labelled directed graph* is a tuple  $(S, E, L)$  where

- $S$  is a set of states,
- $E \subseteq S \times L \times S$  is a set of edges, and
- $L$  is a set of labels.

$\triangleleft$

**Definition 2.13 (Subgraph).** A graph  $(S', E', L')$  is called a subgraph of  $(S, E, L)$  if it is a graph and in addition

- $S' \subseteq S$ ,
- $E' \subseteq E$ , and
- $L' \subseteq L$ .

◁

A path in a graph is a sequence of states,  $m : N \rightarrow S$ , such that each state is connected to its successor by an edge, i.e.  $(m(i), l, m(i + 1)) \in E$  for some  $l \in L$ . The *length* of a path and the definition *finite* and *infinite* paths carry over immediately from the respective definitions for sequences.

We say two states  $s$  and  $s'$  are connected if there is a finite path with whose first state is  $s$  and last state is  $s'$ , or whose first state is  $s'$  and last state is  $s$ . This means that when  $s$  is connected to  $s'$  then  $s'$  is always also connected to  $s$ . In addition, we say  $s$  can reach  $s'$  if there is a finite path starting in  $s$  and ending in  $s'$ . In general, if  $s$  can reach  $s'$ , then this does not imply that  $s'$  can also reach  $s$ , as the graphs are directed. In fact, if  $s$  can reach  $s'$  and  $s'$  can reach  $s$  then the graph is cyclic (if  $s \neq s'$ ).

A finite path is called a *cycle*, if the first and the last state are the same. A graph is called *cyclic* if it contains a cycle, and *acyclic* otherwise.

**Definition 2.14 (Connectedness).** A set states of a graph is called *connected*, if every two states are connected. It is *strongly connected* if every state can reach every other state. ◁

**Definition 2.15 (Label Restricted Graph).** Given a graph  $G = (S, E, L)$  and a label  $a \in L$ , we denote by  $G|_a$  the graph  $(S, E', \{a\})$  with

$$E' = \{(s, a, s') \in E\}$$

◁

**Definition 2.16.** Let  $G = (S, E, L)$  be a graph and  $a$  a label in  $L$ .  $G$  is called

- *finite* if  $|S|$  and  $|E|$  is finite, and *infinite* otherwise,
- *a-finite* if every connected subgraph of  $G|_a$  is finite, and *a-infinite* otherwise,
- *a-acyclic*, if  $G|_a$  is acyclic, and *a-cyclic* otherwise.

◁

## 2.4. Probability Theory

Let  $X$  be some set and let  $\mathcal{P}(X)$  be its powerset.

**Definition 2.17 ( $\sigma$ -Algebra).** A subset  $\Sigma \subseteq \mathcal{P}(X)$  is called a  $\sigma$ -algebra if it satisfies the following conditions:

1.  $\Sigma$  is non-empty.
2. If  $A \in \Sigma$  then also  $X \setminus A \in \Sigma$ .
3. Let  $I$  be a *countable* index set. If for each  $i \in I$   $A_i \in \Sigma$  then also

$$\bigcup_{i \in I} A_i \in \Sigma.$$

&lt;

**Definition 2.18 (Measure).** Let  $\Sigma$  be a  $\sigma$ -algebra. A function  $\mu : \Sigma \rightarrow \mathbb{R}_{\geq 0}$  is called a *measure* if

1.  $\mu(\emptyset) = 0$
2. Let  $I$  be a *countable* index set. Let  $\{E_i\}_{i \in I}$  be a family of pairwise disjoint sets in  $\Sigma$ . Then

$$\mu\left(\bigcup_{i \in I} E_i\right) = \sum_{i \in I} \mu(E_i).$$

&lt;

**Definition 2.19 ((Sub-) Probability Measures).** Let  $\Sigma$  be a  $\sigma$ -algebra over  $X$ . A measure  $\mu$  is called sub-probability measure, or a subdistribution, if  $\mu(X) \leq 1$ . It is called a (full) probability measure, or a distribution, if  $\mu(X) = 1$ . <

We denote sub-probability distributions by Greek letters  $\gamma, \rho, \mu, \nu$  and  $\xi$  as well as indexed variants thereof.

For any countable or finite set  $X$ , its powerset  $\mathcal{P}(X)$  is a  $\sigma$ -algebra. We let  $Subdist(X)$  denote the set of all subdistributions over  $X$ , and  $Dist(X)$  the set of all distributions over  $X$ .

**Notation 2.2.** For singleton sets  $\{x\} \in \Sigma$ , we usually write  $\mu(x)$  instead of  $\mu(\{x\})$ .

Given  $\mu \in Subdist(X)$ , we denote by  $|\mu|$  the overall measure  $\mu(X)$ , by  $Supp(\mu)$  the set  $\{x \in X \mid \mu(x) > 0\}$ , by  $\mu(\perp)$  the value  $1 - \mu(X)$  where  $\perp \notin X$ . Furthermore, we denote by  $\delta(x)$ , where  $x \in X \cup \{\perp\}$ , the *Dirac* distribution such that  $\mu(y) = 1$  for  $y = x$ , 0 otherwise;  $\delta(\perp)$  represents the empty distribution such that  $\mu(X) = 0$ .

If  $\mu_1$  and  $\mu_2$  are subdistributions with  $|\mu_1| + |\mu_2| \leq 1$  we write  $\mu_1 \oplus \mu_2$  to denote the subdistribution  $\mu$  with  $\mu(s) = \mu_1(s) + \mu_2(s)$ . We say  $\mu$  can be split into  $\mu_1$  and  $\mu_2$ , or that  $\mu_1$  and  $\mu_2$  are a *splitting* of  $\mu$ , if  $\mu = \mu_1 \oplus \mu_2$ . Since  $\oplus$  is associative and commutative, we may use the notation  $\bigoplus$  for arbitrary finite sums.

### 2.4.1. Infinite Sum

For a finite index set  $I = \{0, 1, \dots, n\}$ , we let

$$\sum_{i \in I} a_i = a_0 + a_1 + \dots + a_n.$$

For the infinite, but countable index  $I = \mathbb{N}$ , we denote by  $\sum_{i \in I} a_i$  the limit of the sequence  $\langle \sum_{m \in \{0, 1, \dots, i\}} a_m \rangle_{i \in I}$ , if it exists. Throughout the thesis, we only deal with summands  $a_i \in \mathbb{R}_{\geq 0}$ . Therefore, the limit is always uniquely determined, if it exists, independent of the concrete order of the summands, as part of the following lemma.

**Lemma 2.1.** Let  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  be a permutation of the natural numbers. Then,

$$\sum_{i \in \mathbb{N}} a_i = \sum_{i \in \mathbb{N}} a_{\sigma(i)},$$

whenever the limit exists.

We mostly make use of this lemma silently.

### 2.4.2. Convex Combinations

Let  $I$  be a finite or countably infinite index set. Let  $\vec{p} = (p_1, p_2, \dots)$  be a vector with finite or countably many components  $p_i \in [0, 1]$  with  $i \in I$  such that  $\sum_{i \in I} p_i = 1$ . Let  $\vec{\mu} = (\mu_1, \mu_2, \dots)$  be a vector of subdistributions of the same size. We call

$$\bigoplus_{\vec{p}} \vec{\mu} := \bigoplus_{i \in I} p_i \cdot \mu_i$$

the *convex combination* of  $\mu_1, \mu_2, \dots$ , with weights (or coefficients)  $p_1, p_2, \dots$ . It is often convenient to use the notation  $p_1 \cdot \mu_1 \oplus p_2 \cdot \mu_2 \oplus \dots \oplus p_n \cdot \mu_n$  instead of  $\bigoplus_{\vec{p}} \vec{\mu}$ , especially if  $I$  is finite.

We further let

$$\mu_1 \oplus_p \mu_2 := \bigoplus_{(p, 1-p)} (\mu_1, \mu_2).$$

It is helpful to recall that for the  $\oplus$ -operator

$$|\mu_1 \oplus \mu_2| = |\mu_1| + |\mu_2|,$$

and for  $\oplus_p$ , if  $|\mu_1| = |\mu_2|$ ,

$$|\mu_1 \oplus_p \mu_2| = |\mu_1| = |\mu_2|.$$

An analogue property holds for arbitrary (non-binary) convex combinations. Note that for full distributions, this implies that a convex combination of full distribution is always a full distribution.

For two distributions  $\mu_1$  and  $\mu_2$  we define  $\mu = \mu_1 \ominus \mu_2$  as the renormalized non-negative difference between the probability mass of each element in the support of  $\mu_1$  and  $\mu_2$ . Formally, as an intermediate step we define for each  $s \in \text{Supp}(\mu_1)$  a subdistribution  $\mu'$  to be

$$\mu'(s) = \max\{\mu_1(s) - \mu_2(s), 0\}$$

and then normalize  $\mu'$  to a full distribution  $\mu$  by letting  $\mu := \frac{\mu'}{|\mu'|}$ , if  $\mu'$  is not the empty distribution. For notational convenience, for a state  $s$ , we denote by  $\mu \odot s$  the distribution  $\mu \odot \delta(s)$ .

*Notation 2.3.* If  $I$  is finite, we call a convex combination *finitary*, and *infinitary* otherwise.

**Lemma 2.2.**

$$\bigoplus_{i \in I} \bigoplus_{j \in J} \mu_{i,j} = \bigoplus_{j \in J} \bigoplus_{i \in I} \mu_{i,j}$$

**Lemma 2.3.**

$$\bigoplus_{i \in I} (\mu_i \oplus_p \gamma_i) = \left( \bigoplus_{i \in I} \mu_i \right) \oplus_p \left( \bigoplus_{i \in I} \gamma_i \right)$$

**Lemma 2.4.**

$$p \cdot \bigoplus_{i \in I} \mu_i = \bigoplus_{i \in I} p \cdot \mu_i$$

## Convex Sets

**Definition 2.20 (Convex Hull).** Let  $P = \{p_1, \dots, p_n \in \mathbb{R}^k\}$  be a finite set of points in  $\mathbb{R}^k$ . We call  $CHull(P) = \{c \in \mathbb{R}^k \mid \exists c_1, \dots, c_n \geq 0 : \sum_{i=1}^n c_i = 1 \text{ and } c = \sum_{i=1}^n c_i \cdot p_i\}$  the *convex hull* of  $P$ .  $\triangleleft$

$C$  is a *finitely generated convex set*, if  $C = CHull(P)$  for a finite set  $P \subseteq \mathbb{R}^k$ .

**Lemma 2.5** (cf. [CS02, Sec. 2]). Every finitely generated convex set  $C$  has a unique minimal set of generators  $Gen(C)$  such that  $C = CHull(Gen(C))$ .

**Liftings** It is often convenient to consider subdistributions as relations over  $X \times \mathbb{R}_{>0}$  and thus, to explicitly denote a subdistribution  $\mu$  by the relation  $\{(s, p_s) \mid s \in X, p_s = \mu(s)\}$ . Whenever we use the set notation to specify a subdistribution, we use square brackets  $\langle$  and  $\rangle$  instead of curly brackets to denote the set. Thus, to denote the subdistribution  $\mu$  we will write  $\langle (s, p_s) \mid s \in X, p_s = \mu(s) \rangle$ .

Given a relation  $\mathcal{S} \subseteq A \times Subdist(B)$ , we lift  $\mathcal{S}$  to a relation over  $Subdist(A) \times Subdist(B)$  by letting  $\mu \mathcal{L}(S) \gamma$  hold if and only if

$$\gamma = \bigoplus_{a \in Supp(\mu)} \mu(a) \cdot \gamma_a$$

where for each  $a \in Supp(\mu)$  it holds that  $\mathcal{S}(a, \gamma_a)$ . The set  $\mathcal{L}(S)$  is called the *lifting* of  $\mathcal{S}$ .





**Part I.**

**Processes**



## 3. Labelled Transition Systems

Labelled transition systems (LTS) are the most elementary models for concurrent and reactive systems. As the standard model of interleaving semantics, it represents a concurrent execution of sequences of actions as the non-deterministic choice between all possible interleavings of the sequences. It is the original model of concurrency, from which the quantitative models, which are the main focus of this thesis, are based on. This chapter therefore serves as an introductory chapter. It reviews the formal definition of the model, its bisimilarities as well as the standard decision algorithm for bisimilarity. As for the model, the bisimilarities are a blue-print for all bisimilarities introduced later in other chapters.

**Outline and Contributions.** In Section 3.1 we introduce the base model as well as two composition operators, parallel composition and abstraction. The parallel composition features CSP-style synchronization [BB87; Her02; Hoa85]. In Section 3.2 we introduce strong and weak bisimilarity. We also review the experimenter’s perspective on bisimilarity, which is an intuitive justification for bisimilarity being a natural notion of observational equivalence. Section 3.3 presents a decision algorithm for bisimilarity. Section 3.4 concludes.

### 3.1. Model

In every state, certain different activities may occur. When an activity happens, it causes a change of state in the system, thus allowing a different set of activities to occur next. Similar to the classical automata model from theoretical computer science, they consist of a set of states  $S$  and a transition relation representing the single activities. However, their purpose is different from determining a language of accepted words, but rather model the complete behaviour of system via their structure. Every transition is labelled by some symbol, called *action*, which is taken from a set  $Act$  of all possible actions. This allows to represent (and abstractly classify) different behaviours a system may exhibit. Thus, actions are abstract representations of concrete activities. Depending on the concrete systems, the set of actions may describe activities like pressing a button or dropping a can of coke, as well as specific computations and molecule reactions. Different to the classical automata model, labelled transition systems do not have any notion of accepting state.

Being structurally akin to classical automata, labelled transition systems are easy to understand, create and to (algorithmically) analyse, which makes them an ideal modelling and specification formalisms amenable to various automata-based analysis and verification techniques.

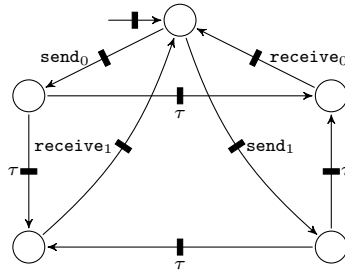
**Definition 3.1 (Labeled Transition Systems).** A *labelled transition systems* is a quadruple  $(S, \bar{s}, Act, \rightarrow)$ , where

- $S$  is a non-empty finite set of states,
- $\bar{s}$  is the initial state, in which the system’s behaviour is considered to start,

- $Act$  is a set of actions, and
- $\rightarrow \subseteq S \times Act \times S$  is a set of transitions.

◁

Whenever the set of states  $S$  and the transition relation of an LTS  $\mathcal{A}$  is finite, then  $\mathcal{A}$  is called *finite*. Throughout this thesis we will only consider finite transition systems, unless mentioned otherwise. The set of actions  $Act$  may be infinite, even for finite automata. In general, it is a proper superset of the actions that actually occur as a label of any transition of  $\mathcal{A}$ . For convenience it is often assumed that all automata under consideration share one common set  $Act$  containing every action occurring in any automaton.



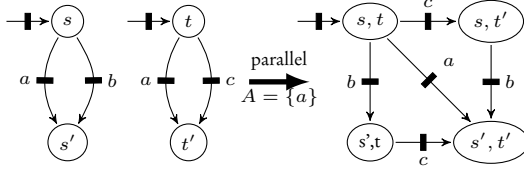
**Figure 3.1.:** A Simple Communication Channel

**Example 3.1.** The labelled transition system depicted in Figure Figure 3.1 represents a noisy channel, where sending and receiving is modelled in one component. In an ideal world, the sending of 0 would always lead to the reception of 0, and accordingly for 1. In this noisy channel version, there exists the chance for faulty transmissions, modelled by internal transitions. When a 0 has been sent, it is non-deterministically decided whether actually 0 is received as expected, or 1 is received erroneously. ◁

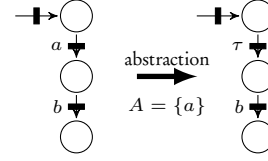
**Notations** We usually denote elements from  $Act$  with letters  $a, b, \dots$ . States from  $S$  are denoted by  $s$  and  $t$  and indexed variants. A transition  $tr = (s, a, s') \in \rightarrow$  is usually denoted by  $s \xrightarrow{a} s'$ . Furthermore, it is said to *leave* from state  $s$ , to be *labelled* by  $a$ , and to *lead* to  $s'$ . We also say that  $s$  enables action  $a$ , that action  $a$  is enabled in  $s$ , and that the transition  $(s, a, s')$  is enabled from  $s$ .

### Parallel Composition and Abstraction

We will now define parallel composition of labelled transition systems. The construction is similar to the cross-product of automata. It perfectly reflects the idea that concurrency is represented by all possible interleavings of the possible executions of the combined automata. In order to express communication, the execution of certain action will be synchronized. For labelled transition systems, different variation of communication exist. The primitive introduced by Milner is usually referred to as CCS-style communication, as it is the original communication primitive used in the *Calculus of Communication Systems* (CCS) [Mil89b].



**Figure 3.2.:** Parallel Composition of two Labelled Transition Systems



**Figure 3.3.:** Abstraction of an Labelled Transition System

CCS-style synchronization envisages that communication between parallel systems always happens between exactly two systems. Synchronization can only happen on complementary actions. To this end, it is assumed that every action has a complement. When two systems synchronize on two complementary actions, for systems that do not participate in the communication, this communication becomes unobservable. In fact, formally, it results in a transition labelled with the internal transition  $\tau$ . As then no other system is able to observe the communication, it cannot join in the synchronization. Another feature of this type of communication is that without further means, two systems that may communicate on two actions do not have to do so. Alternatively, both systems can interleave the two actions without every synchronizing. So to say, communication is not obligatory. Within CCS, however, other operators exist that allow to enforce synchronization.

Another communication primitive that has been adopted by many process calculi is CSP-style communication, which is the original primitive in Hoare's *Communicating Sequential Processes* (CSP)[Hoa85]. While before, communication, i.e. synchronization, could only happen between two systems, we here have broadcast communication, in which as many parallel systems join in as possible. Synchronization here must happen on transitions that are labelled with exactly the same action. When two systems synchronize, it results in the same action being issued again, so that another system can pick up and join the communication. Furthermore, not every action serves as a potential communication action, but at the time of the parallel composition of two processes, the set of actions, which are intended as communication actions, is fixed. One could say that this set of action specifies a communication interface between the two processes. As now communication is restricted to dedicated actions, communication on them is now enforced. This means that if two systems are supposed to synchronized on a certain action, none of them can execute any transition labelled by this action on its own. To restrict broadcast to a specific group of parallel components, in CSP a specialized *abstraction* operator exists, that masks execution of communication actions by the internal action  $\tau$ .

In this thesis, we adopt CSP-style communication for labelled transition system. The main reason is that it is also the standard communication primitive for probabilistic automata and interactive Markov chains, the other two basic models we will base our work on.

In order to denote parallel composition of two LTS, we will write  $\mathcal{A} \parallel_A \mathcal{A}'$ , where  $A$  is an arbitrary set of actions excluding the internal action  $\tau$ . It is notationally helpful to use the symbol  $\parallel_A$  as syntactic sugar so that if  $s$  and  $s'$  are two states, we write  $s \parallel_A s'$  to denote the pair  $(s, s')$ .

**Definition 3.2 (Parallel Composition).**

Given two LTS  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\quad})$  and  $\mathcal{A}' = (S', \bar{s}', Act, \xrightarrow{\quad})$ , their parallel composition  $\mathcal{A} \parallel_A \mathcal{A}'$  is defined as the automaton  $\mathcal{A}^* = (S^*, \bar{s}^*, Act, \xrightarrow{\quad}^*)$  where

- $S^* = S \times S'$
- $\bar{s}^* = \bar{s} \parallel_A \bar{s}'$
- $\xrightarrow{\quad}^*$  is the least relation satisfying

$$\begin{array}{ll} s \parallel_A s' \xrightarrow{a}^* t \parallel_A t' & \text{if } a \in A \text{ and } s \xrightarrow{a} t \text{ and } s' \xrightarrow{a} t' \\ s \parallel_A s' \xrightarrow{a}^* t \parallel_A s' & \text{if } a \notin A \text{ and } s \xrightarrow{a} t \\ s \parallel_A s' \xrightarrow{a}^* s \parallel_A t' & \text{if } a \notin A \text{ and } s' \xrightarrow{a} t'. \end{array}$$

◁

**Example 3.2.** In Figure 3.2, we see an example of the parallel composition of two automata with a synchronization on the common action  $a$ . ◁

During the modelling process, it frequently happens that certain aspects of a system's behaviour should be hidden from the environment. We denote this process by *abstraction*. A typical situation is when we want to model a component of a system that acts like a single device towards its environment, but actually is composed of several submodules that interact internally. For example, if we enter a coin in a beverage vending machine, the coin is first checked by one component for its integrity, and then a second component evaluates the money value of the coin, which is then shown on the display, which is again a third component. As a thirsty user, we are not interested in these details. To us, the complete vending machine appears as one single machine that we interact with. So, even though internal components interact when a coin is inserted, we do not notice the necessary internal communication that go on in order to fulfil their task.

**Definition 3.3 (Abstraction).** Given a LTS  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\quad})$ , its abstraction  $\mathcal{A}|_A = (S', \bar{s}', Act, \xrightarrow{\quad})$  with respect to the set of actions  $A \subseteq Act \setminus \{\tau\}$  satisfies

- $S' = S$
- $\bar{s}' = \bar{s}$
- $\xrightarrow{\quad}$  is the least relation satisfying

$$\begin{array}{ll} s \xrightarrow{a} t & \text{if } a \notin A \text{ and } s \xrightarrow{a} t \\ s \xrightarrow{\tau} t & \text{if } a \in A \text{ and } s \xrightarrow{a} t \end{array}$$

◁

**Example 3.3.** Figure 3.3 shows the abstraction of action  $a$  in a labelled transition system. ◁

## 3.2. Bisimilarities

Bisimulations are co-inductively defined binary relations, that relate states of a concurrent system provided they exhibit the same transitions to the same states up to the bisimulation. If we agree on saying that a bisimulation relation describes *behaviour* by relating two states precisely when they are behaviourally equivalent, we can also say that a bisimulation relates two states if their outgoing transitions agree on their action label and their goal states *up to behavioural identity*. One typically distinguishes between *strong* and *weak* notions of bisimilarity. The latter treat internal transitions, labelled by the special action  $\tau$ , differently from those labelled by an observable action. The former does not assign any special meaning to transitions labelled by  $\tau$ .

### 3.2.1. Strong Bisimilarity

**Definition 3.4 (Strong LTS Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\quad})$  be an LTS. A symmetric binary relation  $\mathcal{R}$  on  $S$  is a *strong LTS bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} s'$  for some  $a \in Act$  and  $s' \in S$  implies  $t \xrightarrow{a} t'$  for some  $t'$  with  $s' \mathcal{R} t'$ .  $\triangleleft$

We write  $s \sim_{\text{LTS}} t$  if some strong LTS bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two LTS  $\mathcal{A}$  and  $\mathcal{A}'$  are strong LTS bisimilar if in their disjoint union the two initial states are strong LTS bisimilar, i.e.  $\bar{s} \sim_{\text{LTS}} \bar{s}'$ . We call  $\sim_{\text{LTS}}$  strong LTS bisimilarity.

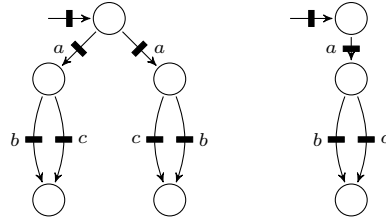


Figure 3.4.: Two Strong LTS Bisimilar LTS

We will now discuss which aspects of a systems structure are deemed observable by virtue of the equivalence classes induced by this bisimilarity.

**Readiness and Failure** The basic building block of observable behaviour are the actions a system can issue, i.e. the inputs it can receive and the outputs it can send. In every moment of its execution, a system may be able (and willing) to react to certain inputs and to produce certain outputs. We say that a system is *ready* to execute this actions, or, alternatively, that those actions are *enabled*. We further say that a system *fails* to execute an action in a specific moment of execution, if it is not able to receive or to send it. If we know the complete set of actions that a system is ready to execute in its next execution step, *and* the complete set of actions it must fail to execute, we can completely characterize its possible behaviour in the next step. Furthermore, if we assume that we know the set of all actions a system could possible execute, it fully suffices to know the set of actions it is ready to execute.

Bisimulation compares whether the set of ready actions of one process is a subset of the ready actions of the other. By its symmetric nature, this actually implies that two bisimilar states must agree on their sets of ready actions.

Most notions of observational equivalence (and also preorders) that are finer than trace equivalence, take readiness and/or failures as their basic building blocks, for instance [BIM95; Phi87; BBK87; MS92a].

**Non-Deterministic Branching Structure** Our second requirement of an observational equivalence (Remark 1.1) was to preserve the non-deterministic branching structure. In fact, this is a distinguishing feature of bisimulation. At the heart of its definition resides the demand that whenever one process branches with a certain transition ( $s \xrightarrow{a} s'$ ) then also its bisimilar counterparts needs to exhibit this branching ( $t \xrightarrow{a} s'$ ) and vice versa. Furthermore, due to its co-inductive nature, in bisimilarity it is possible to understand two processes to be observably behaviourally equivalent precisely when they are bisimilar. Given this, two bisimilar states need to reach bisimilar states again in every branch, which precisely means that they must agree on their future behaviour. With that in mind, it seems natural that two processes are only bisimilar if the points during their execution, where decisions for future behaviour are made i.e. their states, exhibit precisely the same branching structure — up to behaviourally redundant differences, i.e. additional transitions/branchings that lead to bisimilar states. But this is exactly our definition of preservation of the non-deterministic branching structure. In this way, we dare to say that the use of the bisimulation methodology itself immediately implies the preservation of the non-deterministic branching structure.

It is straightforward to show that strong LTS bisimilarity is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two compositions we have defined.

**Theorem 3.1.** Strong LTS bisimilarity  $\sim_{\text{LTS}}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

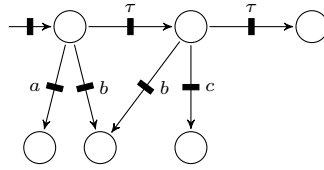
*Remark 3.1* (The Game Perspective On Bisimulation). For certain applications, it is convenient to describe bisimulations as a game, played on the LTS, by two players, also called *challenger* and *defender*. The challenger corresponds to the  $\forall$ -quantified premise of the bisimulation condition, while the defender corresponds to the  $\exists$ -quantified conclusion. Given a pair of states (which is supposedly bisimilar), the challenger first picks any of the two states and then chooses a transition. The defender then needs to find a transition of the remaining state that has the same label as the transition picked by the challenger. Then, the game starts anew with the two successor state. The challenger wins the game if and only if the defender is not able to find a matching transition.

The game formalism is well-suited to refute bisimilarity of two states, as a winning strategy exists for the first player if and only if the states are *not* bisimilar. While we will implicitly make use of the game perspective when we want to show that two systems are not bisimilar in this thesis, we will not rely on it as a formal tool too much. We more focus on the nice intuitive description it provides in terms of a challenger and a defender. This nomenclature is often very handy when discussing differences between different notions of bisimulation.



### 3.2.2. Weak Bisimilarity

In labelled transition systems, explicit internal computations are transitions labelled by the distinguished action  $\tau$ . They usually result from *abstraction*, even though they can also be used directly to model internal computation. As long as internal computations only appear in the form of sequences of internal computation, where all involved states do not have any other enabled transitions, they can be completely ignored from an observational perspective. In other words, deterministic internal computations do not have any influence on the observable behaviour of a system. This is different, if internal transitions are enabled in states that also enable other transitions, regardless whether these transitions are labelled by  $\tau$  or any observable action. Then, internal behaviour may have influence on what is observable for the outside.



**Figure 3.5.:**  $\tau$ -Transitions may be Indirectly Observed

**Example 3.4.** In the LTS of Figure 3.5, the initial state is ready to execute both the actions  $a$  and  $b$ . However, by an internal computation the system can decide to let any further attempts to execute the action  $a$  fail, by taking the only internal transition that is enabled from the initial state. From this state, it can even decide to refuse any further interaction by taking the transition to the state on the very right.  $\triangleleft$

Note that internal behaviour can always only lead to a situation where *less* behaviour is possible. In a weak setting, we say that a transition is ready to be executed from a state, even if it does not immediately originate from this state. It suffices that the state can reach via internal transitions the state from which the transition originates. The reason for this postulate is that internal activity cannot be observed itself, but only by virtue of the changes of internal behaviour it induces. Thus, we cannot distinguish the case where an observable action is executed from a state because a transition labelled by that action originates directly from that state, or the case where before an internal state change has to happen.

**Example 3.5.** In the LTS of Figure 3.5, the initial state is ready to execute actions  $a$ ,  $b$  and  $c$ , even though no transition leaves this state that is labelled by  $c$ . However, with one internal transition this state can reach its right neighbour, which enables  $c$ .  $\triangleleft$

When internal activities should be considered unobservable, we have to adapt the definition of bisimulation slightly. In this setting, a system may perform internal actions unnoticed. This implies that the observation of the execution of an action  $a$  by system  $s$  may allow  $a$  to perform a sequence of unobserved internal actions before and after the obligatory execution of  $a$ . We express this by the following notation.

**Definition 3.5 (Weak Transition).** There exists a *weak transition* from state  $s$  to state  $t$ , denoted by  $s \Longrightarrow t$ , if there exists a finite sequence  $s_0 s_1 \dots s_n$  such that  $s_0 = s$  and  $s_n = t$  and for all  $i < n$  holds:  $s_i \xrightarrow{\tau} s_{i+1}$ . Note that for the case  $n = 0$  no transitions occurs at all.

There exists a *weak transition* from state  $s$  to state  $t$  with label  $a \in Act \setminus \{\tau\}$ , if  $s \Longrightarrow s'$ ,  $s' \xrightarrow{a} t'$  and  $t' \Longrightarrow t$ .  $\triangleleft$

**Remark 3.2.** Every state  $s$  satisfies  $s \Longrightarrow s$  by our definition. This means that the relation  $\Longrightarrow$  basically expresses reachability by internal transitions, including the case that every state can reach itself trivially by taking zero internal transitions. This fact will play an important role for the definition of weak bisimulation.

If  $a \in Act$ , we overload the notation and let  $s \xRightarrow{a} t$  denote  $s \Longrightarrow t$  in the case that  $a = \tau$ . In the case we want to make explicit, that  $s$  performs an internal weak transition, in which actually at least one  $\tau$ -transition occurs, we write  $s \xrightarrow{\tau} \circ \Longrightarrow t$ .

**Definition 3.6 (Weak LTS Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\cdot})$  be an LTS. A symmetric binary relation  $\mathcal{R}$  on  $S$  is a *weak LTS bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} s'$  for some  $a \in Act$  and  $s' \in S$  implies  $t \xRightarrow{a} t'$  for some  $t' \in S$  with  $s' \mathcal{R} t'$ .  $\triangleleft$

We write  $s \approx_{\text{LTS}} t$  if some weak LTS bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two LTS  $\mathcal{A}$  and  $\mathcal{A}'$  are weakly LTS bisimilar if in their disjoint union the two initial states are weakly LTS bisimilar, i.e.  $\bar{s} \approx_{\text{LTS}} \bar{s}'$ . We call  $\approx_{\text{LTS}}$  weak LTS bisimilarity in the following.

It is a well-know fact that the transition  $s \xrightarrow{a} s'$  can be replaced by  $s \xRightarrow{a} s'$  without changing the resulting bisimilarity.

It is straightforward to show that weak LTS bisimilarity is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two compositions we have defined.

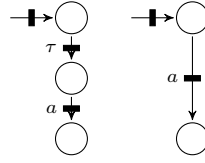
**Theorem 3.2.** Weak LTS bisimilarity  $\approx_{\text{LTS}}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

Formally, the difference between weak bisimulation and strong bisimulation is small. The defender, i.e. state  $t$ , is now allowed to match the transition of the challenger  $s$  by a weak transition, while before, it had to choose a strong transition. This has several consequences on the notion of observable behaviour induced by weak bisimilarity

**Weak Readiness and Failure** As already mentioned, in a weak setting we should consider every action ready to be executed that can be reached via a finite sequence of internal transitions. Accordingly, a state of a system can only fail to execute an action  $a$  if it cannot reach any state with internal transitions that is able to execute  $a$ .

**Weak Branching Structure** In a weak setting, we must rethink what we want to understand to be the non-deterministic branching structure of a system. Due to the presence of internal transitions it may now be the case that a state change occurs in a system completely unnoticed by the system's environment. An example of such a system, together with its bisimilar counterpart without any internal behaviour, can be found in Figure 3.6. Another interesting variant of this specific type of internal behaviour is discussed in Example 3.6. Another possible scenario is that due to an internal state change a system can spontaneously restrict its possibilities to interact with the environment. We have discussed this issue in



**Figure 3.6.:** Internal Steps Not Revealing New Behaviour can be Ignored

Example 3.4. Clearly, state changes of the later type, i.e. restricting interactions, are part of the observable behaviour of a system, and thus must be integrated into our notion of the observable branching structure of a system.

In weak bisimulation, these situations are covered by considering both observable actions and the internal action  $\tau$  in its simulation condition. This ensures that whenever a system can perform an internal state change which results in a change of its interaction pattern, then any weakly bisimilar system must also be able to simulate this change. At the same time, the definition of bisimulation allows to let internal behaviour pass completely unnoticed, if it does *not* influence the interactions in any way. This is ensured by ensuring  $s \Longrightarrow s$  for any state. Thus, whenever a challenger proposes an internal transition that has no observable consequences, the defender may answer this by doing nothing.

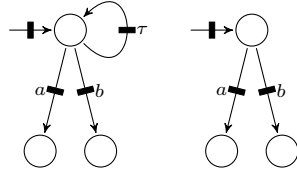
This perspective on the weak branching structure has seen some controversy in the history of weak notions of bisimilarity. In [Gla93; Van94; GW96], *Branching Bisimulation* is introduced, which realizes a stricter notion of weak branching structure, which equates fewer systems. Still, this relation has an appealing mathematical theory, and has been used in practice and theory of concurrent systems. In fact, among all practically relevant behaviour properties of a concurrent system, branching bisimulation and weak bisimulation preserve the same.

**Example 3.6.** Weak bisimulation allows to ignore internal self-loops because we can always simulate them with a trivial weak transition of the state to itself. Note that this complete abstraction also from infinitary internal behaviour has been disputed sometimes and alternative notions of bisimilarity have been proposed that distinguish the ability to perform a finite number of successive internal transitions from the ability to do so with an infinite number, which one usually refers to as *divergence* [Gla93]. In the course of this thesis, the precise treatment of divergent behaviour will play an important role in the context of stochastic timed behaviour.  $\triangleleft$

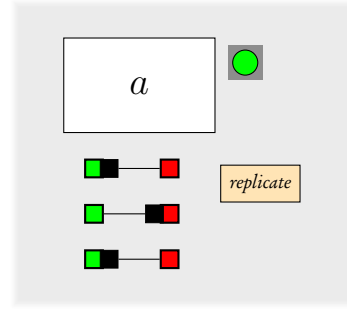
### 3.2.3. The Experimenter's Perspective

Robin Milner has provided a motivation for what observation behaviour of concurrent systems should be in terms of simple *button-pushing experiments* [Mil81]. It can be shown that bisimilarity indeed coincides with the button-pushing experiments we will describe in the following [Gla93]<sup>1</sup>. Even though these consideration intentionally remain at an informal level, they provide interesting insights, that will help us to understand the nature of bisimulation better.

<sup>1</sup>In [Gla93], actually a slightly different, but equivalent, setting of observations is introduced and then formalized in order to establish this result.



**Figure 3.7.:** Weak Bisimulation allows to Ignore Internal Self-Loops



**Figure 3.8.:** Basic Machine

### Button-Pushing Experiments

As before, we want to characterize behaviour in terms of observations, an external observer can make. The observer remains, however, not fully passive, but can provoke reactions of the system by stimulating it via pushing a system's buttons. As the observer can actively provoke reactions of the system, he can issue painful tests, which allow him to derive some information about the way the system is structured. In a way, the observer acts as an observer in the strict sense and as the environment of the system at the same time.

The way the observer interacts with a system is best described by imagining the system as a simple machine, equipped with a set of buttons and a status light [Mil81; Gla93]. For every action a system can possibly exhibit, the machine in Figure 3.8 is equipped with a button that is labelled by the name of the action. This does not include *internal* actions. The observer can issue an experiment by providing continuous pressure to a single button, say button *a*. Now, there are three possible ways the machine can react to the experiment:

1. The machine does not react at all. In this case, we say the test has *failed*.
2. The machine can start to work, which is indicated by the status light. Then, after some while, the button, which the observer is still keeping under pressure, will lock down in order to signal that the machine has *accepted* the experiment *a*. After locking down, the machine will finally stop working. When the observer releases the button, it will unlock again. In order to continue his experimentation, the observer has to press and hold any of the buttons again. Note that in order to issue a sequence of tests with the same button, the observer has to release the button after each acceptance of the machine. In this way, he can clearly distinguish how often the same action has occurred in sequence.
3. The machine starts working, however, it never accepts the experiment by either coming to a stop again before the button locks down, or by working forever without ever accepting the experiment.

We can clearly recognize that internal activity can occur at every point in time during an experiment. It remains unobservable up to the information provided by the status light that goes on during pressing a button, but the experiment has not been accepted, this means that the machine is performing internal activities. The status light is only a means to enable the observer to distinguish the case where an action has only *not yet* been accepted, or where it will completely fail to

be accepted. Details of internal behaviour, however, remain unobservable. Especially the precise number of internal transitions a machine performs when the light goes on cannot be deduced.

Besides a button for every action, and the status light, a machine is equipped with a *replication* button. The button can be pressed at every moment, and generates a (countable infinite) number of exact copies of the machine. All copies as well as the original machine are in exactly the same state the original machine was in before. This replication facility allows the observer to perform multiple tests for the each state of the machine. For example, in this way, he can extensively test which experiments are accepted by a machine in a certain state and which not. Making an arbitrary number of copies of the machine in every state allows to fully explore the (weak) branching structure of a system. Even though from an intuitive perspective this button is unrealistic, it is unavoidable if we want to ensure that testing equivalent process agree on their branching structure.

We finally want to note that one further needs to assume that the observer needs to have the additional power to influence the way the machine resolves non-determinism in such a way that it is guaranteed that the machine cannot constantly avoid to show certain aspects of its behaviour. For example, if the machine can react in a certain state to pushing button *a* followed by button *b* both by accepting *a* and *b*, and in a way where *a* will be accepted, but not *b*, it is guaranteed that, if the observer has two copies of the machine in this very state, he will observe both outcomes of this two step experiment. Milner called this assumption *exposing the machine to different weather conditions*.

### 3.2.4. Summary

We have reviewed the two most well-known notions of weak bisimilarity for labelled transitions systems, Milner's original notions of strong and weak bisimilarity. We have discussed how they characterize *observable behaviour* in terms of readiness and failure, and highlighted how they preserve the non-deterministic branching structure of a system virtually *by definition*, in addition to being congruence relations. Moreover, weak LTS bisimilarity has been designed to be ignorant of the internal structure of a system as much as possible. These facts make weak LTS bisimilarity a suitable candidate for an observational equivalence in the sense of Remark 1.1.

In addition, we have illuminated bisimilarity from an experimenter's perspective and argued that it faithfully captures a notion of observability that is characterized by a set of intuitively reasonable experiments upon systems.

In summary, this underpins that weak LTS bisimilarity is an ideal foundation to build upon our search for candidate notions of observational equivalence in the probabilistic and stochastic realms in this thesis.

## 3.3. Decision Algorithm

The practical value of bisimilarity, be it its strong or its weak variant, critically depends on the ability to algorithmically decide whether two LTS are bisimilar. In the following, we will present such algorithms. In fact, the algorithms solve the more general problem of computing equivalence classes of states with respect to bisimilarity. With this information at hand, it is easy to decide whether two states are bisimilar, namely by checking if they fall into the same

$Decide(\mathcal{A})$
<pre> 1: <math>\mathbb{W} = \{S\}</math>; 2: <b>repeat</b> 3:   <math>\mathbb{W}' = \mathbb{W}</math>; 4:   <math>(\mathcal{C}, a, \mathcal{C}') = FindSplit(\mathbb{W})</math>; 5:   <math>\mathbb{W} = Refine(\mathbb{W}, (\mathcal{C}, a, \mathcal{C}'))</math>; 6: <b>until</b> <math>\mathbb{W} = \mathbb{W}'</math> 7: <b>return</b> <math>\mathbb{W}</math> </pre>

**Figure 3.9.:** Compute partitioning  $\mathbb{W}$  according to  $\approx_{LTS}$ .

$FindSplit(\mathbb{W})$
<pre> 1: <b>for all</b> <math>(s, a, s') \in \mathcal{T}</math> <b>do</b> 2:   <b>for all</b> <math>t \in [s]_{\mathbb{W}}</math> <b>do</b> 3:     <b>if not</b> <math>\exists s'' \in [s]_{\mathbb{W}} : (t, a, s'') \in \mathcal{T}</math> 4:       <b>then</b> 5:         <b>return</b> <math>([s]_{\mathbb{W}}, a, [s']_{\mathbb{W}})</math> 5: <b>return</b> <math>(\emptyset, \tau, \delta(\perp))</math> </pre>

**Figure 3.10.:** Finding a splitter

equivalence class. Bisimilarity of two LTS can then be decided by first combining the two LTS into a single one by constructing their disjoint union and then checking bisimilarity of the initial states in the new LTS.

The algorithms for both the strong and the weak case are based on the idea of successively refining a partitioning of states, initially starting with a safe over-approximation [KS90; PT87; Fer90]. The algorithm starts out with one single partition comprising the whole state space. In each iteration of the refinement loop, a partition is split when it contains at least two states that cannot be bisimilar given the current partitioning. The information, that led to the splitting, called *splitter*, is then used to divide the partition into those states that satisfy the splitter, and those that do not. When no more refinement is possible, it is guaranteed that all partitions satisfy the bisimulation conditions, and thus, are bisimilar. As we start from an over-approximation of bisimilarity, it is guaranteed that the final result is as coarse as possible, and thus agrees with bisimilarity and is not any finer bisimulation relation instead. We summarize the approach in Figure 3.9, which is common to both algorithms, i.e. for strong and weak bisimilarity.

The two algorithms only differ in the way splitters are computed. A splitter consists of

- a spurious equivalence class  $\mathcal{C}$  (i.e. a partition) that should be split because at least two of its states being incompatible with respect to their behaviour *and*
- the source of the incompatibility: an action  $a$  and a partition  $\mathcal{C}'$ , with the property that  $\mathcal{C}$  contains states *do* have a transition labelled by  $a$  to a state in  $\mathcal{C}'$ , and also states that *do not* have such a transition.

When we have been speaking of transitions before, this is only correct for strong bisimilarity. For weak bisimilarity, we actually need to consider weak transitions instead. In fact, this is also the only difference between the two algorithms.

Splitters are sought for in the subroutine *FindSplit*, which we will explain later. The subroutine *Refine* then computes the actual splitting of the spurious partition  $\mathcal{C}$ , given the splitter  $(\mathcal{C}, a, s')$  as a parameter (together with the partitioning  $\mathbb{W}$ ). We will not provide details of this straightforward procedure.

Finding a splitter is also straightforward. Let in the following  $\mathcal{T} = \dashv\!\rightarrow$  in the case of strong bisimilarity, and  $\mathcal{T} = \Rightarrow$  in the case of weak bisimilarity. *FindSplit* returns  $(\emptyset, \tau, \delta(\perp))$  to indicate that no suitable splitter has been found. *Refine* needs to be implemented such that it does not perform any splitting of partitions if it receives this splitter as a parameter. In this way, the outer main loop terminates once *FindSplit* has failed to find further splitters.

Note that in the case of weak bisimilarity the relation  $\mathcal{T} \Longrightarrow$  has to be computed a priori, while for strong bisimilarity  $\longrightarrow$  is immediately available as it is the original transition relation of the input LTS. As a matter of fact, the computation of  $\Longrightarrow$  dominates the complexity of the algorithm for weak bisimilarity. Still, both strong and weak bisimilarity can be decided in polynomial time. The complexity of the above partition refinement approach is in  $\mathcal{O}(m \log(n))$ , where  $m$  is the number of transitions and  $n$  the number of states, if implemented suitably. The complexity of weak bisimilarity is dominated by the computation of the reflexive transitive closure of internal transitions, which is owed to the usage of weak transitions. Naive implementations of transitive closure have cubic complexity. More advanced approaches are able to establish that the computational complexity is below  $\mathcal{O}(n^{2.376})$  [CW87] or, only recently, below  $\mathcal{O}(n^{2.3727})$  [Wil12].

### 3.4. Summary and Discussion

In this chapter, we have introduced labelled transition systems as the most basic model of concurrent and reactive system we will consider in this thesis. We have reviewed strong and weak notions of bisimilarity, and discussed their semantic properties. Finally, we have seen that the relations can be decided by efficient polynomial time algorithm that makes them valuable tools in applications. Weak LTS bisimilarity, moreover, has turned out an ideal candidate for observational equivalence that satisfies all requirements of Remark 1.1. These facts together make weak LTS bisimilarity a valuable blueprint for our search for candidate notions of observational equivalence in the probabilistic and stochastic realms.





## 4. Probabilistic Automata

Probabilistic automata are a probabilistic extension of labelled transition systems. The main difference is that transitions no longer lead deterministically to a single state, but a probabilistic experiment may decide in which concrete state a transition will result, each time it is executed. While LTS are very well-accepted for formal specification and verification for concurrent, reactive and distributed systems in the non-probabilistic settings, nevertheless, a lot of typical examples for formal analysis show the need of analysing probabilistic aspects. Even a simple model of a noisy channel becomes more precise if it includes a probabilistic choice between correct and wrong receipt of the value just sent instead of a non-deterministic one. Thus, probabilistic extensions of LTS, which we will refer to as probabilistic automata, have been introduced in the past in various variants. In this thesis, we resort to a variant often called *simple* probabilistic automata, as introduced by Roberto Segala [SL94].

**Outline and Contributions.** In this chapter, we give an overview of the state-of-the-art theory of simple probabilistic automata, which we will call probabilistic automata in the following. The model itself will be introduced in Section 4.1. We will also provide a short overview over related models and how they differ from simple probabilistic automata. As for LTS, parallel composition and hiding will be defined. The synchronization primitive is in *CSP*-style. In Section 4.2, the standard strong and weak probabilistic bisimilarities will be introduced. Section 4.3 reviews the decision algorithm for weak probabilistic bisimilarity of [CS02]. In Section 4.4, we compare the parallel composition operator that we have introduced to the original parallel composition operator of PA. Furthermore, we discuss variations of weak bisimilarity with alternative definitions of weak transitions, and their semantic properties, concluding the chapter.

### 4.1. Model

The variant of probabilistic automata we use in this thesis finds wide acceptance in the literature [Bai96; BK00; BK97; JY02; KN98; SV99; SV03].

Our definition deviates slightly from the original definition in [Seg95]. The difference is in the way we treat internal actions. We distinguish between a countable set of external actions and the unique internal action  $\tau$ . In the original definition, the set of actions is also partitioned into two disjoint sets of external and internal actions. However, the set of internal actions may contain an arbitrary number of distinguishable actions. Intuitively, the original definition thus allows the modeller to distinguish internal processes by a name, while in our definition, every internal activity is tagged by one single label that does not allow for any further distinction.

This adaption also allows us assume that the set of actions  $Act$  is the same for every PA we consider. This will simplify otherwise tedious technical aspects of many considerations.

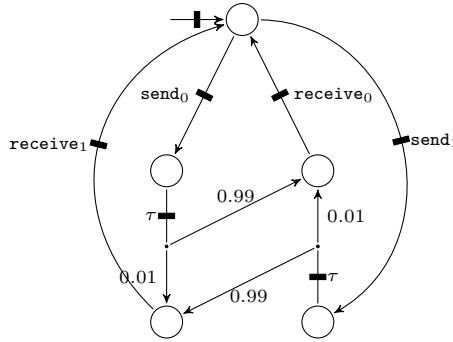
This difference bears no important semantic consequences, while at the same time it allows us to present all model of this thesis in a uniform way. We will make a remark, when technical

differences will arise later, notably in the definition of parallel composition.

**Definition 4.1 (Probabilistic Automaton).** A *Probabilistic Automaton* (PA) is a quadruple  $(S, \bar{s}, Act, \multimap)$ , where

- $S$  is a non-empty countable set of states,
- $\bar{s}$  is the initial state,
- $Act$  is a countable set of actions containing the internal action  $\tau$ ,
- $\multimap \subseteq S \times Act \times Dist(S)$  is a set of probabilistic transitions.

◁



**Figure 4.1.:** Probabilistic Version of a Simple Communication Channel

**Example 4.1.** In the LTS version of the noisy channel, we modelled a transmission error with a non-deterministic choice. The probabilistic transitions improves the model by specifying the probability 0.01 of an erroneous transmission. ◁

We now provide a quick overview over other existing variants of probabilistic automata and how they compare to the variant of Segala. For a more in depth overview, we refer the reader to [SV04], where this section also draws its inspiration from.

The variants of probabilistic automata mainly differ in their transitions. The most basic variant of probabilistic automata are reactive probabilistic automata [GSS95; Gla+90; LS91; LS89; LS92] and generative probabilistic automata [GSS95; Gla+90; CC92; Chr90; CSZ92; Har02; Har99]. In reactive probabilistic automata, actions are treated as inputs provided by the environment. In contrast, generative probabilistic automata generate actions as output. On the level of the transitions, this can be seen by an important difference: transitions in the reactive setting lead from a state *and* an action to a distribution over states, transitions in the generative setting lead from a state to distributions over actions and states. Thus, in the former, the current state of the automaton and the action that it receive as input completely determine its behaviour – up to probabilistic choices; the automaton *reacts* to input with a probabilistic decision. Reactive probabilistic automata are *deterministic* in every moment during execution, once all input actions are fixed. In the generative setting, the automaton makes in each state a probabilistic decision about *both* which state *and* action will occur next. It thus determines probabilistically the action

that is to occur next; it *generates* them. These type of automata offers no non-deterministic decisions.

Another category of probabilistic automata are alternating probabilistic automata [Han91; HJ94]. Here, two types of transitions exist. The so-called probabilistic transitions lead to a distribution over states, and do not involve any action. The so-called non-deterministic transitions lead to a set of pairs of actions and states, representing a non-deterministic decision.

The automata are named alternating because in each state, exactly one type of transition exists. In this way, we can partition the state space according to the type of the outgoing transition into probabilistic states and non-deterministic states. If probabilistic and non-deterministic states alternate strictly along any path, the automaton is called *strictly alternating* [Han91; HJ94; And99; And02].

The variant of probabilistic automata that we focus on in this thesis and that we have defined in Definition 4.1 extends reactive probabilistic automata by full non-determinism. They are sometimes also referred to as *Simple (Segala) probabilistic automata* [Sto02b; Sto02a]. A more general version are *Segala probabilistic automata* [Seg95; SL94] (note the omission of *simple*). There, transitions lead from a state to a distribution over actions and states. In contrast to generative probabilistic automata, in each state there potentially is a non-deterministic choice between the transitions. Note that despite to the coincidence of names, this is *not* the variant of probabilistic automata from Definition 4.1.

We finally want to note that this overview has not been exhaustive, but focuses solely on the variants that will appear in this thesis.

## Parallel composition

We will now define parallel composition of two probabilistic automata. Similar to labelled transition systems, this definition resembles the cross-product of the two automata, with special care taken for synchronous communication. At the same time, we must now specify the probability distribution over the goal states of a transition. Transitions of the parallel composition can either result from the interleaving of the individual transitions of the automata, or from a synchronous execution of two transitions with the same label. In the former case, the transitions are executed independently of the transitions of the other automaton. This means, while one automaton executes its transition, the other automaton stays where it is. Hence, the goal distribution over states remain unchanged. In the latter case, we must determine the probability with which the two automaton reach certain states at the same time. For each pair of possible goal states  $t$  and  $t'$ , this is the probability that the first automaton reaches  $t$  *and* the second automaton reaches  $t'$ . By elementary probability theory, this corresponds to the product of both probabilities, assuming stochastic independence. The following definition formalizes the product of two distributions over states according to our discussion.

**Definition 4.2 (Product of Distributions).** For two distribution  $\mu$  and  $\gamma$  over  $S$ , we define

$$\mu \otimes \gamma := \langle ((s, t) : \mu(s) \cdot \gamma(t)) \mid s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \rangle.$$

◁

In order to denote parallel composition of two Markov automata, we will write  $\mathcal{A} \parallel_A \mathcal{A}'$ , where  $A$  is an arbitrary set of actions excluding the internal action  $\tau$ . It is notationally helpful to use the symbol  $\parallel_A$  as syntactic sugar as follows:

<table><tr><td><math>a</math></td><td><math>b</math></td></tr><tr><td>0.6</td><td>0.4</td></tr></table>	$a$	$b$	0.6	0.4	$\otimes$	<table><tr><td><math>s</math></td><td><math>t</math></td><td><math>u</math></td></tr><tr><td>0.3</td><td>0.2</td><td>0.5</td></tr></table>	$s$	$t$	$u$	0.3	0.2	0.5		
$a$	$b$													
0.6	0.4													
$s$	$t$	$u$												
0.3	0.2	0.5												
<hr/>														
<table><tr><td><math>a, s</math></td><td><math>a, t</math></td><td><math>a, u</math></td><td><math>b, s</math></td><td><math>b, t</math></td><td><math>b, u</math></td></tr><tr><td>0.18</td><td>0.12</td><td>0.3</td><td>0.12</td><td>0.08</td><td>0.2</td></tr></table>			$a, s$	$a, t$	$a, u$	$b, s$	$b, t$	$b, u$	0.18	0.12	0.3	0.12	0.08	0.2
$a, s$	$a, t$	$a, u$	$b, s$	$b, t$	$b, u$									
0.18	0.12	0.3	0.12	0.08	0.2									

**Figure 4.2.:** Product of two Distributions

- If  $s$  and  $s'$  are two states, we write  $s \parallel_A s'$  to denote the pair  $(s, s')$ .
- If  $\mu$  and  $\gamma$  are two distributions over states, we write  $\mu \parallel_A \gamma$  to denote the distribution  $\mu \otimes \gamma$ .

For intuition, it is also convenient to use this notation for pairs of states  $(s, t)$  within the product of two distributions. We write

$$\mu \parallel_A \gamma = \langle (s \parallel_A t : \mu(s) \cdot \gamma(t)) \mid s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \rangle$$

instead of

$$\mu \parallel_A \gamma = \langle ((s, t) : \mu(s) \cdot \gamma(t)) \mid s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \rangle.$$

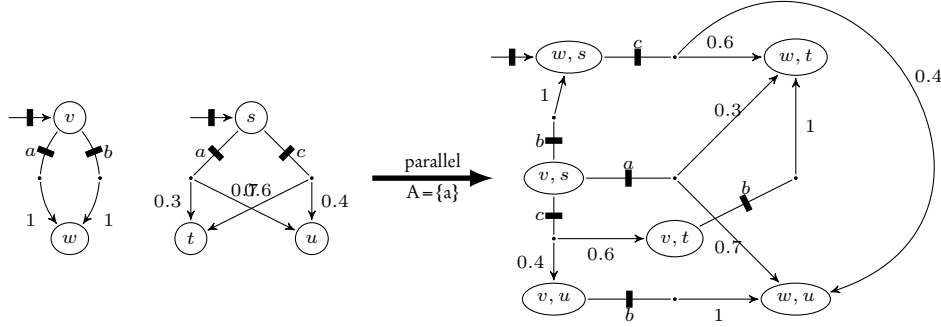
**Definition 4.3 (Parallel composition).** Given to PA  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv\!\!\rightarrow)$  and  $\mathcal{A}' = (S', \bar{s}', \text{Act}, \dashv\!\!\rightarrow')$ , their parallel composition  $\mathcal{A} \parallel_A \mathcal{A}'$  is defined as the automaton  $\mathcal{A}^* = (S^*, \bar{s}^*, \text{Act}, \dashv\!\!\rightarrow^*)$  where

- $S^* = S \times S'$
- $\bar{s}^* = \bar{s} \parallel_A \bar{s}'$
- $\dashv\!\!\rightarrow^*$  is the least relation satisfying

$$\begin{array}{ll} s \parallel_A s' \dashv\!\!\rightarrow^* \mu \parallel_A \mu' & \text{if } a \in A \text{ and } s \dashv\!\!\rightarrow \mu \text{ and } s' \dashv\!\!\rightarrow' \mu' \\ s \parallel_A s' \dashv\!\!\rightarrow^* \mu \parallel_A \delta(s') & \text{if } a \notin A \text{ and } s \dashv\!\!\rightarrow \mu \\ s \parallel_A s' \dashv\!\!\rightarrow^* \delta(s) \parallel_A \mu' & \text{if } a \notin A \text{ and } s' \dashv\!\!\rightarrow' \mu'. \end{array}$$

◁

As we have seen, parallel composition is the straightforward probabilistic extension of parallel composition for labelled transition systems.



**Figure 4.3.:** Parallel Composition of two Probabilistic Automata

**Definition 4.4 (Abstraction).** Given a PA  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\cdot})$ , its abstraction  $\mathcal{A}|_A = (S', \bar{s}', Act, \xrightarrow{\cdot})$  with respect to the set of actions  $A \subseteq Act \setminus \{\tau\}$  satisfies

- $S' = S$
- $\bar{s}' = \bar{s}$
- $\xrightarrow{\cdot}$  is the least relation satisfying

$$\begin{array}{ll}
 s \xrightarrow{a} \mu & \text{if } a \notin A \text{ and } s \xrightarrow{a} \mu \\
 s \xrightarrow{\tau} \mu & \text{if } a \in A \text{ and } s \xrightarrow{a} \mu
 \end{array}$$

◁

## 4.2. Bisimilarity

Probabilistic automata differ from labelled transition systems in that a system may perform a probability experiment within every transition. The outcome of the experiment determines to which state the system proceeds. Formally, this is modelled as a probability distribution over all possible (successor) states. When an observer monitors a system that is modelled as a probabilistic automaton, she can perceive the same basic observations as for LTS: the executable actions. In addition, we must take into consideration that she can in some way observe with what *probability* a specific observation can be made in each point in time.

We will study notions of observation formalized by different established equivalence notions in the strong and in the weak setting.

We defer the full discussion in how far the existing equivalences are suitable candidates of an observational equivalence for PA to Chapter 8. In Section 4.2.3, we will provide an intuitive discussion of the existing relations with respect to the experimenter's perspective on behaviour (echoing Section 3.2.3).

### 4.2.1. Strong Bisimilarity

In probabilistic automata, any transition of a state leads to a probability distribution over states. While in a strong LTS bisimulation the goal states of the challenger's and the defender's transition need to match, in the probabilistic setting we ask that the probabilities of reaching bisimilar states agree. In this section, we will introduce two standard notions of strong bisimilarity on PA, namely strong probabilistic bisimilarity, and strong probabilistic distribution bisimilarity.

Technically, incorporating the probability of reaching specific successor states in the formulation of bisimulation is realized by defining strong bisimulations as equivalence relations and demanding that the distribution of both transitions assign the same probability mass to each equivalence class with respect to the bisimulation. This elementary idea stems from [JL91; LS89; LS91].

We introduce a special notation to express that two distributions over states agree with respect to the probability they assign to the individual equivalence classes induced by an equivalence relation  $\mathcal{R}$ .

**Definition 4.5 (Lifting).** The lifting  $\mathcal{L}(\mathcal{R}) \subseteq \text{Subdist}(X) \times \text{Subdist}(X)$  [LS89] of an equivalence relation  $\mathcal{R}$  on  $X$  is defined as: for  $\mu_1, \mu_2 \in \text{Subdist}(X)$ ,  $\mu_1 \mathcal{L}(\mathcal{R}) \mu_2$  if and only if for each  $C \in X/\mathcal{R}$ ,  $\mu_1(C) = \mu_2(C)$ .  $\triangleleft$

**Lemma 4.1 (Linearity of Liftings).** For any finite or countable infinite index set  $I$  the following holds: if  $\xi_i \mathcal{L}(\mathcal{R}) \nu_i$  for each  $i \in I$  then also

$$\bigoplus_{i \in I} c_i \xi_i \mathcal{L}(\mathcal{R}) \bigoplus_{i \in I} c_i \nu_i.$$

Strong bisimilarity for probabilistic automata is defined as follows.

**Definition 4.6 (Strong Bisimulation).** Let  $(S, \bar{s}, \text{Act}, \dashv\rightarrow)$  be a PA. An equivalence relation  $\mathcal{R}$  on  $S$  is a *strong bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \dashv\overset{a}{\rightarrow} \mu$  for some  $a \in \text{Act}$  and  $\mu \in \text{Dist}(S)$  implies  $t \dashv\overset{a}{\rightarrow} \gamma$  for some  $\gamma' \in \text{Dist}(S)$  with  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .  $\triangleleft$

Below Definition 3.4 on page 27, we discussed the aspects of a system's structure that are deemed observable by virtue of the equivalence classes induced by strong LTS bisimilarity. We resume this discussion for strong bisimilarity for PA. The observations of strong LTS bisimilarity and strong bisimilarity for PA agree with respect to the readiness and failure of actions in the individual states, as well as with respect to the preservation of the non-deterministic branching structure between bisimilar systems. The difference is that the probability distribution of reaching a certain behaviour, expressed as an equivalence class over states with respect to strong LTS bisimilarity, is now taken into account.

However, the consequences of the probabilistic setting are not considered in their full consequence with respect to non-deterministic decisions. It is reasonable to assume that in a probabilistic setting, non-deterministic decisions may not only be resolved by choosing *one* of the

alternatives, as it is the case in LTS, but also by means of a *probabilistic* choice over all alternatives. This perspective is, however, not represented in strong LTS bisimilarity.

An alternative proposal for a strong bisimilarity on PA, strong probabilistic bisimilarity, takes this idea up.

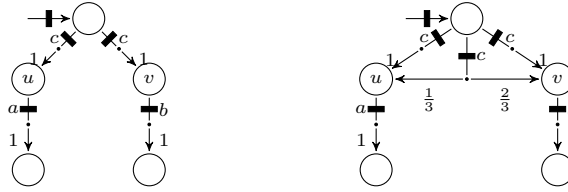
**Definition 4.7 (Strong Combined Transitions).** We write  $s \xrightarrow{a}_c \mu$ , if  $a \in Act$  and there is a finite indexed set  $\{(c_i, \mu_i)\}_{i \in I}$  of pairs of positive real valued weights and distributions such that  $s \xrightarrow{a} \mu_i$  for each  $i \in I$  and  $\sum_{i \in I} c_i = 1$  and  $\mu = \bigoplus_{i \in I} c_i \mu_i$ .  $\triangleleft$

**Example 4.2.** In Figure 4.4, the automaton on the left is missing the middle transition in comparison to the automaton on the right. This middle transition is a convex combination of the two transitions to  $u$  and  $v$ , respectively, with weights  $\frac{1}{3}$  and  $\frac{2}{3}$ . Still, for the automaton on the left, we can justify that there exists a *strong combined transition*  $t \xrightarrow{c}_c \langle (\frac{1}{3}, u), (\frac{2}{3}, v) \rangle$  that can simulate the missing transition.  $\triangleleft$

Note that the non-probabilistic resolution of non-determinism as we find it in LTS, now becomes a special case. Choosing a single transition corresponds to a Dirac distribution, where the chosen transition is assigned probability 1 and all others probability 0.

**Definition 4.8 (Strong Probabilistic Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\cdot})$  be a PA. An equivalence relation  $\mathcal{R}$  on  $S$  is a *strong probabilistic bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} \mu$  for some  $a \in Act$  and  $\mu \in Dist(S)$  implies  $t \xrightarrow{a}_c \gamma$  for some  $\gamma' \in Dist(S)$  with  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .  $\triangleleft$

We write  $s \sim_{PA} t$  if some strong probabilistic bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two PA  $\mathcal{A}$  and  $\mathcal{A}'$  are strong probabilistic bisimilar if in their disjoint union the two initial states are strong probabilistic bisimilar, i.e.  $\bar{s} \sim_{PA} \bar{s}'$ .



**Figure 4.4.:** Two Strong Probabilistic Bisimilar PA

It is straightforward to show that strong probabilistic bisimilarity is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two composition operators we have defined.

**Example 4.3.** The two automata in Figure 4.4 are strong probabilistic bisimilar, but not with respect to strong bisimilarity for PA.  $\triangleleft$

**Theorem 4.1.** Strong probabilistic bisimilarity  $\sim_{PA}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

We have seen that generalizing strong bisimilarity to PA can be done in two variants: a more conservative one, where non-deterministic decisions are resolved by choosing one of the alternative transitions, and one that is more apt for the probabilistic setting, as it allows to resolve non-deterministic choice probabilistically. In fact, the former variant can be considered as a special case of the latter. The probabilistic variant of strong bisimilarity,  $\sim_{\text{PA}}$ , is thus the most general, yet conservative extension of strong bisimilarity to the probabilistic setting.

### 4.2.2. Weak Transitions in a Probabilistic Setting

Any notion of weak bisimulation necessarily relies on weak transitions. Since the formalization of weak transitions for probabilistic automata is technically involved, we dedicate this section to this task, before we begin our discussion of weak bisimilarity for PA in Section 4.2.3 on page 48. Besides the definitions, we also provide a few essential lemmas for weak transitions.

The notion of weak transition we introduce here is more general than the standard definition. The additional expressiveness is not needed in the course of this chapter, but will be elementary for our deliberations in the second part of the thesis. We therefore advise also the reader proficient in probabilistic automata to take special notice of Definition 4.13 and Remark 4.1. This section only focuses on technical aspects of weak transition. Their importance for weak bisimulation and advanced semantic discussions will follow in Section 4.2.3.

In the literature, several approaches of defining weak transitions exist, for example, via probabilistic execution in [Seg95], trees [Des+10], or infinite sum [Den+09]. As shown recently, these formalizations are effectively equivalent [Bre13]. It is thus irrelevant for the resulting bisimulation relations, which formalization is chosen. We mainly will rely on the tree characterization introduced in [EHZ10a]. Except for some of the proofs in the appendix, which use probabilistic executions. In Appendix A, we introduce the necessary definitions.

**Weak Transitions** The definition of weak transitions needs several auxiliary concepts.

**Definition 4.9 (Sequence Order).** For two finite sequences of positive integers  $\sigma, \sigma' \in \mathbb{N}_{>0}^*$  we write  $\sigma \leq \sigma'$  if there exists a (possibly empty) sequence  $\phi$  such that  $\sigma\phi = \sigma'$ .  $\triangleleft$

We will now define labelled trees by means of a set of sequences of positive integers. Each sequence represents a node of the tree by encoding its address. The labelling is realized by a partial function mapping each node to its label.

**Definition 4.10 (Labeled Tree).** A partial function  $\mathcal{T} : \mathbb{N}_{>0}^* \rightarrow L$  is called an (*infinite*) *L-labelled tree*, if it satisfies the following conditions

1. for all  $\sigma, \sigma' \in \mathbb{N}_{>0}^*$ :  $\sigma \leq \sigma'$  and  $\sigma' \in \text{dom}(\mathcal{T})$  implies  $\sigma \in \text{dom}(\mathcal{T})$
2.  $\sigma i \in \text{dom}(\mathcal{T})$  for  $i > 1$  implies  $\sigma(i-1) \in \text{dom}(\mathcal{T})$
3.  $\varepsilon \in \text{dom}(\mathcal{T})$

$\triangleleft$

*Notation 4.1.*

- A node  $\sigma \in \text{dom}(\mathcal{T})$  is called a leaf of  $\mathcal{T}$  if there is no  $\sigma' \in \text{dom}(\mathcal{T})$  such that  $\sigma < \sigma'$ .



- $\varepsilon$  is called the root of  $\mathcal{T}$ .
- For a node  $\sigma$  let  $Children(\sigma) = \{\sigma i \mid \sigma i \in dom(\mathcal{T})\}$ .
- We denote the set of all leaves of  $\mathcal{T}$  by  $Leaf_{\mathcal{T}}$ . The nodes that are no leaves are called inner nodes. We subsume them in the set  $Inner_{\mathcal{T}}$ . If the tree has only one node, the root node, then this node is assumed to be contained in both  $Inner_{\mathcal{T}}$  and  $Leaf_{\mathcal{T}}$ . In any other case the two sets are disjoint.

We consider in the following  $S \times \mathbb{R}_{>0} \times (Act \cup \{\perp\})$ -labelled trees. In principle, they represent a probabilistic execution fragment of a probabilistic automata. The first label component denotes the states along which the execution evolves, the last component denotes the action, and the middle component the accumulated probability of reaching this node from the node the root is labelled with. For a node  $\sigma$  we write  $Sta_{\mathcal{T}}(\sigma)$  for the first component of  $\mathcal{T}(\sigma)$ ,  $Prob_{\mathcal{T}}(\sigma)$  for the second component of  $\mathcal{T}(\sigma)$  and  $Act_{\mathcal{T}}(\sigma)$  for the third component of  $\mathcal{T}(\sigma)$ .

**Notation 4.2.** For a set  $\check{A} \subseteq \multimap$ , a node  $\sigma$ , an action  $a \in Act$  and a distribution  $\mu$  over  $S$ , we write  $\sigma \xrightarrow{a}_{\check{A}} \mu$  to express that  $a = Act_{\mathcal{T}}(\sigma)$ ,  $(Sta_{\mathcal{T}}(\sigma), Act_{\mathcal{T}}(\sigma), \mu) \in \check{A}$  and

$$Prob_{\mathcal{T}}(\sigma) \cdot \mu = \langle (Sta_{\mathcal{T}}(\sigma'), Prob_{\mathcal{T}}(\sigma')) \mid \sigma' \in Children_{\mathcal{T}}(\sigma) \rangle.$$

Whenever  $\check{A}$  is clear from the context, we omit its subscript from the transition arrow.

**Definition 4.11.** Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap)$  be a PA and  $\check{A} \subseteq \multimap$ . A *transition tree*  $\mathcal{T}$  over  $\check{A}$  is a  $S \times \mathbb{R}_{>0} \times (Act \cup \{\perp\})$ -labelled tree that satisfies the following condition:

1.  $Prob_{\mathcal{T}}(\varepsilon) = 1$ ,
2.  $\forall \sigma \in Leaf_{\mathcal{T}} : Act_{\mathcal{T}}(\sigma) = \perp$ ,
3.  $\forall \sigma \in Inner_{\mathcal{T}} \setminus Leaf_{\mathcal{T}} : \exists \mu : \sigma \xrightarrow{Act_{\mathcal{T}}(\sigma)}_{\check{A}} \mu$

◁

Note that the definition implies that  $\sum_{\sigma \in Leaf_{\mathcal{T}}} Prob_{\mathcal{T}}(\sigma) = Prob_{\mathcal{T}}(\varepsilon) = 1$ . We now explain how a transition tree  $\mathcal{T}$  corresponds to a probabilistic execution fragment: it starts from  $Sta_{\mathcal{T}}(\varepsilon)$ , and resolves the non-deterministic choice by executing a probabilistic transition with the action  $Act_{\mathcal{T}}(\sigma)$  at the inner node  $\sigma$ . The second label of  $\sigma$  is then the probability of reaching  $Sta_{\mathcal{T}}(\sigma)$ , starting from  $Sta_{\mathcal{T}}(\varepsilon)$  and following the selected transitions. If the action label of a node  $\sigma$  indicates a timed transition, i.e.  $Act_{\mathcal{T}}(\sigma) = \chi(r)$  for some real number  $r$ , then  $r$  represents the exit rate of  $Sta_{\mathcal{T}}(\sigma)$ . In this case, a child  $\sigma'$  is reached with  $Prob_{\mathcal{T}}(\sigma)$  times the discrete branching probability  $P(Sta_{\mathcal{T}}(\sigma), Sta_{\mathcal{T}}(\sigma'))$  of the underlying exponential distribution.

For two transition trees  $\mathcal{T}$  and  $\mathcal{T}'$ , we say  $\mathcal{T}$  is a prefix of  $\mathcal{T}'$ , written  $\mathcal{T} \leq \mathcal{T}'$  if  $dom(\mathcal{T}) \subseteq dom(\mathcal{T}')$  and  $\forall \sigma \in Inner_{\mathcal{T}'} \setminus Leaf_{\mathcal{T}'} : \mathcal{T}(\sigma) = \mathcal{T}'(\sigma)$  and  $\forall \sigma \in Leaf_{\mathcal{T}'} : \text{either } \mathcal{T}(\sigma) = \mathcal{T}'(\sigma) \text{ or } \sigma \notin Leaf_{\mathcal{T}} \text{ and } Sta_{\mathcal{T}}(\sigma) = Sta_{\mathcal{T}'}(\sigma) \text{ and } Prob_{\mathcal{T}}(\sigma) = Prob_{\mathcal{T}'}(\sigma)$ . An *internal transition tree*  $\mathcal{T}$  is a transition tree where each  $Act_{\mathcal{T}}(\sigma)$  is either  $\tau$  or  $\perp$ .

**Definition 4.12 (Distribution Induced by a Transition Tree).** Let  $\mathcal{T}$  be a transition tree. Then the distribution induced by  $\mathcal{T}$ , denoted by  $\mu_{\mathcal{T}}$ , is defined as

$$\mu_{\mathcal{T}} = \bigoplus_{\sigma \in \text{Leaf}_{\mathcal{T}}} \langle (\text{Sta}_{\mathcal{T}}(\sigma), \text{Prob}_{\mathcal{T}}(\sigma)) \rangle.$$

◁

With the above definitions we are now able to define weak transitions.

**Definition 4.13.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashrightarrow)$  be a PA. For  $s \in S$  and  $\mu$  a full distribution we write

- (1)  $s \xRightarrow{a|\check{A}} \mu$  if  $\mu$  is induced by some transition tree  $\mathcal{T}$  over  $\check{A}$  with  $\text{Sta}_{\mathcal{T}}(\varepsilon) = s$ . In case  $a \neq \tau$ , on every maximal path from the root precisely one node  $\sigma$  is labelled  $\text{Act}_{\mathcal{T}}(\sigma) = a$ . All not yet labelled inner nodes must be labelled by  $\tau$ .
- (2)  $s \xRightarrow{a} \mu$  if  $s \xRightarrow{a|\check{A}} \mu$  and  $\check{A} = \dashrightarrow$ .

For all these transition relations we say that the transition tree that induces  $\mu$  also *induces the transition* to  $\mu$ . ◁

Note that  $s \xRightarrow{\phantom{a}} \delta(s)$  always holds independently of the actual transitions  $s$  can perform.

*Remark 4.1.* Usually, one defines weak transitions without the ability to restrict the set of transitions  $\check{A}$  appearing in the transition tree.

It is thus convenient to use separate terms for the more general and the standard variant. We therefore speak of weak *allowed* transitions if we refer to general weak transitions (Case (1) of Definition 4.13), and of *weak transitions* if we refer to weak transitions following the standard definitions (Case (2)).

**Hyper-Transitions** We lift the notion of transitions of a state to transitions of a distribution over states.

**Definition 4.14 (Hypertransitions).** We say that there is a (weak allowed) *hyper-transition* from  $\rho \in \text{Subdist}(S)$  to  $\nu \in \text{Subdist}(S)$  labelled by  $a \in \text{Act}$ , denoted by  $\rho \xrightarrow{a} \nu$  ( $\rho \xRightarrow{a} \nu$ ), if there exists a family of (weak allowed) transitions  $\{s \xrightarrow{a} \nu_s\}_{s \in \text{Supp}(\rho)}$  ( $\{s \xRightarrow{a} \nu_s\}_{s \in \text{Supp}(\rho)}$ ) such that

$$\nu = \bigoplus_{s \in \text{Supp}(\rho)} \rho(s) \cdot \nu_s.$$

◁

**Combined Transitions** As in the strong case and as discussed, it is reasonable to assume that in a probabilistic setting, internal non-determinism can be resolved probabilistically. Below, we define the notion of *weak combined transitions* [Seg95], which arise as convex combination of a set of weak transitions with the same label. This notion is a straightforward adaption of the related concept for (non-weak) transitions we have introduced before. We only define the concept for weak allowed hyper-transitions, which then carries over to weak allowed transitions immediately.

**Definition 4.15.** We write  $\mu \xRightarrow{a \mid \bar{A}}_c \mu'$ , if  $\mu'$  is an infinitary or finitary convex combination of a family of distributions  $\mu'_i$  with  $\mu \xRightarrow{a \mid \bar{A}} \mu'_i$ .  $\triangleleft$

*Remark 4.2.* If clear from the context, we will often reference all variants of transitions by the term *transition*, regardless of whether they are weak/strong/allowed/hyper transitions.

### Properties of Transitions

**Lemma 4.2 (Composition).** Let  $a \in \text{Act} \cup \{\tau\}$ .

- If  $\mu \xRightarrow{a \mid \bar{A}} \mu'$  and  $\mu' \xRightarrow{\tau \mid \bar{A}} \mu''$  then  $\mu \xRightarrow{a \mid \bar{A}} \mu''$ .
- If  $\mu \xRightarrow{\tau \mid \bar{A}} \mu'$  and  $\mu' \xRightarrow{a \mid \bar{A}} \mu''$  then  $\mu \xRightarrow{a \mid \bar{A}} \mu''$ .

and

- If  $\mu \xRightarrow{a \mid \bar{A}}_c \mu'$  and  $\mu' \xRightarrow{\tau \mid \bar{A}}_c \mu''$  then  $\mu \xRightarrow{a \mid \bar{A}}_c \mu''$ .
- If  $\mu \xRightarrow{\tau \mid \bar{A}}_c \mu'$  and  $\mu' \xRightarrow{a \mid \bar{A}}_c \mu''$  then  $\mu \xRightarrow{a \mid \bar{A}}_c \mu''$ .

By letting  $\bar{A}$  equal the set of all transition, this lemma also applies to ordinary weak (combined) transitions. Furthermore, with letting  $\mu = \delta(s)$  this lemma applies also for sequences of transitions where the first one is not a hyper-, but an ordinary (combined) transition.

This lemma implies that we can compose arbitrary finite sequences of weak (allowed) (combined) (hyper) transitions into one single weak (combined) (hyper) transitions. In other words, every distribution that is reachable by a finite sequences of such transitions is also reachable directly by a single such transition. We often make use of this lemma without explicit notice.

**Lemma 4.3 (Combining Combined Transitions).** Let  $\mu, \mu' \in \text{Dist}(S)$  and  $k \in \mathbb{N}$ . Let  $I$  be a finite or infinitary index set. The hypertransition  $\mu \xRightarrow{a}_c \mu'$  exists *if and only if* there exist distributions  $\mu_i$  and coefficients  $c_i \in [0, 1]$  with  $i \in I$  such that

$$\mu = \bigoplus_{i \in I} c_i \mu_i$$

and distributions  $\mu'_i$  with

$$\mu_i \xRightarrow{a}_c \mu'_i \text{ and } \mu' = \bigoplus_i c_i \mu'_i.$$

In this section, we have introduced weak transitions in terms of labelled transition trees, that correspond to probabilistic execution fragments. As it has been the case for strong transitions, we can incorporate the probabilistic resolution of non-determinism into weak transitions in the form of a convex combination of several weak transitions, leading to weak *combined* transitions. We have also extended weak transition to relations over distributions instead of a state and a distribution, and called the resulting relation weak hypertransitions. They will provide a necessary tool in our development of the bisimilarities for Markov automata. For reasons of convenience, we will refer to hypertransitions as well as transitions. Finally, we have seen several lemmas that deal with recombining weak transitions in various ways, and that will be in permanent use during later proofs.

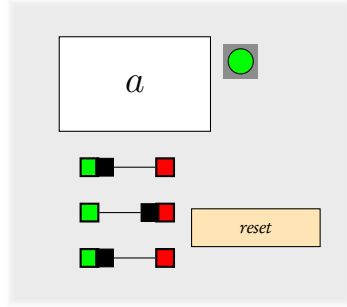


Figure 4.5.: Machine with Reset Button

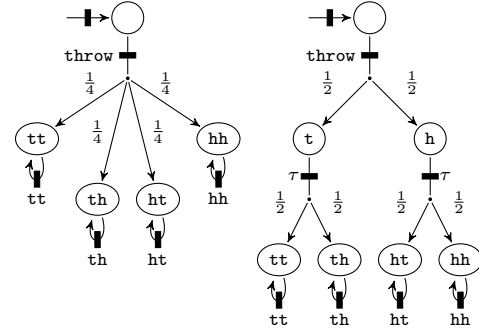


Figure 4.6.: Two Automata with the Same Observable Outcomes

### 4.2.3. Weak Bisimulations and Simulations

As an entry point for our discussion of weak bisimilarities for probabilistic systems, we again take on the perspective of an experimenter, that tries to analyse the observable behaviour of given probabilistic automata by means of different experiments and observations.

From there, we will review important weak process relations, both simulation and bisimulation relations, that implement the experimenter's perspective to different degrees. However, all of them fail to satisfy our requirements for observational equivalences for different reasons. They will, however, serve as a starting point for our investigations in the second part of the thesis, where novel notions of weak bisimilarity for the probabilistic setting will be introduced.

#### The Experimenter's Perspective

Testing scenarios in terms of button-pushing experiments, as we have introduced them in Section 3.2.3 for LTS, have not been investigated extensively for probabilistic systems. While these ideas may have supposedly played an inspirational role in finding various notions of bisimilarity, there is hardly any explicit reference.

Still, testing scenarios for probabilistic systems have been developed and extended gradually over the decades. After the fundamental approach of [LS91; LS89] for the fully probabilistic setting without non-determinism, the first steps in this direction have been undertaken in [SV03; CSV07]. There, a testing scenario based on button-pushing machines is discussed for probabilistic automata. The machine is a strongly reduced version of the machine we have described in Section 3.2.3. It is only equipped with the action buttons, but it misses any further manipulation mechanisms. In particular, a replication is not present. However, the machine has a special *reset* button, that resets the machine (Figure 4.5) to its *initial* state.

In this thought experiment, the concrete probabilities with which experiments succeed are not displayed on the machine, but are determined by the application of statistical methods. The necessary samples are obtained by resetting the machine to its initial state with the reset button over and over again. In this way, it is possible, to determine the exact probabilities with which observations will be made.

**Example 4.4.** Recall the probability experiment from the introduction on page 10 that was illustrated in Figure 1.9. Rolling a four-sided die should be observationally equivalent to flipping

two coins in sequence, as long as only the probability outcomes are presented, and not the technical details how the probabilities have been determined. We illustrate the two situations as a PA in Figure 4.6, which corresponds to Figure 1.9.

The machine on the left, where the die is rolled consists of a single  $\tau$  transition that results in a uniform distribution over the four possible outcomes. The machine on the right splits its coin throwing experiment into two experiments, speak  $\tau$  transitions, where each time a single coin is thrown. At the end, the outcomes, denoted by  $tt$ ,  $th$ ,  $ht$ , and  $hh$  are the same.

In the testing scenario, the two automata are indistinguishable. The fact, that in the automaton on the right includes intermediate states remains transparent. The observer can only see that when the system starts, or when it has been reset, one of the observable actions  $tt$ ,  $th$ ,  $ht$  or  $hh$  occur after some time. The same is true for the automaton on the left. When evaluating the samples, the observer will determine that each of the four outcomes has appeared with probability  $\frac{1}{4}$ .  $\triangleleft$

The testing scenario also allows to observe the *non-deterministic* branching structure of the system under test *in parts*. The reset button corresponds to an unbounded replication button that can be only applied in the initial state. The information revealed about the branching structure in this way is very limited. It does definitely not satisfy our second requirement in Remark 1.1.

In the mentioned papers, it is shown that the behavioural equivalence resulting from this testing scenario is *trace distribution equivalence* [LSV07]. Trace distribution equivalence is the probabilistic counterpart to trace equivalence for the non-probabilistic setting. It basically ensures that the probabilities for each trace to occur are the same in equivalent systems. Note that in the presence of non-determinism, the same trace may have assigned more than one probability number.

From the perspective of our search for an observational equivalence, trace distribution equivalence is not fully satisfactory. Trace distribution does not meet our requirements of a behaviour equivalence for two reasons. First, it does not preserve the branching structure. Second, it fails to be a congruence with respect to parallel composition.

We forego a more formal discussion of trace distribution equivalence, and instead, will next consider a different relation that is based on the same testing scenario, but has the advantage of being a congruence.

**Probabilistic Forward Simulation** The unsymmetric variant of trace distribution equivalence is called probabilistic trace distribution *preorder*. As before, this relation fails to be a congruence. The coarsest *congruence* relation in which probabilistic trace distribution preorder is contained is a relation called probabilistic forward simulation [Seg95; LSV07]. In fact, this relation already posses many of the requirements we demand for an observational equivalence: It satisfies the first requirement of Remark 1.1, namely being a congruence. Also intuitively, it suggests itself to come as close to the third requirement as possible, as it stems from a testing scenario with a very limited set of allowed observations and experiments. Finally, though, the second requirement is not met. As probabilistic forward simulation is a simulation relation it cannot preserve the the non-deterministic branching structure, even when cast to an equivalence relation by taking its symmetric kernel.

In the following we will discuss the formal definition of this relation, as it will play an important role for the further development in the second part of the thesis. We begin with a new

notion of lifting a binary relation between two sets, to distributions over these sets. Note that it differs conceptually from the lifting relation we have introduced in Definition 4.5.

**Definition 4.16.** Let  $\mathcal{R} \subseteq S \times S'$  be a relation. The weighted lifting  $\hat{\mathcal{L}}(\mathcal{R}) \subseteq \text{Dist}(S) \times \text{Dist}(S')$  of  $\mathcal{R}$  relates  $\mu \mathcal{R}' \mu'$  if and only if there exists a function  $w : S \times S' \rightarrow [0, 1]$  such that  $\mu \mathcal{R}' \mu'$  implies:

1.  $w(s, s') > 0$  implies  $s \mathcal{R} s'$ ,
2. for each  $s \in S$ ,  $\sum_{s' \in S'} w(s, s') = \mu(s)$ , and
3. for each  $s' \in S'$ ,  $\sum_{s \in S} w(s, s') = \mu'(s')$ .

◁

This definition splits the distribution  $\mu$  into Dirac subdistributions, proceeds accordingly for  $\mu'$ , and then links them together via  $w$  such that for each state  $s$  in  $\text{Supp}(\mu)$  its probability mass in  $\mu$  is the same as the overall probability mass of the linked subdistributions stemming from  $\mu'$  (Condition 2), and vice versa (Condition 3). Condition 1 ensures that only those subdistributions are linked whose supports are related by  $\mathcal{R}$ .

**Definition 4.17 (Flattening).** For a distribution  $\mu \in \text{Dist}(\text{Dist}(S))$ , let

$$\text{flatten}(\mu) = \bigoplus_{\gamma \in \text{Supp}(\mu)} \mu(\gamma) \cdot \gamma.$$

◁

The flattening operator reduces a distribution over distributions over  $S$  into a distribution over  $S$ , while preserving the accumulated probabilities of the elements in the support of the inner distributions.

**Definition 4.18 (Probabilistic Forward Simulation).** Let  $(S, \bar{s}, \text{Act}, \dashv\dashv)$  be an PA. A relation  $\mathcal{R} \subseteq S \times \text{Dist}(S)$  is called a probabilistic forward simulation if whenever  $s \mathcal{R} \mu_2$  then for all  $a \in \text{Act}$ :

whenever  $s \dashv\dashv \mu'_1$ , then  $\exists \mu'_2 \in \text{Dist}(\text{Dist}(S)) : \mu_2 \xrightarrow{a}_c \text{flatten}(\mu'_2)$  and  $\mu'_1 \hat{\mathcal{L}}(\mathcal{R}) \mu'_2$ . ◁

For a state  $s$  and a distribution  $\mu$ , we write  $s \preceq_{\text{fwd}} \mu$  if  $s \mathcal{R} \mu$  for some probabilistic forward simulation  $\mathcal{R}$ .

**Theorem 4.2.**  $\preceq_{\text{fwd}}$  is the largest probabilistic forward simulation.

The most salient detail of the definition of probabilistic forward simulation is that it relates states to distributions over states. This links the behaviour of a single state in a probabilistic system to a distribution over states and their behaviours. As we will see especially in the second part of this thesis, this concept is key to obtain a relation that is relaxed with respect to probabilistic behaviour. In fact, it is the closest we can come to the experimental perspective with statistic sampling and a reset button.

## Combining the Testing Scenarios

As we have discussed, the probabilistic testing scenario only allows for very limited observations of the non-deterministic branching structure. In our attempt to arrive at a probabilistic testing

scenario, in which also the non-deterministic branching structure is observable, we will now consider a scenario that combines the probabilistic testing scenario with the scenario for LTS.

In the testing scenario for LTS, each system comes equipped with an unbounded replication button, that allows the observer to create an unbounded number of copies of a system in each state. In the probabilistic testing scenario, only a reset button is devised that allows to bring a system back to its initial state. If an unbounded set of copies is created in the initial state with the unbounded replication button, each copy can be treated like a freshly reset system. Therefore, the unbounded replication button can replace the reset button in the combined scenario. From the probabilistic scenario, we inherit the ability of the observer to use sampling of executions and statistical methods in order to derive probability.

In summary, we obtain a scenario in which an observer can apply unbounded replication in every state, and the combination of the unbounded replication button with statistic sampling to insight in the probabilistic aspects of the system.

Consider again the two automata in Figure 4.6. With respect to their non-deterministic branching structure, the two automata agree, as neither of them is subject to any non-probabilistic decision. Therefore, we expect the observations that can be made to agree with those possible in the probabilistic testing scenario. There, the two automata are indistinguishable. Surprisingly, in the combined testing scenario, they can be distinguished, as the points during the execution of the automata, where *probabilistic* decisions are made, become fully observable. In the left automaton only one decision is made, while in the automaton on the right, two decisions are made in sequence.

*Remark 4.3* (Probabilistic Branching Structure). In correspondence with the notion of non-deterministic branching structure of a system, we let the term *probabilistic branching structure* refer to the points during an execution of a system, where *probabilistic* decisions are made what future observable behaviour will be possible or impossible.

The reason why in this testing scenario the probabilistic branching structure becomes observable is rooted in the fact that the sampling, which the observer needs in order to derive probabilities is no longer restricted to executions starting in the initial state, as it was the case in the probabilistic testing scenario. As the unbounded replication button can be applied in every state, also those from which only  $\tau$ -transitions leave, a new set of samples can be generated for execution starting in every state, too.

**Example 4.5.** Recall that the observer can press the replication button in every state and create an unbounded number of copies of the system in that state. When the observer keeps creating copies (by continuous pressure on the button) on the automaton on the right in Figure 4.6, he will also create an unbounded number of copies of state  $t$  and  $h$ , respectively, depending on how the probabilistic choice has been resolved before.

By sampling the executions starting from these copies, he will reckon that the result of the first coin is already determined at the time the copies are made, while the outcome of the second coin throw is still open.

In contrast, for the machine on the left, he will not be able to find such a moment in the execution. Whenever he can create copies of the machine the outcome is either fully determined or not at all determined, depending on whether the atomic event of rolling the dice has yet happened or not.  $\triangleleft$

In summary, this means that the experimenter's ability to make copies of the machine in every

moment of the execution, and his ability of applying statistic sampling from these copies, in combination, enables him to reckon the probabilistic branching structure. However, in accordance to the testing scenario for probabilistic systems, we want the probabilistic branching structure of a system to be as transparent as possible.

**Weak Probabilistic Bisimilarity** On the level of equivalence relations, the combined testing scenario intuitively seems to correspond to weak probabilistic bisimulation [Seg95]. This relation is a straightforward generalization of strong probabilistic bisimilarity to the weak setting. The statistical sampling of the testing scenario manifests as weak combined transitions; the replication button is incorporated in the form of the typical bisimulation conditions. A formal proof for this claim is pending, however.

**Definition 4.19 (Weak Probabilistic Bisimulation).** Let  $(S, \bar{s}, Act, \multimap)$  be a PA. An equivalence relation  $\mathcal{R}$  on  $S$  is a *weak probabilistic bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} \mu$  for some  $a \in Act$  and  $\mu \in Dist(S)$  implies  $t \xRightarrow{a}_c \gamma$  for some  $\gamma' \in Dist(S)$  with  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .  $\triangleleft$

We write  $s \approx_{PA} t$  if some weak probabilistic bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two PA  $\mathcal{A}$  and  $\mathcal{A}'$  are weak probabilistic bisimilar if in their disjoint union the two initial states are weak probabilistic bisimilar, i.e.  $\bar{s} \approx_{PA} \bar{s}'$ .

**Example 4.6 (Discrimination by Weak Probabilistic Bisimilarity).** We will show that weak probabilistic bisimilarity distinguishes the two automata in Figure 4.6 and highlight the similarities in the argumentation with respect to our discussion in Example 4.5, where argued that the two automata are distinguished in the combined testing scenario.

First, we remark that pressing the replication button corresponds to relating two states in a bisimulation. Whenever two states are related in a bisimulation, they can be exposed to arbitrary many experiments, which manifest in the form of the challenger's and defender's transition conditions of bisimulation.

Assume that there exists a bisimulation  $\mathcal{R}$  containing the initial states of the automata. Choosing the left automaton as challenger first, the bisimulation conditions are easily matched for the only transition available. It suffices for the automaton on the right to performing a weak transition to the uniform distribution over the states  $tt$ ,  $th$ ,  $ht$ , and  $hh$ . We now exchange the roles of the challenger and the defender. The challenger proposes the only available transition, labelled by *throw*, to the distribution  $\mu = \langle (\frac{1}{2} : t), (\frac{1}{2} : h) \rangle$ . The defender now must be able to match this behaviour by a transition resulting in a distribution who assigns the same probability to states that are bisimilar to the states in  $\mu$ , i.e. the two distribution must be in the lifting of the bisimulation relation  $\mathcal{R}$ . In the testing scenario, this corresponds to the observation of an action *throw*, followed by pressing the replication button, when the execution has reached  $t$  or  $h$ , respectively. In fact, the only transition of the automaton on the left leads to the uniform distribution over the states  $tt$ ,  $th$ ,  $ht$  and  $hh$ . If, as we assumed, the two automata are bisimilar, at least one of those four states must be related to  $t$ , and  $h$ , respectively. It is, however, easy to verify that none of them satisfies the necessary conditions. The reason is, as before in Example 4.5, that, for instance, the state  $t$  of the right automaton can evolve to either state  $tt$  or  $th$  with probability one half, while the states  $tt$ ,  $th$ ,  $ht$  and  $hh$  of the automaton on the left cannot evolve any further. The analogue argument holds for  $h$ .  $\triangleleft$

As this example illustrates, weak probabilistic bisimilarity does not allow to equate automata that only differ in their probabilistic branching structure. Despite, weak probabilistic bisimilar-



ity has become the *de facto* standard notion of weak bisimilarity for probabilistic automata, as it satisfies the usual properties of bisimilarities. It is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two compositions we have defined.

**Theorem 4.3.** Weak probabilistic bisimilarity  $\approx_{\text{PA}}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

From our requirements for observational congruences (Remark 1.1), weak probabilistic bisimilarity clearly satisfies the first two, being a congruence and preservation of the non-deterministic branching structure. With respect to our third and final requirement, abstraction from internal details, this relation, does not seem to be a good candidate. Thus, the quest for a suitable notion of observational equivalence for probabilistic automata is still open so far.

### The Experimenter's Dilemma

As we have learned, the observers' or experimenters' ability to create an unbounded number of copies of a system *in every moment* of its execution is essential for observing the non-deterministic branching structure. At the same time, statistic sampling is a necessary ability of the experimenter in order to derive probabilities of observable events that he observes. In combination, however, the two abilities lead to unwanted observation, namely the exposure of the probabilistic branching structure. A way out of this dilemma seems only possible at the cost of less simplistic testing scenario with stronger assumptions.

**Example 4.7.** One suggestion to resolve the dilemma is to restrict the use of replicated automata for sampling. For instance, in the automaton on the right of Figure 4.6, the negative interference of replication and statistic sampling could be avoided if pressing the replication button was forbidden during the execution of those internal transitions, where the probabilities of the ultimately observable activities  $tt$ ,  $th$ ,  $ht$  and  $hh$  are computed. Assume that in our testing scenario, we could somehow oblige the system under test to enable the replication button if *and only* if non-deterministic decisions are to be made, or at the beginning of a sequence on internal probabilistic transitions. Then, in Figure 4.6 above, the replication button would only be available in the initial states; we could do our statistic sampling by means of the generated copies of the system *in the initial state*, but we could not use the button in any other state. Thus, the two automata in Figure 4.6 will appear observationally the same. As the replication button would still be available in situation where non-determinism occurs, the observer would still be able to reconstruct the non-deterministic branching structure.  $\triangleleft$

At this point, it seems that testing scenarios are no longer suited as a benchmark in the question, whether a specific relation is a natural candidates for an observational equivalence. Our demand for the preservation of the non-deterministic branching structure and the simultaneous neglect of the probabilistic branching structure seems to defy this approach. We will therefore not deepen this approach further. However, it will serve as a source of inspiration, when we propose novel candidates for an observational equivalence in the second part of the thesis.

#### 4.2.4. Discussion

In this section, we have started with a review of existing notions of strong bisimulation. We then continued with the formalization of weak transitions for probabilistic automata. Finally, we have discussed a testing scenario for probabilistic systems that realizes the observation of probabilistic behaviour by means of statistic sampling and a reset button. The important property of the reset button is that it always resets the system to its initial state. In this way, only the probabilities of traces can be observed. The fine grained probabilistic branching structure, i.e. the precise points in an execution, where probabilistic decisions are made, remain unobservable. We have then considered several notions of process equivalence as well as a preorder relation. All of them miss at least one of the requirements we demanded for observational equivalences. In this regard, an interesting dichotomy has emerged: the relations either allow to make too fine grained observations with respect to the probabilistic branching structure, or fail to preserve the non-deterministic branching structure. A relation that satisfies both requirements seems hard to establish. Our discussion of combining the testing scenarios of probabilistic automata and labelled transition system suggests that this will, in fact, demand novel approaches.

### 4.3. Decision Algorithm

The first known decision algorithm for weak probabilistic bisimilarity [CS02] has a complexity that is exponential in the number of states and transitions. Only recently a polynomial solution of this problem has been given in [HT12]. The core of the problem lies, similar as for weak LTS bisimilarity in determining whether two states that lie in the same supposed equivalence class actually exhibit the same behaviour. For LTS, the behaviour of a state can be described by the action that can be executed and the states that can be reached by weak transitions. For PA, the problem is similar, however, weak transitions reach distributions over states, and furthermore, weak transitions can be combined in an uncountable number of ways. Hence, while for LTS the behaviour of a state is finite, for PA it is infinite. The remarkable contribution of [CS02] has been to show that the uncountably many distributions that can be reached by a weak combined transition can be fully characterized by a finite set of generating distributions. It was further shown that these distributions are all reachable by transitions that are generated by a *determinate* transition tree. This means that on every state, the same successor transition is chosen whenever it is visited along the tree. Only before and after the observable action (if there is any), the transitions can be different. It is then a rather simple exercise to compute the distribution that result from such transitions by solving a linear equation system. In this way, in a finite system, the generators of all distributions reachable from a certain state by a weak combined transitions are finite, and can thus be computed. Yet, the number of determinate transition trees is exponential in the number of transitions, as in order to consider every possible tree, one has to consider all combinations how the single states in the tree are continued in the transition tree.

Then, in [HT12], the following proposition has allowed to overcome the exponential approach.

**Proposition 1** (cf. [HT12, Prop. 3]). *Given a PA  $\mathcal{A}$ , two probability distributions  $\rho_1, \rho_2 \in \text{Dist}(S)$ , two actions  $a_1, a_2 \in \text{Act}$ , two sets  $\check{A}_1, \check{A}_2 \subseteq \rightarrow$  of transitions, and an equivalence relation  $\mathcal{W}$  on  $S$ , the existence of  $\nu_1, \nu_2 \in \text{Dist}(S)$  such that*

$$\rho_1 \xrightarrow{a_1 \mid \check{A}_1}_c \nu_1, \rho_2 \xrightarrow{a_2 \mid \check{A}_2}_c \nu_2, \text{ and } \nu_1 \mathcal{L}(\mathcal{W}) \nu_2$$

*can be checked in polynomial time.*

The proof that this check can be performed in polynomial time relies on the construction of a generalized flow problem, that in turn can be encoded into an LP-problem of polynomial size spanned by the parameters  $\rho_1, \rho_2, a_1, a_2, \check{A}_1, \check{A}_2$ , and  $\mathbf{W}$ . Details are given in [HT12].

The proposition is more general than needed for the purpose of deciding  $\approx_{\text{PA}}$ . For example, the concept of allowed transitions is not needed here. However, this general form will turn out helpful later. A more accessible variant of the proposition is the following corollary, which is the way in which the proposition is actually used for  $\approx_{\text{PA}}$ .

$Decide(\mathcal{A})$
<pre> 1: <math>\mathbb{W} = \{S\};</math> 2: <b>repeat</b> 3:   <math>\mathbb{W}' = \mathbb{W};</math> 4:   <math>(\mathcal{C}, a, \rho) = FindSplit(\mathbb{W});</math> 5:   <math>\mathbb{W} = Refine(\mathbb{W}, (\mathcal{C}, a, \rho));</math> 6: <b>until</b> <math>\mathbb{W} = \mathbb{W}'</math> 7: <b>return</b> <math>\mathbb{W}</math> </pre>

**Figure 4.7.:** Compute partitioning  $\mathbb{W}$  according to  $\approx_{PA}$ .

$FindSplit(\mathbb{W})$
<pre> 1: <b>for all</b> <math>(s, a, \rho) \in \mathcal{T}</math> <b>do</b> 2:   <b>for all</b> <math>t \in [s]_{\mathbb{W}}</math> <b>do</b> 3:     <b>if</b> <math>P(\mathbb{W}, a, \rho, \delta(t))</math> has no solution <b>then</b> 4:       <b>return</b> <math>([s]_{\mathbb{W}}, a, \rho)</math> 5: <b>return</b> <math>(\emptyset, \tau, \delta(\perp))</math> </pre>

**Figure 4.8.:** Finding a splitter

**Corollary 4.1.** *Given a PA  $\mathcal{A}$ , two probability distributions  $\rho_1, \rho_2 \in Dist(S)$ , an action  $a \in Act$ , and an equivalence relation  $\mathbb{W}$  on  $S$ , the existence of  $\nu \in Dist(S)$  such that*

$$\rho_2 \xrightarrow{a}_c \nu \text{ and } \rho_1 \mathcal{L}(\mathbb{W}) \nu$$

*can be checked in polynomial time.*

We denote by  $P(\mathbb{W}, a, \rho_1, \rho_2)$  an instance of the decision problem of Corollary 4.1. The resulting decision algorithm is then a straightforward adaption of the algorithm for  $\approx_{LTS}$ .

We summarize the algorithm in Figure 4.7. It employs the usual partition refinement approach, where partition are refined according to a splitter. A splitter consists of

1. a spurious equivalence class that should be split because at least two of its states are not compatible with respect to their behaviour *and*
2. the reason of the incompatibility: an action and a distribution over states.

Together, they describe a weak transition that can be performed by at least one state in the spurious equivalence class, while at least one other state cannot do so.

The procedure *FindSplit* in Figure 4.8 computes the splitter. For this, it considers every transition  $(s, a, \rho)$  in  $\rightarrow$  and checks whether every state in the equivalence class of  $s$  according to the current partitioning  $\mathbb{W}$  is able to mimic this transition up to  $\mathcal{L}(\mathbb{W})$ . With other words, it exactly checks the bisimulation condition: whenever  $s \xrightarrow{a} \rho$  then  $t \xrightarrow{a}_c \rho'$  and  $\rho \mathcal{L}(\mathbb{W}) \rho'$ .

$\underbrace{t \xrightarrow{a}_c \rho' \text{ and } \rho \mathcal{L}(\mathbb{W}) \rho'}_{\text{computed by } P(\mathbb{W}, a, \rho, \delta(t))}$

## 4.4. Summary and Discussion

The definitions for probabilistic automata in this chapter are all standard, except for the definition of the parallel composition, which slightly deviates from the original parallel composition defined in [Seg95] for technical reasons. It has, however, appeared in the literature before, for instance in [Den+07]. Originally, when two probabilistic automata are composed in parallel, they are forced to synchronize on every external action they share, while the internal actions are interleaved and not synchronized. We generalized this concept by synchronizing precisely those action contained in some set  $A$ , which annotates the parallel composition operator. So, we obtain the original behaviour by letting  $A = Act \setminus \tau$ . We have adapted synchronization behaviour

here for the sake of a uniform presentation relative to the other models, LTS and IMC, that we discuss in this thesis.

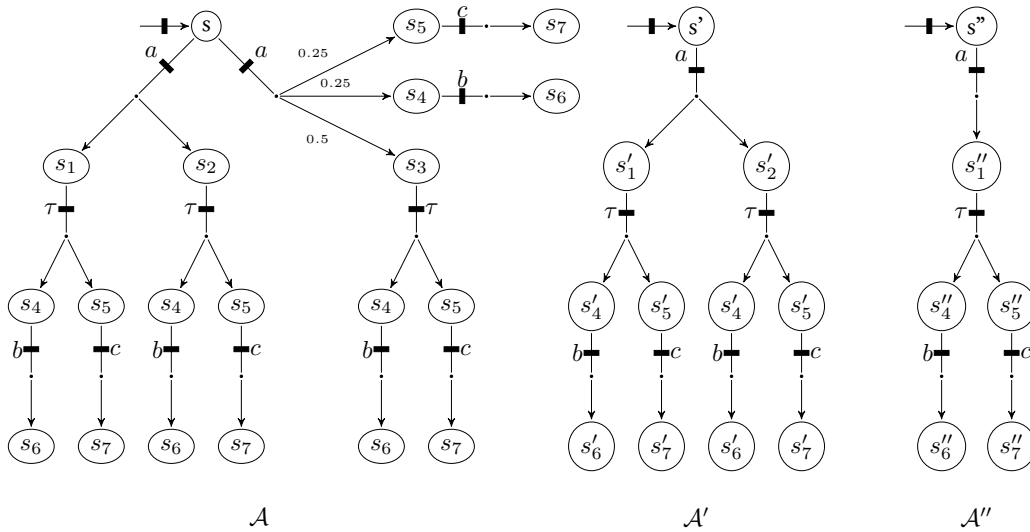
### A Matter of Weak Transitions

The notions of strong and weak bisimilarity we have introduced in this chapter are not the only theoretically possible variations. As we have seen, strong bisimilarity for probabilistic automata comes in two flavour. One, where transitions may be combined and one where ordinary transitions are considered. Both notions enjoy nice properties and are thus reasonable candidates. Furthermore, these are the only two variants in the literature.

For weak bisimulations, there are more places than the usage of probabilistic vs. non-probabilistic transitions, at which the definition can be varied.

One of them is the question, whether the *challenger* may propose only strong transitions or also arbitrary weak transitions. We call the two variants the *strong* and *weak* challenger characterization, respectively. For labelled transition systems and its respective bisimilarities, we know that the two characterizations coincide.

For probabilistic automata, this is not the case. It turns out, that the non-probabilistic variant has several semantic drawbacks. First, as noted for example in [Den05], the strong challenger variant of non-probabilistic weak bisimilarity is *not transitive* (see Figure 4.9). Fortunately, this drawback does not apply to the weak challenger characterization. Differently from what one might expect, it is not a generalization of strong challenger characterization, though. This means, weak challenger non-probabilistic weak bisimilarity fails to equate certain processes that are equated by strong challenger non-probabilistic weak bisimilarity (see Figure 4.9).



**Figure 4.9.:**  $\mathcal{A}$  is strong challenger bisimilar to  $\mathcal{A}'$ , and  $\mathcal{A}'$  is strong challenger bisimilar to  $\mathcal{A}''$ . However,  $\mathcal{A}$  is not strong challenger bisimilar to  $\mathcal{A}''$ . Furthermore,  $\mathcal{A}'$  and  $\mathcal{A}''$  are not weak challenger bisimilar.<sup>1</sup>

<sup>1</sup>The counterexamples are courtesy to Andrea Turrini, who brought it up during a discussion with Pedro D'Argenio

**Example 4.8.** In the following, we write  $\approx_s$  to denote *strong* challenger non-probabilistic weak bisimilarity and  $\approx_w$  to denote *weak* challenger non-probabilistic weak bisimilarity.

To see that  $\approx_s$  is not transitive, we note that  $\mathcal{A} \approx_s \mathcal{A}'$  is justified by a bisimulation with the equivalence classes  $\{s, s'\}$ ,  $\{s_1, s'_1, s_2, s'_2, s_3\}$ ,  $\{s_4, s'_4\}$ ,  $\{s_5, s'_5\}$  and  $\{s_6, s'_6, s_7, s'_7\}$ . Similarly for  $\mathcal{A}' \approx_s \mathcal{A}''$ , the justifying equivalence classes are  $\{s', s''\}$ ,  $\{s'_1, s''_1, s'_2, s''_2\}$ ,  $\{s'_4, s''_4\}$ ,  $\{s'_5, s''_5\}$  and  $\{s'_6, s''_6, s'_7, s''_7\}$ . However,  $\mathcal{A} \not\approx_s \mathcal{A}''$ . Necessarily, any justifying bisimulation needed to contain the pair  $(s, s'')$ . The problem now is that  $s \xrightarrow{a} \mu$  with  $\mu(s_3) = 0.5$ , and  $\mu(s_4) = \mu(s_5) = 0.25$  cannot be matched by  $s''$ . The only transitions of  $s''$  lead to  $\delta(s''_1)$ , i.e. a distribution with only one state. However, all three states in  $\text{Supp}(\mu)$  are in different equivalence classes, as they are not strong bisimilar. Thus,  $\mu$  cannot be matched to  $\delta(s''_1)$ .

We have already argued that  $\mathcal{A}' \approx_s \mathcal{A}''$ . We will now show that  $\mathcal{A}' \not\approx_w \mathcal{A}''$ . Clearly, if this was the case,  $(s', s'')$ , must be contained in any justifying bisimulation. Now  $s' \xrightarrow{a} \mu$  with  $\mu(s'_1) = 0.5$  and  $\mu(s'_4) = \mu(s'_5) = 0.25$ . It is easy to see that all three states in  $\text{Supp}(\mu)$  must necessarily be contained in different equivalence classes. However, the only two weak transitions  $s'$  is capable of, lead to distributions containing at most two states. Hence, the transition of  $s'$  cannot be matched by  $s''$ .  $\triangleleft$

In contrast, the *probabilistic* variants of strong and weak bisimilarity are not affected by this problems. In fact, they satisfy the same properties as bisimilarities for labelled transition systems: both strong and weak probabilistic bisimilarities are transitive, and the strong and the weak challenger characterization of weak probabilistic bisimilarity agree. The probabilistic variant is thus a more satisfying candidate for a natural notion of behavioural equivalence on probabilistic automata.

Another variation of weak bisimulation is rooted in the question, whether weak transitions are allowed to be *infinitary* – as it is the case in the definitions we have presented so far – or whether they are restricted to finitary transitions. In fact, both variants lead to reasonable notions of bisimilarity satisfying all properties well-known from bisimilarities for LTS. Infinitary transitions lead to coarser notions of bisimilarity and are also more widely used in the literature. Throughout this thesis, we have therefore adopted infinitary transitions for all notions of bisimilarity.

Among the many possible variants of weak bisimilarity for probabilistic automata, we have chosen the one that satisfies all relevant properties when compared to weak bisimilarity for LTS, and that is, at the same time, commonly used in the literature.

## Concluding Summary

In this chapter, we have introduced probabilistic automata as a model that augments labelled transition systems by discrete probability decisions. We have defined and reviewed several established notions of process equivalence, both strong and weak, and one preorder. Finally, we shortly reviewed the decision algorithm for PA.

The most important insight of this chapter is that for PA, it is not clear what process relation qualifies as a natural notion of observable behavioural equivalence. To foster our understanding of what should be considered a natural notion of observable behaviour, we have introduced a novel intuitive testing scenario akin to the one well-known for labelled transitions systems, but

---

and the author.

extended to the probabilistic setting. This scenario suggests, that none of the known process relations is actually ideal with respect to formalizing a natural notion of behavioural equivalence on probabilistic automata.





## 5. Interactive Markov Chains

Interactive Markov chains (IMC) extend labelled transition systems by exponentially distributed time delay transitions. It distinguishes between interactive transitions, which coincide with the transitions known from labelled transition systems, and timed transitions, which are labelled by a positive real value. Timed transitions with label  $\lambda$  denote a random time delay that is governed by a negative exponential distribution with parameter  $\lambda$ . In this way IMC are a fully conservative extension of labelled transition systems, in the sense that by dropping the timed transitions, LTS can be fully regained. At the same time, by dropping the interactive transitions, we obtain *CTMC*. The strict separation of interactive transitions and time transitions let IMC stand out among most other compositional models and calculi in a stochastic setting, for example *TIPP* [GHR92], *EMPA* [BG98], and *PEPA* [Hil96]. As summarized in [Cla+07], all of these approaches share a different approach that integrates stochastic time and actions. While in IMC actions happen instantaneous and do not consume time, in the other approaches the actions are decorated by the delay. In this way, actions consume time directly. While this interweaving of action and their duration may be intuitively appealing, it has several semantic drawbacks, as detailed out in [Her02]. In fact, these drawbacks have led to the development of IMC.

Over the decades, IMC have turned out to be very successful models with broad areas of application both in theory and practice-oriented research [Böd+09; Bou+08; BCS07b; BCS07a; BCS10; Boz+09b; Boz+09a; Boz+11; CZM09; Cos+08; Cos+09; HJ08; ZN10; Hav+10; Est+12].

**Outline and Contributions.** In this chapter, we give an overview of the state-of-the-art theory of interactive Markov chains. The model itself will be introduced in Section 5.1. As for LTS, parallel composition and hiding is introduced. The synchronization primitive is the stochastic extension of the operators discussed in Chapter 4 and Chapter 3 as introduced in [Her02]. Also, a notion of uniform representation will be introduced, that allows to express stochastic transitions congruent to the immediate (probabilistic) transitions of PA and IMC. This representation will be especially beneficial in the second part of this thesis, when we strive for a union of PA and IMC in a single model class.

In Section 5.2, strong and weak bisimilarity will be defined. Section 5.3 provides a decision algorithm for IMC. The algorithm is taken from [Her02], however slightly adapted in order to better fit in our comparative setting. We conclude in Section 5.4.

### 5.1. Model

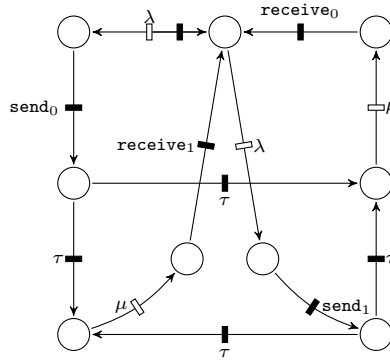
**Definition 5.1 (Interactive Markov Chains).** An *Interactive Markov Chain* (IMC) is a quintuple  $(S, \bar{s}, Act, \multimap, \dashv\dashv)$ , where

- $S$  is a non-empty finite set of states,
- $\bar{s}$  is the initial state,

- $Act$  is a set of actions consisting of immediate actions including the internal action  $\tau$ ,
- $\dashrightarrow \subseteq S \times Act \times S$  is a set of interactive transitions, and
- $\dashrightarrow \subseteq S \times \mathbb{R}_{>0} \times S$  is a multiset of timed transitions.

◁

The attentive reader may have noticed that our definition of interactive Markov chains differs from the definition used in [Her02] by using a *multiset* for the collection of all timed transitions instead of an ordinary set. In fact, this has been a (often adopted) mistake in the original work. When a set is used, the definition of parallel composition, does not yield the expected results.



**Figure 5.1.:** Stochastic Version of a Simple Communication Channel

**Example 5.1.** In the LTS version of the noisy channel, we had to model a reception error with a non-deterministic choice although a correct run of the system with respect to the model would allow to always invert the value sent over the channel.

The timed transitions do not fix this problems as they only quantify the probability of taking a transition over time. ◁

The semantic interpretation of an IMC corresponds to that of an CTMC concerning its stochastic behaviour. A transition  $s \xrightarrow{\lambda} s'$  states that the IMC can change its state from  $s$  to  $s'$  within  $d$  time units with probability  $1 - e^{-\lambda \cdot d}$ . The probability is thus determined by a negative exponential distribution with parameter  $\lambda$ . As this distribution is uniquely determined by  $\lambda$ , it suffices to label stochastic timed transitions with the delay. We observe that the greater  $\lambda$ , the higher the probability to make a transition within  $d$  time units. In this sense, the higher  $\lambda$  is, the faster the transition is. If more than one stochastic timed transition lead to the same state, their rates accumulate. We therefore use

$$\mathbf{R}(s, s') = \sum \llbracket \lambda \mid s \xrightarrow{\lambda} s' \rrbracket$$

to denote the total rate with which  $s$  transitions to  $s'$ . If more than one state can be reached by a stochastic timed transition, a competition between the states exists, called the *race condition*. The probability to move from  $s$  to a particular state  $s'$  within  $d$  time units is given by

$$\frac{\mathbf{R}(s, s')}{\mathbf{R}(s)} \cdot (1 - e^{-\mathbf{R}(s) \cdot d}),$$

where  $\mathbf{R}(s) = \sum_{s' \in S} \mathbf{R}(s, s')$ . The latter is called the *exit rate* of  $s$ . We observe from the formula of the race condition, that again a state  $s$  is left within a short period of time with a higher probability the higher the sum of the rates leaving  $s$ , i.e. its exit rate, is. Once the state is left, it is decided by a discrete probability, which of the potential successor states of  $s$  wins the *race*. The probability that a particular state  $s'$  wins the race is  $\frac{\mathbf{R}(s, s')}{\mathbf{R}(s)}$ .

Immediate transitions are considered to consume no time when executed. This means that no time passes when an immediate transition is taken. If we consider IMC as an open system that awaits communication with its environment, it becomes clear that the point of time *when* an immediate transition is executed cannot be determined without knowing the environment. Therefore, in each state the immediate transition stay enabled until the state is left again, waiting for a potential communication partner. The just said is only fully true for transitions labelled by observable actions. For the internal action  $\tau$  the case is different. Transitions labelled by  $\tau$  are considered to result from a successful communication, or to represent internal implementation details, which we abstract from. As they are completely independent from external influences, and their execution thus only depends on themselves, they can safely be assumed to take place immediately without letting any time pass, as soon as a state is entered. The probability that a stochastic timed transition fires at the same time, i.e. immediately upon activation, is zero. Therefore, it is justifiable that internal transitions take precedence over stochastic timed transitions. This is called the *maximum progress assumption* of interactive Markov chains.

**Definition 5.2.** In any IMC, internal transitions take precedence over stochastic timed transitions. ◁

The maximum progress assumption implies that in states with both outgoing internal transitions and stochastic timed transitions the latter are semantically ineffective. Still, no such restrictions are imposed structurally. Instead, the maximum progress assumption will be formally imposed as a side condition of the bisimilarities defined in Section 5.2.

*Notation 5.1.* We use the predicate  $s \downarrow$  to identify states that do not have any outgoing internal transitions. Such states are usually called *stable*. Later, it will turn out useful to lift the predicate  $\downarrow$  to distributions over states. We do this in the usual pointwise manner, by letting  $\mu \downarrow$  hold if and only if  $s \downarrow$  for each  $s \in \text{Supp}(\mu)$ .

**Uniform Representation** It is often convenient to use a notational variant of timed transition predicate, where all outgoing Markovian timed transitions of one state are represented in contracted form, as a special probabilistic transition labelled with the overall exit rate of the state. Furthermore, as timed transitions only contribute to observable semantic behaviour if they originate from stable states (as formulated by the maximal progress assumption), it is convenient to incorporate the stability directly into this new transition relation. In this way, the new transition predicate only captures those timed transitions that are semantically actually possible due to stability of the originating state. Recall that we denote stability of a state  $s$  as  $s \downarrow$ .

*Notation 5.2* (Uniform Stochastic Timed Transitions).

We write  $s \xrightarrow{\chi(r)} \mu$  for  $r \in \mathbb{R}_{\geq 0}$  if and only if  $s \downarrow$  and  $r = \mathbf{R}(s)$  and

- if  $r = 0$  then  $\mu = \delta(s)$ ,
- if  $r \neq 0$  then for all states  $t$  the distribution  $\mu$  satisfies  $\mu(t) = \frac{1}{r} \cdot \sum_{(s,x,t) \in -\rightarrow} x$ .

We let  $Act^X$  denote the set  $Act \cup \{ \chi(r) \mid r \in \mathbb{R}_{\geq 0} \}$ .

**Definition 5.3 (Uniform Representation of an IMC).** The uniform representation  $\mathcal{A}^U$  of an IMC  $\mathcal{A} = (S, \bar{s}, Act, \rightarrow, -\rightarrow)$  is a quadruple  $(S', \bar{s}', Act^X, \rightarrow', -\rightarrow')$ , where

- $S' = S$ ,
- $\bar{s}' = \bar{s}$ ,
- $\rightarrow' = \rightarrow$ , and
- $-\rightarrow' = \{ (s, \chi(r), \mu) \mid s \xrightarrow{\chi(r)} \mu \}$ .

◁

Uniform representations are very similar to IMC. The main difference is that the set of actions  $Act^X$  now spans  $Act$  and, in principle, all possible rates. To distinguish these special rate labels from rates, though, they are written in the form  $\chi(r)$ . Note that  $\chi(0)$  is also a possible label, while 0 is not a valid rate in an ordinary IMC. It denotes the absence of timed transitions in a *stable* state in the original IMC. The benefit of uniform representations is that now both immediate and timed transitions are represented in a uniform way, namely as transitions leading to a distribution over states. Furthermore, every stable state  $s \in S$  has exactly *one* outgoing *timed transition* that fully describes its timed behaviour. Also, a timed transition is present at a state exactly if the state is stable. Note that the relation  $-\rightarrow'$  is a set, and not a multiset. While the uniform representation of each IMC is unique, it is not possible to recover an IMC from its uniform representation in a unique way. A stochastic timed transitions can be recovered in an infinite number of ways by representing the total rate with which a state is reached by a summation of arbitrary many summands. For instance,  $s \xrightarrow{\chi(5)} \delta(s')$  can be translated to  $s \xrightarrow{2} s'$  and  $s \xrightarrow{3} s'$  and also  $s \xrightarrow{1} s'$  and  $s \xrightarrow{4} s'$  and also  $s \xrightarrow{0.3} s'$ ,  $s \xrightarrow{0.7} s'$  and  $s \xrightarrow{4} s'$ , and so on.

*Remark 5.1.* As the standard transformation of a uniform representation of an IMC into an ordinary IMC we fix the IMC where  $s \xrightarrow{r} t$  if and only if  $s \xrightarrow{\chi(r)} \mu$  with  $\mu(t) = r$  in the uniform representation.

## Parallel composition

Parallel composition for IMC is an extension of the parallel composition operation we have already introduced for LTS. It fully agrees on the immediate transition relation  $\rightarrow$ , where parallel composition results in the interleaving of the transitions of the components. One of the main

reasons why IMC are a mathematically appealing model is that exactly this interleaving structure can be transferred also to the timed transition relation  $\rightarrow$ .

In order to denote parallel composition of two IMC, we will write  $\mathcal{A} \parallel_A \mathcal{A}'$ , where  $A$  is an arbitrary set of actions excluding the internal action  $\tau$ . It is notationally helpful to use the symbol  $\parallel_A$  as syntactic sugar so that if  $s$  and  $s'$  are two states, we write  $s \parallel_A s'$  to denote the pair  $(s, s')$ .

**Definition 5.4 (Parallel composition).** Given two IMC  $\mathcal{A} = (S, \bar{s}, Act, \rightarrow, \rightarrow)$  and  $\mathcal{A}' = (S', \bar{s}', Act, \rightarrow', \rightarrow')$ , their parallel composition  $\mathcal{A} \parallel_A \mathcal{A}'$  is defined as the automaton  $\mathcal{A}^* = (S^*, \bar{s}^*, Act, \rightarrow^*, \rightarrow^*)$  where

- $S^* = S \times S'$
- $\bar{s}^* = \bar{s} \parallel_A \bar{s}'$
- $\rightarrow^*$  is the least relation satisfying

$$s \parallel_A s' \xrightarrow{a}^* t \parallel_A t' \quad \text{if } a \in A \text{ and } s \xrightarrow{a} t \text{ and } s' \xrightarrow{a} t' \quad (\text{sync})$$

$$s \parallel_A s' \xrightarrow{a}^* t \parallel_A s' \quad \text{if } a \notin A \text{ and } s \xrightarrow{a} t \quad (\text{parL})$$

$$s \parallel_A s' \xrightarrow{a}^* s \parallel_A t' \quad \text{if } a \notin A \text{ and } s' \xrightarrow{a} t'. \quad (\text{parR})$$

- $\rightarrow^*$  is the least multirelation satisfying

$$s \parallel_A s' \xrightarrow{\lambda}^* t \parallel_A s' \quad \text{if } s \xrightarrow{\lambda} t \quad (\text{parLM})$$

$$s \parallel_A s' \xrightarrow{\lambda}^* s \parallel_A t' \quad \text{if } s' \xrightarrow{\lambda} t' \quad (\text{parRM})$$

◁

For LTS and PA, our definition of parallel composition marginally deviated from the original definitions. For IMC, this corresponds to the original definition [Her02; HK10].

**Definition 5.5 (Abstraction).** Given a IMC  $\mathcal{A} = (S, \bar{s}, Act, \rightarrow, \rightarrow)$ , its abstraction  $\mathcal{A}|_A = (S', \bar{s}', Act, \rightarrow', \rightarrow')$  with respect to the set of actions  $A \subseteq Act \setminus \{\tau\}$  satisfies

- $S' = S$
- $\bar{s}' = \bar{s}$
- $\rightarrow$  is the least relation satisfying

$$s \xrightarrow{a} t' \quad \text{if } a \notin A \text{ and } s \xrightarrow{a} t$$

$$s \xrightarrow{\tau} t' \quad \text{if } a \in A \text{ and } s \xrightarrow{a} t$$

- $\rightarrow' = \rightarrow$

◁

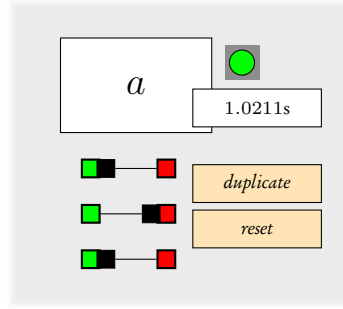


Figure 5.2.: Machine with Reset Button and Measurement of Time

## 5.2. Bisimilarities

In the context of stochastic timed systems, a similar testing scenario as the one we have discussed in the beginning of Section 4.2.3 has been investigated in [WMB05]. The precise underlying model in this scenario are continuous time Markov chains whose transitions are labelled by actions. Note that this model is rather different to the models we are interested in IMC. The interesting aspect, however, is that the machine in this scenario has been extended with one additional display in comparison to the machine of the probabilistic scenario. It provides timing information, that is used to observe and test the time behaviour induced by the Markovian delays of the CTMC.

Interactive Markov chains are a combination of labelled transition systems and continuous time Markov chains. The level of abstraction at which they allow to describe systems allows the same observations we have discussed for labelled transition systems, but extends them by a probabilistic notion of time. As a consequence, an observer can observe the passage of time. More precisely, he can observe a delay of time occurring between two other observations. As by the abstractions imposed by the definition of IMC, these observations cannot be precisely determined time durations, but only probability distributions over time delays, which are governed by an exponential distribution with some total delay rate  $\lambda \in \mathbb{R}_{>0}$ . We may imagine that the observer is able to perform an infinite number of stochastic experiments in every moment (by virtue of its unbounded number of copies of the system) to derive this information.

As for probabilistic automata, stochastic time also introduces a probabilistic choice over successive behaviour after certain events. While in probabilistic automata, every action induces an probabilistic experiment which is performed in order to determine the future behaviour of the system after the immediate transitions has occurred, in interactive Markov chains such a choice only appears after timed transitions, while action transitions always lead to a unique successor state.

As for probabilistic automata, will defer the discussion whether the existing notions of (weak) bisimilarity are reasonable observational equivalences for its model class to Chapter 8.

### 5.2.1. Strong Bisimilarity

**Definition 5.6 (Strong IMC Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\quad}, \dashv\dashv)$  be an IMC. An equivalence relation  $\mathcal{R}$  on  $S$  is a *strong IMC bisimulation*, if whenever  $s \mathcal{R} t$  then

1.  $s \xrightarrow{a} s'$  for some  $a \in Act$  and  $s' \in S$  implies  $t \xrightarrow{a} t'$  for some  $t'$  with  $s' \mathcal{R} t'$ , and
2.  $s \xrightarrow{\chi(x)} \mu$  for some  $x \in \mathbb{R}_{\geq 0}$  implies  $t \xrightarrow{\chi(x)} \gamma$  for some  $\gamma$  and  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .

◁

We write  $s \sim_{\text{IMC}} t$  if some strong IMC bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two IMC  $\mathcal{A}$  and  $\mathcal{A}'$  are strong IMC bisimilar if in their disjoint union the two initial states are strong IMC bisimilar, i.e.  $\bar{s} \sim_{\text{IMC}} \bar{s}'$ .

It is straightforward to show that strong IMC bisimilarity is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two compositions we have defined.

**Theorem 5.1.** strong IMC bisimilarity  $\sim_{\text{IMC}}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

### 5.2.2. Weak Bisimilarity

The following characterization of weak IMC bisimulation incorporates the notion of combined transitions that we have introduced for PA. In fact, replacing the combined transitions in this definition by ordinary transitions does not make a semantic difference. So on its own, this characterization is not particularly interesting. As one of the goals of this thesis is to create a synthesis of both IMC and PA, this characterization turns out to be a necessary step, as it allows to treat PA and IMC weak bisimilarity in a uniform way. We will also use this characterization to derive a straightforward decision algorithm for  $\approx_{\text{IMC}}$  that arises from the corresponding algorithm for  $\approx_{\text{PA}}$  only with minimal adaptations.

Notationally, we will make no distinction between  $s \xrightarrow{a} s'$  and  $s \xrightarrow{a} \delta(s')$  in the following. This allows us to apply the notion of a combined weak transition without further adaptations.

**Definition 5.7 (Weak IMC Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\quad}, \dashv\dashv)$  be an IMC. An equivalence relation  $\mathcal{R}$  on  $S$  is a *weak IMC bisimulation* if and only if whenever  $s \mathcal{R} t$  then

1.  $s \xrightarrow{a} s'$  for some  $a \in Act$  and  $s' \in S$  implies  $t \xRightarrow{a}_c \mu$  for some  $\mu \in \text{Dist}(S)$  with  $\delta(s') \mathcal{L}(\mathcal{R}) \mu$ , and
2.  $s \xrightarrow{\chi(r)} \mu$  for some  $r \in \mathbb{R}_{\geq 0}$  implies  $t \xRightarrow{\chi(r)}_c \gamma$  for some  $\gamma$  and  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .
3. If  $s \downarrow$  then  $t \Rightarrow t'$  for some  $t' \in S$  with  $t' \downarrow$ .

◁

We write  $s \approx_{\text{IMC}} t$  if some weak IMC bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two IMC  $\mathcal{A}$  and  $\mathcal{A}'$  are weak IMC bisimilar if in their disjoint union the two initial states are weak IMC bisimilar, i.e.  $\bar{s} \approx_{\text{IMC}} \bar{s}'$ .

It is straightforward to show that weak IMC bisimilarity is a bisimulation itself and also an equivalence relation. Furthermore, it is a congruence with respect to the two compositions we have defined.

**Theorem 5.2.** Weak IMC bisimilarity  $\approx_{\text{IMC}}$

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

Our definition of weak IMC bisimulation cannot be found in [Her02] in exactly this form. The characterization that is closest stems from Lemma 4.3 of [Her02]. We take this characterization as the original definition of weak IMC bisimulation and show that our characterization coincides with it. We use notations in the following definition deviating from [Her02], which, however, do not affect the underlying semantics.

**Definition 5.8 (Weak IMC Bisimulation according to Lemma 4.3 of [Her02]).** An equivalence relation  $\mathcal{R}$  is a weak IMC bisimulation if  $s \mathcal{R} t$  implies for all  $a \in \text{Act}$  and for all equivalence classes  $\mathcal{C}$  of  $\mathcal{R}$ ,

1.  $s \xrightarrow{a} s'$  implies  $t \xRightarrow{a} t'$  for some  $t'$  with  $s' \mathcal{R} t'$ ,
2.  $s \downarrow$  implies  $t \xRightarrow{} t'$  and  $t' \downarrow$  for some  $t'$  and

$$\sum_{(s,x,s') \in (-\square \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x = \sum_{(t',s,x,t'') \in (-\square \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x$$

◁

**Lemma 5.1.** The resulting bisimilarities of Definition 5.7 and 5.8 agree.

## 5.3. Decision Algorithm

The algorithm we present in the following is an adaption of the algorithm we have seen for probabilistic automata. This is possible, as we can treat immediate transitions and timed transitions in a uniform way owing to the uniform representation of IMC, in which  $\chi(r)$  can be treated like an ordinary action. From this perspective, the only difference between PA and IMC remains then that immediate transitions in IMC end in a Dirac distribution instead of arbitrary distributions over states. Even though convex combinations of transitions are not particularly reasonable semantically in IMC, our characterization of weak IMC bisimilarity makes use of them. Beside others, exactly for the reason that it allows a straightforward adaption of the decision algorithm for PA. Only for the right treatment of stability extra care is needed. Algorithmically, in fact, this check will implicitly happen during the conversion of the IMC to its uniform representation, which is done as a preprocessing step.



$Decide(\mathcal{A})$
1: $\mathcal{A}' = Uniformize(\mathcal{A})$
2: <b>return</b> $Quotient(\mathcal{A}')$

**Figure 5.3.:** Computes partitioning  $\mathbb{W}$  of  $S$  according to  $\approx_{MC}$

$Quotient(\mathcal{A})$
1: $\mathbb{W} = \{S\};$
2: <b>repeat</b>
3: $\mathbb{W}' = \mathbb{W};$
4: $(\mathcal{C}, a, \rho) = FindSplit(\mathbb{W});$
5: $\mathbb{W} = Refine(\mathbb{W}, (\mathcal{C}, a, \rho));$
6: <b>until</b> $\mathbb{W} = \mathbb{W}'$
7: <b>return</b> $\mathbb{W}$

**Figure 5.4.:** Compute partitioning  $\mathbb{W}$  according to  $\approx_{MC}$  on a IMC  $\mathcal{A}$  in uniform representation

Figure 5.3 summarizes the decision algorithm, which receives an arbitrary IMC as input and returns a partitioning  $\mathbb{W}$  of its state space, in which two states  $s$  and  $t$  are contained in the same partition if and only if  $s \approx_{MC} t$ . As a partitioning of a state space implicitly defines an equivalence relation over states, we may use the notation  $s\mathbb{W}t$  to denote that  $s$  and  $t$  lie in the same partition of  $\mathbb{W}$ , and, in general, treat  $\mathbb{W}$  like an equivalence relation whenever convenient without further mentioning.

The preprocessing in *Uniformize* transforms an IMC into its uniform representation. For every state of the IMC it converts existing  $\rightarrow$  transitions into their uniform representation, if the state has no outgoing  $\tau$  transitions, i.e. it is *stable*. Otherwise, it removes all  $\rightarrow$  transitions of this state. If a state  $s$  is stable, but has no outgoing  $\rightarrow$  transitions, the transition  $(s, \chi(0), \delta(s))$  is added to the relation  $\rightarrow$  of the uniform representation. This algorithm can be computed in linear time with respect to the number of states and transitions of the input IMC.

The second step of the decision algorithm is then to compute the quotient with respect to  $\approx_{MC}$  on the uniform representation. The respective algorithm can be found in Figure 5.4. It applies the standard partition refinement approach that we already know from LTS and PA. The procedure *Refine* works exactly as for the other two models. For procedure *FindSplit* we can actually take exactly the same procedure as for PA, with the minimal adaption that we treat  $\chi(r)$  for every  $r \in \mathbb{R}_{\geq 0}$  like an ordinary action.

**Theorem 5.3.** *Decide*( $\mathcal{A}$ ) computes  $\approx_{MC}$  in polynomial time.

*Sketch.* The interesting aspect is to show that stability is preserved according to Definition 5.7, and that timed transitions may only trigger a splitting, if they originate from stable state. The last point is ensured by *Uniformize*, as from all non-stable states the timed transitions are removed. In this way, also the stability condition is ensured (Condition 3). Whenever a state is  $s$  stable, it has a transitions  $\xrightarrow{\chi(r)}$  for  $r \in \mathbb{R}_{\geq 0}$ . A state  $t$  can only be bisimilar to  $s$  if  $t \Rightarrow t' \downarrow$  and  $s \approx_{MC} t'$  by the first condition of bisimilarity. But then, this state  $t'$  must enable a transition  $t' \xrightarrow{\chi(r)}$  (note the identical rates). Thus, checking the second condition of bisimilarity ensures also the stability condition.  $\square$

In [Her02], a more elaborate algorithm is devised that does not rely on techniques developed for PA, and thus can be expected to yield better performance in most cases. The intention behind

presenting an alternative variant has been to allow for a simple direct comparison between the fundamental differences and similarities behind the decision algorithms for the various bisimilarities. This approach turns out to be specifically instructive when we consider decision algorithms for Markov automata in Chapter 10.

## 5.4. Summary and Discussion

This chapter has reviewed the model of interactive Markov chains and its associated notions of bisimilarity. As in the chapters before, we have also reviewed a decision algorithm.

## 6. Comparative Semantics

In mathematics, an algebra is usually considered to be a set of constants and operations defined over a carrier set together with a set of equational laws, also called axioms, that characterize the constants and operators, or more precisely, the way they interact. Algebras are a means to describe mathematical structures in an abstract way without the need to actually devise the concrete objects. Very often, the axioms of an algebra are satisfied by more than one concrete carrier set. In this way, different mathematical structures can be described concisely by means of the axioms of their algebra. Furthermore, if different structures can be characterized by different set of axioms that only differ in few places, their similarities and differences become apparent instantly.

Process algebra aims to describes concurrent processes in terms of an algebra. It allows to describe arbitrary processes in terms of syntactic expressions of a formal language. The expressions of such a language consists of constants and a family of operators. Typical for an algebra, a process algebra is compositional in the sense that arbitrarily complex expressions can be composed from simpler ones. Axioms are used to describe when syntactically different processes should be considered semantically equivalent. The general benefits of algebras immediately carry over to process algebras: the semantic properties of different models of concurrency are expressed in a concise way, but independent of their concrete mathematical structure. For these reasons, process algebra is especially suited for formal comparisons of different models of concurrency. Process theories have been developed that are a process algebra by design, such as for instance ACP [BW90; BPS01; Bae+10]. For most models of concurrency, however, including those we consider in this thesis, formal process languages and axiom sets have been developed in hindsight with the intention to precisely characterize the semantic similarities and differences between the different approaches.

**Outline and Contributions.** The purpose of this chapter is to review the introduced models from a process algebraic perspective, and to unfold the semantic similarities and differences between the various notions of bisimilarity, and also of probabilistic forward simulation. To this end, we will introduce a small algebraic language for each of our models, LTS, PA and IMC in Section 6.1. For each language, we will give a semantics in terms of the respective models. Based on these languages, we will present sound and complete axiomatizations for the respective bisimilarities in Section 6.2, and for probabilistic forward simulation in Section 6.3. This means, that the axioms completely characterize the respective semantic relations. The axioms that are shared by several or all axiomatizations, and the ones that distinguish an axiomatization from another, are then the pivots that manifest the semantic similarities and differences, respectively. Section 6.4 concludes.

## 6.1. Languages

In the following, we will define languages for LTS, PA, and IMC. All languages are rather similar concerning their basic constants and operators. This is not surprising, as the three models are similar in their structures. To focus on the essential differences of the bisimilarities, and to not get lost in intricate technicalities that only provide little insight, we restrict our attention to finite and acyclic systems. In fact, we make one exception to this restriction: we add a special divergent process that is capable of performing an unbounded sequence of internal actions. This process is instrumental to make apparent a specific core difference between weak probabilistic bisimilarity and weak IMC bisimilarity.

### 6.1.1. Syntax

In this section, we introduce the syntax for our process languages. All languages contain as operators basic action prefixing, non-deterministic choice, and a special divergence operator,  $\Delta$ , and as constant the terminated process 0. They differ with respect to the question of what an action precisely is.

**Labelled Transitions Systems** The process language that corresponds to the model of labelled transition systems is a strict subset of the well-known calculus CCS [Mil89b], enhanced by the divergence operator, which does not appear in Milner's original calculus, but has appeared in several endeavours to axiomatize certain more complex notions of bisimilarity e.g. in [LDH05].

**Definition 6.1 (LTS Syntax).** Let  $a \in Act$ , and  $A \subseteq Act$ . We define the language  $\mathbb{LTS}$  by the following grammar in Backus-Naur form:

$$\mathbb{LTS} \ni E, E' ::= 0 \mid a.E \mid E + E' \mid \Delta(E)$$

We write  $\mathbb{LTS}$  for the set of  $\mathbb{LTS}$  expressions. We use  $E, E_1, E_2, F, \dots$  to range over  $\mathbb{LTS}$ .  $\triangleleft$

With 0 we denote the process incapable of any behaviour. In other words, it represents a completely terminated process. The capability of executing an action  $a$  is expressed by the prefix operator “ $a.$ ”. In fact, this is actually a family of unary operators, with one member for each  $a \in Act$ . An expression  $a.E$  then expresses a process that starts with the action  $a$  and then continues as  $E$ . The expression  $E + E'$  denotes non-deterministic choice between two expressions. It may either behave like  $E$  or like  $E'$ , as long as they are able to perform actions. For example  $E + 0$  cannot decide to behave like 0, as 0 does not exhibit any actions. The ability of a process to (exclusively) perform an unbounded number of  $\tau$  actions in uninterrupted succession is denoted by the expression  $\Delta(E)$ . We also call such behaviour *divergence*. Whenever  $E$  can perform an action, divergence can be escaped by this action by continuing as whatever process  $E$  may evolve to.

*Notation 6.1.* Since the operator  $+$  is associative and commutative, we use  $\sum_{i \in I} E_i$  to represent the sum of  $\mathbb{LTS}$ -terms ( $I$  finite).

**Probabilistic Automata** The essential difference between labelled transition system and probabilistic automata is that transitions of the latter have several different successor states, which they reach with a certain probability, instead of one single successor states. In our calculus, this has both semantic and syntactic consequences, as for example the process prefix operator  $a.$  cannot be longer completed by a simple process expression like in  $a.E$ , but has to be followed by a distribution over process expressions. Thus, we need the ability to represent arbitrary distributions syntactically. For this reason, our language for probabilistic processes will provide process expressions as well as distribution expressions. The syntactic structure of the following calculus is strongly inspired by existing calculi for probabilistic processes, e.g. [DPP05; DP07; DP05; Den05; PS04; BS01].

**Definition 6.2 (PA Syntax).** Let  $a \in Act$ ,  $A \subseteq Act$ , and  $p \in [0, 1]$ . We define the languages  $\mathbb{PA}$  of expressions by the following syntax.

$$\mathbb{PA} \ni E, E' ::= 0 \mid a.\mathcal{D} \mid E + E' \mid \Delta(E)$$

where  $\mathcal{D} \in \mathbb{D}$  is a distribution expression with syntax  $\mathcal{D} ::= \bigoplus_{i \in I} p_i E_i$ . If the sum expression has only one operand, we write  $\delta(E)$  instead of  $\bigoplus 1 \cdot E$ .  $\triangleleft$

We write  $\mathbb{D}$  and  $\mathbb{PA}$  for the set of  $\mathbb{D}$ -expressions and  $\mathbb{PA}$  expressions, respectively. We use  $E, E_1, E_2, F, \dots$  to range over of  $\mathbb{PA}$ , and  $\mathcal{D}, \mathcal{D}_1, \dots$  to range over  $\mathbb{D}$  expressions respectively. Misusing notation slightly, we write  $E \in \text{Supp}(\mathcal{D})$  if  $E = E_i$  for some  $E_i$  in  $\bigoplus_{i \in I} p_i E_i = \mathcal{D}$ .

$\text{LTS}$  and  $\mathbb{PA}$  only differ in the prefix operator. Whereas in  $\text{LTS}$  an action leads to the behaviour of a certain process, in  $\mathbb{PA}$  it leads to a distribution over processes, e.g. the process decides probabilistically how to behave afterwards.

The process  $\Delta(E)$  here describes a process which performs an infinite number of  $\tau$  transitions in succession with probability 1. Whenever  $E$  can perform an action, divergence can be escaped by this action by continuing as whatever distribution  $E$  may evolve to.

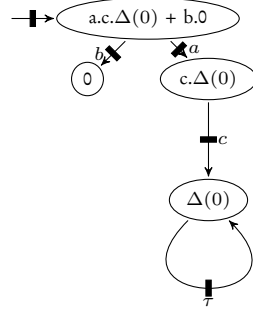
*Notation 6.2.* Since the operators  $+$  is associative and commutative, we use  $\sum_{i \in I} E_i$  to represent the sum of  $\mathbb{PA}$ -terms ( $I$  finite). We let the notational conventions defined for  $\oplus$  in Section 2.4 also apply here.

**Interactive Markov Chains** The following language is taken from [Her02], restricted to the operators essential for our comparison. The language in [Her02] includes the symbol  $\perp$ , instead of  $\Delta$ . Different to  $\Delta$ , it is a process constant and not an unary operator. Both denote some notion of divergence, but  $\perp$  more precisely represents underspecified processes. We do not deal with this aspect here and hence deviate from the original language of [Her02]. In our setting, both operators have the same semantical implications concerning the considered notions of bisimilarity.

**Definition 6.3 (IMC Syntax).** Let  $\lambda \in \mathbb{R}_{>0}$ ,  $a \in Act$ ,  $A \subseteq Act$ . We define the language  $\mathbb{IMC}$  by the following grammar in Backus-Naur form:

$$\mathbb{IMC} \ni E, E' ::= 0 \mid a.E \mid \lambda.E \mid E + E' \mid \Delta(E)$$

$\triangleleft$


 Figure 6.1.: Semantics of  $a.c.\Delta(0) + b.0$ .

We write  $\mathbb{IMC}$  for the set of  $\mathbb{IMC}$  expressions. We use  $E, E_1, E_2, F, \dots$  to range over of  $\mathbb{IMC}$ .

The only difference to our basic process calculus is the inclusion of a new prefix operator, that allows positive real numbers as prefix. In the same way as the action prefix operators denotes an immediate transition labelled by an action, this operator denotes a timed transition labelled by some rate representing the parameter of an exponential distributed time delay.

*Notation 6.3.* Since the operators  $+$  is associative and commutative, we use  $\sum_{i \in I} E_i$  to represent the sum of  $\mathbb{IMC}$ -terms ( $I$  finite).

### 6.1.2. Semantics

We will formalize the intuitive interpretation of the expressions by means of structural operational rules, which generate an special instance of the underlying model, i.e. a labelled transition system, a probabilistic automaton or a interactive Markov chain, respectively. The semantics of each expression is then defined by mapping them onto states in the system. The state space of these systems will be defined as the set of all respective language expressions. In this way, we implicitly define the mapping for each expression  $E$ . We obtain one single automaton for each of the languages  $\mathbb{LTS}$ ,  $\mathbb{PA}$  and  $\mathbb{IMC}$  with infinite state space that covers every expression. The concrete semantics of an expression  $E$  is then obtained in a final step by reducing the infinite system to the subset reachable from state  $E$ .

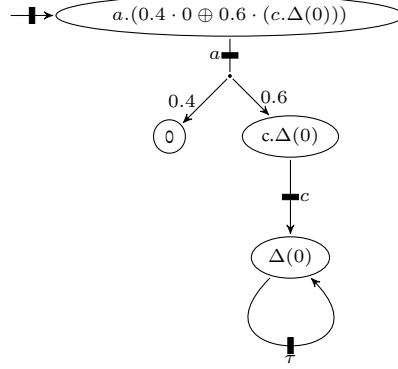
**Labelled Transition Systems** For  $\mathbb{LTS}$  expressions, we will provide rules for action prefix ( $a.$ ), choice ( $+$ ). Termination ( $0$ ) does not require specific rules.

**Definition 6.4 ( $\mathbb{LTS}$  Semantics).**

Let  $\mathcal{A}_{\mathbb{LTS}}^{\bar{s}} = (S, \bar{s}, Act, \rightarrow)$  denote a LTS that parameterized by its initial state  $\bar{s} \in S$ . Let  $S = \mathbb{LTS}$  and let the transition relation  $\rightarrow \subseteq \mathbb{LTS} \times Act \times \mathbb{LTS}$  be the least relation given by the rules in Table 6.1.

The semantics of an arbitrary expression  $E \in \mathbb{LTS}$  is defined as  $\mathcal{A}_{\mathbb{LTS}}^E$ .  $\triangleleft$

As mentioned before, we only consider the reachable part of the semantics underlying an



**Figure 6.2.:** Semantics of  $a.((0.4 \cdot 0) \oplus_{0.4} (c.\Delta(0)))$

expression. In this way, we can draw its relevant parts.<sup>1</sup>

**Example 6.1.** The semantics of  $a.c.\Delta(0) + b.0$  is the LTS in Figure 6.1. The states are labelled by their corresponding expressions. Note that the state  $\Delta(0)$  has an internal self-loop. Recall that we only consider acyclic systems with the notable exception of self-loops of internal transitions that can be attached to any state of the process via the operator  $\Delta$ . This suffices to capture the essence of divergent behaviour.  $\triangleleft$

## Probabilistic Automata

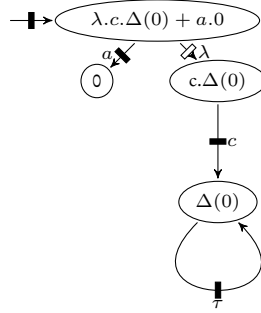
### Definition 6.5 ( $\mathbb{PA}$ Semantics).

Let  $\mathcal{A}_{\mathbb{PA}}^{\bar{s}} = (S, \bar{s}, Act, \rightarrow)$  denote a PA that is parameterized by its initial state  $\bar{s} \in S$ . Let  $S = \mathbb{PA}$  and let the transition relation  $\rightarrow \subseteq \mathbb{PA} \times Act \times \mathbb{PA}$  be the least relation given by the rules in Table 6.2.

The semantics of an arbitrary expression  $E \in \mathbb{PA}$  is defined as  $\mathcal{A}_{\mathbb{PA}}^E$ .  $\triangleleft$

**Interactive Markov Chains** As usual for Markovian calculi, the Markovian transition relation  $\rightarrow$  is a multi-relation, as the precise number of occurrences of the same rate leading to a state determines the stochastic behaviour. For instance, for the expression  $(\lambda).0 + (\lambda).0$  two transitions are generated. Whenever there are  $k$  different proof trees for a transition  $\xrightarrow{\lambda}$  the multi-relation will contain  $k$  such transitions. The need to represent multiplicities, stems from our interpretation of the choice operator “+” in the presence of delays. The delay until the choice is actually resolved is governed by an exponential measure given by the sum of all competing rates. As a consequence, the behaviour of  $(\lambda).0 + (\lambda).0$  should be the same as that of  $(2\lambda).0$ , which has a delay of  $2\lambda$ . In the former expression, however, we represent the delay by two separate transitions labelled by  $\lambda$  that add up to  $2\lambda$ . As the two transitions are between the same two states and labelled by the same label, we formally need to resort to a multi-relation to describe this situation adequately [Göt94; Hil96].

<sup>1</sup>Formally, our definition of  $\mathcal{A}_{\mathbb{PA}}^E$  has an infinite state space, while we will usually treat it like a finite system for obvious reasons. Adapting the definition such that the set of states is effectively reduced to the reachable subset is trivial.


 Figure 6.3.: Reachable Part of the Semantics of  $\lambda.c.\Delta(0) + a.0$ 
**Definition 6.6 (IMC Semantics).**

Let  $\mathcal{A}_{\text{IMC}}^{\bar{s}} = (S, \bar{s}, Act, \multimap, \multimap)$  denote a IMC that is parameterized by its initial state  $\bar{s} \in S$ . Let  $S = \text{IMC}$  and let the transition relation  $\multimap \subseteq \text{IMC} \times Act \times \text{IMC}$  be the least relation given by the rules in Table 6.1, and let the transition relation  $\multimap \subseteq \text{IMC} \times \mathbb{R}_{>0} \times \text{IMC}$  be the least *multi*-relation given by the rules in Table 6.3, where the multiplicity of a transition is determined by the number of derivations that witness its membership.<sup>2</sup>

The semantics of an arbitrary expression  $E \in \text{IMC}$  is defined as  $\mathcal{A}_{\text{IMC}}^E$ . ◁

$$\begin{array}{c}
 \text{prefix} \frac{}{a.E \multimap E} \quad \text{div1} \frac{}{\Delta(E) \xrightarrow{\tau} \Delta(E)} \quad \text{div2} \frac{E \xrightarrow{a} E'}{\Delta(E) \multimap E'} \\
 \\
 \text{choiceL} \frac{E \xrightarrow{a} E'}{E + F \multimap E'} \quad \text{choiceR} \frac{F \xrightarrow{a} F'}{E + F \multimap F'}
 \end{array}$$

Table 6.1.: Operational semantic rules for immediate transitions.

**Theorem 6.1 (Semantic Correspondence).** The semantics of every expression  $E$  of any of the languages  $\text{LTS}$ ,  $\text{PA}$ ,  $\text{IMC}$  is isomorphic to some finite LTS, PA, IMC that is acyclic, except for  $\tau$  self-loops<sup>3</sup> and *vice versa*.

### 6.1.3. Congruence Relations

Due to the strong semantic correspondence between the semantics of process terms and their models (Theorem 6.1), it is not surprising that the notions of bisimilarity we have discussed so

<sup>2</sup> We refer the reader to [Nic+13; BDL13; Bra08] for alternative and arguably more elegant ways to formally deal with potentially multiple occurrences of the same stochastic transition. We here follow the original approach of [Her02].

<sup>3</sup>stemming from the use of  $\Delta$



$$\begin{array}{c}
\text{prefix} \frac{}{a. \bigoplus_{i \in I} p_i E_i \xrightarrow{a} \langle (E_i : p_i) \mid i \in I \rangle} \quad \text{div1} \frac{}{\Delta(E) \xrightarrow{\tau} \delta(\Delta(E))} \\
\text{div2} \frac{E \xrightarrow{a} \mu}{\Delta(E) \xrightarrow{a} \mu} \\
\text{choiceL} \frac{P \xrightarrow{a} \mu}{P + Q \xrightarrow{a} \mu} \quad \text{choiceR} \frac{Q \xrightarrow{a} \mu}{P + Q \xrightarrow{a} \mu}
\end{array}$$

**Table 6.2.:** Operational semantic rules for immediate probabilistic transitions.

$$\begin{array}{c}
\text{prefix-M} \frac{}{\lambda.E \xrightarrow{\lambda} E} \quad \lambda \in \mathbb{R}_{>0} \\
\text{choiceL-M} \frac{E \xrightarrow{\lambda} E'}{E + F \xrightarrow{\lambda} E'} \quad \text{choiceR-M} \frac{F \xrightarrow{\lambda} F'}{E + F \xrightarrow{\lambda} F'}
\end{array}$$

**Table 6.3.:** Operational semantic rules for Markovian transitions.

far can be lifted to process terms in a straightforward way, by calling  $E$  and  $F$  bisimilar if their respective underlying semantic models are bisimilar.

We will define this in a generic way for arbitrary relations over states and arbitrary process languages.

**Definition 6.7.** Let  $S$  be a non-empty set. Let  $\mathbb{L}$  be a language whose expressions  $E, F$  have a semantics  $\llbracket \cdot \rrbracket : \mathbb{L} \rightarrow S$  that maps into  $S$ . A relation  $\mathcal{R} \subseteq S \times S$  is lifted to a relation  $\mathcal{R}'$  over  $\mathbb{L} \times \mathbb{L}$  as follows:

$$E \mathcal{R}' F \text{ if and only if } \llbracket E \rrbracket \mathcal{R} \llbracket F \rrbracket.$$

◁

*Notation 6.4.* In the following, we will not discriminate between a relation  $\mathcal{R}$  and its lifting  $\mathcal{R}'$ .

Axiomatizations can only faithfully characterize an equivalence relation, if this relation is a congruence with respect to all operators of the algebra (or at least those used in any of the axioms). Otherwise, we could not safely replace meta variables by expressions without risking to violate the soundness of the syntactic equations that we derived by means of the axioms.

It is a well-known deficiency of weak bisimilarities that they fail to be a congruence for the non-deterministic choice operator  $+$ .

**Example 6.2.** We consider an example for the most simplest case, labelled transition systems. Consider the two processes  $\tau.a.0$  and  $a.0$ . Obviously, both are weakly LTS bisimilar. However, for example the processes  $c.0 + a.0$  and  $c.0 + \tau.a.0$  are not. We depict their underlying labelled transition systems in Figure 6.4. The reason is that the left process can perform a  $\tau$  transition to

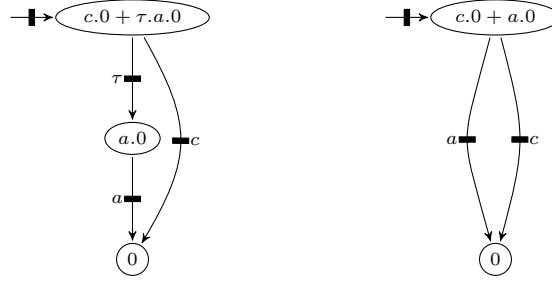


Figure 6.4.: Weak Bisimulation is No Congruence

$a.0$ , but the right process,  $c.0 + a.0$ , cannot perform any  $\tau$  transition. Hence  $c.0 + a.0 \approx_{\text{LTS}} a.0$  were to hold, if their initial states were bisimilar. As this process can now still perform  $b$ , whereas the process  $a.0$  cannot, the two initial process cannot be bisimilar.  $\triangleleft$

The problem is that weak bisimilarity allows to equate processes that only differ in an initial  $\tau$  transition. However, in the presence of non-deterministic choice, such a  $\tau$  transition can spontaneously disable the other summands of the choice. It is well-known that the largest congruence within any of the three notions of weak bisimilarity that we consider here, can be obtained by enforcing that two congruent processes must agree in their ability to perform an initial  $\tau$  transition. This is usually referred to as the *root condition*.

**Definition 6.8 (Root Condition).**

- $E \simeq_{\text{LTS}} F$  holds if and only if  $E \approx_{\text{LTS}} F$  and whenever
  1.  $E \xrightarrow{\tau} E'$  then  $F \xrightarrow{\tau} \circ \implies F'$  and  $E' \approx_{\text{LTS}} F'$ , and
  2.  $F \xrightarrow{\tau} F'$  then  $E \xrightarrow{\tau} \circ \implies E'$  and  $E' \approx_{\text{LTS}} F'$ .
- $E \simeq_{\text{PA}} F$  holds if and only if  $E \approx_{\text{PA}} F$  and whenever
  1.  $E \xrightarrow{\tau} \mu$  then  $F \xrightarrow{\tau} \circ \implies {}_c\gamma$  and  $\mu \mathcal{L}(\approx_{\text{PA}}) \gamma$ , and
  2.  $F \xrightarrow{\tau} \gamma$  then  $E \xrightarrow{\tau} \circ \implies {}_c\mu$  and  $\mu \mathcal{L}(\approx_{\text{PA}}) \gamma$ .
- $E \simeq_{\text{MC}} F$  holds if and only if  $E \approx_{\text{MC}} F$  and whenever
  1.  $E \xrightarrow{\tau} E'$  then  $F \xrightarrow{\tau} \circ \implies F'$  and  $E' \approx_{\text{MC}} F'$ , and
  2.  $F \xrightarrow{\tau} F'$  then  $E \xrightarrow{\tau} \circ \implies E'$  and  $E' \approx_{\text{MC}} F'$ .

$\triangleleft$

**Theorem 6.2.** The relations  $\simeq_{\text{LTS}}$ ,  $\simeq_{\text{PA}}$ , and  $\simeq_{\text{MC}}$  are congruences with respect to all operators of the respective languages.

## 6.2. Axiomatization

The purpose of this section is to compare the different process relations introduced in the preceding chapters, in terms of the sets of axioms that characterize them as precise as possible. Simply

speaking, an axiom is a syntactic law that consist of two process expressions that are joined by the equality sign  $=$ . An *axiomatization* of a process relation  $\simeq$  is a set  $\mathcal{A}$  of axioms, that characterize  $\simeq$  as precisely as possible. This means, that we first of all demand that the axioms in  $\mathcal{A}$  are *sound* with respect to  $\simeq$ . An axiom is called sound with respect to a relation  $\simeq$  if  $E = F$  implies  $E \simeq F$ . Then, if possible, we strive for axiomatizations that are also *complete*. Complete means that whenever  $E \simeq F$ , we can also establish  $E = F$  by the axioms.

Before we investigate the different axiomatization, we need to fix a few formalities.

**Meta Variables and Substitution** While we have so far said that axioms consist of expressions of our languages, this is not true in the strict sense. In fact, we make additional use of *meta variables* that act like place holders that can be instantiated by arbitrary expressions. We denote meta variables by  $E, F, G, H \dots$  as well as primed and indexed variants for LTS, PA, IMC expressions, and  $\mathcal{D}$  and primed and indexed variants for distribution expressions. In fact, we already used meta variables when we defined the languages by means of a grammar.

Strictly speaking, meta variables should not be used in axioms, as they are not part of the expressions of our languages. We use meta variables, however, to express patterns of axioms, which allows us to notate an infinite number of axioms in a finite manner.

**Example 6.3.** In a strict sense,  $E + E = E$  cannot be an axiom as the letter  $E$  is not a valid expression for any of our process languages. We call it, however, a *meta axiom*, as it represents infinitely many axioms, that all follow the same pattern. This meta axiom represents, for instance, the axioms  $0 + 0 = 0$ ,  $a.b.0 + a.b.0 = a.b.0$  and so on.  $\triangleleft$

For simplicity, we will treat meta variables as if they were expressions themselves. So if we speak of expressions in the following, we usually refer also to expressions that include meta variables. We also will call the resulting *meta axioms* simply *axioms* in the following.

For arbitrary expressions  $G$  and  $H$  (possibly containing meta variables themselves), and a *meta variable*  $F$  we write  $(H) \{G/F\}$ , to mean that every occurrence of the meta variable  $F$  in  $H$  is replaced by the expression  $G$ .

**Example 6.4.** The notation  $(E + E) \{a.0 + E/E\}$  stands for  $(a.0 + E) + (a.0 + E)$ .  $\triangleleft$

We denote the simultaneous replacement of variables  $E_1, \dots, E_n$  by the expressions  $H_1, \dots, H_n$  by  $(G) \{H_1, \dots, H_n/E_1, \dots, E_n\}$ . Simultaneous replacement means that all replacements take place at the same moment.

**Example 6.5.** Writing  $(E + F + E) \{a.F, b.0/E, F\}$  stands for  $a.F + b.0 + a.F$ , whereas writing  $((E + F + E) \{a.F/E\}) \{b.0/F\}$  stands for  $a.b.0 + b.0 + a.b.0$ .  $\triangleleft$

**Syntactic Equations and Axiom Instances** A *syntactic equation* is a pair of expressions (including meta variables) joined by the symbol  $=$ . Note that every axiom is a syntactic equation. We say a syntactic equation  $G = G'$  is an *instance* of an axiom  $H = H'$  if there exist *meta variables*  $E_1, \dots, E_n$  and expressions  $H_1, \dots, H_n$  such that  $G = G'$  is identical to

$$(H) \{H_1, \dots, H_n/E_1, \dots, E_n\} = (H') \{H_1, \dots, H_n/E_1, \dots, E_n\}.$$

**Provable Syntactic Rewritability** We say two expressions  $E$  and  $F$  are *provably syntactic rewritable* with respect to the axioms in the set  $\mathcal{A}$ , written as  $\mathcal{A} \models E < F$ , if

1.  $E < F$  is an instance of an axiom in  $\mathcal{A}$ , or
2. there is an expression  $E'$  such that  $\mathcal{A} \models E < E'$  and  $\mathcal{A} \models E' < F$ .

**Notation 6.5** (Provable Syntactic Equality). If the set of axioms  $\mathcal{A}$  is reflexive and symmetric, we write  $E = F$  instead of  $E < F$  and say the two expressions  $E$  and  $F$  are *provably syntactic equivalent*.

As we have discussed, axiomatizations are finite sets, or at least finitely representable sets of syntactic laws that are aimed at precisely characterizing a process relation. Their strength is that by comparing the precise axioms that two axiomatization of different process relations have in common, and in which axioms they differ, we are able to precisely pinpoint semantic differences and similarities between the relations. We will use this power of axiomatizations in the rest of this chapter to analyse and discuss the semantic subtleties of the various process relations we have seen so far.

### 6.2.1. The Axioms

The sound and complete axiomatizations for the calculi together with their respective weak congruences we review in this chapter have been introduced by [Mil89a], [BS01] and [Her02], respectively. For LTS and IMC, the axiomatizations actually extend to arbitrary finite processes including recursion. For PA, the respective extension of the axiom set have been established later by [DPP05]. Also for various other flavors of probabilistic models and notions of process behaviour, that we will not consider in this thesis, sound and complete axiomatizations have been provided [DP07; DP05; Den05; ABW06; PS04; BS01; JS90; LS92; And02; BBS95; Den+08].

Due to the introduction of the special operator  $\Delta$  in the calculi, our axioms slightly deviate from the original axioms from the literature. We will discuss this thoroughly.

To illuminate the differences and similarities between the calculi and their respective weak bisimulation congruences in the following as clear as possible, we present and discuss the axioms not grouped by the congruence they axiomatize, but in such a way that the semantic differences and similarities become apparent immediately by a thematic side-by-side comparison.

**Common Axioms** For equivalence relations, the set of axioms  $\mathcal{A}$  must always be reflexive and symmetric. We will not state this explicitly in the following.

The set of axioms that are shared by all relation is rather small. It is summarized in Table 6.4. Axioms (COM) and (ASS) state that non-deterministic choice is both commutative and associative. The main difference is that syntactically, it is possible to provide a choice between an arbitrary process and the terminated process 0, for example  $E + 0$ , however, semantically, this cannot be distinguished from the process  $E$ , as 0 adds no visible behaviour. Consequently, axiom (NEUT) allows to eliminate 0 from non-deterministic choices.

$$\begin{array}{ll}
 \text{(COM)} & E + F = F + E \\
 \text{(ASS)} & E + (F + G) = (E + F) + G \\
 \text{(NEUT)} & E + 0 = E
 \end{array}$$

**Table 6.4.:** Basic Process Axioms (Common to All Calculi)

LTS, PA	IMC
$(IDEM) \quad E + E = E$	$a.E + a.E = a.E \quad (IDEM-1)$
	$\lambda_1.E + \lambda_2.E = (\lambda_1 + \lambda_2).E \quad (IDEM-2)$

**Table 6.5.:** Idempotency Axioms

**Idempotency** It seems surprising that among the axioms common to all axiomatization, no axiom exists that allows to remove duplicate summands in a non-deterministic choice, for example like  $E + E = E$ . Such a law seems very reasonable, as a non-deterministic choice between identical processes seems to semantically irrelevant. Usually, such an axiom is called an *idempotency* law. In Table 6.5, we can see that for LTS and PA, in fact, this axiom holds.

For IMC, however, this axiom is, in its generality, unsound. The reason is that in IMC, the choice operator not only plays the role of non-deterministic choice, but also the role of stochastic choice between several delays. In the latter, the precise number how often a delay transition occurs makes a difference, as it changes the over delay with which successor states are reached. For instance, the process  $\lambda.0 + \lambda.0$  is behaviourally identical to  $2\lambda.0$ , but not to  $\lambda.0$ .

As a consequence, with respect to immediate transitions, idempotency is valid also for IMC expressions. But with respect to timed transitions, this law is invalid. Hence, in IMC, idempotence is restricted to immediate transitions. For timed transitions, instead, a summation law is introduced that allows to add the rates leading to the same states in accordance with the laws for exponential distributions. We review these differences in Table 6.5. Since in IMC idempotences cannot be represented in a generic way as for the other models, we must also take explicit care for the remaining operator  $\Delta$  and the constant 0. The corresponding law for  $\Delta$  can be found in Table 6.6 ( $(\Delta-6)$ ). For 0, however, no special axiom is needed, since  $(NEUT)$  is enough to eliminate duplicated occurrences of 0.

**Divergence** The second general difference between the weak bisimilarities for LTS and PA, and the IMC weak bisimilarity is the way in which they treat divergence, i.e. the ability of a process to perform an unbounded number of  $\tau$ -transition in succession. In classic process theory, internal transitions are themselves considered as irrelevant for the observable behaviour. Only their indirect influence on observable behaviour is relevant, for example if, by an internal non-deterministic choice, they can decide which observable behaviour is enabled or disabled. Furthermore, in the theories we have considered,  $\tau$  transitions do not represent a fixed quantity of internal activity. Each  $\tau$  transition may stand for zero or any arbitrary number of internal activities. They instead serve as an abstraction mechanism, which is not necessarily coupled with any actual activity inside a system at all. From this perspective, it is only logical if even an infinite sequence of internal transition is not observable, as it does not necessarily correspond to any actual behaviour at all. This viewpoint is represented in LTS weak bisimilarity as well as in probabilistic weak bisimilarity. We find this reflected axiomatically by the axiom  $\Delta(E) = \tau.E$  in Table 6.6, Axiom  $(\Delta)$ . It expresses that  $\Delta$  can be ignored, as long as it is replaced by one single internal transition to ensure congruence with respect to non-deterministic choice (i.e. it maintains the *root condition*).

LTS, PA	IMC
$(\Delta) \quad \Delta(E) = \tau.E$	$\Delta(E) = \tau.\Delta(E) \quad (\Delta.1)$
	$\Delta(\tau.E + F) = \tau.(\tau.E + F) \quad (\Delta.2)$
	$\Delta(\lambda.E + F) = \Delta(F) \quad (\Delta.3)$
	$\Delta(E) + \lambda.F = \Delta(E) \quad (\Delta.4)$
	$\Delta(E + F) = \Delta(E + F) + E \quad (\Delta.5)$
	$\Delta(E) + \Delta(E) = \Delta(E) \quad (\Delta.6)$

**Table 6.6.:** Divergence Axioms

However, this perspective on divergence is not common sense. It has been repeatedly argued that divergence should not be equated with finitary internal behaviour. The argument here assumes that, as internal transitions cannot be influenced by the environment, internal transitions can always preempt any communication by causing a state change before any communication via an observable action can take place. For finite sequences of internal behaviour, the preemptive behaviour *must* finally come to an end, thus the machine cannot refuse communication *ad infinitum*. A diverging machine, however, could decide to deny communication for every, even though communication is still possible. Thus, divergence must be distinguished from finitary internal behaviour. Further finer distinctions in the divergence behaviour are possible. A rather complete overview is given in [Gla93] from the perspective of process relations, while [LDH05] complements this with the axiomatic perspective.

Timed transitions, which are present in IMC, but not in LTS or PA, interfere with internal transition in a yet different way than observable immediate transitions. In the presence of a  $\tau$ -transition, it is supposed that time cannot pass, by the *maximal progress assumption*. As a consequence, timed transitions in states that are not stable do not have any semantic meaning. Similar to our discussion before, the rationale behind this is that immediate transitions cannot be influenced from the environment, and furthermore, as they are merely abstractions, nothing can prevent them from executing without letting time pass before. In the definition of weak IMC bisimilarity (and also of strong IMC bisimilarity), it is thus necessary that a timed transition may only play a role semantically, if it originates from a stable state. Hence, IMC bisimilarities need to distinguish  $\tau$ -behaviour more strictly than the other bisimilarities. From the necessity to preserve stability it follows that divergence cannot be ignored completely. However, with the intention to be still as coarse as possible, weak IMC bisimilarity allows to discard divergence, when the process may escape it by an internal transition. Axiomatically, this discussion is summarized in Table 6.6 in the right column.

**Distribution Axioms** All the axioms that we discussed so far did not contain an explicit reference to distributions, as no prefix term has occurred there. Table 6.7 summarizes the axioms that exclusively deal with distribution expression in the first three lines. They only describe obvious properties. The last axiom of this table, (CC), states that whenever a process exhibits two transitions with the same label, it also implicitly exhibits any of their convex combinations. It syntactically reflects the use of combined transitions in probabilistic weak bisimilarity.

(D-COM)	$a.(\mathcal{D} \oplus_p \mathcal{D}') = a.(\mathcal{D}' \oplus_{1-p} \mathcal{D})$
(D-ADD)	$a.(\mathcal{D} \oplus_p \mathcal{D}) = a.\mathcal{D}$
(D-ASS)	$a.(\mathcal{D}_1 \oplus_p (\mathcal{D}_2 \oplus_{\frac{q}{1-p}} \mathcal{D}_3)) = a.((\mathcal{D}_1 \oplus_{\frac{p}{p+q}} \mathcal{D}_2) \oplus_{p+q} \mathcal{D}_3)$
(CC)	$a.\mathcal{D} + a.\mathcal{D}' = a.\mathcal{D} + a.\mathcal{D}' + a.(\mathcal{D} \oplus_p \mathcal{D}')$

**Table 6.7.:** Distribution Axioms and Convex Combination Axiom

	Non-Probabilistic	Probabilistic	
( $\tau$ -1)	$a.\tau.E = a.E$	$a.(\delta(\tau.E) \oplus_p \mathcal{D}) = a.(\delta(E) \oplus_p \mathcal{D})$	(D $\tau$ -1)

**Table 6.8.:**  $\tau$  - Elimination Axioms

**$\tau$ -Elimination** Except for the axioms involving  $\Delta$ , the axioms we have discussed so far are not specific for the weak aspect of weak bisimilarity. They are also suited to characterize the strong bisimilarity variants. The remaining axioms now treat the characteristics of the weak bisimilarities. In Table 6.8 we find both the probabilistic and the non-probabilistic variant of the simplest  $\tau$ -law, which has been dubbed the ( $\tau$ 1)-law by Milner. Its purpose is to eliminate occurrences of  $\tau$  that are entirely transparent for an observer, as they occur immediately after another transition, and are not part of a non-deterministic choice. In the probabilistic setting, recall that a transition leads to a distribution over states instead of a single state. The idea, however, remains the same here: if any of the probable successor states can only exhibit a  $\tau$  transition as its next behaviour, this transition can be discarded. In Table 6.8, we find the respective axioms for both settings next to each other.

( $\tau$ -2)	$\tau.E = \tau.E + E$
( $\tau$ -3)	$a.(E + \tau.F) = a.(E + \tau.F) + a.F$

**Table 6.9.:** Non-Probabilistic  $\tau$  - Preemption Axioms

To completely characterize weak bisimilarity, two more axioms are needed. Both deal with the preemptive power of  $\tau$ -transitions. The well known laws for non-probabilistic systems are summarized in Table 6.9. In the probabilistic variant, which is summarized in Table 6.10, the first axiom agrees essentially. The only difference is that states are replaced by distributions where necessary. In contrast, the second axiom differs from the corresponding law of the non-probabilistic setting. Replacing states by distributions is not possible in this case, as then the distribution replacing the process term  $E$  would have to occur both after a prefix expression as well as without a preceding prefix expression. The latter construct is, however, syntactically forbidden. The essential idea behind this axiom is that every behaviour that is represented by  $E$

$$\begin{aligned}
(D\cdot\tau\cdot 2) \quad a.\mathcal{D} &= a.\mathcal{D} + a.(\bigoplus_{i \in I} p_i \mathcal{D}'_i) \text{ where } \mathcal{D} = \bigoplus_{i \in I} p_i \delta(E_i + \tau.\mathcal{D}'_i) \\
(D\cdot\tau\cdot 3) \quad \tau.\mathcal{D} &= \tau.\mathcal{D} + a.(\bigoplus_{i \in I} p_i \mathcal{D}'_i) \text{ where } \mathcal{D} = \bigoplus_{i \in I} p_i \delta(E_i + a.\mathcal{D}'_i)
\end{aligned}$$

**Table 6.10.:** Probabilistic  $\tau$  - Preemption Axioms

cannot only be executed after executing  $\tau$ , but can – observationally – also appear to be executed immediately. In the probabilistic setting we can recover this observation by decomposing the distribution  $\mathcal{D}$  that replaces  $E$ , or more precisely, the process expressions in its support, in a sum form  $E_i + a.\mathcal{D}'_i$ . If this sum form then reveals that every process expression in the support enables (at least) one transition labelled with the same action, then the respective goal distributions  $\mathcal{D}'_i$  can be factored out on top level and recombined into  $a. \bigoplus_{i \in I} p_i \mathcal{D}'_i$ . It is important to emphasize that behaviour can only be factored out in this way when all process expressions agree on the action name (here:  $a$ ).

We finally summarize the preceding discussion in a theorem. We first define a sets of axioms for each notion of bisimulation congruence. Let  $\mathcal{A} = \{(COM), (ASS), (NEUT)\}$ . Then we let

$$\begin{aligned}
\mathcal{A}_{\simeq_{LTS}} &= \mathcal{A} \cup \{(IDEM), (\Delta), (\tau\cdot 1), (\tau\cdot 3), (\tau\cdot 2)\} \\
\mathcal{A}_{\simeq_{PA}} &= \mathcal{A} \cup \{(IDEM), (\Delta), (D\cdot\tau\cdot 1), (D\cdot\tau\cdot 2), (D\cdot\tau\cdot 3), \\
&\quad (D\cdot COM), (D\cdot ADD), (D\cdot ASS), (CC)\} \\
\mathcal{A}_{\simeq_{MC}} &= \mathcal{A} \cup \{(IDEM\cdot 1), (IDEM\cdot 2), \\
&\quad (\Delta\cdot 1), (\Delta\cdot 2), (\Delta\cdot 3), (\Delta\cdot 4), (\Delta\cdot 5), (\Delta\cdot 6), (\tau\cdot 1), (\tau\cdot 3), (\tau\cdot 2)\}
\end{aligned}$$

**Theorem 6.3** (*Axiomatizations*).

- $\mathcal{A}_{\simeq_{LTS}}$  is sound and complete for  $\simeq_{LTS}$ .
- $\mathcal{A}_{\simeq_{PA}}$  is sound and complete for  $\simeq_{PA}$ .
- $\mathcal{A}_{\simeq_{MC}}$  is sound and complete for  $\simeq_{MC}$ .

### 6.3. Probabilistic Forward Simulation

After our comparison of the axiomatization of the different weak congruences, we elaborate on a sound and complete axiomatization for probabilistic forward simulation on  $\mathbb{PA}$  and discuss its relation to the axiomatization of  $\simeq_{PA}$ . This will provide a core insight that is of interest in its own right, and also for later deliberations. Probabilistic forward simulation significantly differs from the other relations by being a preorder instead of an equivalence relation. Therefore, we separate the discussion from the other relations.



### 6.3.1. Axioms

The fact that probabilistic forward simulation is defined as a relation over  $S \times \text{Dist}(S)$ , and is thus a binary relation over two different types, turns out to be an inconvenient obstacle for a technically sound axiomatic treatment. To overcome this, we will in the following use the restriction of  $\preceq_{fwd}$  to states.

**Definition 6.9.** Let  $E \preceq'_{fwd} F$  hold if and only if  $E \preceq_{fwd} \delta(F)$ .  $\triangleleft$

*Notation 6.6.* We overload the symbol  $\preceq_{fwd}$  and also use it to denote  $\preceq'_{fwd}$  for the rest of this chapter.

**Precongruence Axioms** We first note that  $\preceq_{fwd}$  already is a congruence relation, with respect to the non-deterministic choice operator  $+$ ; or more precise, a *precongruence*. Thus, different to what we have seen for the other relations on LTS, IMC and PA, the relation is immediately amenable for an axiomatization.

However, as  $\preceq_{fwd}$  is a preorder, and not an equivalence, we use the symbol  $\leq$  in our axiomatization. Although  $\preceq_{fwd}$  is not an equivalence, it is often helpful to write  $E = F$  if the preorder holds from left to right and vice versa.

*Notation 6.7.* We write  $E = F$  if  $E \leq F$  and  $F \leq E$ .

The set of axioms that is sound and complete for  $\simeq_{PA}$  is also sound for  $\preceq_{fwd}$ , and forms the basis for our axiomatization – with one notable exception. As a precongruence is not symmetric, the symmetry axiom, which we did not list explicitly, is not sound with respect to  $\preceq_{fwd}$ . All other elementary axioms that we did not state explicitly, namely, reflexivity and transitivity are also sound for  $\preceq_{fwd}$ .

**Distinguishing axioms** Table 6.11 summarize the axioms that need to be added to the axiomatization of  $\simeq_{PA}$  in order to obtain a *complete* axiomatization of  $\preceq_{fwd}$ .

$$\begin{array}{ll}
 E \leq E + F & (F1) \\
 a.(\mathcal{D} \oplus_p \delta(b.(\mathcal{D}_1 \oplus_q \mathcal{D}_2))) \leq a.(\mathcal{D} \oplus_p (\delta(b.\mathcal{D}_1) \oplus_q \delta(b.\mathcal{D}_2))) & (F2) \\
 a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D}) & (D\cdot\tau\cdot 1)_{MA} \\
 \tau.\delta(E) = E & (D\cdot\tau\cdot 1)_{fwd}
 \end{array}$$

**Table 6.11.:** Additional Axioms for  $\preceq_{fwd}$ .

Axiom (F1) is typical for simulation relations. It allows the right-hand-side process, i.e. the simulating process, to perform arbitrary *additional* activities, as long as every activity of the left-hand-side process is completely preserved. Formally, this is expressed by allowing to add any process  $F$  in non-deterministic choice to the original process  $E$ . This axioms makes clear that simulation relations do not preserve, but only enforce containment of the non-deterministic branching structure of a process. In contrast, axioms (F2) shows that  $\preceq_{fwd}$  also weakens the degree to which the *probabilistic* branching structure is preserved. If we focus our attention to the expressions on the right-hand side of  $\oplus_p$  on both sides of the inequality, (F2) states that every process whose only initial behaviour is an action  $b$  can be simulated by a distribution over different processes all beginning individually with action  $b$ , but with different continuations

afterwards, at least as long as the overall probabilistic behaviour is the same after  $b$  has occurred. In other words, it says that any (binary) splitting of the successor distribution  $\mathcal{D}$  of a process  $b.\mathcal{D}$  can simulate it, as long as the parts of the splittings are individually prefixed by  $b$ . The difference is that the place where a probabilistic decision is made is shifted to an earlier point in time. The reason why the axiom is slightly more complicated and needs the prefixing with  $a$ . is due to the fact that we need a process expression on both sides of the inequality, and can thus not directly talk about distribution expressions as we did before.

**Example 6.6.** Compare the processes  $E = d.a.(\delta(b.0) \oplus_{0.5} \delta(c.0))$  and  $F = d.(\delta(a.b.0) \oplus_{0.5} \delta(a.c.0))$ , where we omitted the symbol  $\delta$  for readability where it is clear from the context. It holds that  $E \preceq_{fwd} F$ . Note that for both process the occurrence of the action  $d$  and  $a$  is 1, and for  $b$  and  $c$ , respectively, it is 0.5. They differ, however, in the moment when it is decided (in a probabilistic sense) which of  $b$  and  $c$  will occur. In  $E$ , the decision happens with the execution of  $a$ , while in  $F$ , this happens already in the preceding execution of  $d$ .  $\triangleleft$

So we clearly see that in the simulating process a probabilistic decision may always be shifted towards an earlier point in time. This is very notable, especially as  $a$  is an observable action. So in this way, the probabilistic aspect of an activity may be shifted towards a preceding (observable) activity.

Axiom  $(D\cdot\tau\cdot I)_{MA}^4$ , in contrast, is less radical. It simply allows to shift internal probabilistic decisions forward (or backward, as the rule induces an equality). Intuitively, this means that whenever parts of an observable probabilistic activity are actually implemented by subsequent internal activities, the latter can be completely ignored semantically. So internal *probabilistic* decisions become transparent to an observer.

Axiom  $(D\cdot\tau\cdot I)_{fwd}$  finally allows to remove a  $\tau$  prefix for processes. This is typical for weak precongruences, which do not need the congruence fix we have applied for the *bisimulation*.

**Comparison with  $\simeq_{PA}$**  Being a simulation relation, it is not surprising that  $\preceq_{fwd}$  is less preserving with respect to the non-deterministic branching structure than the bisimilarity  $\simeq_{PA}$ , as enforced by axiom  $(F\cdot I)$ . Similarly, the idea of shifting probabilistic decisions between observable activities, as expressed by axiom  $(F\cdot 2)$ , is clearly not apt for a bisimulation. An interesting axiom is axiom  $(D\cdot\tau\cdot I)_{MA}$

$$a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D}).$$

It is a generalization of  $(D\cdot\tau\cdot I)$

$$a.(\delta(\tau.E) \oplus_p \mathcal{D}) = a.(\delta(E) \oplus_p \mathcal{D}).$$

$(D\cdot\tau\cdot I)_{MA}$  allows to add or remove arbitrary internal transitions, while  $(D\cdot\tau\cdot I)$  only allows this as long as they do not involve any probabilistic decisions (besides a Dirac decision). In this way, we see that  $\simeq_{PA}$  does not allow any modification of the probabilistic branching structure, while  $\preceq_{fwd}$  is very liberal in this respect. While it is arguably reasonable to forbid anything like  $(F\cdot 2)$  to be sound with respect to any reasonable notion of probabilistic bisimilarity, a similarly definitive answer cannot be given for  $(D\cdot\tau\cdot I)_{MA}$ , as this axiom basically merely allows to implement a single observable probabilistic decision by sequences of several internal *unobservable* probabilistic decisions. Allowing such an axiom seems to confine perfectly with the driving idea behind *weak* bisimilarities, namely, that internal activities should be transparent to the observer as much as possible.

<sup>4</sup>the index MA in the name of the axiom hints at its importance for Markov automata (see Chapter 9)

**Discussion** We claim that our axiomatization of  $\preceq_{fwd}$  is sound and complete. It is strongly inspired by [PS04], where the authors already claim to provide a sound and complete axiomatization for trace distribution precongruence, which coincides with probabilistic forward simulation. However, it seems that the original axiomatization given in [PS04] is technically flawed. The reason is that there, probabilistic forward simulation is treated unchanged as a relation over  $S \times \text{Dist}(S)$ . This leads to axioms like  $E = \delta(E)$ . Clearly, resolving this into  $E \leq \delta(E)$  and  $\delta(E) \leq E$  is unsound, as the latter is a mismatch with respect to the type signature of probabilistic forward simulation. Already the axiom  $E \leq \delta(E)$  is technically problematic, as it allows to derive  $E \leq \delta(\delta(E))$  with two consecutive applications. However,  $\delta(\delta(E))$  is a distribution over distributions over processes, and we can repeat this principle to arrive at arbitrarily nested distributions. In this way, the axiom allows to relate expressions that are not even contained in the algebra.

Still, the cited work has its clear merits in providing the essential insights for an axiomatization. Axiom  $(D\text{-}\tau\text{-}I)_{MA}$  and  $(D\text{-}\tau\text{-}I)_{fwd}$  already appears there, yet, combined in one single axiom,  $\tau.\mathcal{D} = \mathcal{D}$ . It unfortunately suffers from the above mentioned problem in that it equates process expressions and distribution expressions. Our axiomatization aims at healing these formal deficiencies while preserving the original insights. Yet, we leave the formal proof of soundness and completeness as an open problem.

## 6.4. Summary and Discussion

In this chapter, we have compared the different bisimulation relation for LTS, PA and IMC from a process algebraic perspective by means of a comparison of their sound and complete axiomatizations. In the course of this we have highlighted important differences among the bisimilarities that are on the one hand obviously owed to the different settings, such as the presence of timed transitions and probabilities, but also on the other hand, more subtle differences that have to do with the way internal behaviour is treated. Most prominently, we have seen that divergence does not differ from finite sequences of internal behaviour from the perspective of weak LTS bisimilarity and weak probabilistic bisimilarity, whereas for weak IMC bisimilarity, it plays a more pronounced role. In IMC, internal behaviour directly interferes with timed behaviour by the maximal progress assumption. This makes it necessary to distinguish finite sequences of internal behaviour from infinite behaviour, which shows in the axiomatization by refining the single axiom  $\Delta(E) = \tau.E$  for LTS and PA into three more elaborate axioms for IMC.

The arguably most remarkable insight of this chapter has emerged during our comparison of the axiomatizations of weak probabilistic bisimilarity and probabilistic forward simulation. Most differences in the axioms can be explained by the simple fact that the former relation is a bisimulation, while the latter is a simulation and thus naturally less restrictive. Yet, we identified one axiom that is at the core of the axiomatization of probabilistic forward simulation, and that is unsound for weak probabilistic bisimilarity, while it does actually not at all conflict with the principles of bisimulations: axiom  $(D\text{-}\tau\text{-}I)_{MA}$ . We have conjectured that this axiom might even be needed for a fully-fledged extension of the idea of *weak* bisimilarity to a probabilistic setting. In fact, the semantic idea behind this axiom will be our guide line throughout the second part of the thesis. In Chapter 9, we will come back to precisely this conjecture from the axiomatic perspective.



**Part II.**

# **Markov Automata Theory**



## 7. Markov Automata

Developing a model of concurrent and reactive systems that exhibits non-deterministic, probabilistic and stochastically timed behaviour at the same time and providing sound semantic foundations has been an open challenge for decades. In the realm of automata-based models, with their semantic roots in interleaving concurrency semantics, no single such model has emerged so far, to the best of our knowledge. This is especially surprising because

1. the scientific and industrial demand for such a model has been undaunted since many decades,
2. many models covering only one of the two above quantitative aspects have been developed, and do find wide-spread use.

With (Generalized) Stochastic Petri Nets, such a model arguably exists, however, building on the foundations of Petri Nets. Translating the concepts of this model into an automata-based, interleaving setting seems a simple task, at first sight. In fact, importing the structural aspects of such a model is a relatively straightforward task, especially as both probabilistic and stochastic time models are already at hand to provide the necessary inspiration. The real challenge in finding such a model does not lie in the question how to assemble the necessary quantitative aspects. The challenge is to find sound semantic foundations. For Generalized Stochastic Petri Nets, this challenge has not been met, in our opinion. A certain class of *GSPN*, so called *confused* nets, has been willingly denied any semantic treatment. As we shall detail out in Chapter 12, confused models often turn out to be those models where non-determinism has been used as a purposeful modelling means other than expressing concurrent execution. While, in general, non-determinism is highly valued as a powerful tool for abstraction or to allow for freedom of implementation, in the context of *GSPN* it is depreciated as semantic underspecification. As a consequence of the missing semantic foundations, no analysis trajectory for such systems has been developed.

In the realm of automata-based models, the quantitative analysis of systems with *any* kind of non-determinism has been actively studied since more than one decade. Also general quantitative model-checking algorithms for logics like CSL have become available recently [Neu10]. However, very often these advanced and costly analysis methods can be avoided, as the non-determinism present at the model is behaviourally irrelevant with respect to a suitable notion of bisimilarity. Even if reducing a model with respect to bisimilarity cannot always eradicate non-determinism entirely, as it, for instance, may be the result of an intentional modeller decision, it allows to reduce the presence of non-determinism to its necessary minimum. For this reason, bisimilarity and techniques such as quotient construction up to bisimilarity have been at the core of many analysis trajectory of automata-based concurrent and reactive systems both with and without quantitative aspects. Last, but not least, bisimilarity itself is a valuable technique to prove that a system's specification and implementation agree. However, for models combining non-determinism, probabilism and stochastic time, no semantically satisfactory notion of bisimilarity has been known so far. We stipulate that this missing foundation is the reason why in the

realm of automata-based models no combination of both quantitative aspects has been brought up yet.

**Outline and Contributions.** This chapter introduces a novel automata-based model that combines non-determinism, probabilism and stochastic time: Markov automata. This chapter is the first of the second part of this thesis, which is dedicated to the foundations of Markov automata. In this chapter, we develop the foundations on the level of automata structures, and show that the models emerges as the union of PA and IMC. The model itself will be introduced in Section 7.1. We first define a very general model with almost no restriction on the number of states and the structure of the transitions. However, for most of the later development, it will be necessary to restrict the model class to different extents. We define the necessary terms and concepts in this section, too. In Section 7.2, we define parallel composition and abstraction operators on MA. Concluding this chapter, in Section 7.3, we review several well-known (quantitative) models, among them LTS, IMC, PA and their respective sub-models, and evaluate them from the perspective of being sub-models of Markov automata themselves.

## 7.1. Model

Markov automata aim to integrate both probabilistic immediate behaviour as known from PA *and* stochastic timed delay as known from IMC. Quite naturally, we define Markov automata as the straightforward combination of the two models. The reason why this smooth integration is possible is that both models extend labelled transition systems in two orthogonal ways. In particular, both keep the essential idea that concurrency can be modelled as the non-deterministic choice between the interleavings of the (sequential) behaviour of component systems. Concerning the probabilistic automata aspect of Markov automata, this becomes especially visible by the fact that probabilism is only present within individual transitions, while the choice between transition remains non-deterministic. Interactive Markov chains, in turn, do not alter the principles of labelled transition systems, but only extend them in a strictly orthogonal manner by introducing a new transition relation,  $\rightarrow$ , that describes the delay in the form of the parameter of the exponential distribution that characterizes the stochastic delay. So to arrive at Markov automata, it suffices to add  $\rightarrow$  to probabilistic automata, or conversely, to refine IMC immediate transitions with probabilities. Thus, we work with a twofold transition relation  $\Rightarrow$ , containing *immediate probabilistic* transitions and  $\rightarrow$ , containing the Markovian rate transitions.

**Definition 7.1 (Markov Automaton).** A *Markov automaton* (MA) is a quintuple  $(S, \bar{s}, Act, \Rightarrow, \rightarrow)$ , where

- $S$  is a non-empty countable set of states,
- $\bar{s} \in S$  is the initial state,
- $Act$  is a set of actions containing the internal action  $\tau$ ,
- $\Rightarrow \subseteq S \times Act^x \times Dist(S)$  is a set of probabilistic transitions, and
- $\rightarrow \subseteq S \times \mathbb{R}_{>0} \times S$  is a multiset of Markov timed transitions.



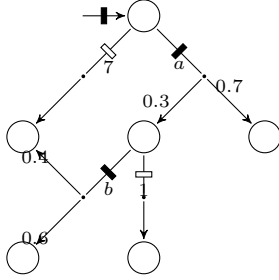
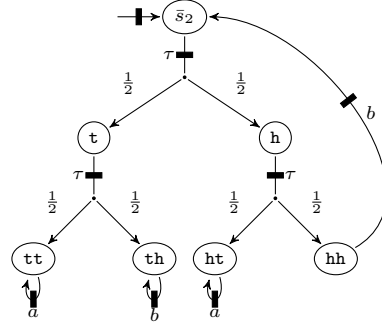


Figure 7.1.: A Markov Automata

Figure 7.2.: Cyclic, but  $\tau$ -Acyclic MA

The sum of the labels of Markov timed transitions leaving a state must be finite, i.e. for each  $s \in S$

$$\sum_{(s,r,s') \in \rightarrow} r < \infty.$$

◁

We assume that  $Act$  contains the atomic action labels (including  $\tau$ ) as we know them from probabilistic automata. Whenever convenient, we use the same notational convention we introduced in Notation 5.2 on page 64, to represent timed transitions as probabilistic transitions labelled by a special action  $\chi(r)$ , where  $r \in \mathbb{R}_{\geq 0}$ . We treat  $\chi(r)$  like an ordinary action. We write  $Act^X$  to denote the set of all immediate actions and all actions of the form  $\chi(r)$ . Note that this set does then not contain actual real numbers (without being encapsulated by  $\chi$ ), since  $Act$  itself is assumed to not contain numbers. Especially for the definition of bisimilarities, this notation will turn out handy. We call this representation of Markov automata their uniform representation.

**Definition 7.2 (Uniform Representation of an MA).** The uniform representation  $\mathcal{A}^U$  of an MA  $\mathcal{A} = (S, \bar{s}, Act, \rightarrow, \rightarrow)$  is a quadruple  $(S', \bar{s}', Act^X, \rightarrow', \rightarrow')$ , where

- $S' = S$ ,
- $\bar{s}' = \bar{s}$ ,
- $\rightarrow' = \rightarrow$ , and
- $\rightarrow' = \{ s, \chi(r), \mu \mid s \xrightarrow{\chi(r)} \mu \text{ in } \mathcal{A} \}$ .

◁

As already discussed in Chapter 5, the uniform representation of a MA *cannot* be translated back to a standard MA in a unique way. Thus, converting an MA into its uniform representation bears a certain loss of information. However, this loss is only syntactical, as all semantically relevant information is fully preserved.

As the interactive transitions of MA are as well probabilistic, we can now use the transition predicate  $\rightarrow$  to uniformly represent timed and immediate transitions. We thus write for a

Markov automaton  $s \xrightarrow{a} \mu$  if and only if either  $a \in Act$  and  $s \xrightarrow{a} \mu$  or  $a = \chi(r)$  and  $s \downarrow$  and  $r = \sum_{(s,x,s') \in \neg\Rightarrow} x$  and for all states  $t$  the distribution  $\mu$  satisfies  $\mu(t) = \frac{\sum_{(s,x,t) \in \neg\Rightarrow} x}{r}$ .

### 7.1.1. Structural Restrictions

While we try to be as general as possible, automata with infinite state space become hard to handle very quickly, if they do not meet an elementary restriction, that allow us to apply probability theoretic standard arguments. This will become apparent later, when notions of weak bisimilarities are discussed. We therefore restrict our attention to different restricted subclasses of infinite MA, or of even finite MA, depending on the context.

If not stated otherwise, we restrict Markov automata to a countable branching structure. As stated already at the definition of MA, we always assume a countable state space and hence, as well countable support for each transition.

A widely-used restriction is to only consider systems that have a finite number of outgoing transitions for each state.

**Definition 7.3 (Image-Finiteness).** An MA  $\mathcal{A} = (S, \bar{s}, Act, \neg\Rightarrow, \neg\Rightarrow)$  is *image-finite*, if for each state  $s \in S$  and every action  $a \in Act$ , the set

$$\{(s, a, \mu) \in \neg\Rightarrow\}$$

is finite. ◁

Note that this restriction does not exclude automata having transitions with countably infinite support. Furthermore, the number of outgoing timed transitions may remain infinite, as well as the total number of outgoing immediate transitions. Image-finiteness restricts only the number of outgoing immediate transitions with identical action label.

**Definition 7.4 (Finite-Support).** We say an MA is *finite-support*, if for each  $s \in S$  the number of outgoing Markovian timed transitions is finite *and* for each transitions  $(s, a, \mu) \in \neg\Rightarrow$ , the support of  $\mu$  is finite. ◁

In the following, we will identify different classes of MA with infinite state space. The next definition is a helpful tool we need to specify the classes.

**Definition 7.5 (Structure Graph).** The structure graph of a MA  $(S, \bar{s}, Act, \neg\Rightarrow, \neg\Rightarrow)$  is a labelled graph  $(S', E, L)$  derived from  $\mathcal{A}$  as follows:

- $S' = S$ ,
  - $L = Act^X$
  - $(s, a, s') \in E$  if and only if  $s \xrightarrow{a} \mu$  for some  $\mu$  with  $\mu(s') > 0$  and  $a \in Act^X$  including  $a = \chi(r)$  for some  $r \in \mathbb{R}_{\geq 0}$ .
- ◁

The structure graph abstracts the structure of an MA by replacing each probabilistic and stochastic transition to a set of transitions. For each state in the support of the original goal

distribution a transition is added. The transition inherits its label from the transition it is derived from. It preserves reachability within an MA, while being entirely forgetful about the concrete probability with which a state is reached.

**Definition 7.6 (Structural Properties).** A Markov automaton is called

1. *finite* if its structure graph is *finite*,
2. *weakly image-finite* if it is *image-finite*, *finite-support* and its structure graph is  $\tau$ -finite,
3.  $\tau$ -(a)cyclic, or (a)cyclic, respectively if its structure graph is  $\tau$ -(a)cyclic, or (a)cyclic, respectively.<sup>1</sup>

◁

Note that a finite Markov automata can still contain infinitely many action in the set  $Act$ . However, only a finite number of them will be actually used as transition labels.

A weakly image-finite MA can be thought of an MA that is locally finite in terms of the states and distributions that can be reached from each state by weak transitions. Note that it may overall have an infinite number of states and transitions.

As for PA, weak transitions of MA are, in principle, tree-like structures similar to an unfolding of the automaton along an execution (cf. Definition 4.11). Different to non-probabilistic models such as LTS and IMC, PA and MA weak transitions may contain infinite sequences of internal transitions as part of these trees. Therefore, (a)cyclicity of transitions and (in)finiteness of the state space, especially for states reachable with internal transitions, make a difference in this setting. For external actions, these differences effectively play no role for most theoretic considerations, as they anyhow occur at most once in every weak transition.

**Example 7.1.** The MA in Figure 7.2 is cyclic, but both  $\tau$ -acyclic and weakly image-finite. The latter two properties ensure that every weak transition corresponds to a finite underlying transition tree. Phrased differently, weakly image-finiteness guarantees that from every state some other state can be reached (by a weak transition) whose outgoing transitions are *all* labelled by external actions, and  $\tau$ -acyclic means that the probability to see the same state twice on a run as long as only internal transitions are executed, is zero. ◁

Weakly image-finite MA guarantee that weak (combined) transitions only lead to distributions with finite support, even though it neither restricts the state space to be finite in general, nor does it disallow infinitary weak transitions. Still, the number of both weak and weak combined transitions from one state may be infinite. However, it is always possible to represent each such transitions by a *finite* convex combination of weak transitions.

**Lemma 7.1.** Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  be a *weakly image-finite* MA. Every *infinitary* convex combination of weak (hyper-)transitions can also be represented by a *finitary* convex combination *and vice versa*.

*Proof.* It is obvious that with infinitary convex combinations we can represent every finitary convex combination by simply letting almost all coefficients be 0.

We now consider the other, less obvious direction. By our restriction to weakly image-finite MA, we only need to consider weak hypertransitions of distributions with *finite* support. Thus,

<sup>1</sup>We refer the reader to Definition 2.16 for the definition of  $\tau$ -(a)cyclicity and  $\tau$ -finiteness.

it suffices to consider the case for weak transitions instead of hypertransitions, as each hypertransition can then be represented as a *finite* convex combination of weak transitions.

Let  $s \in S$ . Whenever  $s \xRightarrow{a}_c \mu$  then, by the definition of weak *combined* transitions, there exists an *countable* number of real values  $c_i \in [0, 1]$  with countable index set  $I$  and weak transitions  $s \xRightarrow{a} \mu_i$  such that

$$\mu = \bigoplus_{i \in I} c_i \mu_i. \quad (7.1)$$

For a fixed state  $s \in S$  and an action  $a \in Act$ , weakly image-finite MA behave like finite state, image-finite and finite-support MA on the subset of states that is reachable with probability greater zero from  $s$  with weak (combined)  $a$ -transitions. Furthermore, by using its uniform representation, they can be treated like PA. We may thus use the following claim, that is a immediate consequence of Propositions 3 and 4 of [CS02], which hold for finite state, image-finite and finite-support PA.

**Claim 7.1.** *The set of distributions  $\mu$  that can be reached by a weak (non-combined) transition  $s \xRightarrow{a} \mu$ , is generated by a finitary convex combination of a finite set of distributions  $s \xRightarrow{a} \nu_k$ . Formally, for every state  $s$  and every action  $a$  there is a finite set of generating distributions  $\nu_k$  with  $s \xRightarrow{a} \nu_k$  (with  $k \in K$ ), such that for every  $\mu$ , with  $s \xRightarrow{a} \mu$  there exist coefficients  $c_k \in [0, 1]$  such that*

$$\mu = \bigoplus_{k \in K} c_k \nu_k.$$

We continue the proof of the lemma. For a given distribution  $\mu$ , we can by Equation 7.1 and above claim, write

$$\mu = \bigoplus_{i \in I} c_i \mu_i = \bigoplus_{i \in I} c_i \bigoplus_{k \in K} c_k^i \nu_k.$$

Note that while the coefficients  $c_k^i$  depend on  $i$ , the generating distributions  $\nu_k$  are the same for each  $i$  (clearly possibly prefixed by a coefficient  $c_k^i = 0$ ). The last term can be further rewritten to

$$\begin{aligned} & \bigoplus_{i \in I} c_i \bigoplus_{k \in K} c_k^i \nu_k \\ &= \bigoplus_{i \in I, k \in K} c_i c_k^i \nu_k \\ &= \bigoplus_{k \in K} \left( \bigoplus_{i \in I} c_i c_k^i \right) \nu_k \end{aligned}$$

As  $K$  is finite, it becomes obvious that we obtain the sought *finitary* convex combination by taking  $\bigoplus_{i \in I} c_i c_k^i$  as coefficients and  $\nu_k$  as distributions. This ends the proof of Lemma 7.1.  $\square$

## Compactness

An important and standard restriction in the context of probabilistic behaviour is to resort to automata, whose transitions define a *compact* space. To define this notion formally, we first introduce a widely used metric on distributions [Des+99].

**Definition 7.7 (Distribution Metric).** For a countable set  $X$  we define the metric  $d$  on  $\text{Dist}(X)$  as  $d(\mu, \gamma) = \sup_{Y \subseteq X} |\mu(Y) - \gamma(Y)|$ .  $\triangleleft$

**Definition 7.8 (Compactness).** A MA  $\mathcal{A}$  is *compact*, if for every state  $s$  and label  $a \in \text{Act}^X$  the set  $\{\mu \mid s \xrightarrow{a}_c \mu\}$ <sup>2</sup> is compact according to the metric  $d$ .  $\triangleleft$

**Lemma 7.2.** Every finite MA is compact.

This lemma has been shown in [Des+10]. The compactness assumption allows us to assume that whenever some state (or distribution) can approximate a distribution with weak transitions arbitrarily closely, then it can also reach the distribution itself by a weak transition.

**Corollary 7.1 (Closedness).** Let  $s \in S$  and  $a \in \text{Act}^X$  in a compact MA. Given any infinite sequence of distributions  $\{\mu_i\}_{i \in I}$  which converges towards the distribution  $\mu$  with respect to the metric  $d$ . If  $s \xrightarrow{a}_c \mu_i$  for every  $i \in I$ , then also  $s \xrightarrow{a}_c \mu$ .

This fact follows immediately, as the set of all distributions reachable from  $s$  is compact and thus closed, if the MA is compact. It will turn out as a helpful tool in many proofs. In fact, many results in this thesis would not hold without this property.

The theorem of Bolzano-Weierstraß can be used to show two useful properties about convergence of sequences distributions.

**Lemma 7.3.** Let  $S$  be a finite set.

1. Every sequence of distribution  $\langle \xi_i \rangle_{i \in \mathbb{N}}$  over  $S$  has a convergent subsequence, i.e. there exists  $I \subseteq \mathbb{N}$  and a distribution  $\xi'$  with

$$\langle \xi_i \rangle_{i \in I} \longrightarrow \xi'.$$

2. Let  $\mathcal{R}$  be a binary relation between distributions. Let  $\langle \mu_i \rangle_{i \in I}$  be a convergent sequence of distributions over  $S$  with limiting distribution  $\mu$ . Let  $\langle \gamma_i \rangle_{i \in I}$  be some sequence of distributions over  $S$  satisfying  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in I$ . Then there exists an infinite subset  $J \subseteq I$  such that  $\langle \mu_i \rangle_{i \in J}$  is still a convergent sequence with limiting distribution  $\mu$ , and  $\langle \gamma_i \rangle_{i \in J}$  is now also a *convergent* sequence with some limiting distribution  $\gamma'$ . Clearly,  $\mu_i \mathcal{R} \gamma_i$  still holds for each  $i \in J$ .

We will frequently apply this lemma to distributions over the set of states  $S$  of a weakly image-finite Markov automata, where all distributions result from weak transitions of a finite number of source distributions. Here, the set of states  $S$  is not necessarily finite, but due to the automaton being weakly image-finite, the set of states relevant for our consideration can be reduced to a finite subset. In these cases, the lemma is then still applicable. We will do so without further mentioning if the necessary preconditions are clear from the context.

The next lemmas are refinements of the crucial insight that if convergent sequences of distributions agree on their behaviour up to some notion of equivalence between states, then also the limiting distributions of the sequences must do so.

**Lemma 7.4.** Let  $\mathcal{R}$  be an equivalence class over  $S$ . Let  $\langle \mu_i \rangle_{i \in I}$  and  $\langle \gamma_i \rangle_{i \in I}$  be two convergent infinite sequences of distributions over  $S$ , with limiting distributions  $\mu$  and  $\gamma$ , respectively. If for every  $i \in I$

$$\mu_i \mathcal{L}(\mathcal{R}) \gamma_i$$

<sup>2</sup>The definition of weak (combined) transitions for Markov automata is completely analogue to the respective definitions for PA and IMC. We do not provide an explicit formal definition.

then also for the limiting distributions

$$\mu \mathcal{L}(\mathcal{R}) \gamma.$$

*Proof.* Assume, in order to arrive at a contradiction, that there exists an equivalence class  $\mathcal{C}$  of  $\mathcal{R}$  such that

$$|\mu(C) - \gamma(C)| = k > 0 \text{ for some } k \in \mathbb{R}_{>0}. \quad (7.2)$$

As the  $\mu_i$  converge against  $\mu$ , for any  $\varepsilon_\mu$  there must exist an index  $l \in I$  such that  $d(\mu_l, \mu) < \varepsilon_\mu$ . Accordingly, as the  $\gamma_i$  converge against  $\gamma$ , for any  $\varepsilon_\gamma$  there must exist an index  $m \in I$  such that  $d(\gamma_m, \gamma) < \varepsilon_\gamma$ .

By the definition of the metric  $d$  it must hold that

$$|\mu(C) - \mu_l(C)| < \varepsilon_\mu \text{ and } |\gamma(C) - \gamma_m(C)| < \varepsilon_\gamma. \quad (7.3)$$

Now choose  $\varepsilon_\mu = \varepsilon_\gamma = \frac{1}{2}k$  and let  $n \in I$  be any index greater than  $\max(l, m)$ . As  $\gamma_n \mathcal{L}(\mathcal{R}) \mu_n$ , it must hold that

$$\gamma_n(C) = \mu_n(C). \quad (7.4)$$

Then,

$$\begin{aligned} |\mu(C) - \gamma(C)| &= |(\mu(C) - \mu_n(C)) - (\gamma(C) - \gamma_n(C))| && \text{by Equation 7.4} \\ &\leq |(\mu(C) - \mu_n(C))| + |(\gamma(C) - \gamma_n(C))| \\ &< \varepsilon_\mu + \varepsilon_\gamma = k && \text{by Equation 7.3.} \end{aligned}$$

This obviously contradicts our initial assumption that  $|\mu(C) - \gamma(C)| = k$ .  $\square$

The next two lemmas state that, under very mild assumptions, a given sequence  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  of distribution can approximate the transitions of its limiting distribution  $\gamma$  and vice versa. In a sense, the limiting distribution of a sequence can perform a transition that is the limit of the transitions of the members, or more precisely, of a subsequence thereof.

**Lemma 7.5** (*Approximative Behaviour I*). Let  $\mathcal{A} = (S, \bar{s}, Act, \dashv\!\!\rightarrow, \dashv\!\!\rightarrow)$  be a weakly image-finite MA. Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be sequences of distributions over  $S$ . Let furthermore

- $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  for some distribution  $\gamma$ , and
- $\gamma_i \xRightarrow{a}_c \gamma'_i$ .

Then, there exists a distribution  $\gamma'$  such that

- $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ , and
- $\gamma \xRightarrow{a}_c \gamma'$ .

In addition, let  $\mathcal{P}$  be an arbitrary predicate over states that is lifted to distributions by letting  $\mathcal{P}(\gamma)$  hold if and only if  $\mathcal{P}(s)$  holds for every  $s \in \text{Supp}(\gamma)$ . Then, if each  $\gamma'_j$  satisfies  $\mathcal{P}(\gamma'_j)$  for  $j \in J$ , then also  $\mathcal{P}(\gamma')$ .

The proof of this lemma can be found in Appendix C.1.

**Lemma 7.6** (*Approximative Behaviour II*). Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  be a Markov automaton. Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be sequences of distribution over  $S$ . Let furthermore

- (i)  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  for some distribution  $\gamma$ , and
- (i)  $\gamma_i \xRightarrow{a}_c \xi_i$  for some  $\xi_i$  in a finitary weak transition, and
- (i)  $\gamma \xRightarrow{a}_c \gamma'$ .

Then, there exist distributions  $\gamma'_i$  such that

- $\gamma_i \xRightarrow{a}_c \gamma'_i$ , and
- $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ .

In addition: if  $\gamma \xRightarrow{a}_c \gamma'$  is a finitary transition, also all  $\gamma_i \xRightarrow{a}_c \gamma'_i$  are finitary transitions.

The proof of this lemma can be found in Appendix C.2.

*Remark 7.1.* If  $a = \tau$ , precondition (ii) can be dropped, since every distributions  $\mu$  satisfies  $\mu \xRightarrow{a}_c \mu$ .

**Lemma 7.7** (*Limit Behaviour*). Given a compact MA  $\mathcal{A}$ , let  $\mathcal{R} \subseteq S \times S$  be an equivalence relation. Let  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  be a convergent sequence of distributions with limiting distribution  $\mu$ . Let  $\rho$  be some distribution and let  $\rho_i$  be a family of distributions such that  $\mu_i \xRightarrow{a}_c \rho_i$  and  $\rho \mathcal{L}(\mathcal{R}) \rho_i$  for all  $i \in \mathbb{N}$ . Then there must exist some  $\rho^\mu$  such that  $\mu \xRightarrow{a}_c \rho^\mu$  and  $\rho \mathcal{L}(\mathcal{R}) \rho^\mu$ .

*Proof.* By Lemma 7.5, we may conclude the existence of a subsequence of the  $\rho_i$  with index set  $J$  that converges against some distribution  $\rho^\mu$  with the desired property that  $\mu \xRightarrow{a}_c \rho^\mu$ . It remains to show that also  $\rho \mathcal{L}(\mathcal{R}) \rho^\mu$ , which follows immediately by Lemma 7.4.  $\square$

*Remark 7.2.* We restrict to *compact* Markov automata throughout the thesis without further explicit mentioning.

## 7.2. Parallel Composition and Abstraction

We will now define parallel composition of two Markov automata. It is the straightforward combination of the corresponding definitions for probabilistic automata and interactive Markov chains. We refer the reader to Section 4.1 and 5.1 for detailed explanations.

**Definition 7.9 (Parallel composition).** Given two MA  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  and  $\mathcal{A}' = (S', \bar{s}', Act, \multimap', \multimap')$ , their parallel composition  $\mathcal{A} \parallel_A \mathcal{A}'$  is defined as the automaton  $\mathcal{A}^* = (S^*, \bar{s}^*, Act, \multimap^*, \multimap^*)$  where

- $S^* = S \times S'$
- $\bar{s}^* = \bar{s} \parallel_A \bar{s}'$

- $\dashv\!\!\rightarrow^*$  is the least relation satisfying

$$s \parallel_A s' \dashv\!\!\rightarrow^* \mu \parallel_A \mu' \quad \text{if } a \in A \text{ and } s \dashv\!\!\rightarrow^a \mu \text{ and } s' \dashv\!\!\rightarrow^a \mu' \quad (\text{sync})$$

$$s \parallel_A s' \dashv\!\!\rightarrow^* \mu \parallel_A \delta(s') \quad \text{if } a \notin A \text{ and } s \dashv\!\!\rightarrow^a \mu \quad (\text{parL})$$

$$s \parallel_A s' \dashv\!\!\rightarrow^* \delta(s) \parallel_A \mu' \quad \text{if } a \notin A \text{ and } s' \dashv\!\!\rightarrow^a \mu'. \quad (\text{parR})$$

- $\dashv\!\!\rightarrow$  is the least multirelation satisfying

$$s \parallel_A s' \dashv\!\!\rightarrow^* t \parallel_A s' \quad \text{if } s \dashv\!\!\rightarrow t \quad (\text{parLM})$$

$$s \parallel_A s' \dashv\!\!\rightarrow^* s \parallel_A t' \quad \text{if } s' \dashv\!\!\rightarrow t' \quad (\text{parRM})$$

◁

**Definition 7.10 (Abstraction).** Given a MA  $\mathcal{A} = (S, \bar{s}, Act, \dashv\!\!\rightarrow, \dashv\!\!\rightarrow)$ , its abstraction  $\mathcal{A}|_A = (S', \bar{s}', Act, \dashv\!\!\rightarrow', \dashv\!\!\rightarrow')$  with respect to the set of actions  $A \subseteq Act \setminus \{\tau\}$  satisfies

- $S' = S$
- $\bar{s}' = \bar{s}$
- $\dashv\!\!\rightarrow$  is the least relation satisfying

$$s \dashv\!\!\rightarrow' \mu \quad \text{if } a \notin A \text{ and } s \dashv\!\!\rightarrow^a \mu$$

$$s \dashv\!\!\rightarrow' \mu \quad \text{if } a \in A \text{ and } s \dashv\!\!\rightarrow^a \mu$$

- $\dashv\!\!\rightarrow' = \dashv\!\!\rightarrow$

◁

## 7.3. Summary and Discussion

In this chapter, we have introduced Markov automata. They are a straightforward combination of probabilistic automata and interactive Markov chains. The fundamental operations parallel composition and abstraction carry over from the base models almost without adaptations.

We now want to review and extend the different kinds of models that Markov automata subsume besides interactive Markov chains and probabilistic automata. The results are all immediate consequences of the according subsumption properties for PA and IMC; still, this overview is of interest on its own as it allows to compare both probabilistic and stochastic model classes in one single uniform framework. In Chapter 8, we will extend this framework with a semantic perspective, when we present an inclusion hierarchy of the canonical notions of bisimilarity for the following models.



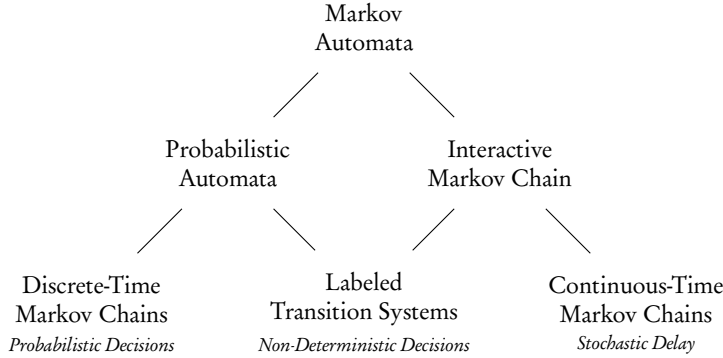


Figure 7.3.: Hierarchy of Sub-Models

In Figure 7.3, the different submodels of Markov automata are ordered in a lattice. Whenever one model lies above another in the lattice, then the former one is a generalization of the latter. We will make precise in the following, how the different models emerge from Markov automata.

We first introduce helpful terminology. Recall that a *Dirac* distribution assigns full probability to a single outcome. We say that  $\rightarrow$  is *Dirac* if the distributions occurring as third components of  $\rightarrow$  are all Dirac. It is *deterministic* if  $s \xrightarrow{a} \mu_1$  and  $s \xrightarrow{a} \mu_2$  implies that  $\mu_1 = \mu_2$ .

1. *Labelled Transition Systems*: If  $\rightarrow = \emptyset$  and  $\rightarrow$  is Dirac, we obtain labelled transition systems.
2. *Discrete-time Markov chains*: If  $\rightarrow = \emptyset$  and  $|Act| = 1$  and  $\rightarrow$  is deterministic, we obtain discrete-time Markov chains (DTMC). Markov chains are unlabelled structures. In this case one usually ignores the single action, and writes it as a triple  $(S, \mathbb{P}, s_0)$  where  $\mathbb{P}$  is called the probability matrix, and is given by  $\mathbb{P}(s, s') := \mu(s')$  provided  $s \xrightarrow{a} \mu$ .
3. *Continuous-time Markov chains*: If  $\rightarrow = \emptyset$  we obtain continuous-time Markov chains (CTMC). It is commonly represented as a triple  $(S, \mathbf{Q}, s_0)$  where  $\mathbf{Q}$  is called the infinitesimal generator matrix, and is given by  $\mathbf{Q}(s, s') := \text{rate}(s, s')$  provided  $s \neq s'$  and  $\mathbf{Q}(s, s) = \text{rate}(s, s) - \text{rate}(s)$ . The latter reflects that in the original mathematical formulation of CTMC it is impossible to make a difference between staying in state  $s$ , and jumping back to  $s$  from  $s$ . Again, CTMC are in general unlabelled, and transitions are labelled only with positive real valued rates. Hence, we require  $Act = \mathbb{R}_{>0}$ .
4. *Probabilistic Automata*: If  $\rightarrow = \emptyset$  we obtain probabilistic automata.
5. *Markov Decision Processes*: If  $\rightarrow = \emptyset$  and additionally  $\rightarrow$  is deterministic, we arrive at Markov decision processes (MDP). MDP can thus be considered as special cases of probabilistic automata.
6. *Interactive Markov chains*: If  $\rightarrow$  is Dirac, we get interactive Markov chains (IMC).

As this hierarchy shows, Markov automata can be applied in many contexts, where any of its simpler submodels has been used so far, but any of the additional modelling aspects Markov

automata provide have been missing for satisfactory results. At the same time, whenever it is possible to simplify a model in a way such that the restrictions of any of the above submodels are met, arbitrary analysis methods known for these submodels can be applied without adaptations. In the next chapter, we will introduce a notion of bisimilarity for Markov automaton, that allows to determine such simplification on a semantic level. In Chapter 12, we will show that in many real-world applications, the reduction will allow to reduce a special class of Markov automata to *CTMC*, even in cases where such reductions have not been possible with existing technology.

However, we want to note that while Markov automata indeed subsume all of these models structurally, one should not generalize properties of Markov automata to its submodels blindly. Most notably, parallel composition for Markov automata agrees with parallel composition for labelled transition systems, probabilistic automata and interactive Markov chains. But all other listed models, except *CTMC*, are not closed under parallel composition, as non-determinism is introduced by this operation.

## 8. Bisimulations

In this chapter, we investigate different notions of bisimulation, both strong and weak, for Markov automata. A special focus will be on the quest of finding a suitable notion of observational equivalence satisfying the three requirements we have proposed in Remark 1.1 on page 7. As PA and IMC are submodels of MA, our findings will have immediate consequences for observational theories of these models. In Section 4.2.3 we have reviewed the state of the art for PA and have concluded, that a satisfactory notion of observational equivalence is not yet available. For IMC, a widely accepted notion of observational equivalence appears to exist in the form of weak IMC bisimilarity.

**Outline and Contributions.** We begin our discussion with the introduction of strong bisimilarity for Markov Automata in Section 8.1. In Section 8.2, we approach the challenge of finding a natural notion of weak bisimulation for Markov automata from different perspectives and make multiple proposals for notions of weak bisimulation. As a core contribution, we introduce a novel notion of weak bisimulation, weak distribution bisimulation, that is weaker than all known notions of weak bisimilarity on probabilistic automata. In the process, we introduce a novel technique for defining bisimulations based on distributions over states instead of directly on states.

Section 8.3 is dedicated to the applicability of this technique to existing notions of bisimulation on PA and IMC, which traditionally are defined as a state-based relation. The question, whether also a state-based characterization of weak distribution bisimulation exist, will be investigated in Section 8.4.

In Section 8.5, we take a look back to the challenge of finding a natural notion of bisimulation, which we posed in Chapter 1, and we discuss, in how far weak distribution bisimilarity meets our criteria.

In Section 8.6, we review related work and compare weak distribution bisimulation to various bisimulation equivalences defined on probabilistic models. Section 8.7 concludes the chapter.

### 8.1. Strong Bisimilarity

We define strong bisimilarity for MA as the straightforward combination of the strong bisimilarity notions of PA and IMC.

**Definition 8.1 (Strong Bisimulation).** Let  $(S, \bar{s}, Act, \twoheadrightarrow)$  be a MA. An equivalence relation  $\mathcal{R}$  on  $S$  is a *strong bisimulation*, if whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} \mu$  for some  $a \in Act^X$  and  $\mu \in Dist(S)$  implies  $t \xrightarrow{a} \gamma$  for some  $\gamma' \in Dist(S)$  with  $\mu \mathcal{L}(\mathcal{R}) \gamma$ .  $\triangleleft$

We write  $s \sim_{\text{MA}} t$  if some strong bisimulation  $\mathcal{R}$  exists with  $s \mathcal{R} t$ . We say that two Markov automata  $\mathcal{A}$  and  $\mathcal{A}'$  are strong bisimilar if in their disjoint union the two initial states are strong bisimilar, i.e.  $\bar{s} \sim_{\text{MA}} \bar{s}'$ .

*Strong probabilistic bisimilarity* is defined akin to the respective definition for PA in Chapter 4, by replacing the single-transition relation  $\xrightarrow{a}$  by its combined variant  $\xrightarrow{a}_c$ . We do not state the full definition explicitly.

Both strong bisimilarities exhibit the usual properties of bisimilarities, such as being a congruence with respect to parallel composition and abstraction.

**Theorem 8.1.** (Probabilistic) strong bisimilarity

1. is the largest bisimulation relation,
2. is an equivalence relation, and
3. is a congruence with respect to parallel composition and abstraction.

**Theorem 8.2.** On the respective submodels, strong (probabilistic) bisimilarity for MA agrees with strong (probabilistic) bisimilarity for PA and IMC.

We do not provide explicit proofs for theorem. In most parts, it follows directly from the definitions, the exception being the fact that the definition of strong IMC bisimulation does not involve combined transitions, i.e. a probabilistic variant of strong bisimilarity does not exist. In fact, substituting combined transitions for non-combined transitions in the definition of strong bisimulation for IMC does not affect the resulting bisimilarity. We have established a corresponding claim for weak IMC bisimilarity, in Lemma 5.1 and the respective proof; an adaption to strong IMC bisimilarity is straightforward.

In summary, this section has shown that defining a notion of strong bisimilarity for Markov automata is straightforward. Its underlying bisimulation emerges as the combination of the respective bisimulations for PA and IMC. In fact, it is fully conservative on the respective submodels.

## 8.2. Weak Bisimilarities

A naïve combination of the respective standard weak bisimilarity notions for probabilistic automata and interactive Markov chain is straightforward to define, along the lines of the last section, where we have proceeded in such way for strong bisimilarity.

Arriving at a bisimilarity that satisfies the three requirements for a suitable notion of observational equivalence is a by far more challenging task. A naïve combination would inherit from weak probabilistic bisimilarity its failure to satisfy the third requirement, namely to abstract from internal details as much as possible. Concretely, as we have discussed in Section 4.2.3, this bisimilarity allows to observe the *probabilistic* branching structure of a process, which is something we consider unnatural. We will hence explore several alternative approaches in order to finally arrive at a satisfactory notion of *observational* bisimilarity for Markov automata.

Most of the following results are only valid for *compact* MA (cf. Section 7.1.1, Definition 7.8). If not stated otherwise, we assume in the following all Markov automata to be compact.

### 8.2.1. Naive Weak Bisimilarity

The straightforward combination of the standard weak bisimilarities for probabilistic automata, namely weak probabilistic bisimilarity, and interactive Markov chains, namely weak IMC bisimilarity, yields the following definition.

**Definition 8.2 (Naïve Weak Probabilistic Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\cdot}, \xrightarrow{\cdot}_c)$  be a MA. An equivalence relation  $\mathcal{R}$  on  $S$  is a *naïve weak probabilistic bisimulation*, if

1. whenever  $s \mathcal{R} t$  then  $s \xrightarrow{a} \mu$  for some  $a \in Act^X$  and  $\mu \in Dist(S)$  implies  $t \xrightarrow{a}_c \gamma$  for some  $\gamma \in Dist(S)$  with  $\mu \mathcal{L}(\mathcal{R}) \gamma$ , and
2. if  $s \downarrow$  then  $t \Longrightarrow t'$  for some  $t' \in S$  with  $t \downarrow$ .

We write  $s \approx_{MA} t$  if there exists a naïve weak probabilistic bisimulation  $\mathcal{R}$  with  $s \mathcal{R} t$ . We say that two Markov automata  $\mathcal{A}$  and  $\mathcal{A}'$  are naïve weak probabilistic bisimilar if in their disjoint union the two initial states are naïve weak probabilistic bisimilar, i.e.  $\bar{s} \approx_{MA} \bar{s}'$ .  $\triangleleft$

The first condition corresponds to the first condition for weak IMC bisimulation and the only condition for weak probabilistic bisimulation on PA. Note that we make use of the uniform presentation for stochastic timed transitions, and immediate probabilistic transitions, using the additional label  $\{\chi(r) \mid r \in \mathbb{R}_{\geq 0}\}$ . We refer the reader to Definition 4.13 and Definition 4.15 for details. The second condition is the stability condition that occurs in weak IMC bisimulation.

It is not hard to show that the bisimilarity resulting from this bisimulation satisfies the usual properties of bisimilarities, such as being an equivalence relation and a congruence with respect to all composition operators.<sup>1</sup> As in the last section, when we discussed strong bisimilarity, these results are inherited from the PA and IMC bisimilarities.

Our discussion in Section 4.2.3 has shown that weak probabilistic bisimilarity is too discriminating with respect to internal details of an automata, thus violating our third requirement for an observational equivalence. More specifically, the relation discerns between differences in the probabilistic branching structure, even in cases where no non-determinism is present and the probabilities of observable actions are the same in either case. This deficiency is inherited by naïve weak probabilistic bisimilarity. In the following sections, we propose weaker notions of bisimilarity for Markov automata and finally present a new candidate as an observational equivalence. It is worth noting that the results presented here carry over to PA almost without adaptations.

### 8.2.2. Relaxed Bisimulation

The three requirements of an observational equivalence (Remark 1.1) are essential for our quest for weaker notions of bisimilarity. We restate them incorporating the results of our discussion in Section 4.2.3.

1. An observational equivalence must be a congruence relation.
2. A reasonable observational equivalence for Markov automata (and also probabilistic automata) must preserve the non-deterministic branching structure of a system. This does

<sup>1</sup>Except for the well-known choice operator, which we have yet not defined so far.

not mean that every occurrence of structural non-determinism in the automaton should be observable. Rather, only then when it has consequences of what observations will be possible in the future. We recall that our characterization of the non-deterministic branching structure refers to the *observable* differences that decisions induce.

3. In those segments of an execution of a process where the non-deterministic branching structure is linear, i.e. no decisions are to be made, the probabilistic branching structure should remain completely unobservable. This means, only the ultimate distribution over consecutive observable events can be determined.

We attempt to guarantee the second requirement by exclusively basing our proposals on the bisimulation methodology. In this section, we will focus on the most challenging point, requirement 3. We begin by an attempt to make the requirement more formal.

**Example 8.1.** In Figure 8.1, the automaton on the left can be obtained from the automaton on the right by *fusing* the probabilities of the internal transitions into a single transition with the same eventual probabilities for the observable actions  $hh$ ,  $ht$ ,  $th$  and  $tt$ . By *fusing* we mean that we replace within the distribution  $\mu = \langle (\textcircled{t} : \frac{1}{2}), (\textcircled{h} : \frac{1}{2}) \rangle$ , which results from the transitions labelled by *throw*, each state by the distributions resulting from their successor transitions, as long as they are only labelled by  $\tau$ , and thus internal.

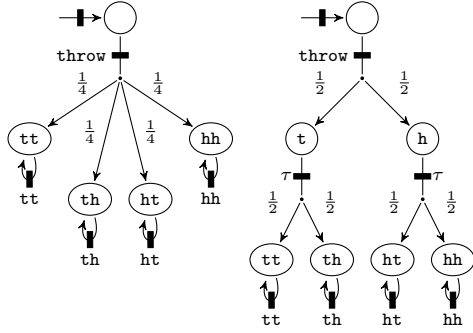
In the example, we replace  $\textcircled{t}$  by the distribution  $\langle (\textcircled{tt} : \frac{1}{2}), (\textcircled{th} : \frac{1}{2}) \rangle$ , and  $\textcircled{h}$  by  $\langle (\textcircled{ht} : \frac{1}{2}), (\textcircled{hh} : \frac{1}{2}) \rangle$ . With the respective probabilities that  $\textcircled{t}$  and  $\textcircled{h}$  have in  $\mu$ , namely  $\frac{1}{2}$  in each case, we obtain the distribution  $\langle (\textcircled{tt} : \frac{1}{4}), (\textcircled{th} : \frac{1}{4}), (\textcircled{ht} : \frac{1}{4}), (\textcircled{hh} : \frac{1}{4}) \rangle$ . This is precisely the distribution we reach in the automaton on the left immediately by the transition labelled with *throw*.  $\triangleleft$

In this example, we have used the notion of fusing of sequences of internal transitions to describe the process of obtaining information about the probabilistic behaviour, when we are not aware of any intermediate states. In terms of the testing scenario, this corresponds to our idea of allowing replication of states and using it for statistic sampling only in the first state in a sequence of internal transitions (cf. Section 4.2.3, page 53).

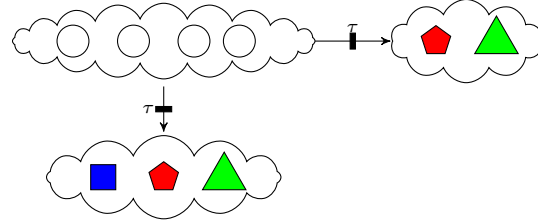
The idea of fusing distribution is not completely new. Generalized Stochastic Petri Nets (*GSPN*), a model that unifies exponentially distributed stochastic timed behaviour and internal probabilistic transitions, use precisely this technique to determine its underlying continuous time Markov chain, which acts as the semantics of a *GSPN*. Obviously, this method is only well defined if non-determinism is absent, as otherwise it is not clear in which way internal transitions shall be fused when a non-deterministic decision point is reached. For this reason, the semantics of *GSPN* is incomplete, in general, as non-determinism can be introduced in the model, but cannot be dealt with semantically. We will provide complete semantics for *GSPN* in Chapter 12 that becomes possible due to the results of this chapter.

**Example 8.2.** Consider the schematic automaton in in Figure 8.2, where cloudy objects denote distributions over states. Circles denote arbitrary states of the automaton and shaped objects denote behaviourally distinguishable states. We abstract from concrete probabilities. Transition arrows between distribution denote hypertransitions.

Clearly, the distributions with the coloured states on the right and on the bottom should not be considered behaviourally equivalent, as the distribution on the right completely lacks  $\blacksquare$ . If



**Figure 8.1.:** Two Automata with the Same Observable Outcomes



**Figure 8.2.:** Process Decides Internally Between Different Distributions

we now try to fuse distributions starting from the distribution on the top-left, we will have to decide which of the two internal transitions to take. No matter for which we decide, we will completely ignore the other distribution, which is, however, behaviourally completely different. Thus, we will change the intended semantics underlying the automaton drastically.  $\triangleleft$

Generalizing fusion to non-deterministic automata is by far not straightforward. The obvious solution is to restrict the fusion to those segments of internal transitions that are structurally free of non-deterministic choices. However, structural non-determinism may be completely irrelevant for the observable behaviour.

**Example 8.3.** In Figure 8.3, we see almost the same situation as before in Example 8.2. However, the distribution on the left contains  $\blacktriangle$  and, instead of  $\blacklozenge$ , another state that can only perform an internal transition to a distribution over  $\blacksquare$  and  $\blacklozenge$ . When we fuse these distributions, i.e. the distribution on the top-right with the one at bottom-right, we obtain a distribution over  $\blacksquare$ ,  $\blacklozenge$  and  $\blacktriangle$ . With suitable concrete probabilities, we then obtain a distribution that is identical to the distribution on the bottom-left.

Now, the top-left distribution has two outgoing internal transitions to two structurally different distributions. Thus, structurally, we see non-determinism here. However, if we apply the fusion process on the right-hand-side distribution that we have described above, we see that the non-determinism is not relevant behaviourally.

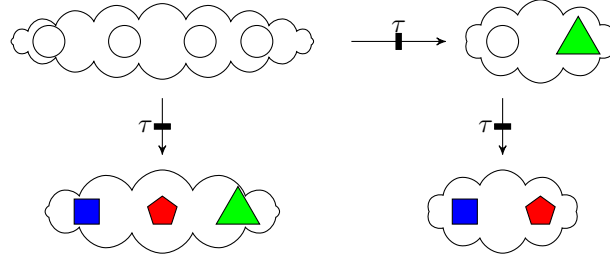
We want to remark that, in general, the situation can be even more complicated than in this example, especially in the presence of recursive behaviour.  $\triangleleft$

We see that a simple structural criterion when to stop the fusion process may not lead to the desired result, and end prematurely. Therefore, we need to investigate semantic criteria of stopping the fusion process.

## A Fusing Transition Relation

We will now investigate how the idea of fusing distributions can be incorporated into a definition of a bisimulation relation in the form of a new notion of weak transition relation.

A relation that describes the result of fusion needs to make statements about the probabilities with which observable behaviour will (finally) occur. To better illustrate the foundations of



**Figure 8.3.:** Structurally Different Distributions Become Behaviourally Identical

our approach, we will ignore internal behaviour for a moment. By the following formula, we express that a distribution  $\mu$  can execute action  $a$  with probability  $p$ , or more precisely, that the probability to be in a state that enables a transition with label  $a$  is  $p$ .

$$\exists \mu_1, \mu_2 : \mu = \mu_1 \oplus_p \mu_2 \wedge (\exists \mu' : \mu_1 \xrightarrow{a} \mu' \wedge \mu_2 \perp_a) \quad (8.1)$$

The first two operands of the logical conjunction states that the probability mass of states in  $\mu$  that enable a transition with action  $a$  is  $p = |\mu_1|$ . By the notation  $\mu_2 \perp_a$  in the last operand, we mean that no state in the support of  $\mu_2$  is able to execute a transition labelled by  $a$ . This operand ensures that the probability  $p$  is *precise* and not only a lower bound.

This formula does not yet allow to make statements about the probabilities of observable behaviour that will *finally* hold, after all internal transitions have been taken.

**Example 8.4.** Consider the uniform distribution  $\mu$  over the states  $t$  and  $h$  in the automaton on the right of Figure 8.1. The final probabilities  $tt$ ,  $th$ ,  $ht$  and  $hh$  to occur is  $\frac{1}{4}$  for each.  $\triangleleft$

To capture this in a formula, it does not suffice to simply replace the strong transition arrow  $\longrightarrow$  by a weak (combined) transition arrow  $\Longrightarrow_c$  in formula 8.1. The reason is that none of the states in the support of a distribution might be immediately able to execute any action besides  $\tau$ .

**Example 8.5 (cont'd.).** Neither  $s$  nor  $t$  in the support of  $\mu$  can execute a transition with label  $hh$ ,  $ht$ ,  $th$ , or  $tt$ , independent of whether we consider strong transitions or weak transitions. Yet, if we fuse the internal behaviour, each of these actions has a probability of  $\frac{1}{4}$ .  $\triangleleft$

Formally, the process of fusion is achieved by performing a weak transition before the distribution  $\mu$  is split. This allows us to replace states in the support of  $\mu$  that have only outgoing internal transitions by the successor distributions. This leads to the following proposal.

$$\exists \mu_1, \mu_2 : \mu \Longrightarrow_c \mu_1 \oplus_p \mu_2 \wedge (\exists \mu' : \mu_1 \xrightarrow{a} \mu' \wedge \mu_2 \perp_a)$$

Unfortunately, this formula does not yet fulfil our intents, as we will illustrate with an example.

**Example 8.6 (cont'd.).** Assume that we replace the action label  $tt$  by  $hh$  in the automaton on the right of Figure 8.1, such that the final probability of executing  $hh$  becomes  $\frac{1}{2}$ . With the last proposed formula, we could wrongly establish that the probability of executing  $hh$  is still  $\frac{1}{4}$ . By the formula, it suffices to find suitable distributions  $\mu_1$  and  $\mu_2$ . We choose  $\mu_2 = \langle (t : \frac{2}{3}), (ht, \frac{1}{3}) \rangle$  and  $\mu_1 = \delta(hh)$ , and  $p = \frac{1}{4}$ . The formula is then satisfied because from  $\mu$ , we can reach the distribution  $\langle (t : \frac{1}{2}), (ht : \frac{1}{4}), (hh : \frac{1}{4}) \rangle$  by a weak transition, where only states  $h$  executes, and



state  $t$  remains unchanged. We verify that this distribution is exactly  $\mu_1 \oplus_p \mu_2$  and that no state in the support of  $\mu_2$  is able to perform a transition labelled by  $\text{hh}$ , including state  $t$ .  $\triangleleft$

To correct this, let  $\mu \rightsquigarrow a$  denote that there exists a distribution  $\mu'$  and a state  $s \in \text{Supp}(\mu')$  such that  $\mu \Longrightarrow_c \mu'$  and  $s \xrightarrow{a}$ . Intuitively, this means that it is *possible* for  $\mu$  to finally execute  $a$  with non-zero probability, if necessary internal transitions are executed before. With this notation at hand, we propose the final generalization of Formula 8.1 to the weak transition setting.

$$\exists \mu_1, \mu_2 : \mu \Longrightarrow_c \mu_1 \oplus_p \mu_2 \wedge \exists \mu' : \mu_1 \xRightarrow{a}_c \mu' \wedge \mu_2 \not\rightsquigarrow a$$

This formalization also offers a reasonable generalization of the fusion process in a non-deterministic setting. As in the deterministic case, the probability of an action to be executed must be maximal, in the sense that it cannot be increased by executing internal transitions. However, for one action, there can be more than one probability, depending on how non-determinism is resolved. Resolution of non-determinism takes places in the weak internal transition before the splitting.

**Notation 8.1.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \multimap, \multimap)$ . Let  $\mu \in \text{Dist}(S)$ . We write

$$\mu \xRightarrow{a|p}_c \mu' \text{ if and only if } \exists \mu_1, \mu_2 : \mu \Longrightarrow_c \mu_1 \oplus_p \mu_2 \wedge \mu_1 \xRightarrow{a}_c \mu' \wedge \mu_2 \not\rightsquigarrow a$$

## A Distribution-Fusing Bisimulation

We will now use this notation to define a novel notion of bisimilarity based on the idea of fusion. As for strong bisimulation and the IMC bisimulations, we treat timed transitions and immediate probabilistic transition in a uniform way by virtue of the  $\chi(r)$  notation. Even though the discussions we have made so far were settled more in a PA setting, the arguments smoothly extend when stochastic timed behaviour is added.

**Definition 8.3 (Relaxed Bisimulation).** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \multimap, \multimap)$  be a MA. A symmetric relation  $\mathcal{R}$  over  $\text{Dist}(S)$  is an relaxed bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$ , for every  $a \in \text{Act}^X \setminus \{\tau\}$  and  $p \in [0, 1]$ :

- (a)  $\mu \Longrightarrow_c \mu'$  implies  $\gamma \Longrightarrow_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b)  $\mu \xRightarrow{a|p}_c \mu'$  implies  $\gamma \xRightarrow{a|p}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (c) if  $\mu \downarrow$  then  $\gamma \Longrightarrow \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  with  $\gamma' \downarrow$ .

$\triangleleft$

We write  $\mu \approx_* \gamma$  if some relaxed bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ . We say that two Markov automata  $\mathcal{A}$  and  $\mathcal{A}'$  are relaxed bisimilar if in their disjoint union the two initial states are relaxed bisimilar, i.e.  $\bar{s} \approx_* \bar{s}'$ .

Relaxed bisimulation is defined as a relation over distributions. In this respect, it differs from all other relations we have considered so far. Bisimulations based on distributions have not been considered much in the scientific community with few exceptions (cf. Section 8.6). To the best of our knowledge, weak notions of bisimilarity based on distributions do not exist. We want to remark that, however, in [FZ14], a notion of strong bisimulation has been defined that is very similar to relaxed bisimulation. We have overcome the technical problem, that not every

distribution over states may be able to perform a transition of a specific action with probability 1, by introducing subprobabilistic transitions  $\xrightarrow{a|p}_c$ . In [FZ14], instead, an additional dead state is introduced, which guarantees input-enabledness.

Defining bisimulation directly as relations over distributions is a natural approach in the context of probabilistic systems. In addition, it allows to abstain from the need to restrict bisimulation relations to have special properties such as being an equivalence relation. In this respect, they are akin to the mathematical elegance of the classic definitions of bisimulation.

The first condition of this definition is the typical bisimulation condition restricted to internal behaviour. Even though they are not observable directly, internal transitions may have consequences that are observable indirectly when they involve non-deterministic decisions. For example, if a system can internally decide to evolve in two different ways, each enabling different observable actions. An example is given in Figure 8.4.

Observable behaviour is treated in the second condition. There, we apply the new transition relation that specifies that some action can occur with maximal probability  $p$ . As we have discussed before, this is the core idea of the fusion process.

The reason why we use two separate conditions for  $\tau$  actions and observable actions is that merging them would counter the idea that internal behaviour is unobservable: if  $\tau$  was included also in the second condition, this would imply that the probability with which internal transitions occur, can be observed, as the bisimulation would then allow to compare two automata with respect to the maximal probabilities of  $\tau$ -occurrences. We, however, claim that it is technically possible to merge the two conditions.

The third condition states that whenever  $\mu$  is stable, then  $\gamma$  must be able to internally reach a distribution that is stable, too. This condition is owed to the maximum progress assumption we inherit from IMC. It is orthogonal to the discussion we made so far, and could in principle be dropped in a setting without stochastic time, for example if  $\mathcal{A}$  corresponds to a probabilistic automaton. We refer the reader to Chapter 5 for a detailed discussion of this condition.

**Example 8.7.** Recall that the two automata in Figure 8.1 are not bisimilar with respect to weak probabilistic bisimilarity. However, as the intuitively observable probabilities of action occurrence is the same, we have argued that they should be related by a suitable observational equivalence.

It is straightforward to define a relaxed bisimulation  $\mathcal{R}$ , that shows that the two automata are indeed relaxed bisimilar. We let

- $\delta(\bar{s}_1) \mathcal{R} \delta(\bar{s}_2)$ ,
- $\langle (\tau\tau : \frac{1}{4}), (\tau h : \frac{1}{4}), (ht : \frac{1}{4}), (hh : \frac{1}{4}) \rangle \mathcal{R} \langle (\tau : \frac{1}{2}), (h : \frac{1}{2}) \rangle$ , and
- $\langle (\tau\tau : \frac{1}{4}), (\tau h : \frac{1}{4}), (ht : \frac{1}{4}), (hh : \frac{1}{4}) \rangle \mathcal{R} \langle (\tau\tau : \frac{1}{4}), (\tau h : \frac{1}{4}), (ht : \frac{1}{4}), (hh : \frac{1}{4}) \rangle$ .

It is straightforward to verify that in each of the related distributions the probability to weakly execute a specific observable action is the same, and that the definition of relaxed bisimulation is satisfied.  $\triangleleft$

It is straightforward to show that relaxed bisimilarity is also the largest relaxed bisimilarity, as arbitrary unions of relaxed bisimulations are again relaxed bisimulations. For state-based probabilistic and/or stochastic bisimulations this proof is in general a little involved, as bisimulations are required to be equivalence relation, and the union of two equivalence relations is not an equivalence in general.

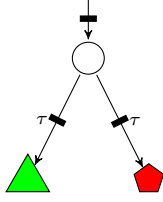
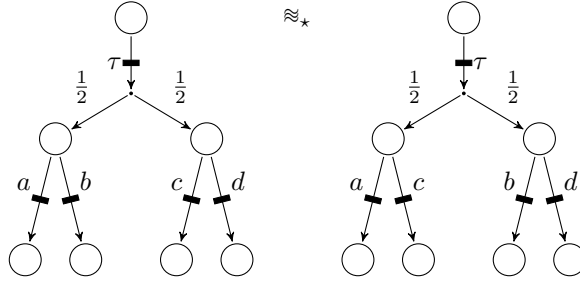
Figure 8.4.:  $\tau$  indirectly observable

Figure 8.5.: Branching Structure Not Preserved

**Theorem 8.3.**  $\approx_*$  is the largest relaxed bisimulation.

### No Candidate for Observational Equivalence

If we take a closer look on relaxed bisimilarity, it turns out that even though the relation is defined following the bisimulation methodology, it does not preserve the non-deterministic branching structure. Furthermore, it fails to be a congruence with respect to parallel composition.

Moreover, even the strong variant of relaxed bisimulation, which we have not stated, does not preserve the non-deterministic branching structure. This is very surprising at first sight, as it seems that exactly this property lies at the heart of every relation defined by the bisimulation methodology. We will illustrate the problem with an example in the following.

**Example 8.8.** Consider the two automata in Figure 8.5. As the probability to execute any of the observable actions  $a, b, c, d$  is  $\frac{1}{2}$ , the two automata are relaxed bisimilar. Clearly, however, the states with the outgoing observable actions do not have the same branching structure. For instance, the automaton on the left has a state with outgoing transition labelled by  $a$  and  $b$ , while no corresponding states exists in the automaton on the right. In fact, none of the states with outgoing observable actions have the same outgoing action labels.

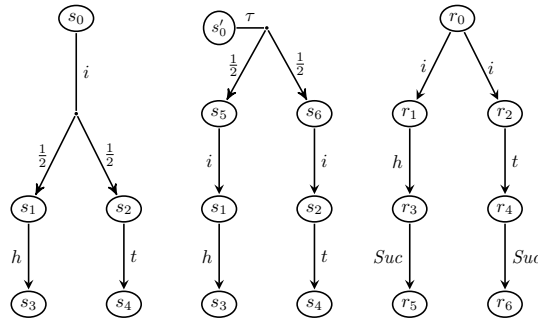
We remark that the difference in the non-deterministic branching structure of the automata in Example 8.8, also causes relaxed bisimilarity to fail at being a congruence with respect to parallel composition.

To continue the example, assume we combine the automata in parallel with another one that can non-deterministically choose between  $a$  and  $b$  (and has no other behaviour), and enforce synchronization upon these two actions. Then the probability that the communication succeeds when combined with the automaton on the left is  $\frac{1}{2}$ , because actions  $a$  and  $b$  are only enabled if the left probabilistic branch is taken in the immediate transition. When combined with the right automaton, the probability is 1. Because no matter how the probabilistic choice of the immediate transition is resolved, either a transition labelled by  $a$ , or a transition labelled by  $b$  is enabled, and thus, the communication can succeed.

If we further assume that the automaton with which we compose in parallel can execute *succ* after he has executed any of its transitions  $a$  or  $b$ , we can easily show that relaxed bisimilarity is even not a congruence with respect to parallel composition. Because then, the probability that *succ* occurs is  $\frac{1}{2}$  in one case, and 1 in the other, name precisely then, when the communication on either  $a$  or  $b$  was successful.  $\triangleleft$

This simple example has shown that relaxed bisimilarity neither preserves the branching structure of processes, nor is a congruence with respect to parallel composition.

Surprisingly, even fully deterministic processes exist that disprove relaxed bisimilarity to be a congruence. In [Eis+15], a distribution-based bisimulation has been developed that is similar to relaxed bisimulation in its intentions, but that succeeds in preserving the non-deterministic branching structure. Even for this relation, the following example is a counterexample to the congruence property. The interesting aspect of this example is that it involves no nondeterministic branching, but solely probabilistic branching.



**Figure 8.6.:**  $s_0$  and  $s'_0$  represent two different ways of tossing a coin and  $r_0$  denotes the guesser.

**Example 8.9 (Probabilistic Counterexample).** The two automata on the left denote the process of tossing a coin and then announcing the result to the environment.

Their difference is that the automata on the very left announces the result immediately after tossing (action  $i$ ), while the automaton in the middle proceeds in two steps. It first tosses the coin internally ( $\tau$ ), and then pretends to toss the coin ( $i$ ), even though it has already been tossed. Note that intuitively from the outside we cannot see a difference between the two automata. Both (pretend to) toss a coin, and with a uniform probability one of the two outcomes head ( $h$ ) or tail ( $t$ ) is announced. It is easy to see that the two automata are indeed relaxed bisimilar. Note, furthermore, that they are deterministic.

The process on the right represents a guesser, who tries to guess the result (before it is announced). The process of guessing happens in the very moment the (pretended) coin toss happens. If the guesser's guess was successful, he can synchronize with the announced result of head ( $h$ ) or tail ( $t$ ) of the tosser, and then execute the action  $Suc$  to signal his success. If he fails, the parallel composition will deadlock. Note that this assumes that synchronization is enforced on  $i$ ,  $h$  and  $t$ .

No communication takes place between the guesser and the tosser. It is thus quite surprising that if we execute the guesser in parallel with the guesser on the left, his maximum probability for guessing correctly is  $\frac{1}{2}$ . Yet, when he plays the game against the automaton in the middle, there is a way to resolve non-determinism in such a way that he guesses correctly with probability 1. So seemingly, he suddenly has a guaranteed winning strategy.

◁

The intuitive explanation for this situation is that the guesser, when put in parallel with  $s'_0$  can base its decision which transition to choose in state  $r_0$  on the state the tosser has reached by performing his internal probabilistic decision, namely either state  $s_5$  or  $s_6$ . This means the state

space that is generated by the parallel composition gives additional information to the guesser which he intuitively should not have. When combined with  $s_0$  no such states exist.

This is an example of a situation where two different modelling purposes of non-determinism interfere in an unexpected way, namely concurrency and implementation freedom. Intuitively, the non-determinism that originates from state  $r_0$  leaves the freedom to implement any guessing strategy. In the automaton that results from the parallel composition of the tosser and the guesser, there is no distinction between the cases where non-determinism results from parallel composition directly, or if it results from different implementation alternatives. So, one possible strategy to resolve non-determinism, often also called a *scheduler*, is one where the information in which state the tosser currently is, is used to devise a winning strategy to resolve the non-determinism of  $r_0$  that has resulted from implementation freedom. However, this strategy is unrealistic, as it is not locally restricted to information the guesser may have, but is based on global knowledge of the complete system.

The phenomenon of unrealistic schedulers is well known and has been studied nuancedly in the literature (see for example [GD07; De 99; GA12] and [AHJ01] in the context of trace distribution preorder. In fact, restricting ourselves to a suitable class of more realistic schedulers suffices to reestablish a congruence result—at least for the bisimulation of [Eis+15].

Unfortunately, restricting schedulers has several drawbacks. On the one hand, this result is not general any longer, and variations in the choice of the scheduler class used in a specific context may invalidate results established by using this bisimulation. On the second hand, realistic schedulers, as for example *distributed schedulers*, are known to quickly introduce undecidability problems in the theory [GD07]. We summarize our discussion in the following remark.

*Remark 8.1.*  $\approx_\star$  is not a congruence with respect to parallel composition, and it does not preserve the non-deterministic branching structure. The failure to be a congruence is separate from its failure to preserve the branching structure. Even for deterministic systems it is not a congruence with respect to parallel composition (cf. Example 8.9).

## Summary

Relaxed bisimilarity has turned out a very weak notion of bisimilarity that satisfies our demand of fusing distributions along sequences of internal transitions. In other words, it allows us to be ignorant of the probabilistic branching structure. It thus satisfies the last requirement of an observational equivalence demanded in Remark 1.1 and Section 8.2.2.

Unfortunately, this came accompanied with a violation of the other two requirements: being a congruence with respect to parallel composition and preserving the non-deterministic branching structure. The two problems have turned out to be independent in so far that even for *deterministic* systems the relation is not a congruence with respect to parallel composition (when combined with a non-deterministic process). While this can be remedied by restricting our theory to certain schedulers, we consider this solution unsatisfactory; mainly, because the bisimulation then loses its wide and general applicability as an observational equivalence. We therefore strive for a different notion of bisimilarity that is oblivious of the probabilistic branching structure up to the point where it interferes with essential properties of classic bisimilarity.

### 8.2.3. Weak Distribution Bisimilarity

This section introduces a novel notion of bisimulation, weak distribution bisimulation that is as oblivious of the probabilistic branching structure as possible, while still preserving the usual properties of bisimilarities. As we will show, weak distribution bisimilarity satisfies all properties that we demanded of a reasonable observational equivalence in Chapter 1. Technically, the definition of this bisimulation will be similar to relaxed bisimulation. It is also defined as a relation over distributions and uses a similar approach to express that a certain behaviour occurs with a certain probability. The difference is that it has stricter conditions on when a splitting of distributions may take place.

We also introduce minor variations of the relation, namely its stability insensitive counterpart as well as the induced simulation relation. They will prove to be a helpful tool in our later formal comparison with existing relations in Section 8.3. We conclude this section with a series of elementary properties of weak distribution bisimilarity, most notably that it is indeed the largest weak distribution bisimulation. More advanced properties will be developed in course of the succeeding sections.

**Definition 8.4 (Weak Distribution Bisimulation).** Let  $(S, \bar{s}, Act, \multimap, \multimap)$  be a MA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a weak distribution bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^X$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Dist(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma \xRightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c) if  $\mu \downarrow$  then  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Dist(S)$  with  $\gamma' \downarrow$ .

We write  $\mu \approx \gamma$  if some weak distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ . We say that two Markov automata  $\mathcal{A}$  and  $\mathcal{A}'$  are weak distribution bisimilar if in their disjoint union the two initial states are weak distribution bisimilar, i.e.  $\bar{s} \approx \bar{s}'$ .  $\triangleleft$

**Notation 8.2.** For two states  $s$  and  $t$ , we write  $s \approx_\delta t$  if  $\delta(s) \approx \delta(t)$ .

This reduction of  $\approx$  to a relation on states is especially helpful for comparisons with existing bisimilarities, which are usually defined as relations over states.

At this point, the reader might be interested in a comparison of the formal differences between relaxed bisimulation and weak distribution bisimulation. To eliminate inessential differences between the definitions, we provide a comparison of the *weak challenger* characterizations of both relations later in this chapter. Arriving at the weak challenger characterization of weak distribution bisimulation, however, needs several additional results before. The interested reader may proceed immediately to page 120 for the comparison.

**Variations** We note that from our definition of bisimilarity it is straightforward to derive the related *simulation* relation by dropping the symmetry condition. We will use this relation for a formal comparison with probabilistic forward simulation in Section 8.3.

**Definition 8.5 (Weak Distribution Simulation).** Let  $(S, \bar{s}, Act, \multimap, \multimap)$  be a MA. A relation  $\mathcal{R}$  over  $Dist(S)$  is a weak distribution simulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^X$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \xRightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c) if  $\mu \downarrow$  then  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  with  $\gamma' \downarrow$ .

We write  $\mu \preceq \gamma$  if some weak distribution simulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ . We say that two Markov automata  $\mathcal{A}$  and  $\mathcal{A}'$  are weak distribution similar if in their disjoint union the two initial states are weak distribution similar, i.e.  $\bar{s} \preceq \bar{s}'$ .  $\triangleleft$

**Stability Insensitive Variants** The maximal progress assumption does not play a role in the purely probabilistic setting, but is a crucial aspect of compositionality in the stochastic-timed setting (cf. Chapter 5). Thus, the standard notions of PA equivalence do not incorporate any condition similar to Condition (c) in the preceding definitions, while IMC equivalences share them.

To allow for better comparison with the process relations for PA, we introduce stability insensitive variants of the above definitions. They only differ in the removal of the third condition.

We note that removal is only semantically effective in the non-stochastic setting. In the stochastic setting, the notation  $\xrightarrow{\chi(r)}$  implicitly encodes stability. With this notation in mind, it is straightforward to see that Condition (a) already subsumes Condition (c). The reason why we state it in Definitions 8.4 and 8.5 at all, is merely to make the treatment of stability (i.e. maximum progress) more explicit to the reader.

Therefore, we define the stability insensitive relations only on PA, where we let  $a \in \text{Act}$  instead of  $a \in \text{Act}^\chi$ .

**Definition 8.6 (Weak Stability Insensitive Distribution Bisimulation).** Let  $(S, \bar{s}, \text{Act}, \xrightarrow{\cdot})$  be a PA. A symmetric relation  $\mathcal{R}$  over  $\text{Dist}(S)$  is a weak stability insensitive distribution bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in \text{Act}$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \xRightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ .

We write  $\mu \approx^\dagger \gamma$  if some weak stability insensitive distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ . We say that two PA  $\mathcal{A}$  and  $\mathcal{A}'$  are weak stability insensitive distribution bisimilar if in their disjoint union the two initial states are weak stability insensitive distribution bisimilar, i.e.  $\bar{s} \approx^\dagger \bar{s}'$ .  $\triangleleft$

**Definition 8.7 (Weak Stability Insensitive Distribution Simulation).** Let  $(S, \bar{s}, \text{Act}, \xrightarrow{\cdot})$  be a PA. A relation  $\mathcal{R}$  over  $\text{Dist}(S)$  is a weak stability insensitive distribution simulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in \text{Act}$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \xRightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ .

We write  $\mu \preceq^\downarrow \gamma$  if some weak stability insensitive distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ . We say that two PA  $\mathcal{A}$  and  $\mathcal{A}'$  are weak stability insensitive distribution similar if in their disjoint union the two initial states are weak stability insensitive distribution similar, i.e.  $\bar{s} \preceq^\downarrow \bar{s}'$ .  $\triangleleft$

Used in a different context and using different notation, this relation has actually been introduced in [Den+09] under the name *testing preorder*, however without the accompanying bisimulation relation.

## Elementary Properties

**Lemma 8.1 (Union).** Let  $\mathcal{R}$  and  $\mathcal{R}'$  be two weak distribution bisimulations. Then  $\mathcal{R} \cup \mathcal{R}'$  is a weak distribution bisimulation.

This lemma establish the existence of a unique largest weak distribution bisimulation, which then necessarily coincides with  $\approx$ .

**Corollary 8.1.**  $\approx$  is the largest weak distribution bisimulation.

**Theorem 8.4 (Equivalence Relation).**  $\approx$  is an equivalence relation.

*Proof.* Symmetry follows from the definition; reflexivity is straightforward. For transitivity, we will prove that the relation  $\mathcal{R} = \{(\mu, \xi) \mid \exists \gamma : \mu \approx \gamma \wedge \gamma \approx \xi\}$  is a weak distribution bisimulation. Consider an arbitrary pair  $(\mu, \xi) \in \mathcal{R}$ . Hence, by definition,  $\mu \approx \gamma \approx \xi$  for some distribution  $\gamma$ .

**Case (a):** Let  $\mu \xrightarrow{a} \mu'$ . As  $\mu \approx \gamma$  and  $\approx$  is a weak distribution bisimulation, there must exist  $\gamma'$  such that  $\gamma \xrightarrow{a} \gamma'$  and  $\mu' \approx \gamma'$ . Now similarly, as  $\gamma \approx \xi$ , there must exist  $\xi'$  such that  $\xi \xrightarrow{a} \xi'$ , and  $\gamma' \approx \xi'$ . It immediately follows that  $(\mu', \xi') \in \mathcal{R}$ .

**Case (b):** Let  $p$  be an arbitrary real number in  $[0, 1]$  and  $\mu_1$  and  $\mu_2$  two distributions such that  $\mu = \mu_1 \oplus_p \mu_2$ . As  $\mu \approx \gamma$  there exist  $\gamma_1$  and  $\gamma_2$  such that  $\gamma \xRightarrow{c} \gamma_1 \oplus \gamma_2$  and  $\mu_i \approx \gamma_i$  for  $i \in \{1, 2\}$ . Let  $\gamma' = \gamma_1 \oplus_p \gamma_2$ . As  $\gamma \xRightarrow{c} \gamma$  holds and since  $\gamma \approx \xi$ , we first use Condition (a) to derive that there exists  $\xi'$  such that  $\xi \xRightarrow{c} \xi'$  and  $\gamma' = \xi'$ . Second, using Condition (b), we can infer that there exist  $\xi_1$  and  $\xi_2$  such that  $\xi' \xRightarrow{c} \xi_1 \oplus_p \xi_2$  and  $\gamma_i \approx \xi_i$  for  $i \in \{1, 2\}$ . Then by Lemma 4.2, also  $\xi \xRightarrow{c} \xi_1 \oplus_p \xi_2$  directly. Again, this is enough to show that  $(\mu_i, \xi_i) \in \mathcal{R}$  for  $i \in \{1, 2\}$ .

**Case (c):** Assume  $\mu \downarrow$ . Since  $\mu \approx \gamma$ , there exists  $\gamma'$  such that  $\gamma \xRightarrow{c} \gamma'$  and  $\gamma' \downarrow$ . As  $\gamma \xRightarrow{c} \gamma'$  and  $\gamma \approx \xi$ , there exists  $\xi'$  such that  $\xi \xRightarrow{c} \xi'$  and  $\gamma' \approx \xi'$ . As  $\gamma' \downarrow$ , there must exist  $\xi''$  such that  $\xi' \xRightarrow{c} \xi''$  and  $\xi'' \downarrow$ . By Lemma 4.2, this satisfy to derive  $\xi \xRightarrow{c} \xi''$  and  $\xi'' \downarrow$ .

□

We finally state a lemma that will be helpful in many proofs. Intuitively, it means that distributions that can mutually reach each other by means of weak internal transitions, are all weak distribution bisimilar.

**Lemma 8.2.** If  $\mu \xRightarrow{c} \gamma \xRightarrow{c} \mu'$  and  $\mu' \approx \mu$  then also  $\mu \approx \gamma \approx \mu'$ .

The lemma is straightforward to prove. Analogue results are standard for most weak bisimilarities.



### Bisimulation-Up-To-Splitting

We now introduce a proof technique that reduces the number of pairs of bisimilar pairs of distributions we need to provide explicitly, when establishing that two distributions are bisimilar.

**Notation 8.3** (Composite Relation). Given a relation  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$ , we let  $\mathcal{R}^\oplus$  be the relation satisfying  $\mu \mathcal{R}^\oplus \gamma$  if and only if there exist a finite or countable infinite index set  $J$ , distributions  $\mu_i$  and  $\gamma_i$  as well as positive real numbers  $p_i \in \mathbb{R}_{>0}$  for each  $i \in J$ , satisfying

- $\sum_{i \in J} p_i = 1$ ,
- $\mu = \bigoplus_{i \in J} p_i \mu_i$  and  $\gamma = \bigoplus_{i \in J} p_i \gamma_i$ , and
- $\forall i \in J : \mu_i \mathcal{R} \gamma_i$ .

A composite relation relates distributions if they can be split into subdistributions in such a way that the subdistributions can be related by the original relation.

**Lemma 8.3.** If  $\mathcal{R}$  is a weak distribution bisimulation then also  $\mathcal{R}^\oplus$  is a weak distribution bisimulation.

*Proof.* Let in the following  $\mu = \bigoplus_{i \in J} p_i \mu_i$  and  $\gamma = \bigoplus_{i \in J} p_i \gamma_i$ .

We begin with Condition (a) of Definition 8.4. Assume that  $\mu \xrightarrow{a} \nu$ . By the definition of hyper-transition, we know that there is a defining  $a$ -transition for each state  $s \in \text{Supp}(\mu)$ . This immediately implies that, for each  $s_j^i \in \text{Supp}(\mu_i)$  there must exist a transition  $s_j^i \xrightarrow{a} \nu_j^i$  such that

$$\nu = \bigoplus_i p_i \bigoplus \mu_i(s_j^i) \nu_j^i.$$

Let  $\nu^i = \bigoplus \mu_i(s_j^i) \nu_j^i$ . By the definition of hyper-transition,  $\mu_i \xrightarrow{a} \nu^i$ . As  $\mu_i \mathcal{R} \gamma_i$  and  $\mathcal{R}$  is a weak distribution bisimulation, there must exist  $\xi^i$  with  $\gamma_i \xrightarrow{a}_c \xi^i$  and  $\nu^i \mathcal{R} \xi^i$ . By Lemma 4.3 we obtain that  $\gamma \xrightarrow{a}_c \bigoplus_{i \in J} p_i \xi^i$ . Immediately,  $\bigoplus_{i \in J} p_i \nu^i \mathcal{R}^\oplus \bigoplus_{i \in J} p_i \xi^i$  follows, as  $\nu^i \mathcal{R} \xi^i$ .

For Condition (b), let  $\mu^a \oplus_p \mu^b = \mu$  be a splitting of  $\mu$ . Then for each of the distributions  $\mu_i$ , there must exist a splitting  $\mu_i^a \oplus_{q_i} \mu_i^b = \mu_i$  satisfying

$$\bigoplus_{i \in J} \frac{p_i q_i}{p} \mu_i^a = \mu^a \text{ and } \bigoplus_{i \in J} \frac{(1-p_i)(1-q_i)}{1-p} \mu_i^b = \mu^b.$$

As  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in J$ , and  $\mathcal{R}$  is a weak distribution bisimulation there must exist weak combined hyper-transitions  $\gamma_i \xRightarrow{a}_c \gamma_i^a \oplus_{q_i} \gamma_i^b$  with  $\mu_i^a \mathcal{R} \gamma_i^a$  and  $\mu_i^b \mathcal{R} \gamma_i^b$  ( $\star$ ). Then, by Lemma 4.3,

$$\gamma \xRightarrow{a}_c \bigoplus_{i \in J} p_i (\gamma_i^a \oplus_{q_i} \gamma_i^b)$$

which, by standard algebraic arguments, is equivalent to

$$\underbrace{\bigoplus_{i \in J} \frac{p_i q_i}{p} \gamma_i^a}_{:= \gamma_a} \oplus_p \underbrace{\bigoplus_{i \in J} \frac{(1-p_i)(1-q_i)}{1-p} \gamma_i^b}_{:= \gamma_b}.$$

We immediately see

$$\begin{aligned}\mu_a &= \bigoplus_{i \in J} \frac{p_i q_i}{p} \mu_i^a \mathcal{R}^\oplus \bigoplus_{i \in J} \frac{p_i q_i}{p} \gamma_i^a = \gamma_a \\ \mu_b &= \bigoplus_{i \in J} \frac{(1-p_i)(1-q_i)}{1-p} \mu_i^b \mathcal{R}^\oplus \bigoplus_{i \in J} \frac{(1-p_i)(1-q_i)}{1-p} \gamma_i^b = \gamma_b,\end{aligned}$$

as  $\mu_i^a \mathcal{R} \gamma_i^a$  and  $\mu_i^b \mathcal{R} \gamma_i^b$  for each  $i \in J$  by  $\star$ .

Finally Condition (c) is similarly to, but simpler than Condition (a), using  $\mu \downarrow$  if and only if  $\mu_i \downarrow$  for each  $\mu_i$ .  $\square$

This lemma effectively means that we can use a composite relation over relatively basic distributions in order to establish bisimilarity of complex distributions. Phrased differently, this means that  $\approx$  is compositional with respect to  $\oplus$ .

**Corollary 8.2** (Linearity).

$$\approx = \approx^\oplus$$

*Proof.* This follows with the fact that  $\approx$  is the largest weak distribution bisimulation (Corollary 8.1) and Lemma 8.3.  $\square$

If we want to prove two distributions bisimilar, it suffices to show that they are contained in some bisimulation-up-to-splitting. The definition of bisimulation-up-to-splitting is identical to that of weak distribution bisimulation, except that it replaces  $\mathcal{R}$  in every condition of the form  $\mu \mathcal{R} \gamma$  by the composite relation, yielding  $\mu \mathcal{R}^\oplus \gamma$ .

**Definition 8.8.** Let  $(S, \bar{s}, Act, \multimap, \dashv)$  be a MA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a bisimulation-up-to-splitting, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^x$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xrightarrow{a}_c \gamma'$  for some  $\gamma' \in Dist(S)$  and  $\mu' \mathcal{R}^\oplus \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma \xrightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R}^\oplus \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c) if  $\mu \downarrow$  then  $\gamma \xRightarrow{} \gamma'$  for some  $\gamma' \in Dist(S)$  with  $\gamma' \downarrow$ .

$\triangleleft$

**Lemma 8.4.** Whenever  $\mathcal{R}$  is a bisimulation-up-to-splitting then  $\mathcal{R} \subseteq \approx$ .

*Proof.* We will establish that  $\mathcal{R}^\oplus$  is a weak distribution bisimulation. The result then follows from  $\mathcal{R} \subseteq \mathcal{R}^\oplus$ .

The proof that  $\mathcal{R}^\oplus$  is a weak distribution bisimulation follows the line of the proof of Lemma 8.3. There, we showed that  $\mathcal{R}^\oplus$  is a weak distribution bisimulation if  $\mathcal{R}$  is a weak distribution bisimulation. For this proof,  $\mathcal{R}$  is only a bisimulation-up-to-splitting, and not a bisimulation. The crucial insight is, however, that the definition of bisimulation-up-to-splitting and weak distribution bisimulation are almost identical, with the notable exception that  $\mathcal{R}$  is replaced by  $\mathcal{R}^\oplus$  in Conditions (a) and (b). Therefore, it suffices to adapt the proof by replacing every argument of the form

“ $\bigoplus_{i \in J} p_i \nu^i \mathcal{R}^\oplus \bigoplus_{i \in J} p_i \xi^i$  follows, as  $\nu^i \mathcal{R} \xi^i$ ”

by an argument of the form

“ $\bigoplus_{i \in J} p_i \nu^i \mathcal{R}^\oplus \bigoplus_{i \in J} p_i \xi^i$  follows, as  $\nu^i \mathcal{R}^\oplus \xi^i$ ”.

The only difference between the two lines is the addition of  $\oplus$  at the relation symbol  $\mathcal{R}$  on the right. Unfortunately, this implication does not hold immediately. It needs the following additional argument: Let  $\nu^i \mathcal{R}^\oplus \xi^i$  be justified by the index set  $K^i$ , weights  $q_k^i$  and distributions  $\nu_k^i$  and  $\xi_k^i$  with

- $\sum_{k \in K^i} q_k^i = 1$ ,
- $\nu^i = \bigoplus_{k \in K^i} q_k^i \nu_k^i$  and  $\xi^i = \bigoplus_{k \in K^i} q_k^i \xi_k^i$ , and
- $\forall k \in K^i : \nu_k^i \mathcal{R} \xi_k^i$ .

Then, we can justify  $\bigoplus_{i \in J} p_i \nu^i \mathcal{R}^\oplus \bigoplus_{i \in J} p_i \xi^i$  with help of the index set  $L := \{l_{i,k} \mid i \in J, k \in K^i\}$ , weight  $r_{i,k} := p_i q_k^i$  and the distributions  $\nu_{l_{i,k}} = \nu_k^i$  and  $\xi_{l_{i,k}} = \xi_k^i$  for  $i \in J$  and  $k \in K^i$ . We first note that

$$\begin{aligned}
 & \bigoplus_{i \in J} p_i \nu^i \\
 &= \bigoplus_{i \in J} p_i \bigoplus_{k \in K^i} q_k^i \nu_k^i \\
 &= \bigoplus_{i \in J} \bigoplus_{k \in K^i} p_i q_k^i \nu_k^i \\
 &= \bigoplus_{i \in J} \bigoplus_{k \in K^i} r_{i,k} \nu_k^i \\
 &= \bigoplus_{l_{i,k} \in L} r_{l_{i,k}} \nu_{l_{i,k}}
 \end{aligned}$$

and that, with a completely analogue argument,  $\bigoplus_{i \in J} p_i \xi^i = \bigoplus_{l_{i,k} \in L} r_{l_{i,k}} \xi_{l_{i,k}}$ . Then we note that by assumption for all  $i \in J$ :  $\forall k \in K^i : \nu_k^i \mathcal{R} \xi_k^i$  holds, and thus also  $\nu_{l_{i,k}} \mathcal{R} \xi_{l_{i,k}}$ . □

## Weak Challenger Characterization

**Definition 8.9 (Weak Challenger Characterization).** A symmetric relation  $\mathcal{R} \subseteq \text{Dist}(S) \times \text{Dist}(S)$  is called a weak challenger weak distribution bisimulation if whenever  $\mu \mathcal{R} \gamma$  then

- (a') if  $\mu \xRightarrow{a}_c \mu'$  then  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b') for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu \xRightarrow{p}_c \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \xRightarrow{p}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c') if  $\mu \xRightarrow{c}_c \mu'$  and  $\mu' \downarrow$  for some  $\mu'$  then  $\gamma \xRightarrow{c}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  with  $\gamma' \downarrow$ .

◁

**Lemma 8.5 (Characterizations).** On weakly image-finite MA, two distribution  $\mu$  and  $\gamma$  are weak distribution bisimilar *if and only if* they are related by some weak challenger weak distribution bisimulation.

The proof of this lemma can be found in Appendix D.1.

The weak challenger characterization drastically simplifies several proofs. We will use it in the following without further explicit mentioning. It is, however, important to note that we have only shown the validity of the lemma for weakly image-finite MA. Note that this restriction merely is a consequence of the proof we provided, and not necessarily needed in general. We are not aware of a counterexample that would definitely establish that the two characterizations disagree for general MA.

### Comparison with Relaxed Bisimilarity

We continue with a technical discussion of the formal differences in the definitions of the two distribution bisimulations. Using the characterization in Lemma 8.5 as the basis of our comparison, we see that the differences between relaxed bisimulation and weak distribution bisimulation mainly lie in the stricter splitting condition, and the uniform treatment of internal and observable action in the definition of weak distribution bisimulation. Condition (c) of relaxed bisimulation and Condition (c) of Definition 8.4 fully coincide. For internal transitions, Condition (a) of relaxed bisimulation agrees with Condition (a) of weak distribution bisimulation. For observable actions (including the stochastic timed action  $\chi(r)$ ), the special transition relation  $\xRightarrow{a|p}_c$  of relaxed bisimulation (Condition (b)), is reflected by both Conditions (a) and (b) of weak distribution bisimulation. For convenience, we recall the definition of  $\xRightarrow{a|p}_c$  as stated in Notation 8.1.

$$\mu \xRightarrow{a|p}_c \mu' \text{ if and only if } \underbrace{\exists \mu_1, \mu_2 : \mu \xRightarrow{a|p}_c \mu_1 \oplus_p \mu_2}_{\phi_1} \wedge \underbrace{\exists \mu' : \mu_1 \xRightarrow{a}_c \mu'}_{\phi_2} \wedge \underbrace{\mu_2 \not\xrightarrow{a}}_{\phi_3}.$$

Subformula  $\phi_2$  corresponds to Condition (a) of weak distribution bisimulation. It reflects the core idea of bisimulations, that a certain behaviour may occur from a distribution. The important differences are linked to the other two subformulas, which we will now discuss in detail.

The splitting condition contained in the notation  $\xRightarrow{a|p}_c$  (subformula  $\phi_1$ ) is stricter than the one of weak distribution bisimulation. There, it is demanded that for every splitting of one of two bisimilar distributions, there is a splitting of the other distribution, such that the distributions of the splitting are bisimilar again. For relaxed bisimulation nothing similar is demanded. The splitting is only used to allow to isolate states that can exhibit a certain observable action, which are gathered in distribution  $\mu_1$ , or  $\gamma_1$ , respectively. The other half of the splitting,  $\mu_2$  ( $\gamma_2$ ), is completely disregarded. Thus, its specific behaviour is lost for further considerations in the bisimulation. In contrast, as both  $\mu_1$  and  $\mu_2$  are demanded to be bisimilar again to  $\gamma_1$  and  $\gamma_2$ , respectively for weak distribution bisimulation, the behaviour of both splitting parts will be fully considered by subsequent checks of the bisimulation conditions. In summary, for both notions of bisimulation the splitting condition is necessary to observe that a certain behaviour occurs with a specific probability. The crucial difference is hereby that while weak distribution bisimu-

lation saves both parts of the splitting for further behavioural analysis, this aspect is completely discarded for relaxed bisimulation.

Another important difference is the complete lack of subformula  $\phi_3$  in weak distribution bisimulation. This lack allows to split distributions arbitrarily. The important consequence is that weak distribution bisimulations allow to consider also behaviour of arbitrary states in isolation, while relaxed bisimulation only allows to consider complete distributions. Thus, weak distribution bisimilarity combines the purely distribution-based view of relaxed bisimilarity with the purely state-based view of the classical state-based bisimilarities as for example weak probabilistic bisimilarity.

These in parts quite subtle differences between relaxed bisimulation and weak distribution bisimulation lead to the effect that weak distribution bisimilarity enjoys many favourable properties, which relaxed bisimilarity did not have, among them the congruence property for parallel composition. We will discuss these properties in the following sections.

## Semantic Insights

We present two properties of weak distribution bisimilarity that are needed for various proofs, and that are interesting in their own right.

**A Kind of Stuttering Condition** A weak transition is called a stuttering transition, if all states that are passed along the transition by internal transitions belong to the same equivalence class with respect to bisimilarity (cf. [GW96]).

In the definition of weak distribution bisimilarity, also a stuttering phenomenon occurs, even though with respect to distributions instead of states.

This can be best seen in Condition (b) of weak distribution bisimulation as expressed in the formulation of Lemma 8.5 (the weak challenger characterization).

**Lemma 8.6.** Let  $\mathcal{R}$  be a weak distribution bisimulation with  $\mu \mathcal{R} \gamma$ . Let  $\mu \Rightarrow_c \mu_1 \oplus_p \mu_2$  for arbitrary  $\mu_1$  and  $\mu_2$ . According to Condition (b) of Lemma 8.5, there exist  $\gamma_1$  and  $\gamma_2$  with  $\gamma \Rightarrow_c \gamma_1 \oplus \gamma_2$  and  $\mu_1 \mathcal{R} \gamma_1$  and  $\mu_2 \mathcal{R} \gamma_2$ .

In addition,

$$\mu \approx \mu_1 \oplus \mu_2 \approx \gamma_1 \oplus \gamma_2 \approx \gamma.$$

As typical for a stuttering condition, the start and goal distributions of the weak transitions  $\mu \Rightarrow_c \mu_1 \oplus_p \mu_2$  and  $\gamma \Rightarrow_c \gamma_1 \oplus \gamma_2$  are weak distribution bisimilar. This implies, that no change of equivalence class occurs in between. This is especially interesting, when we recall that Condition b essentially captures the idea of fusion of distribution. This lemma then illustrates that the fusion process never continues beyond a point where behaviour is changed.

Note that in the lemma we use an arbitrary weak distribution bisimulation  $\mathcal{R}$  to relate  $\mu$  and  $\gamma$ , but  $\approx$  in the conclusion of the lemma. In fact, this claim does not hold in general, if we replace  $\approx$  in the last equation by the weak distribution bisimulation  $\mathcal{R}$ .

*Proof.* As  $\mathcal{R}$  is a bisimulation, according to Condition (b) of Lemma 8.5,  $\mu_1 \mathcal{R} \gamma_1$  and  $\mu_2 \mathcal{R} \gamma_2$ . This immediately implies  $\mu_1 \approx \gamma_1$  and  $\mu_2 \approx \gamma_2$ . With Corollary 8.2  $\mu_1 \oplus \mu_2 \approx \gamma_1 \oplus \gamma_2$  follows. The remaining equalities follow by transitivity of  $\approx$ .  $\square$

Note, however, that weak distribution bisimulation is *not* akin to branching bisimulation ([GW96]), as no stuttering condition is applied on the simulating weak transitions.

**Converse Linearity** Corollary 8.2 states that two distributions are bisimilar if they can be split into families of distributions whose members are pairwise bisimilar. This is especially interesting as it allows to deduce bisimilarity of distribution from pairwise bisimilarity of the states in their support.

The following lemma allows to converse this statement in some sense. It states that, in principle, if  $\mu \approx \gamma$ , we can for an arbitrary splitting of a distribution  $\mu$  find a splitting of  $\gamma$ , such that the components of the splitting are pairwise bisimilar. However, such a splitting of  $\gamma$  cannot be found immediately, in general. Instead,  $\gamma$  must be allowed to perform a weak internal transition first.

**Lemma 8.7** (*Inverse Linearity*). Let  $\bigoplus_{i \in J} p_i \mu_i$  be a splitting of  $\mu$  with a possibly infinite index set  $J \subseteq \mathbb{N}$ . If  $\mu \approx \gamma$  then there exist  $\gamma_i$  for each  $i \in J$  such that

$$\gamma \Rightarrow_c \bigoplus_{i \in J} p_i \gamma_i \text{ and } \mu_i \approx \gamma_i.$$

The proof of this lemma can be found in Appendix D.2.

## Summary

In this section, we have introduced weak distribution bisimilarity with the goal in mind to achieve the relational coarseness of relaxed bisimilarity by splitting distributions, while at the same time maintaining the crucial properties of an observational equivalence (Remark 1.1).

So far, we have shown that bisimilarity can be characterized both with weak and strong challenger transitions, as it is the case for most notions of bisimilarity. We have also presented an up-to-proof technique that reduces the number of distribution pairs that one has to actually compare in a bisimulation proof. Finally, we presented two properties of weak distribution bisimilarity that provided first insights in its semantic structure.

However, we have not yet answered the essential question whether weak distribution bisimilarity satisfies the properties of an observational equivalence. We will answer this question in the following section.

### 8.2.4. Relaxed Bisimilarity Congruence

So far, establishing an observational equivalence relation on Markov automata has turned out to be a challenging task: While relaxed bisimilarity is arguably suited to formalize observations in the probabilistic and stochastic setting as a bisimilarity *being ignorant of a systems internal structure as much as possible*, it has failed to be a congruence relation for parallel composition and non-deterministic choice (cf. Remark 8.1). Thus, while clearly satisfying the last requirement of Remark 1.1, it fails its first.

In this section, we will prove that weak distribution bisimilarity, which is only slightly finer than relaxed bisimilarity, succeeds at being a congruence for all operators. Finally, we will establish that weak distribution bisimilarity is indeed the coarsest congruence relation included in relaxed bisimilarity.

With this result at hand, weak distribution bisimilarity becomes a most promising candidate for a canonical observational equivalence for Markov automata, fully satisfying at least two out of the three requirements of Remark 1.1. As we will argue in Section 8.5, in fact, it meets all three requirements.

### Congruence Properties of Weak Distribution Bisimilarity

This section summarizes the congruence result for abstraction and parallel composition. Among them, the congruence result for parallel composition is technically the most challenging to establish.

**Theorem 8.5.**  $\approx$  is a congruence with respect to abstraction:

$$\mathcal{A} \approx \mathcal{A}' \implies \mathcal{A}|_A \approx \mathcal{A}'|_A$$

The proofs follows well-known patterns (see for instance [Mil89b; Her02]).

**Theorem 8.6.**  $\approx$  is a congruence with respect to parallel composition:

$$\mathcal{A} \approx \mathcal{A}' \implies \mathcal{A} \parallel_A \mathcal{A}^* \approx \mathcal{A}' \parallel_A \mathcal{A}^*$$

The proof of this lemma can be found in Appendix D.3.

## Relaxed Bisimulation Congruence

Weak distribution bisimilarity has been designed to keep the idea behind relaxed bisimilarity, while being a congruence with parallel composition. In the following, we will first establish that  $\approx \subseteq \approx_*$ , showing that weak distribution bisimilarity is indeed a refinement of relaxed bisimilarity. As a second step, we will improve upon this result by showing that it is an optimal refinement in the sense that  $\approx$  is the coarsest congruence contained in relaxed bisimilarity.

**Refinement** We now discuss two examples that illustrate the differences between weak distribution bisimilarity and relaxed bisimilarity.

The following is an example of two processes where both bisimulations agree. The interesting aspect here is to see in what respect the used witness bisimulation relations differ for the respective notion of bisimilarity.

**Example 8.10.** Recall that the two automata from Figure 8.1 are relaxed bisimilar, as we have shown in Example 8.7. We show by providing a suitable weak distribution bisimulation  $\mathcal{R}$  that they are also weak distribution bisimilar.

- $\delta(\bar{s}_1) \mathcal{R} \delta(\bar{s}_2)$ ,
- $\langle (\mathbf{tt} : \frac{1}{4}), (\mathbf{th} : \frac{1}{4}), (\mathbf{ht} : \frac{1}{4}), (\mathbf{hh} : \frac{1}{4}) \rangle \mathcal{R} \langle (\mathbf{t} : \frac{1}{2}), (\mathbf{h} : \frac{1}{2}) \rangle$ , and
- $\langle (\mathbf{tt} : \frac{1}{4}), (\mathbf{th} : \frac{1}{4}), (\mathbf{ht} : \frac{1}{4}), (\mathbf{hh} : \frac{1}{4}) \rangle \mathcal{R} \langle (\mathbf{tt} : \frac{1}{4}), (\mathbf{th} : \frac{1}{4}), (\mathbf{ht} : \frac{1}{4}), (\mathbf{hh} : \frac{1}{4}) \rangle$ ,
- $\delta(\mathbf{t}) \mathcal{R} \langle (\mathbf{tt} : \frac{1}{2}), (\mathbf{th} : \frac{1}{2}) \rangle$ ,
- $\delta(\mathbf{h}) \mathcal{R} \langle (\mathbf{ht} : \frac{1}{2}), (\mathbf{hh} : \frac{1}{2}) \rangle$ .

The bisimulation  $\mathcal{R}$  differs from the one that we have used in Example 8.7 by the two last items, which are not necessary to establish the relaxed bisimulation of Example 8.7. The additional two pairs are needed in  $\mathcal{R}$  in order to satisfy the splitting condition, Condition (b) of Definition 8.4 for the pair

$\langle (\mathbf{tt} : \frac{1}{4}), (\mathbf{th} : \frac{1}{4}), (\mathbf{ht} : \frac{1}{4}), (\mathbf{hh} : \frac{1}{4}) \rangle \mathcal{R} \langle (\mathbf{t} : \frac{1}{2}), (\mathbf{h} : \frac{1}{2}) \rangle$ , when the distribution  $\langle (\mathbf{t} : \frac{1}{2}), (\mathbf{h} : \frac{1}{2}) \rangle$  is split into  $\delta(\mathbf{t}) \oplus_{\frac{1}{2}} \delta(\mathbf{h})$ . All ther conditions are straightforward to establish.  $\triangleleft$

We now illustrate a typical situation where relaxed bisimilarity and weak distribution bisimilarity do not agree.



**Example 8.11.** Let us now consider the two automata with initial states  $s_0$  and  $s'_0$  of Figure 8.6, which represent two ways to perform a fair coin toss. While they are relaxed bisimilar, they are not weak distribution bisimilar. The unique internal transition of  $s'_0$  leads to a uniform distribution over the states  $s_5$  and  $s_6$ , which we will call  $\mu$  in the following. In the intuition, the former of these states represents the situation where it is already determined that the coin toss (denoted by action  $i$ ), will yield the result *head* (action  $h$ ), while the latter represents the opposite situation where *tail* (action  $t$ ) is already determined. If now  $s_0$  and  $s'_0$  would actually be weak distribution bisimilar, this would mean that also  $s_0$  and  $\mu$  must be weak distribution bisimilar. Then,  $\mu$  can be split into  $\delta(s_5) \oplus \frac{1}{2} \delta(s_6)$ . Clearly, the two distributions of the splitting are not weak distribution bisimilar. Now, state  $s_0$  cannot perform any internal transitions. Thus, if the two automata were weak state bisimilar, we must be able to split  $\delta(s_0)$  into two *not* weak distribution bisimilar distributions, matching  $\delta(s_5)$  and  $\delta(s_6)$ , respectively. Obviously, this is impossible. Hence, the two different implementations of a tosser are not weak distribution bisimilar.  $\triangleleft$

The two situations in the last two examples seem very similar. However, while the first was weak distribution bisimilar, the second was not. What is the difference between these situations? In the first example, observable behaviour precedes the fusion of probabilities induced by internal transitions, while in the second example, the fusion process precedes the observable behaviour. Phrased differently, in the first situation, the observable transition originates from a single state and only the distribution over states that is reached by the transition is subsequently refined with internal transition. In the second situation, first internal transitions lead to a distribution over several (non-bisimilar) states, which together exhibit a certain behaviour, which they are not capable of performing in isolation. Notably, it is precisely the second situation that shows that relaxed bisimilarity is not a congruence with respect to parallel composition.

When we have discussed the idea of fusing distributions of sequences of internal transitions, we have stressed that the fusion process may not continue beyond a point where non-deterministic decisions between different behaviours need to be made. Recall that this may already be the case in the presence of a single outgoing internal transition from a state. The decision is then to stay at the state, or to continue along the transition. When the state has (observable) behaviour that the (distribution over) states reached by the internal transition cannot exhibit, taking the transition or not is indeed a non-deterministic decision between two behaviours. In Figure 8.7 we see a simple example of such situations. The initial state may perform a transition labelled by  $a$  to the state  $\blacktriangle$ . With an internal transition, it reaches a state that has both an  $a$  and a  $b$ -transition to  $\blacktriangle$ . The last state, that is reached by an internal transition from this state, has again only an  $a$ -transition to  $\blacktriangle$ . The internal transition from the first to the second state has no influence on the observable behaviour, as the behaviour of the first state is fully preserved by the second. In contrast, the third state exhibits less behaviour than the second, as it loses the ability to perform a  $b$ -transition. Thus, the internal transition from the second to the third state has consequences on the observable behaviour. Hence, the fusion process (which is trivial here, as all distributions are Dirac) will proceed from the first to the second state, but *must not* continue to the third. In fact, weak distribution bisimulations conform with this demand. Due to the simplicity of the example, it is possible to identify state and distribution of equivalent behaviour immediately from the structure. In turn, weak distribution bisimilarity is able to identify distributions of identical behaviour (in the sense that they are then bisimilar) during the fusion process. It can thus continue the fusion process even across structural non-determinism, as long as it is not behavioural.

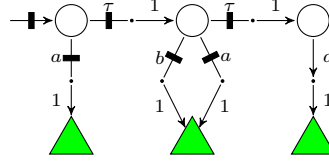


Figure 8.7.: Preserving Transitions

Now we formally state and prove the precise relationship between weak distribution bisimilarity and relaxed bisimilarity, which we have so far only discussed intuitively.

**Theorem 8.7.**  $\approx \subseteq \approx_\star$

*Proof.* The strictness follows immediately from Example 8.11. We now show that  $\approx$  is also an relaxed bisimulation. Both relations are symmetric by definition. The last condition of both bisimulations agrees by definition. The first condition of relaxed bisimulation agrees with the first condition of the weak challenger characterization of weak distribution bisimilarity. Note that we can freely switch between the standard characterization and the weak challenger characterization even though the respective characterization in general induce different bisimulation relations. Weak distribution bisimilarity is a bisimulation relation according to both characterizations.

Now consider the last condition, Condition (b). Assume  $\mu \xRightarrow{a} [p]\mu'$ . By definition of the transition relation, there exists a distribution  $\mu_1 \oplus_p \mu_2$  with

- (i)  $\mu \Longrightarrow \mu_1 \oplus \mu_2$ , and
- (ii)  $\mu_1 \xRightarrow{a}_c \mu'$ , and
- (iii)  $\mu_2 \not\rightarrow a$ .

First, (i) and  $\mu \approx \gamma$  imply that by Condition (a) of the weak challenger characterization of weak distribution bisimulation,  $\gamma \Longrightarrow_c \gamma_1 \oplus_p \gamma_2$  for some distributions  $\gamma_1$  and  $\gamma_2$ , and  $\mu_1 \approx \gamma_1$  and  $\mu_2 \approx \gamma_2$ . Second, (ii) and  $\mu_1 \approx \gamma_1$  imply with Condition (a) that  $\gamma_1 \xRightarrow{a} \gamma'$  with  $\mu' \approx \gamma'$ . Finally, we will establish that  $\mu_2 \not\rightarrow a$  and  $\mu_2 \approx \gamma_2$  imply that also  $\gamma_2 \not\rightarrow a$ . Assume the contrary, i.e. there exists  $\gamma'_1 \oplus_p \gamma'_2$  for some distributions  $\gamma'_1, \gamma'_2$  and real number  $p$  with  $\gamma_2 \Longrightarrow_c \gamma'_1 \oplus_p \gamma'_2$  and  $\gamma'_1 \xRightarrow{a}$ . Since  $\mu_2 \approx \gamma_2$ , thus also  $\mu_2 \Longrightarrow_c \mu'_1 \oplus_p \mu'_2$  for some distributions  $\mu'_1$  and  $\mu'_2$  with  $\mu'_1 \approx \gamma'_1$ . The last statement immediately implies that  $\mu'_1 \xRightarrow{a}$ , which contradicts the assumption that  $\mu_2 \not\rightarrow a$ .  $\square$

**Relaxed Bisimilarity Congruence** We will now show that weak distribution bisimilarity is indeed the coarsest congruence contained in relaxed bisimilarity with respect to parallel composition.

**Definition 8.10 (Relaxed Bisimulation Congruence).** Let  $\simeq_\star$  denote the coarsest relaxed bisimulation that is in addition a congruence with respect to parallel composition ( $\parallel_A$ ), i.e.  $\simeq_\star$  satisfies Definition 8.3 and whenever  $s \simeq_\star t$  then also  $s \parallel_A G \simeq_\star t \parallel_A G$  for  $A \subseteq \text{Act}$  and  $G \in \mathbb{MA}$ .  $\triangleleft$

**Theorem 8.8.**  $\approx = \simeq_\star$ .

The proof of this theorem can be found in Appendix D.4.

With this result, we see that weak distribution bisimilarity also satisfies the first requirement for being considered an observational equivalence for Markov automata (Remark 1.1). Relaxed bisimilarity has been designed with the third requirement in mind, i.e. being as ignorant of the internal structure of a probabilistic and stochastic system as possible. As weak distribution bisimilarity is the *coarsest* refinement of relaxed bisimilarity satisfying also the first requirement, it is valid to say that weak distribution bisimilarity also fully satisfies the third requirement.

Before we finally discuss in Section 8.5 in how far weak distribution bisimilarity also satisfies the second requirement, i.e. preservation of the non-deterministic branching structure, we will shed light on the novel approach of distribution-based bisimulations from two different perspectives.

## 8.3. Revisiting State-Based (Bi)Simulations

In this section, we show that the distribution-based approach is general enough to be also applied to the well-known (bi)similarities for PA and IMC. From this perspective, it will become clear that the distribution-based approach is in fact the natural approach to defining bisimulation on stochastic and probabilistic systems, and that the state-based approach is only a special case of it.

### 8.3.1. PA and IMC

To recover the standard state-based bisimulations for PA and IMC, it suffices to adapt the second condition of weak distribution bisimulation: We do no longer allow to match a distribution splitting by first performing a weak internal transition and then to split. Instead, we demand to split immediately.

In this section, we will show that this new bisimulation is a distribution-based characterization of the established weak bisimulation for PA and IMC. This shows that the technique of distribution splitting is powerful enough to express liftings of a state-based equivalence relation  $\mathcal{R}$  to distributions, which we have denoted by  $\mathcal{L}(\mathcal{R})$ . In other words, state-based probabilistic bisimulations can be easily reformulated in the distribution-based setting. In addition, our formulation relieves us from the burden to demand that a bisimulation must be an equivalence relation.<sup>2</sup>

Finally, we will show that probabilistic forward simulation coincides with weak distribution simulation.

**Definition 8.11 (Semi-Weak Distribution Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\cdot}, \dashv\!\!\rightarrow)$  be a MA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a weak distribution bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^X$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Subdist(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma = \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c) if  $\mu \downarrow$  then  $\gamma \Rightarrow \gamma'$  for some  $\gamma' \in Dist(S)$  with  $\gamma' \downarrow$ .

<sup>2</sup>Other characterizations of weak bisimulation for PA that do not demand probabilistic bisimulations to be equivalence relations are known. See, for instance, [Bai98].

We write  $\mu \overset{\circ}{\approx} \gamma$  if some semi-weak distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ .  $\triangleleft$

As before for weak distribution bisimilarity, we will also consider the simulation induced by this bisimilarity. It is obtained from Definition 8.11 by dropping the symmetry condition. We denote the resulting bisimilarity by  $\overset{\circ}{\preceq} \dagger$ .

The above definition is an exact copy of Definition 8.4 *except* for the omission of the weak transition in the second condition. While it is not obvious how to translate Definition 8.4 in a state-based setting, we try to make a first step towards this direction with this new definition. The crucial point is to understand the nature of distribution splitting. Splitting a distribution  $\mu$  means separating it into two arbitrary distribution  $\mu_1$  and  $\mu_2$  such that  $\mu = \mu_1 \oplus_p \mu_2$  for some  $p \in [0, 1]$ . While splitting  $\mu$  into itself  $\mu \oplus_p \mu$  is formally possible, it does not contribute to our understanding. The value of splitting lies in its ability to single out the individual states in the support of the distribution. It is easy to see that for every distribution  $\mu$  it holds that  $\mu = \delta(s) \oplus_{\mu(s)} (\mu \ominus s)$ . By repeatedly applying this splitting, it is clear that in this way, we can successively expose every single state in the support of  $\mu$ .

As a consequence, when two distributions are bisimilar by virtue of  $\mu \mathcal{R} \gamma$ , then splitting allows us to ensure that for every state  $s$  in  $\text{Supp}(\mu)$  there exists a subdistribution  $\gamma'$  of  $\gamma$  such that  $\delta(s) \mathcal{R} \gamma'$ . As  $\mathcal{R}$  is symmetric, we can use this argument now for  $\gamma'$  and infer that every state  $t$  in  $\text{Supp}(\gamma)$  must satisfy  $\delta(s) \mathcal{R} \delta(t)$ . We can formulate this in a notation familiar from weak probabilistic bisimulation and weak IMC bisimulation, if we let  $\mathcal{R}_\delta$  be defined as  $s \mathcal{R}_\delta t$  if and only if  $\delta(s) \mathcal{R} \delta(t)$ , i.e. the state-wise reduction of  $\mathcal{R}$ . Then, it must hold that  $\mu \mathcal{L}(\mathcal{R}_\delta) \gamma$ , which in other words means that to every equivalence class of states with respect to  $\mathcal{R}_\delta$ , the same probability mass must be assigned to by both  $\mu$  and  $\gamma$ . In this way, splitting of distributions induces a state-based relation from a distribution-based relation.

The intuitive discussion from the last paragraph immediately leads us to the theorem that the largest semi-weak distribution bisimulation,  $\overset{\circ}{\approx}$ , is basically a reformulation of weak probabilistic bisimulation on Markov automata. The only technical difference is that semi-weak distribution bisimulation is additionally aware of timed transitions and demands the addition stability condition also owed to the timed setting of MA. Abstracting from these circumstantial details, the two relations coincide. To state this formally, we first define semi-weak stability insensitive distribution bisimulation.

**Definition 8.12 (Semi-Weak Stability Insensitive Distribution Bisimulation).** Let  $(S, \bar{s}, \text{Act}, \rightarrow)$  be a PA. A symmetric relation  $\mathcal{R}$  over  $\text{Dist}(S)$  is a *semi-weak stability insensitive distribution bisimulation*, if for every  $a \in \text{Act}$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma = \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ .

We write  $\mu \overset{\circ}{\approx}^\dagger \gamma$  if some semi-weak stability insensitive distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ .  $\triangleleft$

Again, we denote the respective induced simulation relation by  $\overset{\circ}{\preceq} \dagger$ .

Both (bi)similarities can also be characterized by using a weak combined hyper-transitions in the first condition instead of a strong transition.

**Definition 8.13 (Weak Challenger Characterization of Semi-Weak Distribution Bisimulation).** Let  $(S, \bar{s}, Act, \multimap, \multimap)$  be a MA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a weak distribution bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^X$

- (a)  $\mu \xRightarrow{a}_c \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Subdist(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma = \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ , and
- (c) if  $\mu \xRightarrow{a}_c \mu'$  and  $\mu' \downarrow$  then  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Dist(S)$  with  $\gamma' \downarrow$ .

We write  $\mu \overset{\circ}{\approx} \gamma$  if some semi-weak distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ .  $\triangleleft$

**Definition 8.14 (Weak Challenger Characterization of Semi-Weak Stability Insensitive Distribution Bisimulation).** Let  $(S, \bar{s}, Act, \multimap, \multimap)$  be a PA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a *semi-weak stability insensitive distribution bisimulation*, if for every  $a \in Act$

- (a)  $\mu \xRightarrow{a}_c \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in Dist(S)$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma = \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ .

We write  $\mu \overset{\circ}{\approx}^\dagger \gamma$  if some semi-weak stability insensitive distribution bisimulation  $\mathcal{R}$  exists with  $\mu \mathcal{R} \gamma$ .  $\triangleleft$

**Lemma 8.8 (Characterizations II).** On weakly image-finite MA, two distribution  $\mu$  and  $\gamma$  are semi-weak distribution bisimilar *if and only if* they are related by some weak challenger semi-weak distribution bisimulation.

The proof follows along the line of the proof of Lemma 8.5, which is the according lemma for weak distribution bisimulation.

**Theorem 8.9.**  $\overset{\circ}{\approx}^\dagger$  is a semi-weak stability insensitive distribution bisimulation and an equivalence relation.

Let  $s$  and  $t$  be two states of some PA  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$ . Let  $s \overset{\circ}{\approx}^\dagger_\delta t$  if  $\delta(s) \overset{\circ}{\approx}^\dagger_\delta \delta(t)$

**Theorem 8.10.** On *finite* PA, it holds that

$$\overset{\circ}{\approx}^\dagger_\delta = \approx_{PA}$$

The proof of the theorem can be found in Appendix D.5. Our proof strategy relies on an inductive argument, which explains the restriction to finite PA. However, we claim that the theorem also holds for infinite MA.

With this theorem, we can understand the state-based weak probabilistic bisimilarity as special cases of a distribution-based bisimilarity. It is a simple exercise to repeat this reduction for interactive Markov chains and weak IMC bisimulation. Indeed, it can be shown that on IMC, semi-weak distribution bisimilarity and weak IMC bisimilarity coincide. As IMC incorporate stochastic time, the stability condition is needed for the two relations to match.

**Theorem 8.11.** On IMC, it holds that

$$\overset{\circ}{\approx}_{\delta} = \approx_{\text{MC}}$$

For Markov automata, we obtain such identities between  $\overset{\circ}{\approx}$  and  $\approx_{\text{MA}}$ . We recall that  $\approx_{\text{MA}}$  was the straightforward combination of  $\approx_{\text{PA}}$  and  $\approx_{\text{MC}}$ . So, this result is an immediate consequence of the last two theorems.

**Theorem 8.12.** On finite MA,

$$\overset{\circ}{\approx}_{\delta} = \approx_{\text{MA}}.$$

In this section, we have shown that the classical state-based definitions of PA and IMC bisimulations can be also obtained from the distribution-based definition by slightly adapting the splitting condition. We have thus identified a novel and at the same time natural way of defining probabilistic and stochastic bisimulation that is on the one hand able to comprise existing notions of bisimulation, and on the other hand gives us a powerful tool to define a novel and coarser notion of bisimulation with only little adaption.

### 8.3.2. Probabilistic Forward Simulation revisited

For probabilistic automata it has been less clear than for other models whether bisimulations are the semantically right way to define observational equivalences as discussed in Chapter 4. Trace distribution preorder has been introduced as the weakest reasonable semantics that treats probabilistic behaviour in a way that corresponds with an intuitive notion of what can be observed externally. It is coarser than bisimulation based relations for PA. And it is not only coarser with respect to the branching structure of systems, but also with respect to the probabilistic branching structure. With respect to the branching structure, this is arguably *not* a very desirable property, as it does not preserve deadlocking behaviour under parallel composition. Even more, trace distribution preorder is even not a congruence with respect to parallel composition. With respect to the *probabilistic* branching structure, however, the coarse nature of trace distribution preorder makes it an attractive candidate for the coarsest reasonable process relation. Probabilistic forward simulation is the coarsest congruence with parallel composition included in trace distribution. Still, being a simulation relation, it does not preserve deadlocking behaviour.

So we are in a situation where we have two competing proposals for an ideal observational equivalence: weak probabilistic bisimilarity on one side, and probabilistic forward simulation on the other side.

- Weak probabilistic bisimilarity preserves deadlocking behaviour, and comes with all the other benefits of bisimulations (cf. Chapter 1). However, its treatment of probabilistic behaviour is arguably too fine.
- Probabilistic forward simulation lacks the nice properties unique to bisimulations, but treats probabilistic branching in the coarsest reasonable way.

We thus strive for a relation that combines the best of these relations: A bisimulation that deals with probabilistic branching structure the way probabilistic forward simulation does. We claim that weak distribution bisimilarity is the relation we have searched for. It is a bisimulation and comes with all its benefits, and at the same time, it is ignorant of the probabilistic branching structure of a process as far as possible, without violating compositionality with respect to parallel

composition. In fact, it turns out that the simulation induced by weak distribution bisimilarity,  $\preceq$  coincides with probabilistic forward simulation in a divergence free setting. Using  $\preceq^\dagger$ , the stability-insensitive variant of  $\preceq$ , we are able to establish the coincidence of weak distribution bisimilarity and probabilistic forward simulation, up to the minor difference that the former relates distributions, while the latter relates a state to a distributions. Technically, we compensate for this minor difference by comparing  $\preceq^\dagger$  to  $\hat{\mathcal{L}}(\preceq_{fwd})$ , the lifting of  $\preceq_{fwd}$  to distributions (cf. Section 4.2.3).

**Theorem 8.13.**

$$\mu \preceq^\dagger \gamma \text{ if and only if } \exists \gamma' : \text{flatten}(\gamma') = \gamma \wedge \mu \hat{\mathcal{L}}(\preceq_{fwd}) \gamma'$$

The proof can be found in Appendix D.6.

**Conclusion** We have seen that the well-known probabilistic and stochastic bisimilarities can be defined directly on distributions with an only slightly stricter variant of the splitting condition of weak distribution bisimulation. One remarkable benefit of this formulation is that it no longer requires weak probabilistic bisimulation and weak IMC bisimulation to be equivalence relations. However, the maybe most compelling insight of recasting the well-known bisimulations in distribution-based characterizations is that, first, it is has now become apparent that weak distribution bisimulation is in principle a rather obvious generalization of these well-known bisimulation relations, and second, that recovering this obvious generalization in the state-based formulation seems virtually impossible.

This impression is intensified by our analogue considerations for probabilistic forward simulation. As it has turned out, this relation is indeed the state-based realization of weak distribution simulation. To the best of our knowledge, in the literature no approach of transforming this definition into a *bisimulation* variant has been known before.

## 8.4. A State-Based Characterization

Naively speaking, notions of bisimulation based on distributions are harder to establish than state-based notions for the simple reason that the former potentially become much larger than the latter. For state-based bisimulations, at most the set of reachable states must be considered and suitably related via bisimilar pairs. In contrast, for distribution-based notions, when a concrete bisimulation is to be constructed, potentially an uncountable number of distributions must be considered, even for finite states systems. This turns out to be especially problematic when it comes to algorithmic treatments of bisimilarity, for example, in the context of verification of systems and state space minimization by bisimulation quotienting. Standard partition refinement approaches usually applied in this context seem infeasible here, as even for finite state space, the problem space (i.e., the reachable distributions) is uncountable. We will in the following develop a state-based characterization of weak distribution bisimulation to overcome the algorithmic limitations and to shed more light on the exact semantic differences between this notion of bisimilarity and naïve weak probabilistic bisimulation, which stands in place for the well-known state-based bisimilarities of Chapter 4 and 5.

We begin with some intuitive discussion of the distinguishing feature of weak distribution bisimulation in terms of observability of states. As discussed in Chapter 1, each notion of bisimilarity is coupled with a specific intuitive notion of observability. Ideally, the bisimilarity is a

perfect formalization of observability. As we have discussed in detail in this chapter, weak distribution bisimilarity is as close to such an ideal bisimilarity for Markov automata as we can get, as any coarser variation cannot be a congruence with respect to parallel composition. The reason why it is coarser than the standard bisimilarities appears intrinsically linked to the fact that it is defined over distributions rather than states. While the ability of the observer to make an unbounded copy of a system *in every moment*, was naturally linked to the notion of states in transitions systems with probabilistic and stochastic behaviour, for probabilistic and stochastic models like Markov automata or probabilistic automata, it is naturally linked to the notion of distributions over states, and especially the ability to split them. Similar to stochastic processes, a probabilistic process is in a certain state only with a certain *probability* in every moment (although the presence of non-determinism complicates the situation here). To respect this, it seems unavoidable that bisimilarity is based on distributions.

In this section, we will discuss how weak distribution bisimulation can still be expressed as a state-based relation. We will see, however, that this characterization needs surprising extensions to the standard state-based approach of formulating bisimulations.


### 8.4.1. A First Intuitive Approach

As a first approach, we try to distinguish states according to their potential to contribute to the observable behaviour of a system. As it turns out, certain states determine the behaviour in a crucial way, while others can in principle be eliminated without changing the observable behaviour. The state-based characterization will use this observation to restrict the challenger to only propose transitions to distributions over states that matter for the observable behaviour.

#### A Behavioural Classification of States

We call a state  $s$  *behaviourally redundant*<sup>3</sup>, if it can reach via internal transitions a distribution  $\mu$  over states, which is observation equivalent to  $\delta(s)$ , i.e. no distinction can be made between them by virtue of any experiment the button-pushing machine allows. Furthermore, the experimenter cannot perceive that any change has happened at all, as the transition itself is unobservable, too. If  $s$  is *not* behavioural redundant, we also call it *behaviourally pivotal*. Phrased differently, a state  $s$  is behaviourally pivotal if and only if it either has no outgoing internal transitions at all or *every* internal transition allows for different observations than  $\delta(s)$  itself allows (note that  $\delta(s)$  always allows for at least the observation that any of its internally reachable successor states or distributions allow for).

**Example 8.12.** The most simple examples of behaviourally redundant states are states whose only outgoing transition is an internal transition. State  $s$  in Figure 8.8 is a typical example.

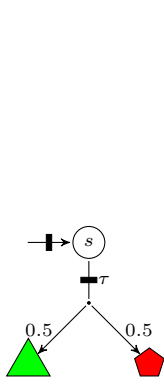
More intricate examples are states that do have outgoing external transitions, but which can be all simulated by a single successor distribution reachable via internal transitions. State  $t$  of Figure 8.9 is such an example. The  $a$ -transitions can be mimicked by the internally reachable distribution over the states  $u$  and  $v$ , since both have an outgoing  $a$  transition to the same state .

Note that a state is only behaviourally redundant, if there is an internal successor distribution that can compensate for *all* behaviours of the state at the same time. It does not suffice if every behaviour can be compensated for by *some* internal successor distribution.

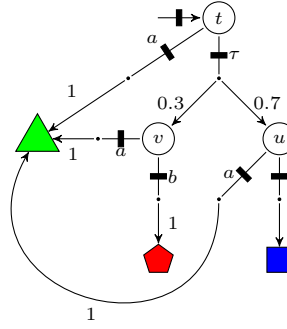
---

<sup>3</sup>Behaviourally redundant states are a generalization of the idea *vanishing* markings known from *GSPN*. We discuss this in Chapter 12.

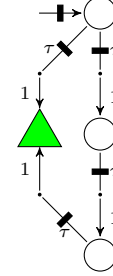




**Figure 8.8.:** Redundant State Without External Transitions



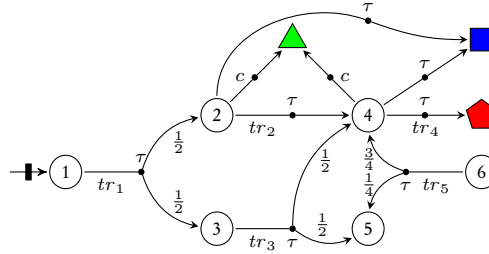
**Figure 8.9.:** Redundant State With External Transition



**Figure 8.10.:** Redundant States and Internal Non-Determinism

It is also instructive to note that the presence of multiple internal transitions does neither immediately imply that the state is pivotal nor that it is redundant. It only depends on whether one of the corresponding successor distributions is able to mimic all behaviours of the others. Figure 8.10 exemplifies this. All states except the green one are redundant.

◁






**Figure 8.11.:** What is Pivotal Here?




**Example 8.13 (Behaviourally Pivotal States).** Consider our running example, which we repeat in Figure 8.11. Assume that all non-round states of the PA induce pairwise distinct behaviour (for example each state can only perform a unique observable action). Then state ④ is behaviourally pivotal, since none of its internal successor distributions  $\delta(\blacksquare)$  and  $\delta(\blacktriangle)$  can behaviourally match the other, and thus cannot preserve the behaviour of  $s$ . Trivially, also ⑤ is behaviourally pivotal, since it has no successors. In contrast, all other states are *not* behaviourally pivotal, as for each of them the behaviour is fully preserved by one of its respective  $\tau$ -successor distributions. In particular, state ② is not behaviourally pivotal since its behaviour is fully preserved by  $\delta(\textcircled{4})$  via transition  $\textcircled{2} \xrightarrow{\tau} \delta(\textcircled{4})$ .

◁

## Characterization

We will now explain how the notion of behaviourally pivotal states is linked to a state-based characterization of weak distribution bisimulation.

Consider the probability distribution  $\mu = (\frac{3}{4}\delta(\textcircled{4})) \oplus (\frac{1}{4}\delta(\textcircled{5}))$  over behaviourally pivotal states and the distribution  $\gamma = (\frac{1}{2}\delta(\textcircled{2})) \oplus (\frac{1}{2}\delta(\textcircled{3}))$  over behaviourally redundant states. The observer will not be able to distinguish between the two distributions, as the respective probabilities for the possible observations are the same for both distributions: with probability  $\frac{3}{4}$  we will see either , , or  $c$  followed by , and with probability  $\frac{1}{4}$  we see no further behaviour, i.e. termination, which is represented by  $\textcircled{5}$ . In  $\mu$  the probabilities, that a certain observation can be made next, are *directly* determined by the probability  $\mu$  assigns to the states causing the respective behaviours, i.e.  $\frac{3}{4}$  for  $\textcircled{4}$  and  $\frac{1}{4}$  for  $\textcircled{5}$ . None of those states can perform any further internal transition without reducing the possible observations, as they are observationally pivotal. Similarly as for naïve weak probabilistic bisimulation, the individual states in the support of  $\mu$  immediately determines the observable behaviour of  $\mu$ . For  $\gamma$ , it is still the case that we can make the same observations, as  $\gamma \xrightarrow{\tau} \mu$  with help of the two transitions  $\textcircled{2} \xrightarrow{\tau} \delta(\textcircled{4})$  and  $\textcircled{3} \xrightarrow{\tau} (\frac{1}{2}\delta(\textcircled{4})) \oplus (\frac{1}{2}\delta(\textcircled{4}))$ .

However, considering state,  $\mu$  and  $\gamma$  are quite different. While the observable behaviour of  $\textcircled{2}$  happens to agree with the observable behaviour of  $\textcircled{4}$ , the behaviour of  $\textcircled{3}$  is unique. No other state allows the observation of , , and  $c$  followed by  with probability  $\frac{1}{2}$  and the observation of termination again with probability  $\frac{1}{2}$ . So considering state-wise behaviour,  $\mu$  and  $\gamma$  are clearly different. So, a direct state-based comparison of the distribution seems impossible.

The reason why this fails is that in a direct state-wise comparison of  $\mu$  and  $\gamma$  we attempt to compare behaviourally pivot states with behaviourally redundant states. As we are only interested in observations, we should only compare states in our bisimulation that are behaviourally pivotal. It is, however, not reasonable to demand that the bisimulation conditions only have to be met by pivotal states, as then likely every two behaviourally redundant states would be considered bisimilar, no matter how they evolve further. Thus, our approach will not be to ignore behaviourally redundant states, but instead to demand that every redundant state  $s$  must be able to reach a distribution  $\mu$  over pivotal states, that fully preserves the behaviour of  $s$ , and that, furthermore,  $\mu$  is able to satisfy any bisimulation condition in lieu of  $s$ .

### 8.4.2. Formalization

The most difficult aspect in the formalization of the above approach is stating that a behaviourally *redundant* state  $s$  should be replaced by distribution  $\mu$  over *pivotal* states, which fully preserve the behaviour of these states. This is not obvious to achieve, and needs the following three steps:

**Challenge 1** We need means to identify states as behaviourally redundant/pivotal

**Challenge 2** We need to be able to relate every behaviourally redundant state to a distribution over pivotal states that can enact in place of it.

**Challenge 3** The concepts of *observationally redundant* and *observationally pivotal* states essentially depend on our concept of observational equivalence, which is exactly what we intend to formally characterize in terms of bisimulation. At the same time, we have elicited that our formalization of bisimulation should be based on precisely these two concepts. Thus, we need to define two mutually dependent concepts.

We solve these challenges as follows:

1. To approach the first challenge, instead of identifying states as redundant/pivotal, we will use the analogue concept of *preserving transitions*. We call a internal transition  $(s, \tau, \mu) \in \xrightarrow{\tau}$  preserving (and only those) if  $\mu$  fully preserves the observable behaviour of  $s$ . Then, a state is behaviourally redundant if and only if it has an originating preserving transition.
2. Given a set of preserving transitions, we implement the second challenge by relating a redundant state  $s$  to a distribution  $\mu$ , which is reachable from  $s$  via a weak transition that only uses preserving transitions, and which is maximal with this property. Before we give further intuitive rationale for this approach, it is helpful to provide a formal definition of the last notion. Instead of relating a state  $s$  to a distribution  $\mu$ , we consider the more general case where we relate a distribution  $\xi$  over (redundant) states to a distribution  $\mu$  over pivotal states. We will denote this as  $\xi \xRightarrow{P} \mu$ .

**Definition 8.15 (Maximal Transition).** Given a MA  $\mathcal{A}$ , a set  $P$  of internal transitions and an equivalence relation  $\mathcal{R}$  on states, we write  $\mu \xRightarrow{P, \mathcal{R}} \mu^*$  to denote a weak combined transition from  $\mu$  to  $\mu^*$  with the following properties:

- a)  $\mu \xRightarrow{\tau \downarrow P}_c \mu^*$ .
- b) whenever  $\mu^* \xRightarrow{\tau \downarrow P}_c \mu'$ , then  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$ .

◁

In the following, the equivalence relation  $\mathcal{R}$  will always be clear from the context. We will thus usually omit it.

The first clause ensures that only preserving transitions may be used to reach  $\mu^*$  starting from  $\mu$ , while the second clause formalizes the maximality condition mentioned before, which intuitively expresses that the transition is of maximal length. In arbitrary infinite-state systems, it is not reasonable to assume that  $\mu^*$  always exists. We will effectively reduce our considerations in the following to MA that are weakly image-finite. For MA, that are both weakly image-finite and  $\tau$ -acyclic, the maximality condition can be simplified to the statement that  $\mu^*$  cannot perform any further weak allowed combined transitions, i.e.  $\mu^* \downarrow$ . In the presence of  $\tau$ -cycles, we apply our generalized statement. In non-probabilistic transition systems, our maximality condition is equivalent to stating that a state is maximal, if states reachable from it, must lie on a  $\tau$ -cycle. Our definition generalizes this to systems with probabilistic transitions and a maximal distribution  $\mu^*$ . The precise role the equivalence relation  $\mathcal{R}$  plays here will become clear in the following discussion. We already want to anticipate, however, that we will let  $\mathcal{R}$  be the bisimulation relation we actually want to define. Intuitively, we can then rephrase this clause to state that a distribution  $\mu^*$  is maximal if arbitrary weak allowed combined transitions always lead to an equivalent distribution with respect to (state-based) bisimulation. If we recall that any two states that lie on one  $\tau$ -cycle in non-probabilistic transition systems are also weakly LTS bisimilar, it becomes clear that  $\mu^*$  can be called maximal if it does not change up to a bisimulation  $\mathcal{R}$ .

3. The final challenge we need to solve that the preceding definitions depend on a relation  $\mathcal{R}$ , which intuitively should already be a bisimulation that captures our intuitive notion of

observability in a formal way. At the same time, as we have discussed, our formal definition of  $\mathcal{R}$  will need to rely on the preceding definitions. Thus, the preceding definitions and our definition of bisimulation will be mutually dependent and intertwined.

The following definitions will formalize what it means that a distribution and a state are behaviourally equivalent and when a transition is preserving. We parameterize them by an equivalence relation  $\mathcal{R}$ , which will turn out to be a bisimulation relation in the end.

**Definition 8.16.** A distribution  $\gamma$  is said to simulate  $s$  with respect to a set of internal transitions  $P \subseteq \neg\tau \rightarrow$  and an equivalence relation  $\mathcal{R}$ , if whenever  $s \xRightarrow{a}_c \mu'$  then there exist  $\mu, \xi'$  and  $\xi$  such that  $\mu' \xRightarrow{P} \mu$  and  $\gamma \xRightarrow{a}_c \xi'$  and  $\xi' \xRightarrow{P} \xi$ , and  $\mu \mathcal{L}(\mathcal{R}) \xi$ .  $\triangleleft$

This definition will be the essence of the following two definitions, whose purpose it is to intertwine the definition of a bisimulation relation  $\mathcal{R}$  and the definitions of the set of preserving transitions  $P$  finally in Definition 8.19, the definition of weak state bisimulation. A graphical representation of Definition 8.16, and how it is applied in the following two definitions, can be found in Figure 8.12.

**Definition 8.17.** Let  $\mathcal{A} = (S, \bar{s}, Act, \neg\tau \rightarrow, \neg\tau \rightarrow)$  be given. A set of internal transitions  $P \subseteq \neg\tau \rightarrow$  is called *preserving with respect to an equivalence relation*  $\mathcal{R} \subseteq S \times S$  if for every  $(s, \tau, \gamma) \in P$  the distribution  $\gamma$  simulates  $s$  with respect to  $P$  and  $\mathcal{R}$ .  $\triangleleft$

The following definition captures the typical scheme of bisimulation extended by the idea that the challenger may only propose a transition to a distribution  $\mu$ , if this distribution contains exclusively behaviourally pivotal states, i.e. it is maximal with respect to  $\xRightarrow{P}$ .

**Definition 8.18.** An equivalence relation  $\mathcal{R} \subseteq S \times S$  is called *preserving with respect to*  $P$  if

- a) whenever  $s \mathcal{R} t$ , then  $\delta(t)$  simulates  $s$  with respect to  $P$  and  $\mathcal{R}$
- b) if  $s \downarrow$  then  $t \xRightarrow{P} \gamma$  for some  $\gamma \in Dist(S)$  with  $\gamma \downarrow$ ;

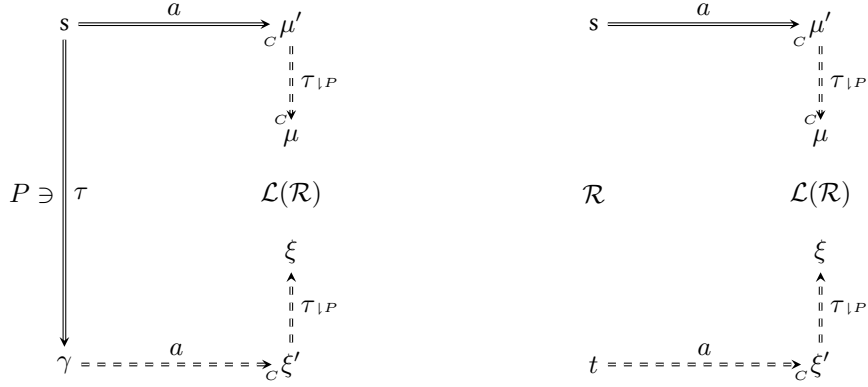
$\triangleleft$

Having solved the three challenges, we are now ready to combine them into a definition of weak state bisimulation. In fact, a weak distribution bisimulation will not solely be a binary relation of distributions, but such a binary relation combined with a set of internal transitions. The former represents the bisimulation relation in the usual sense, while the second is a set of preserving transitions. Note that since the two notions depend on each other, it is necessary to define them in one single instance.

**Definition 8.19 (Weak State Bisimulation).** Let  $\mathcal{A} = (S, \bar{s}, Act, \neg\tau \rightarrow, \neg\tau \rightarrow)$  be a Markov automaton. An *equivalence relation*  $\mathcal{R}$  on  $S$ , together with a set of internal transitions  $P$  is called a *weak state bisimulation pair*, if

- $\mathcal{R}$  is preserving with respect to  $P$ , and
- $P$  is preserving with respect to  $\mathcal{R}$ .

$\triangleleft$



**Figure 8.12.:** Schematic view on the essence of Definition 8.17 and 8.18. In both cases, we have a simulation condition according to Definition 8.16, but with slightly different premises.

We call the first component,  $\mathcal{R}$ , of a weak state bisimulation pair a *weak state bisimulation* and the elements of the second component,  $P$ , its *preserving transitions*. We write  $s \approx_s t$  if there exists a state-based weak distribution bisimulation relating  $s$  and  $t$  (and its respective set of respecting transitions). As usual, we lift this relation to a relation between automata by requiring that the initial states of two automata are related in their disjoint union.

**Example 8.14 (State-Based Weak Distribution Bisimulation).** Consider again the MA depicted in Figure 8.11 and suppose that states  $\color{red}\blacklozenge$ ,  $\color{green}\blacktriangle$ , and  $\color{blue}\blacklozenge$  are not weak bisimilar. The equivalence relation  $\mathcal{R}$  whose non-singleton classes are  $\{\textcircled{1}, \textcircled{6}\}$  and  $\{\textcircled{2}, \textcircled{4}\}$  is a state-based weak distribution bisimulation, given  $P = \{tr_1, tr_2, tr_3, tr_5\}$ .

Checking the step condition for the pair  $(\textcircled{2}, \textcircled{4})$  is trivial, so let us focus on the pair  $(\textcircled{1}, \textcircled{6})$ . Each weak combined transition  $\textcircled{6} \xRightarrow{\tau}_c \nu$  enabled by  $\textcircled{6}$ , can be matched by  $\textcircled{1}$  by reaching  $\nu_{tr_5}$  (via preserving transitions  $tr_1, tr_2$ , and  $tr_3$  chosen with probability 1) and then behaving as in  $\textcircled{6} \xRightarrow{\tau}_c \nu$ . If we stay in  $\textcircled{6}$  with non-zero probability, then we remain in  $\textcircled{1}$  with the same probability and the lifting condition is satisfied.

Now, consider the weak transition  $\textcircled{1} \xRightarrow{\tau}_c \nu$  enabled by  $\textcircled{1}$ , where  $\nu = \{(\textcircled{2} : \frac{1}{2}), (\textcircled{3} : \frac{1}{2})\}$  (this is actually the ordinary transition  $tr_1$ ).  $\textcircled{6}$  has no way to reach  $\nu$  so it needs help of  $\textcircled{1}$  to match such a transition:  $\textcircled{6}$  performs the transition  $\textcircled{6} \xRightarrow{\tau}_c \gamma$  where  $\gamma = \{(\textcircled{4} : \frac{3}{4}), (\textcircled{5} : \frac{1}{4})\}$ , i.e., it performs  $tr_5$ , while  $\nu$  reaches  $\gamma$  by the preserving weak hyper transition  $\nu \xRightarrow{\tau|P}_c \gamma$  by choosing with probability 1 preserving transitions  $tr_2$  from  $\textcircled{2}$  and  $tr_3$  from  $\textcircled{3}$  and then stopping.

The transition  $\textcircled{1} \xRightarrow{\tau}_c \nu$  is not the only weak combined transition enabled by  $\textcircled{1}$ . It enables, for instance, the weak combined transition  $\textcircled{1} \xRightarrow{\tau}_c \rho$  where  $\rho = \{(\color{blue}\blacklozenge : \frac{1}{2}), (\textcircled{3} : \frac{1}{2})\}$ .  $\textcircled{6}$  matches this transition by enabling  $\textcircled{6} \xRightarrow{\tau}_c \phi$  where  $\phi = \{(\color{blue}\blacklozenge : \frac{1}{2}), (\textcircled{4} : \frac{1}{4}), (\textcircled{5} : \frac{1}{4})\}$  that can be reached from  $\rho$  by the preserving weak hyper transition  $\rho \xRightarrow{\tau|P}_c \phi$  obtained by performing no transitions from  $\color{blue}\blacklozenge$  and choosing  $tr_3$  (that is preserving) with probability 1 and then stopping. There are several other transitions enabled by  $\textcircled{1}$  that can be matched in a similar way.  $\triangleleft$

Weak state bisimilarity is unusual for two reasons.

- It is defined in terms of two mutually dependent co-inductive definitions: one describing a binary relation between states, and one defining a subset of the transition relation  $\xrightarrow{\tau}$ .
- Different from all other notions of bisimilarity we have defined, it exclusively makes use of weak combined transitions  $\xRightarrow{a}_c$ , instead of the relation  $\xrightarrow{a}$ , where  $a \in Act^X$ . While one would expect that the first occurrence of such a transition can be safely replaced by  $\xrightarrow{a}$ , this is not the case. We will discuss this more deeply in the following referring to it as the *strong challenger* characterization.

We have discussed and motivated the first point already thoroughly. Yet, it leaves the open questions whether it is possible to give a purely state-based characterization of weak distribution bisimilarity, that does not deviate so much from the standard characterizations of bisimilarity.

The second point reveals an surprising deficiency of this characterization, that cannot be overcome easily. The obvious benefit of a strong challenger characterization is that it would drastically reduce the number of proof obligation we have, when we want to establish that a certain relation is a weak state bisimulation. Currently, theoretically an uncountable number of cases have to be checked with the original definition, that result from the uncountable number of different distributions reachable by a weak combined transition, in general. In practice, as we will see in Chapter 10, this number is still exponential in the number of transitions of a finite MA. With a strong transition characterization only the distributions directly reached by outgoing transitions of each state have to be considered.

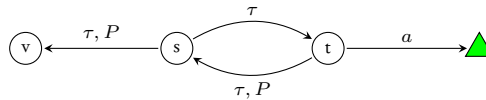
Unfortunately, the strong challenger characterization is incorrect. We will discuss the reasons in the following more thoroughly. Before, we need to introduce the concept of *maximal end components (mecs)* [Alf97]. Intuitively, they are the probabilistic equivalent to cycles in standard labelled transition systems:  $\tau$ -cycles where it is possible to return to each state of the cycle with probability 1.

**Definition 8.20 (Maximal End Components).** Given a MA  $\mathcal{A}$  with set of states  $S$ , a *maximal end component (mec)* is a maximal set  $C \subseteq S$  such that for each  $s, t \in C$ :  $s \xRightarrow{a}_c \delta(t)$  and  $t \xRightarrow{a}_c \delta(s)$ .  $\triangleleft$

The definition stems from [Alf97]. The set of all *mecs* is a disjoint partitioning of  $S$ . Thus, the relation  $=_{mec}$ , where  $s =_{mec} t$  if and only if  $s$  and  $t$  lie in the same *mec*, is an equivalence relation on states.

All states that lie in the same *mec* can mutually reach each other with  $\tau$  transitions with probability 1<sup>4</sup>. It is thus straightforward to show that such states are weak distribution bisimilar.

**Lemma 8.9.**  $s =_{mec} t$  implies  $s \approx_\delta t$ .



**Figure 8.13.:** A simple MA with *mecs*.

<sup>4</sup>Note that *mecs* are not necessarily bottom strongly connected components, as a *mec* may well be escaped by  $\tau$  transitions.

**Example 8.15 (Strong Challenger Characterisation is Broken in the Presence of Mecs.).**

Consider the automaton above. All transitions in this example are Dirac transitions. We label two transitions with  $\tau, P$  in order to express that they are elements of  $P$ , the set of supposedly preserving transitions considering the strong challenger characterization. Note that, however, the transition from  $\textcircled{s}$  to  $\textcircled{v}$  is not a preserving transition in the sense of the original definition with respect to any bisimulation relation  $\mathcal{R}$ , since  $\textcircled{s}$  can reach  $\blacktriangle$  with a weak  $a$  transition, whereas  $\textcircled{v}$  cannot perform an  $a$  transition at all. However, all conditions of the strong challenger characterization are satisfied. The only non-preserving strong transition  $\textcircled{s}$  can perform is the one to  $\textcircled{t}$ . Now it is enough that  $\textcircled{t}$  can reach  $\textcircled{v}$  via preserving transitions, by using  $t \xrightarrow{\tau} \delta(s)$  and  $s \xrightarrow{\tau} \delta(v)$ . For completeness, it is easy to check that the transition from  $\textcircled{t}$  to  $\textcircled{s}$  satisfies the conditions to be a preserving transition. With this result, it is straightforward to construct two bisimulations  $\mathcal{R}_1$  and  $\mathcal{R}_2$  (satisfying the strong challenger characterization), where  $\mathcal{R}_1$  is the reflexive, transitive and symmetric closure of the relation containing only the pair  $(\textcircled{v}, \textcircled{s})$  and  $\mathcal{R}_2$  accordingly containing  $(\textcircled{s}, \textcircled{t})$ . It is easy to check that for both  $\mathcal{R}_1$  and  $\mathcal{R}_2$  our choice of preserving transitions satisfies the strong challenger characterization. If this characterization now indeed was equivalent to  $\approx_\delta$ , the restriction of  $\approx$  to states, then also  $\textcircled{v} \approx_\delta \textcircled{s}$  and  $\textcircled{s} \approx_\delta \textcircled{t}$  would hold and thus, by transitivity, also  $\textcircled{v} \approx_\delta \textcircled{t}$ . But clearly, this cannot hold, as  $\textcircled{t}$  can perform an  $a$ -transition while  $\textcircled{v}$  cannot.  $\triangleleft$

A similar phenomenon is well-known from the theory of term rewriting systems [BN98]. Without going into details, we will sketch the important facts. For our discussion, it suffices to consider a rewriting system as ordinary graph or a (unlabelled) transition system. The transition relation of this graph then called a rewrite relation. For convenience, we denote it by  $\longrightarrow$ . The reflexive-transitive closure of  $\longrightarrow$ , which allows to do an arbitrary number of rewrite steps, including zero, is denoted by  $\Longrightarrow$ . With these relation, we can define *confluence* and *local confluence* as follows:

- A rewrite system is called *confluent*, if whenever  $s \Longrightarrow t_1$  and  $s \Longrightarrow t_2$  then there exists  $t$  such that  $t_1 \Longrightarrow t$  and  $t_2 \Longrightarrow t$ .
- A rewrite system is called *locally confluent*, if whenever  $s \longrightarrow t_1$  and  $s \longrightarrow t_2$  then there exists  $t$  such that  $t_1 \Longrightarrow t$  and  $t_2 \Longrightarrow t$ .

The only difference in the two definition is that the first two times the rewriting relation is used it appears in the form of its reflexive-transitive closure  $\Longrightarrow$ , while in the definition of local confluence it is directly the rewrite relation  $\longrightarrow$ .

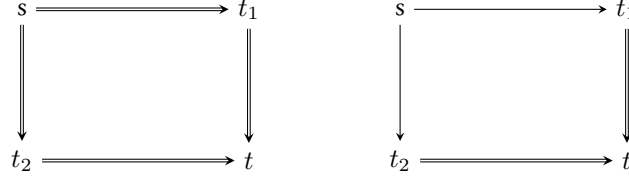
A schematic view on this relation can be found in Figure 8.14, with confluence on the left and local confluence on the right. If we compare this to Figure 8.12, and we replace  $\Longrightarrow$  by  $\Longrightarrow_c$  and the respective action labels, it becomes apparent that confluence strongly resembles the schematic structure of especially the definition of preserving relation in Figure 8.12, if we abstract from all further details. Furthermore, the schematic diagram of local confluence then corresponds to the strong challenger characterization of weak state bisimulation, which we obtain from Figure 8.12 when we replace the undashed double arrows by normal arrows, which then stand for the relation  $\xrightarrow{a}$ .

It becomes now evident that confluence is structurally similar to our original definition of weak state bisimulation, while local confluence is similar to its strong challenger characterization. The interesting insight that we now gain from rewriting theory [BN98] comes from the following theorem.

**Theorem.** *In an acyclic rewrite system, confluence and local confluence coincide.*

In cyclic systems this is not the case in general. In fact, the typical counterexample has precisely the shape of the MA of Figure 8.13, pruned from all labels.

The above theorem shows that, at least for systems without *mecs*, the strong challenger characterization coincides with our original characterization. However, in its full generality, this claim has resisted our proof attempts.



**Figure 8.14.:** Schematic view on Confluence and Local Confluence

We will now establish that  $\approx_s$  indeed coincides with  $\approx_\delta$ .

**Theorem 8.14.** On weakly image-finite MA,  $\approx_s = \approx_\delta$ .

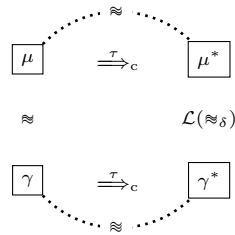
The proof of this theorem is rather involved. It can be found in Appendix D.7 in full length. The core insight of the proof, which is needed to show that  $t \approx_\delta t'$  implies  $t \approx_s t'$  is semantically interesting in its own respect.

**Theorem 8.15.** Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  be a compact Markov automaton, and  $\mu, \gamma \in \text{Dist}(S)$ . If  $\mu \approx \gamma$  then there exist  $\mu^*$  and  $\gamma^*$  such that  $\mu \Rightarrow_c \mu^*$ ,  $\gamma \Rightarrow_c \gamma^*$ ,

$$\mu \approx \mu^* \quad \mathcal{L}(\approx_\delta) \quad \gamma^* \approx \gamma.$$

The proof of this lemma can be found in Appendix D.8. It is by far not obvious and heavily depends on the restriction to compact Markov automata. We deepen this aspect in Section 8.4.3. For general systems, the distributions  $\mu^*$  and  $\gamma^*$  may not exist.

While weak distribution bisimilarity compares whole distributions of states, this theorem allows to resort to a state-based comparison, if only a “long enough” trace of internal transitions is followed from two bisimilar distributions.



**Figure 8.15.:** Graphical Representation of Theorem 8.15

Figure 8.15 illustrates the theorem graphically. We see that instead while  $\mu \approx \gamma$  does not imply  $\forall C \in S/\approx_\delta : \mu(C) = \gamma(C)$  is not true in general, we can reach with internal transitions



only two distributions  $\mu^*$  and  $\gamma^*$  satisfying  $\forall C \in S/\approx_\delta : \mu^*(C) = \gamma^*(C)$ . Furthermore, the distributions  $\mu^*$  and  $\gamma^*$  are still bisimilar to  $\mu$  and  $\gamma$ , respectively. Intuitively speaking, from an observers perspective, we cannot distinguish  $\mu$  and  $\mu^*$ , and  $\gamma$  and  $\gamma^*$ , respectively. Thus, without knowledge of the precise internal structure of the MA under consideration, distribution-based bisimilarity can always be explained from or reduced to a state-based rationale.

Even though we have not investigated it formally, we conjecture that for general (non weakly image-finite) MA,  $\approx_s \subseteq \approx_\delta$  holds. We leave the formal proof for future work.

### 8.4.3. Beyond Compactness

In Theorem 8.15, we have emphasized that we restrict our attention to compact Markov automata.<sup>5</sup> As the mentioned theorem is crucial to the proof of the main contribution of this section, Theorem 8.14, our discussion has immediate consequences also for this theorem.

In the following, we will present an example of a non-compact automaton *violating* Theorem 8.15. However, there also exist non-compact automata *satisfying* the theorem, as illustrated in Example 8.16. The theorem does thus not necessarily only hold for compact automata. We have not investigated more precise criteria when an automaton satisfies the theorem, though. As non-compact automata are theoretically and practically hard to handle, we do not consider this a challenge of primary importance, and leave it as further work.

**Example 8.16 (Not Compact Automaton Satisfying Theorem 8.15).** Consider the infinite MA that consists of an infinite sequence of states  $s_0, s_1, s_2, \dots$  with  $s_i \xrightarrow{T} \delta(s_{i+1})$  for all  $i \in \mathbb{N}$ . This automaton is obviously not compact. Still, it is obvious that  $\delta(s_i) \approx \delta(s_j)$  for every  $i, j \in \mathbb{N}$ .  $\triangleleft$

We will now construct two automata whose initial distributions are weak distribution bisimilar. However, no two states in the automata will be pairwise weak distribution bisimilar.

Throughout the following discussion, we will completely ignore the presence of the initial state in Figure 8.16b. We will speak of its two successor states  $\blacktriangle$  as the initial states, or initial distribution, of the automaton. The initial state is not necessary for our train of thought, and only there because by our definitions, MA do not allow to have an initial distribution.

The layouting of the automata in the figure clearly shows that the automata differ in the number of state contained in each level: while the first automaton has  $2^{2l}$  states in the  $l$ -th level, the second has  $2^{2l+1}$  states (for  $l \in \mathbb{N}$ ).

Each state of the automaton is constructed in a similar way. Besides the internal transitions connecting a state to its successors with a uniform distribution, we add observable transitions, labelled by  $a_1, a_2, \dots, b_1, b_2, \dots$ . Note that only the initial state of the automaton in Figure 8.16a has no outgoing observable transition. We let all observable transitions end in a single state  $\blacksquare$ , which is a terminal state. Introducing this additional state is not strictly necessary. We could as well have added self-loops to the states. It only simplifies the drawing. Note that each state will, in general, have more than one observable outgoing transition, labelled by different labels. In fact, each node will have the same observable transitions as *each* of its direct predecessor *and* one additional observable transition, that is labelled with a fresh action, i.e. that has not yet appeared at any transition of states in lower levels of the tree. However, it may recur at the transition of neighbouring states at the same level (and clearly at higher levels, as they copy the transition).

<sup>5</sup>see Section 7.1.1 for a definition of compactness

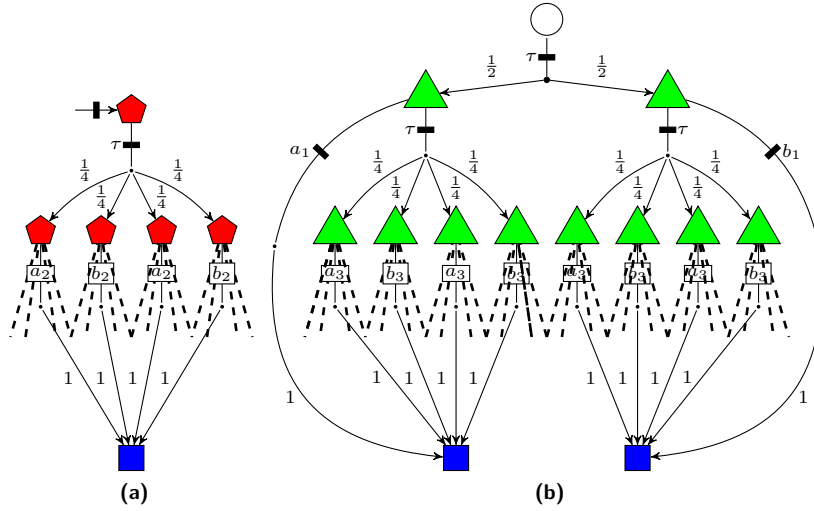


Figure 8.16.: Structure of the Two Automata

Note that in Figure 8.16 we only draw observable transition when they occur the first time. Please also note that the names of the actions (or more precisely their indices) follow a specific pattern that is not fully obvious.

We will now formally define the pattern of observable transitions and their action labels leaving each state. For the formal construction, it is helpful to first imagine a *binary* balanced tree starting with a single root node, from which we will derive the structure of the two automata in Figure 8.16. We depict this binary tree in Figure 8.17.

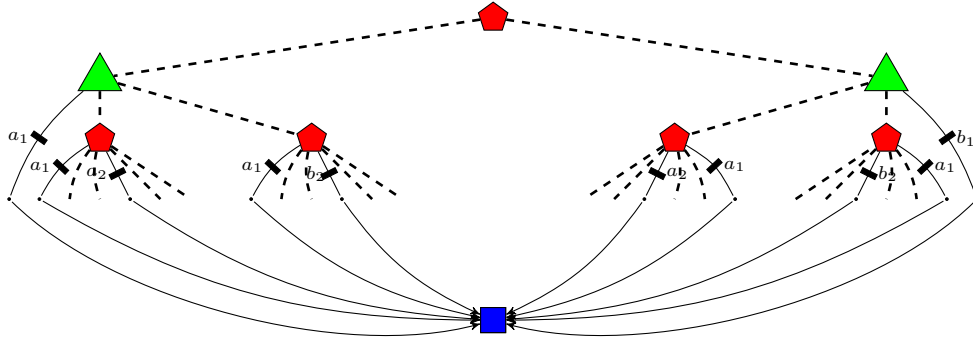


Figure 8.17.: Intermediate Construction Tree

We will then obtain the two quartary trees from this tree by identifying every even level with the first quartary tree, and every odd level with the second quartary tree. Again, we ignore the actual initial state of the automaton in Figure 8.16b. In the binary tree, we index the states names with sequences of digit from  $\{0, 1\}$ , which denote their address in the tree. The  $i$ -th digit

in the sequence tells us whether we need to go left (0) or right (1) when moving from level  $i$  to level  $i + 1$ , in order to reach the desired state. So, for example,  $\varepsilon$  denotes the root node, and the sequence 000 denotes the left most state in the third level, and 001 its neighbour in the same level. 00 denotes their common ancestor in level 2. We now add observable transitions to the states according to an inductively defined rule. Let in the following  $\sigma$  denote a sequence of the binary digits, and  $|\sigma|$  its length (with  $|\varepsilon| = 0$ ).

1. The state with index  $\varepsilon$  has no outgoing observable transition.
2. The state with index  $\sigma i$  has all observable transitions the state with index  $\sigma$  has, plus a transition labelled by  $a_{|\sigma i|}$  if  $i = 0$  and  $b_{|\sigma i|}$  if  $i = 1$ .

Let all transitions lead to the dedicated state  $\blacksquare$ . It is crucial that all external actions  $a_i$  and  $b_i$  are pairwise different. We see from the inductive definition that every state inherits all transitions from its predecessors in the binary tree, and adds one new transitions. Depending on whether it is a left or a right successor, this transition is labelled by an indexed variant of action  $a$  or  $b$ .

As indicated before, we now construct the two quartary trees out of this binary tree. Let the state with index  $\varepsilon$  be the initial state of the first quartary tree (Figure 8.16a, and let the two states with index 0 and 1 be initial states of the second tree (Figure 8.16b). Now, for each state (in both trees), we add the following transitions and the induced states: if the state  $s_\sigma$  (with index  $\sigma$ ) is in one of the quartary trees, then

$$s_\sigma \xrightarrow{\tau} \delta(s_{\sigma 00}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 01}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 10}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 11}).$$

The crucial property of these two automata is that their initial distributions are weak distribution bisimilar, while none of the trees' states are pairwise weak distribution bisimilar.

**Proposition 2.**

$$\delta(s_\varepsilon) \approx \delta(s_0) \oplus_{\frac{1}{2}} \delta(s_1),$$

but

$$s_\sigma \not\approx_\delta s_{\sigma'}$$

for  $\sigma \neq \sigma'$ .

*Proof.* The reason why no two states are bisimilar is that they always differ in the set of external actions that they can execute with probability greater zero. For notational convenience, we denote states by their indices in the following discussion. Let  $\sigma$  and  $\sigma'$  be two different states. Let  $\rho$  be their longest common prefix. First assume, that none of the two indices is a prefix of the other. Without loss of generality, let then  $\rho 0$  be a prefix of  $\sigma$ , and  $\rho 1$  a prefix of  $\sigma'$ . By construction, every state can exhibit every action one of its ancestors could. Also by construction then both  $\sigma$  and  $\rho 0$  can exhibit  $a_{|\rho|+1}$ , and both  $\sigma'$  and  $\rho 1$  can exhibit  $b_{|\rho|+1}$ . Again by construction, neither can  $\rho 0$  exhibit  $b_{|\rho|+1}$ , nor can  $\rho 1$  exhibit  $a_{|\rho|+1}$ . As with every increase of the index length, only actions with increased index are added to the outgoing transitions of states, this ensures that in turn neither  $\sigma$  can exhibit  $b_{|\rho|+1}$  nor  $\sigma'$  can exhibit  $a_{|\rho|+1}$  directly. The same holds — for the same reason — for every state that is reachable via weak transitions from  $\sigma$  and  $\sigma'$ , respectively. Now assume, that without loss of generality,  $\sigma$  is a prefix of  $\sigma'$ , and that  $\sigma 0$  is also a prefix of  $\sigma'$ . Otherwise, the following holds with completely symmetric arguments. The probability with which  $\sigma 0$  may execute  $a_{|\rho|+1}$  is 1, as it has directly an outgoing transition with this labelled. By

the same arguments as in the last case, this property is inherited also to  $\sigma'$  by construction of the automata, as this state is a successor of  $\sigma 0$ . In contrast,  $\sigma$  has no direct outgoing transition labelled by  $a_{|\sigma|+1}$ , as its index is shorter. Its only possibility to perform  $a_{|\sigma|+1}$  is to first perform weak transitions in order to reach a distribution over states that can directly perform  $a_{|\sigma|+1}$ . Its only internal transition, however, leads to  $\delta(\sigma 0) \oplus_{\frac{1}{2}} \delta(\sigma 1)$ . By construction, neither  $\sigma 1$  nor any of its successor states is capable of executing  $a_{|\sigma|+1}$ . Only  $\sigma 0$  (and its successors) are, as we have just argued. Thus, the maximum probability with which  $\sigma$  can exhibit  $a_{|\sigma|+1}$  is  $\frac{1}{2}$ . As the respective probability for  $\sigma'$  was 1, we conclude that  $\sigma$  and  $\sigma'$  cannot be weak distribution bisimilar.

Establish bisimilarity of  $\delta(s_\varepsilon) \approx \delta(s_0) \oplus_{\frac{1}{2}} \delta(s_1)$ , we again use an index  $\sigma$  as a short-hand for the state  $s_\sigma$  whenever convenient. Consider the relation  $\mathcal{R}$  that is the symmetric closure of the relation containing all pairs  $(\delta(\sigma), \delta(\sigma 0) \oplus_{\frac{1}{2}} \delta(\sigma 1))$ .

By construction,  $\delta(s_\varepsilon) \mathcal{R} \delta(s_0) \oplus_{\frac{1}{2}} \delta(s_1)$ . We now need to show that  $\mathcal{R}$  is a bisimulation-up-to-splitting. Consider an arbitrary pair  $(\delta(\sigma), \delta(\sigma 0) \oplus_{\frac{1}{2}} \delta(\sigma 1))$ . Assume that  $\sigma \xrightarrow{x_l} \delta(Q)$ . By  $Q$ , we denote the dedicated state to which we assumed that all external transitions lead to. By construction of our automata, it must be the case that  $l \leq |\sigma|$  and  $x \in \{a, b\}$ , depending on the concrete  $\sigma$ . By construction (of the binary tree), it immediately follows that also  $\delta(\sigma 0) \oplus_{\frac{1}{2}} \delta(\sigma 1) \xrightarrow{x_l} \delta(Q)$ , as transitions are inherited to all states that have indices with  $\sigma$  as a prefix. This covers all possible external transitions of  $\delta(\sigma)$ . The only internal transition of  $\delta(\sigma)$  leads to  $\delta(s_{\sigma 00}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 01}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 10}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 11})$ . As  $\delta(\sigma 0) \mathcal{R} \delta(s_{\sigma 00}) \oplus_{\frac{1}{2}} \delta(s_{\sigma 01})$  and  $\delta(\sigma 1) \mathcal{R} \delta(s_{\sigma 10}) \oplus_{\frac{1}{2}} \delta(s_{\sigma 11})$ ,  $\delta(\sigma) \oplus_{\frac{1}{2}} \delta(\sigma 1) \mathcal{R}^\oplus \delta(s_{\sigma 00}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 01}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 10}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 11})$  follows. This satisfies Condition (a) of bisimulation. As  $\delta(\sigma)$  is a Dirac distribution, Condition (b) is easy to satisfy. As all states, except  $Q$  are guaranteed to diverge, obviously also Condition (c) is satisfied.

Now for the symmetric case, for Condition (a), we first note that  $\delta(\sigma 0) \oplus_{\frac{1}{2}} \delta(\sigma 1) \xrightarrow{x_l} Q$  only if both  $\delta(\sigma 0) \xrightarrow{x_l} Q$  and  $\delta(\sigma 1) \xrightarrow{x_l} Q$ . By construction, this implies that  $l \leq |\sigma|$ , and thus clearly also  $\delta(\sigma) \xrightarrow{x_l} Q$ . The argument for the internal transitions follow similarly as in the last case when we additionally note that  $\delta(\sigma) \implies_c \delta(s_{\sigma 00}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 01}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 10}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 11})$  and  $\delta(s_{\sigma 00}) \oplus_{\frac{1}{2}} \delta(s_{\sigma 01}) \mathcal{R} \delta(s_{\sigma 0})$  and  $\delta(s_{\sigma 10}) \oplus_{\frac{1}{4}} \delta(s_{\sigma 11}) \mathcal{R} \delta(s_{\sigma 1})$ . Precisely this argument also essentially satisfies to deal with Condition (b). Condition (c) is again trivial.  $\square$

#### 8.4.4. Concluding Remarks

In this section, we have provided a state-based characterization of weak distribution bisimulation for weakly image-finite Markov automata. This definition cannot be generalized to arbitrary  $\tau$ -infinite automata, while maintaining its compliance to weak distribution bisimilarity. On arbitrary ( $\tau$ -infinite) Markov automata, weak distribution bisimilarity is able to equate distributions that weak state bisimilarity cannot. The reason is that it crucially depends on the possibility to reach a distribution over pivotal states from every (redundant) state with certain internal transitions (preserving transitions). The existence of such distributions is not guaranteed for arbitrary systems. As we have seen, the pathological representatives of this class connect an infinite number of pairwise non-bisimilar but observationally redundant states with internal transitions without ever reaching observationally pivotal states.

Even though the state-based characterization is not equivalent to the distributions-based bisimilarity, it agrees on all finite-state systems and even a broad range of practically relevant infinite-state systems, namely weakly image-finite automata.

## 8.5. Challenge Completed

In the end of Chapter 1 we have presented the open challenge of finding the coarsest reasonable notion of observational equivalence on probabilistic and stochastic systems, satisfying the three requirements from Remark 1.1, which we further refined in Section 8.2.2. In summary, they have been

1. congruence
2. preservation of the non-deterministic branching structure
3. abstraction from internal details, especially the *probabilistic* branching structure

In the course of this chapter, we have seen that weak distribution bisimilarity satisfies the first and the last requirement. It now remains to discuss in how far it also satisfies the second requirement.

As we have not stated this requirement formally, our argumentation will remain informal and thus is not meant to be a definite answer in a formal sense. In our discussion of strong LTS bisimilarity and weak LTS bisimilarity in Chapter 3, we have argued that using the bisimulation methodology in the definition of an equivalence relation may already be considered a warrant for this relation to preserve the non-deterministic branching structure. In contrast, in our discussion of page 111 we have showed that relaxed bisimilarity does not preserve the branching structure although it follows the bisimulation methodology. This dichotomy is rooted in the fact that relaxed bisimilarity has imposed the bisimulation condition upon a complete distribution, instead over single states, as this is usually the case. In Section 8.4, we have seen that instead weak distribution bisimilarity can also be phrased based on states. In this definition, the usual bisimulation condition on states, i.e. the mutual simulation (up to bisimilarity) of each transition is fully preserved. It comes in the form of Definition 8.16, which we recall in the following.

**Definition** (Recapitulation of Definition 8.16). A distribution  $\gamma$  is said to simulate  $s$  with respect to a set of internal transitions  $P \subseteq \xrightarrow{\tau}$  and an equivalence relation  $\mathcal{R}$ , if whenever  $s \xrightarrow{a}_c \mu'$  then there exist  $\mu, \xi'$  and  $\xi$  such that  $\mu' \xrightarrow{P} \mu$  and  $\gamma \xrightarrow{a}_c \xi'$  and  $\xi' \xrightarrow{P} \xi$ , and  $\mu \mathcal{L}(\mathcal{R}) \xi$ .

Even though not fully obvious from this definition, we also recall that in the final definition of weak state bisimilarity (Definition 8.19), the simulation condition is applied only in the case where  $\gamma$  is a Dirac distribution, and thus effectively a state. The simulation condition is thus mutually applied on states, as it is also the case for all other well-known state-based bisimulations.

Without doubt, the simulation condition of Definition 8.16 is unlikely more involved than the simple condition of the well-known bisimulation, for instance weak probabilistic bisimilarity. The precise difference is the inclusion of the extra internal transitions with the predicate  $\xrightarrow{P}$ . Therefore, if weak state bisimilarity failed to preserve the non-deterministic branching structure of bisimilar systems, its cause must be rooted in this transition predicate. The set of transitions  $P$ , however, is chosen in such way, that it precisely contains these transitions, where behaviour is preserved fully. Hence, also the non-deterministic branching structure is preserved.

We finally want to note that weak distribution bisimilarity, and thus also weak state bisimilarity, coincides with weak LTS bisimilarity on LTS. This clearly illustrates that the addition of  $\xRightarrow{P}$  does not interfere with the way the non-deterministic branching structure is perceived by the bisimulations.

In summary, this establishes that weak distribution bisimilarity indeed satisfies all three requirements of Remark 1.1. We have thus reached a successful ending for our quest for the canonical observational equivalence on Markov automata. Even more, as Markov automata subsumes probabilistic automata and interactive Markov chains, we have also characterized the canonical observational equivalence for these model classes.

## 8.6. Related Work

Notions of bisimulation that are based on distribution are rare and have seemingly not appeared in the literature before 2008 in a strong and deterministic setting, and then 2010 for a weak and fully non-deterministic setting. The breakthrough contributions here were [DHR08] and [EHZ10a], respectively. [Den+09] were the first to use this idea for simulation relations.

The idea of fusing distributions along sequences of internal transitions has, however, already been implemented in 2001 in a notion of branching bisimulation for alternating probabilistic automata [AB01]. We will discuss these and other notions in detail in the following, and relate it to our work.

### 8.6.1. Distribution Bisimilarity on Rabin-style Probabilistic Automata

The insight that it is somehow natural to define bisimilarities for probabilistic systems over distributions instead of states is not completely new. In [DHR08], a distribution based notion of bisimilarity has been introduced for Rabin-style probabilistic automata. To discriminate this bisimilarity and its corresponding bisimulations, simply baptized *bisimulation* in [DHR08], from the variation distribution-based notions of bisimilarity in this thesis, we will call it in the following Doyen-style distribution bisimilarity, after the name of the first author. While this bisimilarity and weak distribution bisimilarity share the same essential idea, they differ vastly in both their technical details and semantic intentions.

First, Rabin-style probabilistic automata are a simpler model than the one we consider. It has no internal non-determinism *and* it is input-enabled, i.e. for every state, there exists exactly one transition for each action label. In our setting, we can translate this to the following restricted kind of PA: Let  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\cdot})$  be an internal deterministic probabilistic automaton that is input-enabled, i.e. for every  $a \in Act$  and  $s \in S$  there exists a  $\mu \in Dist(S)$  with  $s \xrightarrow{a} \mu$ . In addition,  $\tau \notin Act$ . We will refer to such automata as Rabin-style PA for the rest of this section.

In this way, the most challenging aspects that we had to solve in defining weak distribution bisimulation do not occur in Rabin-style automata, among them the question how to treat transition labels that are only enabled in some, but not in all states in the support of a transition, and, after all, how to deal with internal non-determinism. In addition, Doyen-style distribution bisimilarity is a *strong* bisimilarity, i.e., it does not treat internal transitions in a special way. As

we have discussed thoroughly in this chapter, however, the semantical treatment in this regard has been the most challenging part of our work.

Concerning semantic intentions, weak distribution bisimilarity has been designed to capture behaviour that is essentially characterized by the probabilities with which actions can occur in the various states of an automaton. The semantic intention of Doyen-style distribution bisimilarity is instead to characterize behaviour in terms of probabilistic traces, where individual states and their branching structure play only a minor role. More leaning towards the classic concept of automata theory, a Rabin-style probabilistic automaton can accept *words*, i.e. traces, with a certain probability, if there is a non-zero probability to reach an *accepting state*. Two automata are then considered equivalent, if they have the same probabilities for acceptance of traces. Doyen-style distribution bisimilarity can be shown to coincide with this equivalence. To achieve this, Doyen-style distribution bisimulations have the additional condition that two probability distributions can be bisimilar only if the respective probabilities to be in any accepting state are the same. In our setting and using our notation, we can rephrase Doyen-style distribution bisimulation in the following definition.

**Definition 8.21 (Doyen-style distribution bisimulation).** Let  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\cdot})$  be a Rabin-style PA. Furthermore, let  $\mathcal{F} \subseteq S$  be a set of *accepting states*.

Then, a symmetric relation  $\mathcal{R} \subseteq Dist(S) \times Dist(S)$  is a Doyen-style distribution bisimulation if whenever  $\mu \mathcal{R} \gamma$  then

- $\mu(\mathcal{F}) = \gamma(\mathcal{F})$
- whenever  $\mu \xrightarrow{a} \mu'$  then  $\gamma \xrightarrow{a} \gamma'$  and  $\mu' \mathcal{R} \gamma'$ .

As usually, we define Doyen-style distribution bisimilarity as the largest Doyen-style distribution bisimulation.  $\triangleleft$

Note that this definition crucially depends on the set of accepting states  $\mathcal{F}$ . If we set  $\mathcal{F} = \emptyset$  or  $\mathcal{F} = S$  then any two distributions can be shown to be bisimilar.

A direct comparison of Doyen-style distribution bisimilarity and weak distribution bisimilarity is not possible, as the latter is not aware of the concept of accepting state. To allow technically for a comparison, we can replace the set of accepting states by adding a self-loop to every accepting state, labelled by a fresh distinguished action  $\dagger$ , which we assume not to be in  $Act$ . Formally, we add a transition  $s \xrightarrow{\dagger} \delta(s)$  for every  $s \in S$ . Clearly, this is a bijective transformation of Rabin-style probabilistic automaton with accepting states to a Rabin-style PA. In this way, we can compare Doyen-style distribution bisimilarity on Rabin-style probabilistic automata to weak distribution bisimilarity on the transformed automata.

**Theorem 8.16.** Doyen-style distribution bisimilarity is strictly coarser than weak distribution bisimilarity (on the transformed probabilistic automaton).

The subset relation is not hard to see. For the strictness, we provide a counterexample:

**Example 8.17.** We consider a simple automaton with  $Act = \{a\}$  with the following states and transitions:

- let  $t$  be a state with  $t \xrightarrow{a} \delta(t)$  and  $t \notin \mathcal{F}$ .
- let  $f$  be a state with  $f \xrightarrow{a} \delta(f)$ , but  $f \in \mathcal{F}$ .

- let  $s, s'_1$  and  $s'_2$  be states with transitions  $s \xrightarrow{a} \langle (f : \frac{1}{4}), (t : \frac{3}{4}) \rangle$ ,  $s'_1 \xrightarrow{a} \langle (f : \frac{1}{2}), (t : \frac{1}{2}) \rangle$ , and  $s'_2 \xrightarrow{a} \delta(t)$ . Let none of these states be in  $\mathcal{F}$ .

Consider the distributions  $\mu = \delta(s)$  and  $\gamma = \langle (s'_1 : \frac{1}{2}), (s'_2 : \frac{1}{2}) \rangle$ . It holds that both  $\mu \xrightarrow{a} \langle (f : \frac{1}{4}), (t : \frac{3}{4}) \rangle$  and  $\gamma \xrightarrow{a} \langle (f : \frac{1}{4}), (t : \frac{3}{4}) \rangle$ . It is easy to verify that  $\mu$  and  $\gamma$  are thus Doyen-style distribution bisimilar. However, they are not weak state bisimilar. The reason is that  $\gamma$  can be split into the distributions  $\delta(s'_1)$  and  $\delta(s'_2)$ , which are obviously not weak distribution bisimilar, as their only transition leads to two different distributions over pairwise non-bisimilar states  $f$  and  $t$ .  $\triangleleft$

This example shows that weak distribution bisimilar is still more distinctive concerning the behaviour of individual states in the support of a distribution, while Doyen-style distribution bisimilarity is exclusively concerned with whole distribution behaviour. In fact, it is not hard to see that it coincides with relaxed bisimilarity. Therefore, both allow to relate distributions that provide the same probability to accepting states, or states with  $\dagger$  transitions, respectively, irrespectively of the concrete additional behaviours of the individual states.

**Theorem 8.17.** Doyen-style distribution bisimilarity coincides with relaxed bisimilarity (on the transformed probabilistic automaton).

*Proof Sketch.* As  $\tau \notin \text{Act}$ , only the second clause of Definition 8.3 is relevant. For the same reason, weak transitions coincide with strong transitions then. As we have no internal non-determinism, combined strong transition and strong transitions coincide, too. Finally, as every state  $s$  is input-enabled,  $s \xrightarrow{a|_p}_c \mu$  holds for some  $\mu$ , i.e. the execution probability of  $a$  is 1. Taking all these consideration into account, we see that  $\mu \xrightarrow{a|_p}_c \mu'$  for  $p \in [0, 1]$  if and only if  $\mu \xrightarrow{a} \mu'$ . This holds for all actions  $a \neq \dagger$  and distributions  $\mu, \mu'$ . For  $\dagger$ ,  $\mu \xrightarrow{\dagger|_p}_c$  holds in the transformed automaton if and only if  $\mu(t) = p$  in the original Rabin-style probabilistic automaton.  $\square$

To the best of our knowledge, this bisimilarity has been the only known existing characterization of a probabilistic bisimilarity defined on distributions before the introduction of weak distribution bisimilarity in [EHZ10a]. We have shown, that it in fact coincides with relaxed bisimilarity. Like the latter, it thus fails to be a congruence relation and to preserve the non-deterministic branching structure.

In [FZ14], a strong distribution-based bisimulation has been introduced that uses similar ideas as relaxed bisimulation. In the literature, it was the first attempt ever made to bridge the gap between equivalences on Rabin-style probabilistic automata and bisimulations on PA.

### 8.6.2. Probabilistic Branching Bisimilarities

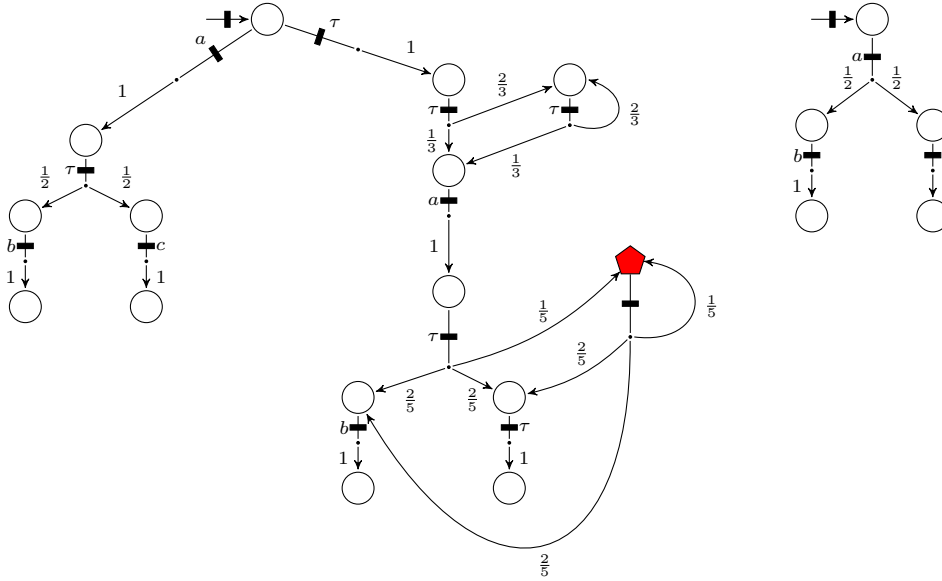
Branching bisimilarity for labelled transition systems [GW96] is a stricter variant of weak LTS bisimilarity. It additionally demands that all states that lie on sequences of internal transitions that occur before an observable action must all belong to the same equivalence class. Thus, a weak transition may leave the equivalence class of the initial state at most at the moment when the observable transition is taken, but not as long as internal transitions are taken. This notion of bisimilarity preserves the non-deterministic branching structure of process in a stricter sense




than weak bisimilarity. As we have noted in Section 3.2.2, this strict form of preservation is not necessary in order to preserve the properties we aim for. Therefore, we have not investigated branching bisimilarities in this thesis so far.

Several notions of branching bisimilarity for probabilistic automata exist. The first proposal for Segala-style probabilistic automata, as we consider them in this thesis, stems from [Seg95; SL95]. Being defined akin to weak probabilistic bisimilarity concerning its treatment of probabilistic behaviour, it does not allow to fuse probability distributions along sequences of internal transitions. The same holds for delay branching bisimulation [Sto02b].

For the alternating probabilistic automata of Hansson [Han91], rooted probabilistic branching bisimulation has been introduced. It has been first introduced in [AB01; And02] for deterministic systems, and then extended to non-deterministic systems in various slightly different variations in [ABW06; AW06; AGT08; AG09]. In [AGT12], the coarsest congruence contained in probabilistic branching bisimulation is established. The advantage of these variant of branching bisimilarity over the others is that it does allow to fuse sequences of internal behaviour in a similar way than weak distribution bisimilarity. However, being also a relation over states, this is only the case for the alternating model, where so called probabilistic states exist that represent distributions. In this way, relating non-deterministic states and distributions, i.e. probabilistic states, comes natural. Still, a translation of this notion of bisimilarity to Segala-style probabilistic automata is so far not at hand and is also not obvious to obtain.



**Figure 8.18.:** Branching Bisimilar Automata in Distribution-Based Settings

**Example 8.18.** In [AGT12], the PA in Figure 8.18 are used to show that in the non-alternating model (i.e. the one we consider here), no notion of branching bisimulation exists that is able to equate the two automata. The authors suppose that this might be due to the fact that the state  is not bisimilar to any of the states in the PA on the right, no matter which notion of branching bisimulation is given.  $\triangleleft$

It is a straightforward exercise to establish that the two automata in Figure 8.18 are indeed weak distribution bisimilar. Thus, it seems promising to give a novel definition of probabilistic branching bisimulation for PA based on distributions. We claim that weak distribution bisimilarity can be changed into a notion of branching bisimilarity on Segala-style probabilistic automata with only few adaptations. This relation will both maintain ability to fuse probabilities along sequences of internal transitions like weak distribution bisimilarity, and preserve the branching structure of a process in a stricter sense.

**Definition 8.22 (Branching Distribution Bisimulation).** Let  $(S, \bar{s}, Act, \xrightarrow{\cdot}, \dashv\!\!\rightarrow)$  be a finite MA. A symmetric relation  $\mathcal{R}$  over  $Dist(S)$  is a branching distribution bisimulation, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in Act^X \setminus \{\tau\}$

- (a)  $\mu \xrightarrow{\tau} \mu'$  implies either  $\mu' \mathcal{R} \gamma$  or  $\gamma \xrightarrow{\tau}_c \gamma''$  for some  $\gamma'' \in Dist(S)$  with  $\gamma \mathcal{R} \gamma''$  and  $\gamma'' \xrightarrow{\tau}_c \gamma'$  and  $\mu' \mathcal{R} \gamma'$ ,
- (b)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xrightarrow{\tau}_c \gamma''$  for some  $\gamma'' \in Dist(S)$  with  $\gamma \mathcal{R} \gamma''$  and  $\gamma'' \xrightarrow{a}_c \gamma'$  and  $\mu' \mathcal{R} \gamma'$ ,
- (c) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in Dist(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in Dist(S)$  such that  $\gamma \xrightarrow{\tau}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ .

◁

Branching distribution bisimulation differs from weak distribution bisimulation in the transition conditions (the first two), but not in the splitting condition. The transition conditions are typical for branching bisimulations. They demand that if the defender needs to take internal transitions then the starting distribution and the goal distribution must exhibit the same behaviour, i.e. they are bisimilar.

**Example 8.19 (Example 8.18 continued).** The two automata in Figure 8.18 are branching distribution bisimilar. ◁

The translation of this notion of branching bisimilarity on Segala-style probabilistic automata to (rooted) probabilistic branching bisimilarity on the alternating model is not obvious. As the goal of this thesis is to investigate weak bisimilarities that are as coarse as possibly, we leave further investigations here for further work, as any variant of probabilistic branching bisimilarity is necessarily strictly coarser than weak distribution bisimilarity, as it preserves a stricter notion of the branching structure.

### 8.6.3. Probabilistic Barbed Congruence

The only weak notion of bisimulation on distributions we are aware of besides weak distribution bisimulation has been developed in [DH11; DH12] independently. It is also a notion of bisimilarity on Markov automata, and, as it has been shown in [DH12], their notion of bisimilarity coincides with our notion except for minor details. The differences are mainly caused by the usage of a slightly different underlying notion of compositionality, which ensures the maximal progress property by definition, and thus, makes a semantic treatment of divergent behaviour within the definition of bisimulation dispensable. Technically, also differences exist in the definition of weak (hyper)transitions. However, as shown in [Bre13], these notions actually agree.

Another contribution of [DH11; DH12] is the definition of a novel notion of barbed congruence. Barbed congruence has been introduced by Milner and Sangiorgi [MS92b] [MS92b] as an alternative approach to characterization observable behaviour of processes. The idea is to equip the observer with a minimal set of observations, the so called *barbs* she can make of a process and the state it is currently in, for example, the currently enabled actions. In contrast to bisimulation, it is then not demanded that two related processes can match any of each others transitions. Instead, only for internal activities such a bisimulating behaviour is demanded. This property is also called *reduction-closedness*. Thus, in summary, related process states must have the same readiness and be reduction-closed. While on the first sight, this approach is ignorant of behaviour that happens *after* observable actions have been executed, this is not the case: The fact that barbed congruence is a congruence relation and by the reduction-closedness, the application of abstraction or CCS-style parallel composition reveals such behaviour by converting observable behaviour into internal behaviour. In [San95] it has been shown that observational bisimulation and barbed congruence actually coincide.

For probabilistic systems [DD07] introduces a notion of barbed congruence and shows that it coincides with a notion of bisimulation, that is essentially identical to weak probabilistic bisimilarity. Most notably, this barbed congruence is defined on states and the barbs are readiness (and failure) of actions as for the original notion of barbed congruence. The novelty of [DH11; DH12] lies in the use of probabilistic barbs  $\Downarrow_a^{\geq p}$  and that is defined over distributions over states. A distribution  $\mu$  satisfies a barb  $\Downarrow_a^{\geq p}$ , i.e.  $\mu \Downarrow_a^{\geq p}$  if  $\mu \Longrightarrow_c \mu'$  for some  $\mu'$  and the overall probability of states that enable an  $a$ -transition in  $\mu'$  is greater than  $p$ .

It is finally shown in [DH11; DH12] that barbed congruence and their variation of weak distribution bisimilarity coincide for Markov automata. This seconds from an alternative perspective our claim from Section 8.2.4 that weak distribution bisimilarity can be considered a canonical notion of observational equivalence for Markov automata. In fact, our proof idea for Theorem 8.8 is strongly inspired by the proof of [DH11; DH12].

## 8.7. Summary and Discussion

During our quest for a suitable notion of bisimilarity for Markov automata, we have encountered a multitude of different notions of bisimilarities and similarity, both well-known as well as novel. Several of them have turned out to be strongly related and to coincide in several of the submodels of Markov automata. Most notably, we have seen that the established standard notions of weak bisimilarity can as well be expressed as distribution-based weak bisimilarities, which are a genuine contribution of this chapter.

More concretely, we have seen that notions of bisimilarity considering distribution-behaviour instead of single state behaviour are rare in the literature. We have discussed contributions that share the distribution based approach with weak distribution bisimilarity, however on distinctively different underlying settings. While these differences make it impossible to compare the relations directly, we have investigated translations of the relations to our setting. It has turned out that both relations significantly differ from our weak distribution bisimilarity. Doyen-style distribution bisimilarity turns out to coincide with relaxed bisimilarity.

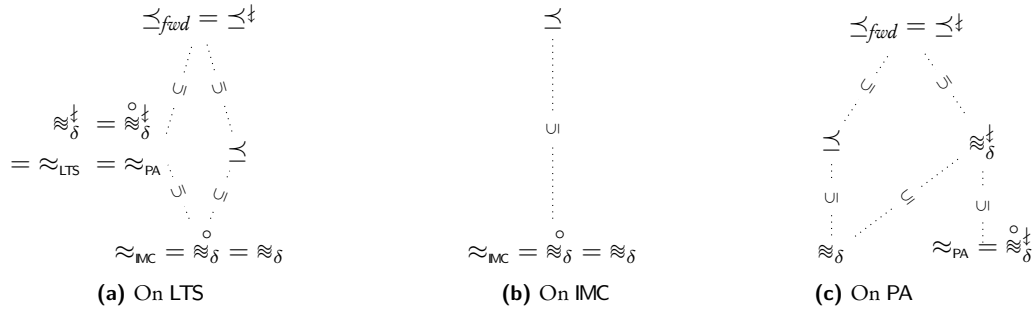
The other thread of research we reviewed focuses on alternating probabilistic automata, which again are less general than PA in that with every transition special so called probabilistic states are reached, which, in principle, encode distributions of states. By this, relating distributions (expressed as special states) comes natural. As alternating probabilistic automata are known to

differ fundamentally from PA, such that there is no faithful translation known between them, we could not undertake any direct comparison between known notions of process equivalence with the notions from this thesis. However, it inspired the definition of a novel notion of branching distribution bisimilarity. Yet, we have not pursued further investigations of its properties, since it is a stricter notion of process equivalence than weak distribution bisimilar.

While it has not been in the focus of this chapter, for completeness we remark that the findings of Section 8.3, that the well-known notions of bisimilarity can be equivalently rephrased as relations over distributions have been first shown in [EHZ10b] for the weak case, while the case for strong variants of bisimilarity has been discussed in [Hen12].

We will now summarize the findings of this chapter in the form of three subset-inclusion lattices in Figure 8.19. For each of the three prominent submodels, LTS, IMC, and PA, we consider a separate lattice. On reason for this separation is that on different submodels, different notions coincide. For example, on LTS, most equivalences coincide, as it only has a very elementary notion of transition, as compared to the probabilistic transitions of PA. Another reason for the separation is that not every notion of (bi)similarity can be reasonably deployed in every submodel. For example, notions that do not respect the maximal progress assumption (marked by  $\dagger$ ), or that are completely ignorant of timed transitions, are not considered in the IMC hierarchy.

Distribution-based relations only appear in the form of their reductions to states. For example, instead of weak distribution bisimilarity,  $\approx$ , we consider  $\approx_\delta$ . Such reductions are always marked by a subscript  $\delta$ . Finally, a circle  $\circ$  above a relation refers to the *semi*- variation. For example,  $\overset{\circ}{\approx}_\delta$  refers to semi-weak distribution bisimilarity (reduced to states).



**Figure 8.19.:** Inclusion Hierarchy of the Weak (Bi)Similarities

We end this summary with some remarks about specific well-known bisimilarities that have not been mentioned explicitly so far.

On *labelled transitions systems*,  $\approx_\delta$ ,  $\overset{\circ}{\approx}_\delta$  and  $\approx_{IMC}$  coincide with a notion of bisimilarity called *stability respecting bisimilarity* [Gla93].

On *discrete-time Markov chains*, which are submodels of Markov Automata (see Section 7.3)  $\approx_\delta$  and  $\overset{\circ}{\approx}_\delta$  coincide with weak bisimilarity [BH97; Bai+05]. We claim that  $\preceq$  and  $\overset{\circ}{\preceq}$  can be adapted such that they both coincide with weak similarity for state-labelled DTMC [Bai+05].

On *continuous-time Markov chains*, which are as well submodels of Markov Automata,  $\approx_\delta$  and  $\overset{\circ}{\approx}_\delta$  coincide with lumping equivalence [HHK02], due to the fact that our weak transitions do not affect timed transitions. In contrast, for the same reason, we see no obvious way to adapt  $\preceq$

and  $\overset{\circ}{\preceq}$  to match weak similarity for *state-labelled* CTMC, as introduced in [Bai+05].

On *interactive Markov chains*  $\approx_\delta$  and  $\overset{\circ}{\approx}_\delta$  coincide with weak bisimilarity  $\approx_{\text{IMC}}$  [Her02]. A yet weaker variant has been introduced in [Bra02]. However, the difference is orthogonal to the issues we have focused on here. We are confident, that our novel relation can be adapted in such a way that the core aspect of [Bra02] becomes available also for Markov automata.

To the best of our knowledge, no weak similarity relations for IMC have been introduced in the literature so far, so the one jointly induced by  $\overset{\circ}{\preceq}$  and  $\preceq$  is new. In [HK10], a *strong* simulation relation has been introduced, though.



## 9. Axioms

This chapter aims to establish a sound and complete axiomatization for weak distribution bisimilarity on the fragment of finite and recursion-free MA processes. It thus continues our effort to compare the precise semantics of the models discussed in this thesis, which we started in Chapter 6.

**Outline and Contributions.** We begin with defining a suitable calculus that allows us to express Markov automata processes by terms of an algebraic language in Section 9.1. As weak bisimilarities in general, also  $\approx_\delta$  fails to be a congruence with respect to the non-deterministic choice operator  $+$ . We thus need to introduce the coarsest congruence contained in weak distribution bisimilarity. In Section 9.2, we will then provide a set of axioms for finite processes without recursion, which we will prove sound and complete in Section 9.3 and Section 9.4. As we will show, in principle, the axioms for weak distribution bisimilarity are for the most part common with the axioms of weak probabilistic bisimilarity and weak IMC bisimilarity. The fact that weak distribution bisimilarity is a significantly coarser relation will be expressed by the addition of one single axiom. Besides, we will see several new axioms that result from a syntactical simplification we made in the algebra that allows us to treat immediate and stochastic behaviour in a uniform way. While not strictly necessary, it allows for a more convenient notation. In addition, we will also provide sound and complete axiomatization for the simulation relation included in weak distribution bisimilarity, which is itself already a congruence. Section 9.5 is dedicated to the discussion which challenges arise when one attempts to generalize the axiomatization to finite process *with* recursive behaviour. Section 9.6 concludes the section.

### 9.1. Calculus

The *calculus of Markov processes* unifies the approach of the calculi of interactive Markov chains and of probabilistic process. Our intentions behind this calculus is solely to serve as the formal basis for our comparative semantics, and to show the basic mechanics behind an algebra for Markov automata. Our language does by no means aim to be as refined or expressive as independent developments that have appeared in the meanwhile [Tim+12; Tim13]. These algebras usually come with more elaborate applications and are often extended with orthogonal aspects, such as data. Yet, it is clear that they share the same elementary ideas that stem from the algebras of the ancestors of Markov automata, which we discussed in Chapter 6.

Let for this chapter the set  $Act^X$  denote the set  $Act \cup \{\chi(r) \mid r \in \mathbb{R}_{>0}\}$ . It differs from the prior meaning in that  $\chi(0)$  is not included in the set.

**Definition 9.1 (MA Syntax).** Let  $a \in Act^X$ ,  $A \subseteq Act$ , and  $p \in [0, 1]$ . We define the languages MA of expressions by the following syntax.

$$\mathbb{MA} \ni E, E' ::= 0 \mid a.\mathcal{D} \mid E + E' \mid \Delta(E)$$

where  $\mathcal{D} \in \mathbb{D}$  is a distribution expression with syntax  $\mathcal{D} ::= \bigoplus_{i \in I} p_i E_i$ . If the sum expression has only one operand, we write  $\delta(E)$  instead of  $\bigoplus 1 \cdot E$ .  $\triangleleft$

We write  $\mathbb{D}$  and  $\mathbb{MA}$  for the set of  $\mathbb{D}$ -expressions and  $\mathbb{MA}$  expressions, respectively. We use  $E, E_1, E_2, F, \dots$  to range over  $\mathbb{MA}$ , and  $\mathcal{D}, \mathcal{D}_1, \dots$  to range over  $\mathbb{D}$  expressions respectively. Misusing notation slightly, we write  $E \in \text{Supp}(\mathcal{D})$  if  $E = E_i$  for some  $i \in I$ .

Because in the syntax of our calculus, the symbol  $a$  ranges over the set  $\text{Act}^X$ , we allow expressions such as  $\chi(r).\mathcal{D}$  for any  $r \in \mathbb{R}_{>0}$ . Except for this addition, the syntax of this calculus equals the syntax of  $\mathbb{PA}$ , our calculus for  $\mathbb{PA}$ .

Similar as before,  $\chi(r)$  represents a sum of individual rates leading to a distribution of goal states. If there is only one summand of this kind in an  $\mathbb{MA}$  expression, it coincides with the total exit rate of the state represented by that expression. Instead of adding this special parameterized action, we could have as well adopted the approach of  $\mathbb{IMC}$ , where a real valued prefix such as  $\lambda.E$  denotes that with rate  $\lambda$  the current state transitions to state  $E$ . Both notations can be converted in each other immediately in the case where  $\mathcal{D}$  corresponds to a Dirac distribution:  $\chi(\lambda).\delta(E)$  can be written as  $\lambda.E$  with  $\lambda \in \mathbb{R}_{>0}$  and vice versa. As we will show later, the case of a general  $\mathcal{D}$  can be reduced to the last case by an axiomatic rewriting. Essentially, our notation can thus be considered a conservative generalization of the  $\mathbb{IMC}$ .

The benefit of our approach is that it allows us to deal with immediate transitions and timed transitions in a uniform way. Both now obey a structure where an action prefix is followed by a *distribution* expression. Without this adaption, we would have needed to present several of the following axioms twice: once for immediate prefix expressions and once for delay prefix expressions.

As a consequence of this decision, some of the axioms that we inherit from the axiomatization of  $\mathbb{IMC}$  need adaption. However, as we will see, all adaptations are straightforward.

*Notation 9.1.* Where convenient, we write  $\lambda.\mathcal{D}$  instead of  $\chi(\lambda).\mathcal{D}$  for  $\lambda \in \mathbb{R}_{>0}$ .

*Notation 9.2.* Since the operator  $+$  is associative and commutative, we use  $\sum_{i \in I} E_i$  to represent the sum of  $\mathbb{MA}$ -terms ( $I$  finite). We apply the notational conventions defined for  $\bigoplus$  in Section 2.4 also here.

### 9.1.1. Congruence

Being a weak bisimulation,  $\approx_\delta$  is not a congruence with respect to non-deterministic choice  $+$ . We apply the usual fix.

**Definition 9.2.**  $E \simeq_{\mathbb{MA}} F$  holds if and only if  $E \approx_\delta F$  and whenever

1.  $E \xrightarrow{\tau} \mu$  then  $F \xrightarrow{\tau}_c \circ \Longrightarrow_c \gamma$  and  $\mu \approx \gamma$ , and
2.  $F \xrightarrow{\tau} \gamma$  then  $E \xrightarrow{\tau}_c \circ \Longrightarrow_c \mu$  and  $\mu \approx \gamma$ .

$\triangleleft$

**Theorem 9.1.**  $\simeq_{\mathbb{MA}}$  is the coarsest congruence contained in  $\approx_\delta$ .



### 9.1.2. Semantics

We will formalize the intuitive interpretation of  $\mathbb{MA}$  expressions by means of structural operational rules. We define a semantics for each expression of  $\mathbb{D}$  and  $\mathbb{MA}$ , by mapping them onto probability distributions over expressions and a transition system, respectively. The state space of the transition system is defined as the set of all  $\mathbb{MA}$  expressions. Each expression  $E \in \mathbb{MA}$  thus appears somewhere in this transition system. Its semantics is therefore determined naturally by the state space reachable from the expression/state.

For  $\mathbb{MA}$ -terms, we define two transition relations, one for actions, and one to represent the impact of time. We will provide rules for prefix ( $a.$ ) with  $a \in Act^X$ , choice ( $+$ ) and divergence ( $\Delta$ ). Termination ( $0$ ) does not require specific rules.

The transition relations  $\multimap$  and  $\multimap$  are derived by the SOS semantics in Definition 9.3. There, a unifying notation is used in the form of  $\longrightarrow$ , where  $\xrightarrow{a}$  with  $a \in Act$  stands for  $\xrightarrow{a}$  and  $\xrightarrow{\lambda}$  with  $\lambda \in \mathbb{R}_{>0}$  stands for  $\xrightarrow{\lambda}$ .

**Definition 9.3.**

1. The transition relation  $\multimap \subseteq \mathbb{MA} \times Act \times Dist(\mathbb{MA})$  is the least relation given by the rules in Table 9.1.
2. The transition relation  $\multimap \subseteq \mathbb{MA} \times \mathbb{R}_{>0} \times Dist(\mathbb{MA})$  is the least *multi*-relation given by the rules in Table 9.1, where the multiplicity of a transition is determined by the number of different derivations that witness the membership of this transition.

◁

$$\begin{array}{c}
 \text{prefix} \frac{}{a. \bigoplus_{i \in I} p_i E_i \xrightarrow{a} \langle (E_i : p_i) \mid i \in I \rangle} \quad a \in Act^X \\
 \\
 \text{choiceL} \frac{P \xrightarrow{a} \mu}{P + Q \xrightarrow{a} \mu} \quad \text{choiceR} \frac{Q \xrightarrow{a} \mu}{P + Q \xrightarrow{a} \mu} \\
 \\
 \text{diverge1} \frac{}{\Delta(E) \xrightarrow{a} \Delta(E)} \quad \text{diverge2} \frac{E \xrightarrow{a} E'}{\Delta \xrightarrow{a} E'}
 \end{array}$$

**Table 9.1.:** Operational semantic rules for  $\mathbb{MA}$ .

In our Markov automata calculus  $\mathbb{MA}$ , the prefix operator is followed by a distribution expression independent from whether it represents an immediate action ( $a \in Act$ ) or a timed delay ( $\chi(r)$ ). In contrast, in Markov automata themselves, rate transitions lead to states as inherited from IMC. This difference implies that the transition systems induced by  $\mathbb{MA}$  terms are not fully isomorphic to Markov automata. In fact, this is only the case when each distribution expression

in a  $\mathbb{MA}$  term corresponds to a Dirac distribution. Thus, while every  $\mathbb{MA}$  has an isomorphic correspondent  $\mathbb{MA}$  term, it is not true vice versa. However, each  $\mathbb{MA}$  term can be uniquely mapped to a term isomorphic to a  $\mathbb{MA}$  by the following transformation rule, where every occurrence of distribution expression  $\mathcal{D} = \bigoplus_{i \in I} p_i E_i$  is replaced by a Dirac distribution expression:

$$\chi(r) \cdot \bigoplus_{i \in I} p_i E_i = \sum_{i \in I} \chi(r_i) \cdot \delta(E_i)$$

with  $r_i = r \cdot p_i$ . This rule follows immediately from  $(IDEM-2)_{\mathbb{MA}}$ , which we will introduce later and show to be sound with respect to  $\simeq_{\mathbb{MA}}$ .

In summary, while the set of transition systems induced by  $\mathbb{MA}$  terms and the set of Markov automata are not fully identical up to isomorphism, their correspondence is strong enough to consider  $\mathbb{MA}$  a reasonable process algebra for Markov automata.

## 9.2. Axioms

For a larger part, the axioms of this chapter are identical to the axioms from Chapter 6 for PA and IMC. This is not surprising, as these formalisms form the basis of Markov automata. In fact, most differences result from our technical decision to use a uniform syntax for immediate and timed transitions in  $\mathbb{MA}$ . In this way, several axioms for IMC, dealing with timed transitions, are not sound any longer, and need, mostly straightforward, adaptations. We summarize the axioms that remain unchanged in Table 9.3.

(COM)	$E + F = F + E$
(ASS)	$E + (F + G) = (E + F) + G$
(NEUT)	$E + 0 = E$
(D-COM)	$a.(\mathcal{D} \oplus_p \mathcal{D}') = a.(\mathcal{D}' \oplus_{1-p} \mathcal{D})$
(D-ADD)	$a.(\mathcal{D} \oplus_p \mathcal{D}) = a.\mathcal{D}$
(D-ASS)	$a.(\mathcal{D}_1 \oplus_p (\mathcal{D}_2 \oplus_{\frac{q}{1-p}} \mathcal{D}_3)) = a.((\mathcal{D}_1 \oplus_{\frac{p}{p+q}} \mathcal{D}_2) \oplus_{p+q} \mathcal{D}_3)$
(CC)	$a.\mathcal{D} + a.\mathcal{D}' = a.\mathcal{D} + a.\mathcal{D}' + a.(\mathcal{D} \oplus_p \mathcal{D}') \quad (a \in Act)$
(D- $\tau$ -2)	$a.\mathcal{D} = a.\mathcal{D} + a.(\bigoplus_{i \in I} p_i \mathcal{D}'_i)$ where $\mathcal{D} = \bigoplus_{i \in I} p_i \delta(E_i + \tau.\mathcal{D}'_i)$
(D- $\tau$ -3)	$\tau.\mathcal{D} = \tau.\mathcal{D} + a.(\bigoplus_{i \in I} p_i \mathcal{D}'_i)$ where $\mathcal{D} = \bigoplus_{i \in I} p_i \delta(E_i + a.\mathcal{D}'_i)$

**Table 9.2.:** Standard Axioms

In Table 9.2 axioms for idempotency and divergence are missing. These axioms need a special

treatment due to the fact that our language is no longer fully compatible syntactically with the algebras  $\mathbb{PA}$  and  $\mathbb{IMC}$ . We provide them thus separately in Table 9.3. They follow along the lines of the axiomatization for  $\mathbb{PA}$  and  $\mathbb{IMC}$  (cf. Section 6.2.1) with some necessary adaptations: For  $\mathbb{PA}$ , the idempotency law  $E + E = E$  was sufficient. In the presence of timed transitions this law becomes unsound, as it would allow to write for example  $1.\delta(0) + 1.\delta(0) = 1.\delta(0)$ . By the additivity of rates, however, it follows that in this equation we incorrectly add the rates. A correct equation would be  $1.\delta(0) + 1.\delta(0) = 2.\delta(0)$ . Therefore, similar to the  $\mathbb{IMC}$  setting, we need to distinguish between idempotency in the context of immediate transitions and of timed transitions. In the latter case, idempotency actually becomes more an additivity law. We use  $(IDEM-1)_{\text{MA}}$  and  $(IDEM-2)_{\text{MA}}$  for these two cases. Note that we cannot directly use the corresponding axioms for  $\mathbb{IMC}$ , as there the prefix operator is followed by a process expression, but not a distribution expression.

$$\begin{aligned} a.\mathcal{D} + a.\mathcal{D} &= a.\mathcal{D} & (IDEM-1)_{\text{MA}} \\ \lambda_1.\mathcal{D} + \lambda_2.\mathcal{D}' &= (\lambda_1 + \lambda_2).(\mathcal{D} \oplus_{\frac{\lambda_1}{\lambda_1 + \lambda_2}} \mathcal{D}') & (IDEM-2)_{\text{MA}} \end{aligned}$$

**Table 9.3.:** Idempotency Axioms

The axioms for divergence, summarized in Table 9.4, are obtained from the corresponding axioms for  $\mathbb{IMC}$  by replacing all occurrences of process expressions by distributions expressions.

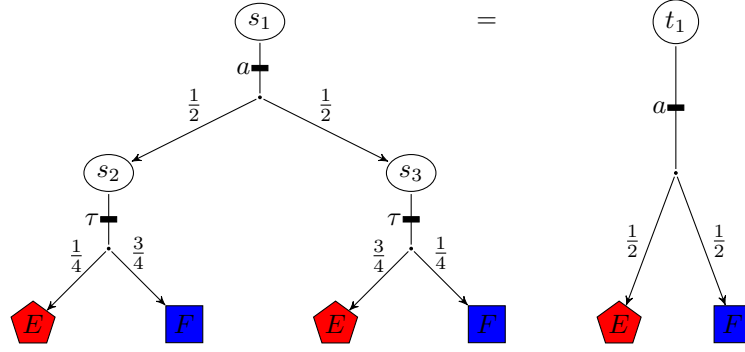
$$\begin{aligned} \Delta(E) &= \tau.\delta(\Delta(E)) & (\Delta-1)_{\text{MA}} \\ \Delta(\tau.\mathcal{D} + F) &= \tau.\delta((\tau.\mathcal{D} + F)) & (\Delta-2)_{\text{MA}} \\ \Delta(\lambda.\mathcal{D} + F) &= \Delta(F) & (\Delta-3)_{\text{MA}} \\ \Delta(E) + \lambda.\mathcal{D} &= \Delta(E) & (\Delta-4)_{\text{MA}} \\ \Delta(E + F) &= \Delta(E + F) + E & (\Delta-5)_{\text{MA}} \\ \Delta(E) + \Delta(E) &= \Delta(E) & (\Delta-6)_{\text{MA}} \end{aligned}$$

**Table 9.4.:** Divergence Axioms

The axiom that actually distinguishes  $\simeq_{\text{MA}}$  from the other congruences, particularly from  $\simeq_{\text{PA}}$ , is  $(D-\tau-1)_{\text{MA}}$  (Table 9.5). It expresses the distinguishing property of weak distribution bisimilarity to fuse sequences of internal transitions. This axiom has already appeared under the same name in our axiomatization of  $\preceq_{\text{fwd}}$  in Section 6.3, with the only differences that the latter is an axiom over distributions. This is no very surprising, as weak distribution bisimilarity has turned out to be the notion of bisimulation containing the probabilistic forward simulation kernel (cf. Theorem 8.13).

$$a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D}) \quad (D-\tau-1)_{\text{MA}}$$

**Table 9.5.:**  $\tau$  - Axioms ( $a \in \text{Act}^X$ )

Figure 9.1.: A Typical Instance of  $(D\text{-}\tau.1)_{\text{MA}}$ 

**Example 9.1.** Figure 9.1 shows a typical instance of an equivalence between two MA that can be proven with axiom  $(D\text{-}\tau.1)_{\text{MA}}$ . The MA terms corresponding to the automaton on the left hand side is

$$a.(\underbrace{\delta(\tau.(E \oplus_{\frac{1}{4}} F))}_{\mathcal{D}'} \oplus_{\frac{1}{2}} \underbrace{\delta(\tau.(E \oplus_{\frac{3}{4}} F))}_{\mathcal{D}}).$$

For better readability we wrote  $E$  instead of  $\delta(E)$  and accordingly for  $F$ . By  $(D\text{-}\tau.1)_{\text{MA}}$  this is equal to

$$a.(\underbrace{(E \oplus_{\frac{1}{4}} F)}_{\mathcal{D}'} \oplus_{\frac{1}{2}} \underbrace{\delta(\tau.(E \oplus_{\frac{3}{4}} F))}_{\mathcal{D}}).$$

By axiom  $(D\text{-}COM)$ , we can swap  $\mathcal{D}$  and  $\mathcal{D}'$ , which allows us to apply  $(D\text{-}\tau.1)_{\text{MA}}$  again. By swapping the distributions back again this leads to

$$a.(\underbrace{(E \oplus_{\frac{1}{4}} F)}_{\mathcal{D}'} \oplus_{\frac{1}{2}} \underbrace{(E \oplus_{\frac{3}{4}} F)}_{\mathcal{D}}).$$

With multiple applications of  $(D\text{-}COM)$ ,  $(D\text{-}ADD)$  and  $(D\text{-}ASS)$ , we finally obtain

$$a.(E \oplus_{\frac{1}{2}} F),$$

which is the process term corresponding to the MA on the right-hand side of the Figure 9.1.  $\triangleleft$

We have now summarized all axioms that together establish an axiomatization of  $\simeq_{\text{MA}}$ . We collect them in the set  $\mathcal{A}_{\text{MA}}$ , which shall contain all axioms in Table 9.2, 9.3, 9.4 and 9.5. We will show in the remainder of this chapter that  $\mathcal{A}_{\text{MA}}$  is sound and complete for  $\simeq_{\text{MA}}$  for finite and recursion-free processes.

### 9.3. Soundness

Most axioms are known to be sound for  $\simeq_{\text{PA}}$ , which implies their soundness also for  $\simeq_{\text{MA}}$ . Only one axiom needs further considerations.

**Lemma 9.1 (Soundness).** All axioms in  $\mathcal{A}_{\text{MA}}$  are sound with respect to  $\simeq_{\text{MA}}$ .

*Proof Sketch.* Axiom  $(D\tau\cdot 1)_{\mathbb{MA}}$  is the only new axiom and remains to be discussed. As both sides of

$$a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D})$$

have only one initial transition labelled by  $a$ , the only challenging part is to show that their successor distributions are weak distribution bisimilar, i.e.  $\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D} \approx \mathcal{D}' \oplus_p \mathcal{D}$ . As witness relation, we use the identity relation on distributions plus the pair  $(\tau.\mathcal{D}', \mathcal{D}')$ . Checking the necessary conditions is then a simple exercise.  $\square$

## 9.4. Completeness

In this section we will show that the set  $\mathcal{A}_{\mathbb{MA}}$  of axioms is complete for acyclic  $\mathbb{MA}$  processes. Formally, we will establish that whenever two acyclic processes are congruent,  $E \simeq_{\mathbb{MA}} F$ , then we can rewrite one into the other only by use of the axioms from  $\mathcal{A}_{\mathbb{MA}}$ , i.e.  $E =_{\mathbb{MA}} F$ .

For notational convenience, we usually keep the application of axioms  $(D\text{-}COM)$ ,  $(D\text{-}ASS)$ ,  $(D\text{-}ADD)$  and  $(CC)$  implicit in the following. This allows us to represent distribution expressions always in the form  $\bigoplus_{i \in I} p_i E_i$  where  $I$  is a finite index set and  $p_i \in [0, 1]$  with  $\sum_{i \in I} p_i = 1$ .

Acyclic processes are exactly the expressions of  $\mathbb{MA}$  without variables and recursive behaviour. For  $a_i \in Act$ , we use the notation  $\sum \{ a_i.\mathcal{D}_i \mid i \in I \}$  for a finite index set  $I$  as an abbreviation of the finite summation of the summands  $a_i.\mathcal{D}_i$  (with duplicated summands removed). This is justified by the axioms  $(COM)$ ,  $(ASS)$  and  $(IDEM\cdot 1)_{\mathbb{MA}}$ .

**Definition 9.4 (Sum Form).** A process  $E$  is in *sum form* if it is either of the form

1.  $\chi(r). \bigoplus_{j \in J} p_j E_j + \sum \{ a_i. \bigoplus_{j \in J_i} p_j E_j \mid i \in I \}$ ,  
where  $a_i \neq \tau$  for all  $i \in I$ , or
  2.  $\sum \{ a_i. \bigoplus_{j \in J_i} p_j E_j \mid i \in I \} + \sum \{ \Delta(\sum \{ b_i. \bigoplus_{j \in J_i} q_j F_j \mid i \in I_k \}) \mid k \in K \}$ ,  
where  $b_i \neq \tau$  for all  $i \in I_k$  and  $k \in K$ , or
  3. 0, and
- each  $E_j$  and  $F_j$  is again in sum form, and
  - $a_i, b_j \in Act$ , i.e. they are not timed actions, and
  - all  $E_j$  and  $F_j$  that are part of the same distribution expression are pairwise different with respect to  $\equiv$ .

$\triangleleft$

Note that in Case 2 summands of the form  $\Delta(0)$  are included, as  $I_k$  may be the empty set.

In the following, we will call the processes  $E_j$  and  $F_j$  occurring in  $E$  the successor processes of  $E$ .

By axioms  $(COM)$  and  $(ASS)$ , the specific order of the summands is not important. We will therefore always assume the summands to be ordered in a way suitable for our presentation.

Furthermore, by axiom  $(IDEM-1)_{MA}$ , for the summands in  $\sum$  we can assume that  $E_i = E_j$  implies  $i = j$ . Thus, in order to show that two processes in sum form are syntactically identical, it suffices to check that every summand of one process also is a summand of the other process and vice versa.

The proof for completeness proceeds by induction over the longest sequence of actions a process  $E$  can perform. We count actions resulting from prefix expressions. Furthermore, actions resulting from divergence  $\Delta$  will count as one. We formalize this characteristic as  $\mathcal{D}(E)$ , by an inductive definition along the structure of  $E$ .

$$\begin{aligned}\mathcal{D}(0) &= 0 \\ \mathcal{D}(\Delta(E)) &= 1 + \mathcal{D}(E) \\ \mathcal{D}(a.\mathcal{D}) &= 1 + \mathcal{D}(\mathcal{D}) \text{ for } a \in Act^x \\ \mathcal{D}(E + F) &= \max\{\mathcal{D}(E), \mathcal{D}(F)\} \\ \mathcal{D}(\delta(E)) &= \mathcal{D}(E) \\ \mathcal{D}(\mathcal{D} \oplus_p \mathcal{D}') &= \max\{\mathcal{D}(\mathcal{D}), \mathcal{D}(\mathcal{D}')\}\end{aligned}$$

A direct inductive proof of completeness seems impossible. Instead, we need to strengthen the induction hypothesis by several unobvious statements. We need to introduce further definitions first.

**Definition 9.5.** We call a process  $E$  *saturated*, if whenever  $E \xrightarrow{a} \mu$  with  $a \in Act$  then also  $E \xrightarrow{a} \mu$  and also each successor process of  $E$  is saturated. In addition, if  $E \xrightarrow{\tau} \mu = \delta(\Delta(F))$  for some  $F$ , then  $\Delta(F)$  shall also be a summand of  $E$ .  $\triangleleft$

This definition states that every weak transition of  $E$  is also represented by a strong transition. This means that if we look at a process in sum form, we can immediately recognize every possible (non-convex) observable behaviour that might happen next, only from looking at its summands. In the special case that a weak transition leads to a Dirac distribution over a  $\Delta$  expression, we add the expression as summand since  $\Delta$  expressions are a special syntactical form of an internal transition, which should be immediately recognizable on top-level of the process structure.

**Definition 9.6.** We call a process  $E$  *weak saturated*, if either  $E$  is saturated or  $E$  is of the form  $\tau.\mathcal{D}$  and every process  $E' \in Supp(\mathcal{D})$  is saturated.  $\triangleleft$

Note that the difference between saturated and weak saturated is only in the initial shape of a process. In both cases, the successor processes must be saturated. If  $E$  is weak saturated, however,  $E$  itself may not be saturated, if it has only one initial transition.

**Definition 9.7.** We call a process  $E$  *convex reduced*, if for every action  $a \in Act$  the set  $\mathcal{T}_a = \{\mu \mid E \xrightarrow{a} \mu\}$  of  $a$ -successor processes, is the (unique) minimal set of generators of its set of convex combinations  $\{\bigoplus_{\mu \in \mathcal{T}_a} p_\mu \cdot \mu \mid \sum_{\mu \in \mathcal{T}_a} p_\mu = 1\}$ , and each successor process of  $E$  is convex reduced.  $\triangleleft$

It is a well-known fact that every convex set has a minimal set of generators. This minimal set is furthermore unique. We will make use of this fact in our completeness proof by reducing the transitions of a process to exactly this unique set of generators.

**Definition 9.8.** We call a process  $E$  *tangible*, if whenever  $E \xrightarrow{\tau} \mu$  then either  $\delta(E) \not\approx \mu$  or  $E = \Delta(F)$  for some process  $F$  and  $\mu = \delta(E)$  — and every successor process of  $E$  is tangible.  $\triangleleft$

**Definition 9.9.** We call a process  $E$  *weak tangible*, if either  $E$  is tangible, or  $E$  is of the form  $\tau.\mathcal{D}$  and every  $E' \in \text{Supp}(\mathcal{D})$  is tangible.  $\triangleleft$

Note that the difference between tangible and weak tangible is only in the initial shape of a process. In both cases, the successor processes must be tangible. If  $E$  is weak tangible, however,  $E$  itself may not be tangible, if it has only one initial transition.

**Lemma 9.2.** Every acyclic process can be rewritten into sum form by use of the axioms (COM), (ASS), (NEUT), (IDEM-1)<sub>MA</sub>, (IDEM-2)<sub>MA</sub>, and (D-ADD), and the divergence axioms ( $\Delta\cdot 1$ )<sub>MA</sub>, ( $\Delta\cdot 2$ )<sub>MA</sub>, ( $\Delta\cdot 3$ )<sub>MA</sub>, ( $\Delta\cdot 4$ )<sub>MA</sub>, and ( $\Delta\cdot 5$ )<sub>MA</sub>.

*Proof.* The proof proceeds by structural induction.

$E \equiv 0$ : 0 is in sum form by definition.

$E \equiv \Delta(F)$ : By induction,  $F$  is in sum form. If the structure of  $F$  follows Case 1, we eliminate the summand  $\chi(r) \cdot \bigoplus_{j \in J} p_j E_j$  by applications of ( $\Delta\cdot 3$ )<sub>MA</sub>. If the structure of  $F$  follows Case 2, we need to eliminate the  $\Delta$ , since then  $F$  has a summand of the form  $\tau.\mathcal{D}$  or of the form  $\Delta(H)$ . Let us first assume that  $F \equiv \tau.\mathcal{D} + G$ . By ( $\Delta\cdot 2$ )<sub>MA</sub>, we rewrite  $E = \Delta(F) = \Delta(\tau.\mathcal{D} + G)$  to  $\tau.\delta(\tau.\mathcal{D} + G) = \tau.\delta(F)$ . Since  $F$  is by induction in sum form, so is  $\tau.\delta(F)$ .

Now, we consider the case where  $F \equiv \Delta(H) + G$ . We show how reduce this case to the last case. First, we rewrite  $F$  with ( $\Delta\cdot 1$ )<sub>MA</sub> to  $\tau.\delta(\Delta(H)) + G$ . Hence  $E = \Delta(\tau.\delta(\Delta(H)) + G)$ , which is an instance of the last case.

$E \equiv a.\mathcal{D}$ : Follows immediately by induction and the definition, and possibly an application of (D-ADD): Assuming by induction that all processes in  $\text{Supp}(\mathcal{D})$  are in sum form, we reduce  $\mathcal{D}$  by (D-ADD) whenever necessary, and then we immediately see that  $E$  is an instance of Case 1 or 2, depending on whether  $a = \lambda$  or not.

$E \equiv F + G$ : By induction,  $F$  and  $G$  are in sum form. If both are instances of Case 2, we are done. If both are instances of Case 1, we apply (IDEM-2)<sub>MA</sub> to merge the timed actions. If without loss of generality  $F \equiv 0$ , i.e. an instance of Case 3, we eliminate the instance by (NEUT). The interesting case occurs when without loss of generality  $F$  is an instance of Case 1 and  $G$  is an instance of Case 2. The problem occurs when in  $F + G$  we have both a timed actions, stemming from  $F$ , and  $\tau$  or  $\Delta$ , stemming from  $G$ . First assume that  $G \equiv \tau.\mathcal{D} + H$ . We will reduce this case to the other case, where  $G \equiv \Delta(L) + H$ , and then show how to eliminate the semantically superfluous timed action from  $F$ .

First, we show the transformation.

$$\begin{aligned}
 \tau.\mathcal{D} &= \tau.\delta(\tau.\mathcal{D}) && \text{by } (D\cdot\tau\cdot 1)_{MA} \\
 &= \tau.\delta(\tau.\mathcal{D} + 0) && \text{by } (NEUT) \\
 &= \Delta(\tau.\mathcal{D} + 0) && \text{by } (\Delta\cdot 2)_{MA} \\
 &= \Delta(\tau.\mathcal{D}) && \text{by } (NEUT)
 \end{aligned}$$

Now we have reduced the problem to the case, where  $G \equiv \Delta(L) + H$  ( with  $L \equiv \tau.\mathcal{D}$ ). The more general case where  $G \equiv \Delta(L) + H$  for an arbitrary  $L$  is treated with  $(\Delta\cdot 4)_{\mathcal{MA}}$ , which allows us to eliminate any summands  $\lambda.\mathcal{D}'$  occurring in  $F$ .

□

**Definition 9.10 ( $\Delta$ -Expandedness).** We say a process is  $\Delta$ -expanded, if every subexpression of the form  $\Delta(E+F)$  or  $\Delta(E)$  occurs as a summand of a summation, where also  $E$  is a summand, except if  $E \equiv 0$ . ◁

**Lemma 9.3.** Every process  $E$  in sum form can be transformed into a process in sum form that is in addition  $\Delta$ -expanded with help of axiom  $(\Delta\cdot 5)_{\mathcal{MA}}$ .

*Proof.* We can apply the axiom on every  $\Delta$  subterm containing summands different from 0. Since these summands cannot be  $\Delta$  expressions themselves, or  $\lambda$  prefix expression, this transformation will not hurt any of the conditions of sum form. □

**Lemma 9.4.** Every process  $E$  in sum form can be transformed into a saturated process in sum form by the axioms in  $\mathcal{A}_{\mathcal{MA}}$ .

*Proof.* We proceed by induction over  $\mathcal{D}(E)$ . By induction, we may assume that every successor process of  $E$  is already saturated. Let  $E \xRightarrow{a} \mu$  with  $a \in \text{Act}$ . Then, we distinguish two cases:

1. there exists  $\mu'$  such that  $E \xrightarrow{\tau} \mu'$  and  $\mu' \xRightarrow{a} \mu$ .
2. there exists  $\mu'$  such that  $E \xrightarrow{a} \mu'$  and  $\mu' \Rightarrow \mu$ .

Without loss of generality, we assume that in the second case  $\mu' \neq \mu$ , because otherwise, we are already done.

As all processes contained in the support of  $\mu'$  are by induction hypothesis already saturated, it holds that either

1.  $\mu' \xrightarrow{a} \mu$ , or
2.  $\mu' \xrightarrow{\tau} \mu$ ,

depending on the case we consider. As all the processes in the support of  $\mu'$  are in sum form, our argumentation so far lets us derive that  $E$  must contain a summand with the following structure, with possibly empty index sets  $I$  and  $J$ :

$$\begin{aligned}
 & \overbrace{1. \tau. ((\bigoplus_{i \in I} c_i \delta(a.\mathcal{D}_i) \oplus_p (\bigoplus_{j \in J} d_j \delta(E_j + a.\mathcal{D}'_j)))}^{\mu'} \text{ with} \\
 & \qquad \qquad \qquad \mu = (\bigoplus_{i \in I} c_i \mathcal{D}_i \oplus_p (\bigoplus_{j \in J} d_j \mathcal{D}'_j)) \\
 & \overbrace{2. a. ((\bigoplus_{i \in I} c_i \delta(\tau.\mathcal{D}_i) \oplus_p (\bigoplus_{j \in J} d_j \delta(E_j + \tau.\mathcal{D}'_j)) \oplus_q \mathcal{D}'')}^{\mu'} \text{ with} \\
 & \qquad \qquad \qquad \mu = (\bigoplus_{i \in I} c_i \mathcal{D}_i \oplus_p (\bigoplus_{j \in J} d_j \mathcal{D}'_j) \oplus_q \mathcal{D}'')
 \end{aligned}$$



Note that this case distinction only is complete since we can, by Lemma 9.3, assume without loss of generality, that  $E$  is  $\Delta$ -expanded. Hence, we do not need to treat any cases where relevant expressions occur inside a  $\Delta$ -expression.

We proceed with Cases 1 and 2. The idea is now to apply  $(D\text{-}\tau\text{-}3)$  to the first case and Axiom  $(D\text{-}\tau\text{-}2)$  to the second case in order to add an additional summand  $a.\mu'$  to  $E$ . In neither case we cannot proceed directly, as syntactically, the axioms are not applicable. For the first case, it suffices to apply Axiom  $(NEUT)$  followed by Axiom  $(COM)$  to all subexpressions of the form  $a.\mathcal{D}_i$ . We then arrive at

$$\tau.\left(\left(\bigoplus_{i \in I} c_i \delta(0 + a.\mathcal{D}_i) \oplus_p \left(\bigoplus_{j \in J} d_j \delta(E_j + a.\mathcal{D}'_j)\right)\right)\right)$$

We now apply Axiom  $(D\text{-}\tau\text{-}3)$  to obtain

$$\begin{aligned} &\tau.\left(\left(\bigoplus_{i \in I} c_i \delta(0 + a.\mathcal{D}_i) \oplus_p \left(\bigoplus_{j \in J} d_j \delta(E_j + a.\mathcal{D}'_j)\right)\right)\right) \\ &\quad + a. \underbrace{\left(\bigoplus_{i \in I} c_i \delta(\mathcal{D}_i) \oplus_p \left(\bigoplus_{j \in J} d_j \delta(\mathcal{D}'_j)\right)\right)}_{\mu} \end{aligned}$$

For the second case, we proceed completely analogously, for subexpressions  $\tau.\mathcal{D}_i$ , but need additional rewriting for the distribution expressions  $\mathcal{D}''$ , which did not appear in the preceding case. With Axiom  $(D\text{-}\tau\text{-}1)_{MA}$ , we can prefix  $\mathcal{D}''$  by a  $\tau$ , and then again apply Axioms  $(NEUT)$  followed by Axiom  $(COM)$  as before to finally arrive at

$$a.\left(\left(\bigoplus_{i \in I} c_i \delta(0 + \tau.\mathcal{D}_i) \oplus_p \left(\bigoplus_{j \in J} d_j \delta(E_j + \tau.\mathcal{D}'_j)\right) \oplus_q \delta(0 + \tau.\mathcal{D}'')\right)\right).$$

Now, the expression is in the right shape to apply Axiom  $(D\text{-}\tau\text{-}2)$  to obtain the desired result as before. Finally, in both cases we undo our changes done to the original summand. If we proceed in this way for every summand of  $E$ , we obviously arrive at a saturated process at the end, that is still in sum form.

For the last case, namely that  $E \xrightarrow{\tau} \mu = \delta(\Delta(F))$  for some  $F$ , which demands the addition of the summand  $\Delta(F)$ , we first use our results from before. Hence assume without loss of generality that  $E \xrightarrow{\tau} \mu'$  and  $\mu' \implies \delta(\Delta(F))$ . Then as we have shown before, we can saturate  $E$  such that also  $E \xrightarrow{\tau} \delta(\Delta(F))$ , i.e.  $E$  has a summand of the form  $\tau.\delta(\Delta(F))$ . Now by  $(\Delta\text{-}1)_{MA}$  we can remove the trailing  $\tau$  and obtain the desired summand  $\Delta(F)$ .  $\square$

**Lemma 9.5.** Every saturated process  $E$  in sum form can be transformed into a convex reduced and saturated process in sum form by the axioms in  $\mathcal{A}_{MA}$ .

Again, we can turn every such process in a process that is in addition  $\Delta$ -expanded. We will make use of this lemma in the future without further notice.

**Lemma 9.6.** Every saturated and convex reduced process  $E$  in sum form can be transformed into a saturated and convex reduced process in sum form that is in addition  $\Delta$ -expanded with help of axiom  $(\Delta\text{-}5)_{MA}$ .

*Proof.* Follows as for Lemma 9.3, but with the additional remark that adding an expression from inside a  $\Delta$  expression as a summand does not break saturatedness nor convex reducedness, since every expression reachable now with a weak or strong transition has been reachable with a weak transition before.  $\square$

The next lemma is key to our completeness proof. We need to deviate from the standard approach, in which we would first show that every process can be rewritten into some type of normal form, and then show, that the normal forms of two bisimilar processes always coincide. Surprisingly, a sequential approach to prove these two lemmas does not seem possible. Instead, we need to establish both results at the same time, as in the inductive step, the two results mutually rely on each other.

In the following, we write  $E \equiv F$ , if  $E = F$  only by the axioms (COM) and (ASS).

**Definition 9.11 (Normal Form).** We say a process is in *normal form*, if it is weak saturated, weak tangible, convex reduced,  $\Delta$ -expanded, and in sum form.  $\triangleleft$

**Lemma 9.7.** The following statements hold:

1. Every process can be rewritten into normal form with the axioms in  $\mathcal{A}_{\text{MA}}$ , and
2. For two processes  $E$  and  $F$  in normal form,  $E \simeq_{\text{MA}} F$  implies  $E \equiv F$ .

With this lemma, it is straightforward to establish completeness, Theorem 9.2. The proof of this lemma can be found in Appendix E.

**Theorem 9.2.** Let  $E$  and  $F$  be two MA processes. Then  $E \simeq_{\text{MA}} F$  implies  $E =_{\mathcal{A}_{\text{MA}}} F$ .

## 9.5. Open Challenges

Our axiomatization has only covered finite systems without recursive behaviour. This has enabled us to proceed for most proofs with induction over the longest sequence of actions a process can potentially perform. For finite processes *with* recursion, a different approach is needed, as a process may perform action sequences of infinite length.

Already for standard labelled transitions systems and weak bisimilarity, extending axiomatizations to the setting with recursive processes is more involved and requires, besides several new axioms, a completely new proof idea, based on process equations. This principle idea has been successfully applied to several notions of bisimilarity [Mil89a; Her02; BS01; DP07; DP05; LDH05; ABW06; Eis07; GW96]. Surprisingly, an adaption for weak distribution bisimilarity seems impossible, and has resisted our attempts so far. In order to stimulate further research, we record our insights here why the standard approach seems to fail.

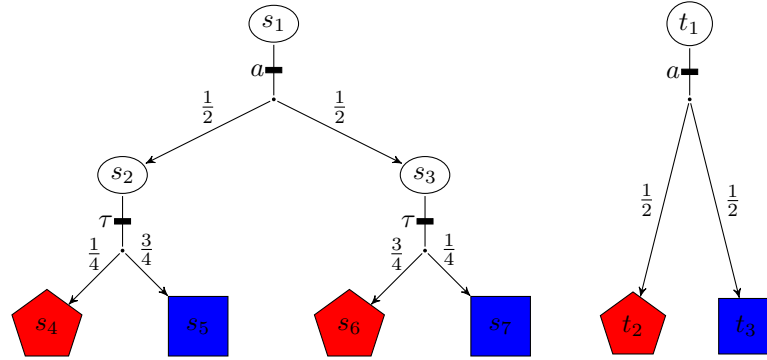
The core idea of proving completeness for finite-state systems *with* recursion is to use that two process algebraic expressions  $E$  and  $F$  are provably equal if there is an equation system over processes expressions that has a unique solution and which is provably satisfied by both systems. In this context, equation systems always are of the form  $X = E$ , where the left-hand side only consists of a variable and the right-hand side  $E$  is an arbitrary expression without recursion (usually in some standard form).

We say that an expression  $E$  provably satisfies an equation system  $X_1 = G_1, \dots, X_n = G_n$ , if for every equation  $X_i = G_i$  in the system there are process expressions  $E_i$  such that  $E_i = (G_i) \{E_1/X_1\} \{E_2/X_2\} \dots \{E_n/X_n\}$  can be proven by the axioms, and, in addition  $E_1 = E$ .

A major proof step is then to construct a single equation system, that is satisfied by both process expressions  $E$  and  $F$ . Therefore, first  $E$  and  $F$  are assumed (or rewritten) to be saturated. The idea is then to introduce one equation for every pair  $(E_i, F_j)$  of bisimilar processes expressions reachable from  $E$  and  $F$ , respectively. Every equation is of the form  $X_{i,j} = \sum a.X_{i',j''}$ . The left-hand side is a variable and its right-hand side is a (finite) summation over actions followed by a variable, and possibly unbound variables, which is something not of importance for our discussion. Intuitively, a variable  $X_{i',j'}$  represents the combined behaviour of expression  $E_i$  and  $F_j$ . Therefore, a summand  $a.X_{i',j'}$  stands on the right-hand side exactly if  $E_i \xrightarrow{a} E_{i'}$  and  $F_j \xrightarrow{a} F_{j'}$  and  $E_{i'}$  is bisimilar to  $F_{j'}$  (for  $a \neq \tau$ ) or if  $E_i \xrightarrow{\tau} E_{i'}$  and  $E_{i'}$  is already bisimilar to  $F_{j'}$  and  $F_{j'} = F_j$  or  $F_j \xrightarrow{\tau} F_{j'}$  and  $E_i$  is already bisimilar to  $F_{j'}$  and  $E_{i'} = E$ . As  $E_i$  and  $F_j$  are bisimilar and they are saturated, whenever one of them can do an  $a$ -transition to a state, the other one can follow with an  $a$ -transition to a bisimilar state. Thus, it is almost immediate by construction that  $E_i$  and  $F_j$  both are provable solutions of  $X_{i,j}$ , as every next step behaviour is reflected in their structure. The only crux is to deal with the possibility, that, from the perspective of either  $E_i$  or  $F_j$ , in certain equations an additional  $\tau$  transition may be present, which has been introduced by the other process expression. Structurally, this is reflected by the presence of a summand  $\tau.X_{i',j}$  or  $\tau.X_{i,j'}$ , where only one of the indices changes. For example, in an equation with left-hand side  $X_{i,j}$ , if  $E_i \xrightarrow{\tau} E_{i'}$ , it may be possible, that  $F_j$  cannot match this transition with an explicit transition, even though  $E_i$  and  $F_j$  are bisimilar. Still, following the above construction, the summand  $\tau.X_{i',j}$  will be part of the right-hand side of the equation. To deal with this problem, the proof strategy is then to not show directly that  $F_j$  satisfies the equation, but instead use  $\tau.F_j$  as a solution. By Axiom  $(\tau-2)$ , we can rewrite this to  $\tau.F_j + F_j$ , which means that structurally,  $F_j$  is fully preserved *and* in addition, the missing summand  $\tau.F_j$  is then present. Showing that this expression is a solution of the equation is then straightforward. Furthermore, whenever  $\tau.F_j$  is replaced in a right-hand side expression (of another equation), it must be in a summand of the form  $a.X_{l,j}$  for some  $l$ . Then by Axiom  $(\tau-1)$ ,  $a.\tau.F_j$  can be reduced to  $a.F_j$  and the additional  $\tau$  has been removed.

The crucial insight of this discussion is that the proof depends on the fact that every equation with left-hand side  $X_{i,j}$  is defined in a way that allows to almost immediately see that both  $E_i$  and  $F_j$  provably satisfy the equation  $X_{i,j}$ . If  $E_i$  ( $F_j$ ) should fail to immediately satisfy a corresponding equation, then it is guaranteed that  $\tau.E_i$  ( $\tau.F_j$ ) does so.

For weak distribution bisimilarity, an analogous construction is not obvious. The reason is that two bisimilar process expressions may reach states that are not bisimilar to any state of the other process expression. For all other known bisimilarities, this not the case.



**Figure 9.2.:** Bisimilar Automata With Non-Bisimilar States

**Example 9.2.** In Figure 9.2, the states  $s_1$  and  $t_1$  are weak distribution bisimilar. However, the states  $s_2$  and  $s_3$  are not bisimilar to any of the states, given that  $\color{red}\blacklozenge$  and  $\color{blue}\blacksquare$  are non-bisimilar states. The reason is that the distributions the two states reach by their only internal transition are unique. In contrast,  $\delta(s_1)$ ,  $\delta(t_1)$ ,  $\langle (s_2, \frac{1}{2}), (s_3, \frac{1}{2}) \rangle$  and  $\langle (\color{red}\blacklozenge, \frac{1}{2}), (\color{blue}\blacksquare, \frac{1}{2}) \rangle$  are all weak distribution bisimilar.  $\triangleleft$

It seems that the problem is rooted in the fact that weak distribution bisimilarity allows to equate distributions whose supports contain states that are not bisimilar. While being the unique feature of this bisimilarity addressing chiefly needed abstraction possibilities, we have also repeatedly observed it to be the cause of unexpected phenomena (see for example Section 8.4.3).

**Example 9.3 (cont'd).** The equation system for the automata in Figure 9.2 is not straightforward to construct. It is clear that, due to the saturation property, the first equation will contain one summand for the  $a$  transition. This summand is also rather easy to construct, as it the goal transition only involves states of equivalence class  $\color{red}\blacklozenge$  and  $\color{blue}\blacksquare$ . Hence, so far, we can conclude that the first equation must be of the following shape.

$$X_{1,1} = a. \left\langle (X_{4,2}, \frac{1}{8}), (X_{5,3}, \frac{3}{8}), (X_{6,2}, \frac{3}{8}), (X_{7,3}, \frac{1}{8}) \right\rangle + ?$$

However, it is hard to see how to construct the missing  $\tau$  transition of  $s_1$ . We cannot write something like  $\tau. \langle (X_{2,1}, \frac{1}{2}), (X_{3,1}, \frac{1}{2}) \rangle$  for  $?$ , as this would mean that  $s_2$  and  $t_1$ , and also  $s_3$  and  $t_1$ , respectively, should be bisimilar. But this is not the case, as we know.  $\triangleleft$

A solution in the concrete example is, to completely remove states  $s_2$  and  $s_3$  by the use of Axiom  $(D\text{-}\tau\text{-}I)_{MA}$ . However, in more complex examples this might not always obviously be possible, as states with redundant transitions like  $s_2$  and  $s_3$  might still have other outgoing transitions, that prevent a naive application of this axiom. One approach could be to remove such transitions whenever possible by something like an inverse saturation procedure, that prunes out transition whenever they can also be reconstructed via a sequence of preceding internal transitions. However, it is by far not obvious how to identify the candidates for the removal only in an axiomatic way.

## 9.6. Summary and Discussion

This chapter has provided a sound and complete axiomatization of weak distribution bisimilarity, or more precisely, the coarsest congruence with respect to choice, contained in this relation. We have based our axiomatization on a simple algebraic language for MA that is basically the straightforward combination of the algebras for IMC and PA, however with a slight adaption of the notion of delay prefix that allows a syntactically uniform treatment of timed, and non-timed transitions, and thus reduced the number of axioms slightly.

The axiomatization *syntactically* highlights that the semantic core of weak distribution bisimilarity is its ability to fuse internal transition distributions, as expressed by this axiom:

$$a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D}) \quad (D\text{-}\tau\text{-}I)_{\text{MA}}$$

In fact, it is the only axiom that is not also valid for IMC and PA axiomatization. This axiom is also pivotal for axiomatizations of probabilistic forward simulation, the only other relation that allows for fusion of internal transition distributions (and which is, as we have already discussed, in fact only the simulation variant of weak distribution bisimulation).

The completeness proofs for our axiomatization have shown to be non-trivial already for system without recursive behaviour. While we conjecture that the axiomatization can be extended to recursive processes, we have failed to establish a proof of completeness for any such extended set of axioms. We have discussed the problems that arise and also various routes we have explored in vain. Yet we hope that our insights may prove useful for future proof attempts.



## **Part III.**

# **Markov Automata Algorithmics**





## 10. Decision Algorithms

In this chapter, we investigate algorithms that decide whether two states of an automaton are equivalent with respect to the various notions of bisimilarity we have introduced.

We restrict our consideration to finite MA. For strong bisimilarity and naïve weak probabilistic bisimilarity, we can rely on existing decision algorithms with minimal adaptations [Her02; BEM00; CS02]. The only challenge is to combine treatment of immediate probabilistic transitions with the treatment of timed transitions. For naïve weak probabilistic bisimilarity, the algorithm is indeed a straightforward combination of the respective algorithms we have presented in Chapter 4 and 5.

For weak distribution bisimilarity, the case is harder. We will base our algorithm on deciding weak state bisimilarity instead of weak distribution bisimilarity. If we want to recover an algorithm for weak distribution bisimilarity that relates a pair of distributions  $(\mu, \nu)$  instead of states, we can introduce two fresh states from which a transition with a fresh label goes to distribution  $\mu$ , respectively  $\nu$ , and then check the bisimilarity of these two states with the algorithm tailored to the state-based setting.

**Outline and Contributions.** Establishing a decision algorithm for weak distribution bisimilarity comes with several challenges. Section 10.1 explains the challenges we have met and tries to provide helpful insights for future developments. In Section 10.2, we present a decision algorithm for weak distribution bisimilarity and discuss its runtime complexity. Section 10.3 concludes the chapter. The results this chapter stem from [Eis+13b].

### 10.1. Challenges

For decision algorithms of bisimilarities it is natural to pursue a partition refinement approach (cf. Sections 3.3, 5.3, and 4.3, or for further reference, [CS02; Her02; KS90; PLS00]). For weak distribution bisimilarity this seems not feasible in a straightforward manner. Instead, the standard approach needs to be adapted drastically. In the following, we will shed light on the challenges we met and provide some explanation on alternative approaches that we have explored but finally abandoned.

**Distribution- vs. State-Based Bisimilarity** A direct partition refinement approach would need to refine a partitioning over the set of all distributions over the state space, as weak distribution bisimilarity is defined naturally over distributions. However, this seems infeasible, since even for finite state spaces, the set of all distributions is infinite. Clearly, it seems possible to reduce the set of all distribution to only those distributions reachable by (sequences of) weak combined hypertransitions from the initial distributions. While this set is still infinite, it is possible to represent the set of distributions reachable from a state via weak combined transitions, by an exponential number of distributions (cf. [CS02]). In our case, however, we would need to

consider distributions reachable from distributions (and so on), and not states. Thus, we face an exponential complexity with considerably large constants.

Another obstacle is the second condition of weak distribution bisimulation that demands to check –at least if implemented naïvely– if for every splitting of the challenger distribution the defender can reach a distribution that may be split accordingly. The number of splittings that need to be checked is (naïvely) again infinite. All of this potential infiniteness must be countered by some suitable finite representations. The bisimulation conditions for state-based characterization of weak distribution bisimulation, weak state bisimulation, are, in turn, simpler to deal with by finite representations. Being state-based, it again suffices to compute a partitioning for a finite set. Moreover, there is no explicit condition dealing with distributions any longer. However, now, much of the difficulty has shifted towards determining a suitable computation of the set  $P$  of *preserving transitions*, that plays a key role in the state-based characterization (cf. Chapter 8).

In summary, different to the standard partition refinement approach, our algorithm does not only need to compute a state space partitioning  $\mathbb{W}$  but, at the same time, needs to compute the set of preserving transitions  $P$ .

**Challenging Refinement.** The still open challenge with this approach is to compute the set of preserving transitions  $P$  in a time-efficient way. It seems natural to take an iterative refinement approach here, too, so that both the tentative set of preserving transitions  $P$  and the tentative state space partitioning  $\mathbb{W}$  are refined at the same time. So starting with the set of all internal transitions as the initial value for  $P$  and the set  $\{S\}$  as initial state space partitioning  $\mathbb{W}$ , we would first refine  $\mathbb{W}$  (if possible), and then remove those transitions from  $P$ , which do no longer satisfy the necessary conditions with respect to the current tentative partitioning  $\mathbb{W}$ . Then, in the next round, we again try to refine  $\mathbb{W}$  with respect to the refined set  $P$ . We continue this until none of the two sets changes. The following example shows that this approach may lead to erroneous results.

**Example 10.1.** Let  $s \xrightarrow{\tau} \blacktriangle$  and  $s \xrightarrow{\tau} \blacklozenge$  and also  $t \xrightarrow{\tau} \blacktriangle$  and  $t \xrightarrow{\tau} \blacklozenge$ . Let furthermore both  $s$  and  $t$  be equipped with a self-loop labelled by action  $a$ .

Assume that states  $\blacktriangle$  and  $\blacklozenge$  are not weak distribution bisimilar. Then, clearly, none of the transitions is preserving. However,  $s$  and  $t$  are obviously weak distribution bisimilar. Assume the transition  $t \xrightarrow{\tau} \blacktriangle$  has been eliminated from the candidate set  $P$  already, but  $s \xrightarrow{\tau} \blacktriangle$  has not. Furthermore, assume that  $\mathbb{W}$  has already been refined enough to recognize that  $s$  and  $t$  are not in the same class as  $\blacklozenge$  and  $\blacktriangle$ , and that the two are in separate classes as well.

Then, when during the next refinement loop, which refines the partitioning  $\mathbb{W}$ , we check whether  $s$  and  $t$  can simulate their respective  $a$  transition. Now, while  $t \xRightarrow{P} \delta(t)$ , this is not the case for  $s$ , as  $s \xrightarrow{\tau} \blacktriangle$  is still contained in  $P$ . Instead  $s \xRightarrow{P} \blacktriangle$  (assuming that  $\blacktriangle$  cannot perform any further relevant  $\tau$  transitions). Now, the condition of weak state bisimulation demand that  $\delta(t) \mathcal{L}(\mathbb{W}) \blacktriangle$ . Unfortunately, this is not satisfied. As a consequence  $s$  and  $t$  are falsely considered to be *not* bisimilar.  $\triangleleft$

Clearly, this example is not a concrete counterexample against this approach. It merely demonstrates the theoretic possibility that the interleaved refinement of the two sets may interfere negatively. By this we want to emphasize that any algorithmic approach that iteratively refines both sets in a somehow interleaved manner, needs careful soundness considerations. In fact, we failed to prove that the interleaved refinement approach actually yields correct results, even for various variations of this approach. We, for instance, tried to remove as many transitions as pos-

sible in one iteration, instead of only some, with the intent that the situation as in Example 10.1 will not occur. However, we could then not prove that both the transitions from  $s$  and  $t$  become invalidated as being preserving in the same iteration step.

Our proof attempts neither lead us to the derivation a concrete counterexample. We thus conjecture that computing the relation  $P$  with a refinement approach should be possible. However, great care has to be taken in the verification of such an approach.

## 10.2. Algorithm

The algorithm we propose follows the brute-force approach to circumvent the problem we have discussed before. In Figure 10.1, which provides a top-level perspective on the algorithm, we see in Lines 4 to 7, that the quotienting procedure, *Quotient- with respect to pres* (Line 5), is repeated for all conceivable sets of preserving transitions, i.e. for all subsets of the set of all internal transitions,  $D(\tau)$ . During its refinement steps, *Quotient- with respect to pres* repeatedly checks whether  $P$  still satisfies the necessary conditions with respect to the current partitioning  $\mathbb{W}$ . If at some point it fails to do so, the procedure returns  $\emptyset$  to indicate that  $P$  was not a suitable choice for the set of preserving transitions.

In order to guarantee that finally the coarsest relation is computed, we join the results of all refinements in Line 6. Procedure *Join* computes the partitioning that results from the union of the two partitionings. Treating  $\mathbb{W}$  and  $\mathbb{W}'$  as equivalence relations over  $S$ , it computes the reflexive, transitive and symmetric closure of  $\mathbb{W} \cup \mathbb{W}'$ . Thus, when *Quotient- with respect to pres* returns  $\emptyset$  in order to indicate that no weak distribution bisimulation exists for the current candidate set  $P$ , this outcome will not alter  $\mathbb{W}'$ .

$Decide(\mathcal{A})$
1: $\mathcal{A}' = \text{Uniformize}(\mathcal{A})$ 2: $\mathcal{A}' = \text{Quotient-under-mec}(\mathcal{A}')$ 3: $\mathbb{W} = \emptyset$ 4: <b>for all</b> $P \subseteq D(\tau)$ <b>do</b> 5: $\mathbb{W}' = \text{Quotient- with respect to pres}(\mathcal{A}', P)$ 6: $\mathbb{W} = \text{Join}(\mathbb{W}, \mathbb{W}')$ 7: <b>return</b> $\mathbb{W}$

**Figure 10.1.:** Computing the partitioning with respect to weak distribution bisimilarity

In Lines 1 and 2, preprocessing steps are performed. First, the timed behaviour of MA, inherited from IMC, is treated in exactly the same way as in the algorithm for IMC that we presented in Chapter 5. In particular, the input MA is transformed into its *uniform representation* before further treatment. We recall that this allows us to treat a MA like a PA extended by special actions  $\chi(r)$  for  $r \in \mathbb{R}_{\geq 0}$ . We refer the reader to Chapter 5, Definition 5.3 for further details.

The second preprocessing step, *Quotient-under-mec*, is needed to correctly compute matching weak transitions later in the quotienting process when searching splitters. The procedure transforms every input MA into a *mec*-contracted MA. In other words, it computes the quotient MA with respect to  $=_{mec}$ . Thus, the core of the partition refinement approach relies on its input MA being *mec*-contracted. Lemma 8.9 guarantees the soundness of this preprocessing step.

The quotienting itself, procedure *Quotient- with respect to pres*, will be discussed in detail in Section 10.2.1.

### 10.2.1. The Algorithmic Core

The partition refinement occurs in procedure *Quotient- with respect to pres*, which is parameterized by  $P$ , the current candidate for the set of preserving transitions. The procedure itself is entirely unsurprising except for a consistency check performed in procedure *Consistency*: During each refinement iteration of  $\mathbb{W}^1$  in *Quotient- with respect to pres*, we check whether the currently assumed set  $P$  still satisfies Definition 8.17. If it does not, we stop refining and immediately return  $\mathbb{W} = \emptyset$ . As we have already mentioned, after each call of *Quotient- with respect to pres*, in procedure *Decide* the returned partitioning  $\mathbb{W}$  is joined with the previously computed partitioning  $\mathbb{W}'$ .

<i>Quotient- with respect to pres</i> ( $\mathcal{A}, P$ )	
1:	$\mathbb{W} = \{S\};$
2:	<b>repeat</b>
3:	$\mathbb{W}' = \mathbb{W};$
4:	<b>if not</b> <i>Consistency</i> ( $P, \mathbb{W}$ ) <b>then</b>
5:	<b>return</b> $\emptyset$
6:	$(C, a, \rho) = \text{FindSplit}(\mathbb{W}, P);$
7:	$\mathbb{W} = \text{Refine}(\mathbb{W}, (C, a, \rho));$
8:	<b>until</b> $\mathbb{W} = \mathbb{W}'$
9:	<b>return</b> $\mathbb{W}$

As for LTS, PA and IMC, the core of the quotienting process is to find splitter and refine partition accordingly. This is again done in the procedures *FindSplit* and *Refine*, respectively. While the principle ideas are standard and follow along the lines of the decision algorithm for PA and weak probabilistic bisimilarity, the technical execution and establishing soundness is highly non-trivial. We will discuss this in the rest of this section.

**Bisimulation Core Conditions.** The core condition of weak probabilistic bisimilarity is that whenever one of two distributions related by a bisimulation relation performs a transition, the other distribution can perform a transition with the same label and the resulting distributions of both transitions are again related by the lifting of the bisimulation relation.

To derive a polynomial time routine that is suited to check the core bisimulation condition, we have used Proposition 1 (page 54), which we will recall in the following, trivially adapted from PA to MA.

<sup>1</sup>For the rest of this section, we will treat  $\mathbb{W}$  both as a set of partitions and as an equivalence relation, wherever convenient, without further mentioning.

**Proposition** (Recapitulation of Proposition 1). *Given a MA  $\mathcal{A}$ , two probability distributions  $\rho_1, \rho_2 \in \text{Dist}(S)$  such that  $|\rho_1| = |\rho_2| > 0$ , two actions  $a_1, a_2 \in \text{Act}^X$ , two sets  $\check{A}_1, \check{A}_2 \subseteq \rightarrow$  of transitions, and an equivalence relation  $\mathbb{W}$  on  $S$ , the existence of  $\nu_1, \nu_2 \in \text{Subdist}(S)$  such that*

$$\rho_1 \xrightarrow{a_1 \upharpoonright \check{A}_1}_c \nu_1, \rho_2 \xrightarrow{a_2 \upharpoonright \check{A}_2}_c \nu_2, \text{ and } \nu_1 \mathcal{L}(\mathbb{W}) \nu_2$$

*can be checked in polynomial time.*

For weak state bisimulation, we actually have two core conditions:

1.  $\mathbb{W}$  is preserving with respect to  $P$ , and
2.  $P$  is preserving with respect to  $\mathbb{W}$ .

If we inspect them more closely, we see that they are indeed similar to the core condition of weak probabilistic bisimilarity. In fact, they are similar enough to allow us to reuse Proposition 1. We will explain the details in the following. First, let us recall the essential definitions for convenience.<sup>2</sup>

**Definition** (Recapitulation of Definition 8.17). Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv\!\!\rightarrow, \dashv\!\!\rightarrow)$  be given. A set of internal transitions  $P \subseteq \dashv\!\!\rightarrow$  is called *preserving with respect to an equivalence relation  $\mathbb{W} \subseteq S \times S$*  if for every  $(s, \tau, \gamma) \in P$  the distribution  $\gamma$  simulates  $s$  with respect to  $P$  and  $\mathbb{W}$ .

**Definition** (Recapitulation of Definition 8.18). An equivalence relation  $\mathbb{W} \subseteq S \times S$  is called *preserving with respect to  $P$*  if

1. whenever  $s \mathbb{W} t$ , then  $\delta(t)$  simulates  $s$  with respect to  $P$  and  $\mathbb{W}$
2. if  $s \downarrow$  then  $t \implies \gamma$  for some  $\gamma \in \text{Dist}(S)$  with  $\gamma \downarrow$ ;

The *core* of both conditions is that distribution  $\gamma$  or state  $t$ , respectively, must be able to *simulate* state  $s$  with respect to  $P$  and  $\mathbb{W}$ . Concretely, the term *simulating* here refers to Definition 8.16.

**Definition** (Recapitulation of Definition 8.16). A distribution  $\Psi$  is said to *simulate*  $s$  with respect to a set of internal transitions  $P \subseteq \dashv\!\!\rightarrow$  and an equivalence relation  $\mathbb{W}$ , if whenever  $s \xrightarrow{a}_c \Lambda$  then there exist  $\mu, \xi'$  and  $\xi$  such that  $\Lambda \xRightarrow{P} \mu$  and  $\Psi \xrightarrow{a}_c \xi'$  and  $\xi' \xRightarrow{P} \xi$ , and  $\mu \mathcal{L}(\mathcal{R}) \xi$ .

If we replace  $\Psi$  by  $\gamma$ , we obtain the core of Definition 8.17, and if by  $\delta(t)$ , the core of Definition 8.18. In the following, we will write  $\text{Sim}(s, \mathbb{W}, P, a, \Lambda, \Psi)$ , if  $\mu, \xi'$  and  $\xi$  can be found that satisfy the conditions of Definition 8.16.

Checking the predicate  $\text{Sim}$  algorithmically will be at the core of our algorithmic treatment. Even though the structures of this definition and of Proposition 1 are somewhat similar, it is not completely straightforward to see how the latter can be used to algorithmically check  $\text{Sim}$ : the proposition is not directly capable of dealing with the special transition relation  $\xRightarrow{P}$ . This is due to the fact that, by its definition, the special transition relation  $\xRightarrow{P}$  is more complex than simply restricting a weak transition to some allowed transitions. This means, it cannot obviously be expressed by  $\xrightarrow{\upharpoonright \check{A}}_c$  for some set  $\check{A}$  of suitable allowed transitions, as used within Proposition 1.

<sup>2</sup>Compared to our original definitions, we here have replaced  $\mathcal{R}$  by  $\mathbb{W}$ . As mentioned before, we treat  $\mathbb{W}$  as a binary relation whenever convenient.

**Computing the Core Condition.** As we have seen, the most challenging aspect is to compute the transition predicate  $\Longrightarrow$ . Everything else can easily be computed with help of Proposition 1. Recall that  $\mu \xRightarrow{P} \mu^*$  holds if and only if

1.  $\mu \xRightarrow{\tau \mid P}_c \mu^*$ .
2. whenever  $\mu^* \xRightarrow{\tau \mid P}_c \mu'$ , then  $\mu^* \mathcal{L}(\mathbb{W}) \mu'$ .

Computing  $\text{Sim}(s, \mathbb{W}, P, a, \Lambda, \Psi)$  now relies on two steps:

- Instead of working on the partitioning  $\mathbb{W}$  directly, we transform it into an equivalence  $\mathbb{W}'$  as follows: whenever for a state  $s \in S$ , a transition  $(s, \tau, \mu) \in P$  exists for some distribution  $\mu$ , we remove  $s$  from its equivalence class with respect to  $\mathbb{W}$ , and isolate it, such that  $s \mathbb{W} s'$  implies  $s' = s$ . For all states  $t$  and  $t'$  that do not have this property, we maintain that  $t \mathbb{W}' t'$  if and only if  $t \mathbb{W} t'$ . More formally, if  $\mathcal{C}$  is the equivalence class of  $\mathbb{W}$  containing  $s$ , we let  $\mathcal{C}' = \mathcal{C} \setminus \{s\}$  be the class replacing  $\mathcal{C}$  in  $\mathbb{W}'$ , and we add the singleton equivalence class  $\{s\}$  to  $\mathbb{W}$ .

It is easy to see that this operation is at most polynomial with respect to the number of transitions and states of the MA.

- We then apply Proposition 1 with
  - $a_1 = \tau, a_2 = a$ ,
  - $\rho_1 = \Lambda$  and  $\rho_2 = \Psi$ , and
  - $\check{A}_1 = P, \check{A}_2 = P \cup \xrightarrow{a}$  and  $\mathbb{W} = \mathbb{W}'$ .

Using  $\mathbb{W}'$  instead of  $\mathbb{W}$  now guarantees that whenever the proposition determines a match  $\nu_1 \mathcal{L}(\mathbb{W}') \nu_2$ , then the states related by  $\mathbb{W}'$  are either

- (i) identical, or
- (ii) cannot perform any further transitions in  $P$ .

While this is not exactly the meaning of  $\Longrightarrow$ , it is already quite close.

**Theorem 10.1.** There exist two distribution  $\nu_1$  and  $\nu_2$  with  $\nu_1 \mathcal{L}(\mathbb{W}') \nu_2$  according to Proposition 1 if and only if  $\nu_1 \xRightarrow{P} \nu_1^*$  and  $\nu_2 \xRightarrow{P} \nu_2^*$  and  $\nu_1^* \mathcal{L}(\mathbb{W}) \nu_2^*$ .

*Proof. Case 1* We first show this logical equivalence from left to right. Hence, assume  $\nu_1 \mathcal{L}(\mathbb{W}') \nu_2$ . Recall that each equivalence classes of  $\mathbb{W}'$  is either a subset of an equivalence class of  $\mathbb{W}$  and each state in this class cannot perform any transition in  $P$ , or it is a singleton-set. Let  $\mathcal{C} = \{s\}$  be an arbitrary class satisfying the second case. Let  $\mu_s$  be the distribution that satisfies  $s \xRightarrow{P} \mu_s$  and every state in  $\mu_s$  cannot perform any further transition in  $P$ . The distribution  $\mu_s$  must exist as we can assume the underlying MA to be *mec*-contracted. Let us write  $\nu_1$  and  $\nu_2$  as

$$p_1 \delta(s_1) \oplus \cdots p_n \delta(s_n) \oplus q \bar{\nu}_1 \text{ and } p_1 \delta(s_1) \oplus \cdots p_n \delta(s_n) \oplus q \bar{\nu}_2,$$

where the state  $s_1$  to  $s_n$  are exactly the states that lie in their own respective (singleton-set) equivalence classes, and let  $\bar{\nu}_1$  and  $\bar{\nu}_2$  denote the respective missing part of  $\nu_1$  and  $\nu_2$ . As

$\nu_1 \mathcal{L}(\mathbb{W}') \nu_2$ , it is guaranteed that we can indeed split  $\nu_1$  and  $\nu_2$  in this way. With the transitions  $s \xrightarrow{P} \mu_s$  we have introduced above, and the fact that all states in  $\text{Supp}(\bar{\nu}_1)$  and  $\text{Supp}(\bar{\nu}_2)$  cannot perform any further transition in  $P$ , it follows that

$$\nu_1 \xrightarrow{P} p_1\delta(\mu_1) \oplus \cdots p_n\delta(\mu_n) \oplus q\bar{\nu}_1 \text{ and } \nu_2 \xrightarrow{P} p_1\delta(\mu_1) \oplus \cdots p_n\delta(\mu_n) \oplus q\bar{\nu}_2.$$

Now it is immediate to check that the two distributions reached in this way are related with respect to  $\mathcal{L}(\mathbb{W})$ . To see this, we first note that trivially for every  $\mu_i$  with  $i \in \{1, \dots, n\}$ ,  $\mu_i \mathcal{L}(\mathbb{W}) \mu_i$ , and by the choice of  $\bar{\nu}_1$  and  $\bar{\nu}_2$ , from  $\bar{\nu}_1 \mathcal{L}(\mathbb{W}') \bar{\nu}_2$ —which we can assume—we can also conclude  $\bar{\nu}_1 \mathcal{L}(\mathbb{W}) \bar{\nu}_2$ , as the equivalence classes of  $\mathbb{W}$  and  $\mathbb{W}'$  agree on the states in  $\text{Supp}(\bar{\nu}_1)$  and  $\text{Supp}(\bar{\nu}_2)$ .

**Case 2** For the other direction of the logical equivalence, assume that the proposition cannot establish  $\Lambda = \rho_1 \xrightarrow{a_1 \downarrow \bar{A}_1} \nu_1$ ,  $\Psi = \rho_2 \xrightarrow{a_2 \downarrow \bar{A}_2} \nu_2$  for some distributions  $\nu_1$  and  $\nu_2$  with  $\nu_1 \mathcal{L}(\mathbb{W}') \nu_2$ , but  $\Lambda \xrightarrow{P} \Lambda'$  and  $\Psi \xrightarrow{P} \hat{\Psi}$  for some distributions  $\Lambda'$  and  $\hat{\Psi}$  with  $\Lambda' \mathcal{L}(\mathbb{W}) \hat{\Psi}$  holds.

This can only be the case if in  $\text{Supp}(\Lambda') \cup \text{Supp}(\hat{\Psi})$  there are states  $s$  that satisfy the second condition of the definition of  $\xrightarrow{P}$ , i.e. whenever  $s \xrightarrow{\tau \downarrow P} \mu'$ , then  $\delta(s) \mathbb{W} \mu'$ , and *at the same time* can still perform transitions in  $P$ . The second statement must hold, because *only* with respect to states that satisfy this statement,  $\mathbb{W}'$  and  $\mathbb{W}$  differ at all. Now assume that  $s \mathbb{W} t$ . Now if  $s$  and/or  $t$  are such a state, by the second condition of the definition of  $\xrightarrow{P}$ , whenever  $s \xrightarrow{\tau \downarrow P} \mu_s$ , then  $\delta(s) \mathcal{L}(\mathbb{W}) \mu_s$ , and accordingly for  $t$ . As we work in a *mec*-contracted MA, we can choose  $\mu_s$  in such a way that none of the state in its support can perform any further transition in  $P$ . For  $t$ , we can find a distribution  $\mu_t$  with according properties. By the transitivity of  $\mathbb{W}$ , from  $\mu_s \mathcal{L}(\mathbb{W}) s \mathbb{W} t \mathcal{L}(\mathbb{W}) \mu_t$  we can infer that  $\mu_s \mathcal{L}(\mathbb{W}) \mu_t$ . As furthermore, all states in the supports of  $\mu_s$  and  $\mu_t$  cannot perform any further transition from  $P$ , it must also hold that  $\mu_s \mathcal{L}(\mathbb{W}') \mu_t$ . This means that even though from  $\Lambda' \mathcal{L}(\mathbb{W}) \hat{\Psi}$ , we cannot immediately conclude  $\Lambda' \mathcal{L}(\mathbb{W}') \hat{\Psi}$ , it is the case that from both  $\Lambda'$  and  $\hat{\Psi}$  we can continue with transitions from  $P$  in such a way that the resulting distributions are related by both  $\mathcal{L}(\mathbb{W})$  and  $\mathcal{L}(\mathbb{W}')$ . The latter clearly contradicts our assumption. Thus, we arrived at a contradiction, and we have proven the claim.  $\square$

$\text{FindSplit}(\mathbb{W}, P)$	
1:	<b>for all</b> $(s, a, \rho) \in \mathcal{T}$ <b>do</b>
2:	<b>for all</b> $t \in [s]_{\mathbb{W}}$ <b>do</b>
3:	<b>if</b> $\text{Sim}(s, \mathbb{W}, P, a, \rho, \delta(t))$ has no solution <b>then</b>
4:	<b>return</b> $([s]_{\mathbb{W}}, a, \rho)$
5:	<b>return</b> $(\emptyset, \tau, \delta(\perp))$

**Figure 10.2.:** Finding Splitters with core computation  $\text{Sim}(s, \mathbb{W}, P, a, \rho, \delta(t))$

**Algorithmic Context.** The remaining parts of the algorithm are now easy to understand. Following the same line as for instance [CS02], *Quotient with respect to pres* makes use of a sub-

procedure *Refine*, which actually creates a finer partitioning, as long as there is a partition containing two states that violate the bisimulation condition, which is checked for in procedure *FindSplit*. More precisely, as in [CS02], procedure *Refine* divides partition  $\mathcal{C}$  into two new partitions according to the discriminating behaviour, which has been identified by *FindSplit* before. We do not provide *Refine* explicitly. The two procedures *FindSplit* and *Consistency* mainly rely on *Sim*, but are straightforward otherwise. As we have discussed, *Sim* can be decided in polynomial-time by using Proposition 1 and some smart preprocessing.

<i>Consistency</i> ( $P, \mathbb{W}$ )	
1:	<b>for all</b> $(s, \tau, \rho) \in P$ <b>do</b>
2:	<b>for all</b> $(s, a, \nu) \in \mathcal{T}$ <b>do</b>
3:	<b>if</b> <i>Sim</i> ( $s, \mathbb{W}, P, a, \nu, \rho$ ) has no solution <b>then</b>
4:	<b>return false</b>
5:	<b>return true</b>

**Figure 10.3.:** Consistency check, with core computation *Sim*( $s, \mathbb{W}, P, a, \nu, \rho$ )

**Complexity Summary.** Most aspects of the algorithm have a polynomial asymptotic worst-case runtime. Concerning the preprocessing steps, procedure *Uniformize* is already known to be polynomial from the corresponding algorithm for weak IMC bisimilarity (cf. Section 5.3). It can be computed in linear time with respect to the number of states and transitions of the input MA. Procedure *Quotient-under-mec* can be efficiently computed according to [CH11]. The quotienting core, procedure *Quotient- with respect to pres* is again polynomial. The main loop (Lines 2 to 8) is repeated at most as often as the partitioning  $\mathbb{W}$  can be split, which is at most the number of states minus one. The procedures *Consistency*, *FindSplit* and *Refine*, which are executed inside this loop, have themselves polynomial complexity. *Refine* is known to be polynomial from the corresponding algorithms for the PA and IMC bisimilarities. Both *FindSplit* and *Consistency* have a nested loop with the computational check for predicate *Sim* at their inner body. The number of executions of *Sim* can be bound by the number of transitions times the number of states for *FindSplit*, and by the number of transitions to the square for *Consistency*. As we have already argued, *Sim* itself can be checked in polynomial time. In total *Quotient- with respect to pres* thus has a runtime polynomial in the number of states and transitions.

Unfortunately, an overall polynomial runtime is prevented by the computation of the set of preserving transitions,  $P$ . As we have discussed in Section 10.1, we currently cannot iteratively refine  $P$ . We therefore adopted a brute force approach, which considers all subsets from the set of internal transitions, which is of size in  $\mathcal{O}(2^{|\rightarrow|})$ . Once an (provably correct) algorithm is found, where also the computation of the set  $P$  is iterative, the overall complexity of the algorithm can turn polynomial.

### 10.3. Summary and Discussion

In this chapter, we have developed a decision algorithm for weak distribution bisimilarity, based on an algorithm for weak state bisimilarity. This algorithm can be considered as the nucleus for



extending the compositional specification and reasoning means in use for IMC to the more expressive MA setting. The presented algorithm uses worst-case exponential time and polynomial space. In [SS14] an alternative approach is presented that is based on the computation of vanishing states rather than preserving transitions. This algorithm shares the exponential worst-case runtime with the approach presented here.

We were unable to show that the exponential bound is, in fact, hard. On the contrary, we think that a polynomial algorithm exists. During our investigations, we failed at proving the correctness of several polynomial-time aspirant algorithms, while we failed, at the same time, to establish a superpolynomial lower bound. Finding a provably correct polynomial-time decision algorithm for weak distribution bisimilarity remains an open challenge.



# 11. Minimal Normal Forms

Markov automata, and its submodels (see Section 7.3), form the backbone of successful model checkers such as PRISM [KNP11], SCOOP [Guc+13], or IscasMC [Hah+14], enabling the analysis of randomized concurrent systems. Despite the remarkable versatility of this approach, its power is limited by the state space explosion problem, and several approaches are being pursued to alleviate that problem.

In related fields, a favourable strategy is to minimize the system – or components thereof – to the quotient under bisimilarity. This can speed up the overall model analysis or turn a too large problem into a tractable one [Che+96; HK00; Kat+07; Bar+11; HK10]. Both, strong and weak bisimilarity are used in practice, with weaker relations leading to greater reduction. This approach has not been explored in the context of Markov automata, and even in the probabilistic automata setting support is fragmentary. Chapter 10 provides a base functionality, a decision algorithm, but this is not a minimisation algorithm. This chapter therefore focuses on a seemingly tiny problem: Does there exist a *unique minimal* representative of a given probabilistic automaton with respect to weak bisimilarity? And can we compute it? In fact, this turns out to be an intricate problem. We nevertheless arrive at a polynomial time algorithm, if the quotient of the MA is assumed given.

Notably, minimality with respect to the number of states of a probabilistic automaton or Markov automaton does not imply minimality with respect to the number of transitions. And further minimization is possible with respect to transition fanouts, the latter referring to the number of target states of a transition with non-zero probability. The minimality of an automaton thus needs to be considered with respect to all the three characteristics: number of states, of transitions and of transitions' fanouts.

As a byproduct, our results provide tailored algorithms for strong probabilistic bisimilarity on PA and strong and weak bisimilarity on labelled transition systems.

On general Markov automata and weak distribution bisimilarity, the situation is more complex, though. As we will see, minimizing state space, number of transitions *and* fanout at the same time is not possible. In the worst case, a minimization of the first two, may lead to a polynomial growth of the latter. Still, this minimization can lead to *drastically* smaller automata with respect to state space and number of transitions than the other notions of bisimilarity for MA and PA. Therefore, analysis algorithm that are applied on minimized automata, may thus strongly profit from the minimization depending on the concrete factor dominating their run-time complexity.

In the following discussion, we will present minimization and decidability results for all models we have considered in this thesis, namely labelled transitions systems, probabilistic automata, interactive Markov chains and Markov automata. As we have discussed in Section 7.3, all of these models are subsumed by Markov automata. To avoid redundancies in the presentation, we will technically use Markov automata as the only model in the following discussion. Whenever it is necessary or instructive to restrict to any of the above mentioned specific submodels, however, we will call the underlying MA an LTS, a PA or an IMC, respectively. We refer the reader to Section 7.3 for a discussion which restrictions on general MA are necessary to obtain each of the

different submodels.

The notions of bisimilarity we discuss will range over all the ones we have introduced in this thesis, both strong and weak, with the exception of naïve weak probabilistic bisimilarity, which we have shown to not be a particularly attractive relation. We let in the following the symbol  $\asymp$  range over any bisimilarity in

$$\mathbb{B} = \{\sim_{\text{LTS}}, \sim_{\text{PA}}, \sim_{\text{IMC}}, \approx_{\text{LTS}}, \approx_{\text{PA}}, \approx_{\text{IMC}}, \sim_{\text{MA}}, \approx_{\text{MA}}, \approx_{\delta}\}$$

unless stated differently.

**Outline and Contributions.** In Section 11.1, we define several preorder relations between MA that make precise in which respect one automaton is structurally smaller than the other. We consider various structural aspects, namely number of states, number of transitions and number of states in the support of distributions. Reduction relations will be presented in Section 11.2 together with their time complexity, for the whole spectrum of different MA submodels and the respective bisimilarities. In Section 11.3 normal forms will be defined, based on the reduction relations introduced before. As we will see, for all bisimilarities, with the notable exception of  $\approx_{\delta}$ , the normal forms will be minimal with respect to all three characteristics. We conclude the chapter in Section 11.4.

This chapter is an extended version of [Eis+13c].

## 11.1. Structural Preorders

The size of an automaton is usually expressed in terms of the size of the set of states  $|S|$  and the size of the transition relations  $|\dashv\!\!\!\rightarrow| + |\dashv\!\!\!\rightarrow|$  of the automaton. The complexity of algorithms working on probabilistic automata often depends exactly on those two metrics. A less commonly considered metric is the number of target states of a transition reached with a probability greater than zero. Especially in practical applications it is known that the first two of these metrics – the number of states and transitions of an automaton – can be drastically reduced while preserving its behaviour with respect to some notion of bisimilarity. In contrast, the last metric is usually considered a constant, but in some cases it can be reduced as well. We will formalize these three metrics by means of three preorder relations, thus allowing us to define the notion of *minimal automata* up to bisimilarity.

To capture the last of the three metrics, we introduce the following definition.

*Notation 11.1.*  $D := \dashv\!\!\!\rightarrow \cup \dashv\!\!\!\rightarrow$

**Definition 11.1 (Transition Fanout).** For a distribution  $\mu \in \text{Dist}(S)$  we define  $\|\mu\| = |\text{Supp}(\mu)|$ . For a set of transitions  $T$  we define  $\|T\| = \sum_{(s,a,\mu) \in T} \|\mu\|$ .  $\triangleleft$

**Definition 11.2 (Size Preorders).** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv\!\!\!\rightarrow, \dashv\!\!\!\rightarrow)$  and  $\mathcal{A}' = (S', \bar{s}', \text{Act}, \dashv\!\!\!\rightarrow', \dashv\!\!\!\rightarrow')$  be two MA, and let  $\asymp$  be any notion of bisimilarity. We write

- $\mathcal{A} \preceq^{|\mathcal{S}|} \mathcal{A}'$  if  $\mathcal{A} \asymp \mathcal{A}'$  and  $|S| \leq |S'|$ ,
- $\mathcal{A} \preceq^{|\mathcal{D}|} \mathcal{A}'$  if  $\mathcal{A} \asymp \mathcal{A}'$  and  $|D| \leq |D'|$ , and
- $\mathcal{A} \preceq^{\|\mathcal{D}\|} \mathcal{A}'$  if  $\mathcal{A} \asymp \mathcal{A}'$  and  $\|\dashv\!\!\!\rightarrow\| \leq \|\dashv\!\!\!\rightarrow'\|$ .

◁

We let from now on  $\preceq$  range over  $\preceq^{[S]}$ ,  $\preceq^{[D]}$ , and  $\preceq^{[ID]}$  for  $\asymp \in \mathbb{B}$ , unless mentioned otherwise. It is easy to verify that these relations are preorders.

**Definition 11.3 ( $\preceq$ -Minimal Automata).** We call a MA  $\mathcal{A}$   $\preceq$ -minimal, if whenever  $\mathcal{A}' \preceq \mathcal{A}$  for some MA  $\mathcal{A}'$ , then also  $\mathcal{A} \preceq \mathcal{A}'$ . ◁

**Lemma 11.1 (Existence of  $\preceq$ -Minimal Automata).** For every MA  $\mathcal{A}$  there exists a MA  $\mathcal{A}'$  such that  $\mathcal{A}' \asymp \mathcal{A}$  and  $\mathcal{A}'$  is  $\preceq$ -minimal.

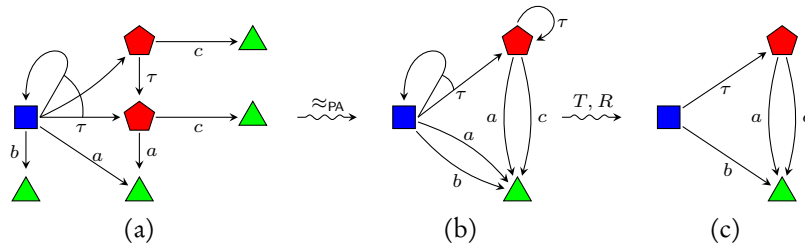
The proof of this lemma is actually as trivial as we might expect. The base reason is that each of the metrics maps on a well-founded set.

*Proof.* Every MA must consist of at least one state and at least 0 transitions. Therefore, for every equivalence class up to  $\asymp$ , its smallest representative (with respect to states) must consist of at least one state of 0 transitions. Hence, every smallest representative must have the same number of states, or transitions, respectively, as there cannot be a infinite sequence of automata with a strictly decreasing number of states or transitions. As we do not demand that this representative is unique, this suffices to conclude that at least one smallest representative must exist. This suffices to establish the result for  $\preceq^{[S]}$  and  $\preceq^{[D]}$ . Note that 0 transitions implies also a fanout of size 0. Thus, the result immediately also follows for  $\preceq^{[ID]}$ . ◻

For each of the preorders considered, the proof of this lemma exploits that for every automaton the respective metric is a finite natural number and at least 0.

As each relation  $\preceq$  is a preorder, minimal automata are not necessarily unique. For example, two  $\preceq^{[S]}$ -minimal automata  $\mathcal{A}$  and  $\mathcal{A}'$  with  $\mathcal{A} \asymp \mathcal{A}'$  may differ in the underlying set of states, and/or transitions. This will be investigated in Section 11.3.

## 11.2. Reductions



**Figure 11.1.:** (a) Example PA, (b) Quotient reduction. (c) Rescaling of convex-transitive reduction.

In this section, we introduce and formalize several methods to reduce the size of an automaton. Except for the first method, quotient reduction, the methods are especially tailored towards one or two distinct notions of bisimilarity. Two bisimilarities, weak IMC bisimilarity and weak distribution bisimilarity, stand out from the others as they preserve stability. Therefore, the

reductions that will be applied both for any of the two bisimilarities and any of the other bisimilarities, need to be defined in two variants. One that does preserve stability, and one that does *not*. The latter is necessary in order to achieve *optimal* reductions.

We summarize the properties of the reductions at the end of this section. We will further discuss the time-complexity of the reductions.

### 11.2.1. Quotient Reduction

**Definition 11.4 (Quotient Automaton).** Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  be an arbitrary MA and  $\mathcal{P}(S) = \{C \mid C \subseteq S\}$ . Given an equivalence relation  $\approx$  on  $S$ , we define the *quotient* MA  $[\mathcal{A}]_{\approx}$  with respect to  $\approx$  as the reachable fragment of the MA  $(S/\approx, [\mu_0]_{\approx}, Act, [\multimap]_{\approx}, [\multimap]_{\approx})$  where

- (i) the equivalence class mapping  $[\cdot]_{\approx} : S \rightarrow \mathcal{P}(S)$  is defined for every  $s \in S$  as  $[s]_{\approx} = \{s' \mid s' \approx s\}$ . We lift it to distributions in the usual way by letting  $[\mu]_{\approx}(C) = \sum_{s \in C} \mu(s)$  for all  $C \in S/\approx$ .
- (ii)  $S/\approx = \{[s]_{\approx} \mid s \in S\}$ ,
- (iii)  $[\multimap]_{\approx} = \{([s]_{\approx}, a, [\mu]_{\approx}) \mid (s, a, \mu) \in \multimap\}$
- (iv)  $(c, \lambda, c') \in [\multimap]_{\approx}$  if and only if  $(s, \lambda, s') \in \multimap$  for fixed representatives  $s \in c$  and  $s' \in c'$  for all equivalence classes  $c$  and  $c'$ .

◁

For the last line, recall that  $\multimap$  is a multi-set. Thus, a transition may be contained more than once if this is the case for the representatives. Also note that this definition relies on  $\approx$  guaranteeing the usual conditions for Markovian transitions, i.e. that they have the same total exit rate and that the sum of the rates leading to an equivalence class of states is the same. This is always the case for the equivalence relations we consider.

**Definition 11.5 (Quotient Reduction).** We write  $\mathcal{A} \rightsquigarrow \mathcal{A}'$  if  $\mathcal{A}' = [\mathcal{A}]_{\approx}$ .

◁

Figure 11.1(b) shows the result of applying Definition 11.5 to weak probabilistic bisimilarity and the PA in Figure 11.1(a).

### 11.2.2. Convex Reduction

In essence, strong probabilistic bisimilarity  $\sim_{\text{PA}}$  enhances standard bisimilarity by the idea that the observable behaviour of a system is closed under convex combinations of transitions. Using this fact, we minimize the number of transitions in a PA by replacing the transitions of each state by a unique and *minimal* set of generating transitions.

Recall that by Lemma 2.5, every finitely generated convex set  $C$  has a unique minimal set of generators  $\text{Gen}(C)$ , such that  $C = \text{CHull}(\text{Gen}(C))$ . Thus, the lemma immediately guarantees the optimality of our approach with respect to  $\sim_{\text{PA}}^{[D]}$ .

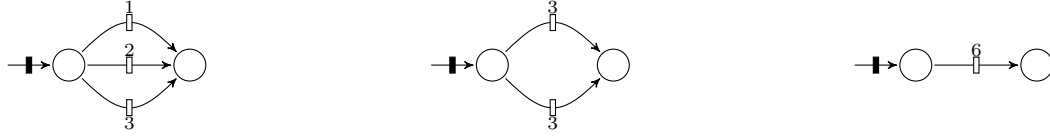


Figure 11.2.: Three equivalent Interactive Markov Chains

**Definition 11.6 (Convex Reduction).** Let  $\mathcal{A}$  be a MA. We write  $\mathcal{A} \xrightarrow{C} \mathcal{A}'$  if the automaton  $\mathcal{A}'$  differs from  $\mathcal{A}$  only by replacing the set  $\multimap$  by the set  $\multimap'$ , where

$$(s, a, \gamma) \in \multimap' \text{ if and only if } \gamma \in \text{Gen}(\text{CHull}(\{\mu \mid (s, a, \mu) \in \multimap\})).$$

◁

### 11.2.3. Markovian Summation

In MA, and also IMC as a special case. Markovian transitions are labelled by real numbers, which denote the rate with which a state change occurs. The higher the rate to transition from one state  $s$  to another state  $t$ , the more likely it is that the state  $s$  will actually transition to  $t$ . At the same time, the higher the rate, the faster the transition will likely occur. Therefore, it does not only matter that a Markovian transition from  $s$  to  $t$  exists, but also the precise number. While the label on interactive transition has the qualitative aspect to indicate which action occurs when the transition is executed, the real number with which interactive transitions are labelled has a quantitative meaning. Therefore, the more transition lead into one state, the higher is the total transition rate into this state. Thus,  $\multimap$  must allow to differentiate how many transitions from one state to another exist, even if all transitions are labelled by the same number. We therefore defined it as a multi-relation. However, the same total transition rates from one state into another are not uniquely represented by  $\multimap$ .

**Example 11.1.** In Figure 11.2, we see three times the same situation, in which  $s$  reaches  $t$  with total rate 6. However, in the automaton on the left side, this is represented by three transitions with rate 1, 2 and 3, while the middle automaton represents this as two transitions with rate 3. The automaton on the right, finally, comes with only a single transition, labelled by 6. The sum of the respective transitions is the same in each case, and thus semantically the three automata are equivalent. ◁

Clearly, the representation on the right hand-side is the most compact. Arriving at this representation from the other two is straightforward, as we only need to sum up the individual transitions.

**Definition 11.7 (Markovian Summation).** Let  $\mathcal{A}$  be a MA. We write  $\mathcal{A} \xrightarrow{\Sigma} \mathcal{A}'$  if the automaton  $\mathcal{A}'$  differs from  $\mathcal{A}$  only by replacing the set  $\multimap$  by the set  $\multimap'$ , where

$$(s, \lambda, t) \in \multimap' \text{ if and only if } \lambda = \sum_{(s, \lambda', t)} \lambda' \cdot \multimap(s, \lambda', t).$$

Note that inside the summation, we use the representation of multi-set  $\multimap$  as a mapping into natural numbers, to determine how often a certain transition occurs in  $\multimap$ .  $\triangleleft$

### 11.2.4. Maximal-Progress Cut

For the purpose of constructing a minimal representation of systems with Markovian transitions, the important difference of Markovian transitions  $\multimap$  compared to interactive transitions  $\multimap$  is that whenever they originate from an unstable state  $s$ , i.e.  $s \xrightarrow{\tau}$ , then the maximal progress assumption dictates that these transition are actually semantically ineffective (cf. Chapter 5). Thus, Markovian transitions originating from unstable states can always be *cut off* without changing the behaviour of the system.

**Definition 11.8 (Maximal-Progress Cutting).** Let  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap)$  be an IMC. We write  $\mathcal{A} \xrightarrow{M} \mathcal{A}'$  if  $\mathcal{A}' = (S', \bar{s}', Act, \multimap', \multimap')$  such that  $S = S'$ ,  $\bar{s} = \bar{s}'$ ,  $\multimap = \multimap'$ , and  $(s, \lambda, t) \in \multimap'$  if and only if  $(s, \lambda, t) \in \multimap$  but *not*  $s \xrightarrow{\tau}$ .  $\triangleleft$

### 11.2.5. Convex-Transitive Reduction

Just like strong probabilistic bisimilarity, all weak probabilistic bisimilarity variants we consider embody the idea that the observable behaviour of a system is closed under convex combinations. Yet, this has to be interpreted for weak transitions. Finding a minimal set of generators turns out to be harder in this setting, as the behaviour of each state  $s$  no longer only depends on (convex combinations of) single step transitions leaving  $s$ . Instead, reachable distributions are now characterized by arbitrarily complex transition trees, and the convex combinations of the result distributions. The number of different transition trees – and also their induced distributions – with identical root is in general infinite. Yet, the convex set resulting from the convex combinations of the infinite number of induced transitions is known to be finitely generated [CS02]. However, in the worst case, the set of generators contains an exponential number of elements with respect to the number of transitions. In Section 11.2.2, in the case of strong probabilistic bisimilarity, we have used the set of generators to obtain a minimal representation of the automaton. With weak transitions, naively adopting this approach will not lead to the desired minimization due to the above mentioned possibly exponential blow-up. We therefore take inspiration from the standard approach followed in transitive reduction of order relations. Intuitively, this is the opposite of the transitive closure operation, and is achieved by removing transitions that can be reconstructed from other transitions by transitivity. We extend this idea by removing transitions as long as they can be reconstructed from other transitions by transitivity *and* convex combinations. In this spirit, we propose a simple algorithm that iteratively removes transitions, as long as their target distribution can also be reached by a weak combination of other transitions. Similar to transitive reduction on order relations, this reduction algorithm has polynomial complexity.

We will later show that this reduction leads to a minimal result with respect to  $\approx_{LTS}^{[D]}$  and  $\approx_{PA}^{[D]}$ , if applied on a model that a priori has been subjected to a quotient reduction. However, as we will see, for  $\approx$ , the case is more complicated, and this reduction alone is not effective to achieve minimality.



**Definition 11.9 (Convex-Transition Reduction Preorder).**

Given the MA  $\mathcal{A} = (S, \bar{s}, Act, \multimap, \multimap')$  and  $\mathcal{A}' = (S', \bar{s}', Act, \multimap', \multimap')$ , we write  $\mathcal{A} \subseteq_D \mathcal{A}'$  if and only if  $\multimap = \multimap'$ ,  $S = S'$ ,  $\bar{s} = \bar{s}'$ , and  $\multimap \subseteq \multimap'$  and for each transition  $(s, a, \mu) \in \multimap'$  there exists a weak combined transition  $s \xRightarrow{a}_c \mu$  in  $\mathcal{A}$ .  $\triangleleft$

**Definition 11.10 ( $\subseteq_D$ -Minimal Automata).** We call a MA  $\mathcal{A} \subseteq_D$ -minimal, if whenever  $\mathcal{A}' \subseteq_D \mathcal{A}$  for some MA  $\mathcal{A}'$ , then also  $\mathcal{A} \subseteq_D \mathcal{A}'$ .  $\triangleleft$

**Lemma 11.2 (Existence of  $\subseteq_D$ -Minimal Automata).** For every MA  $\mathcal{A}$  there exists a MA  $\mathcal{A}'$  such that  $\mathcal{A}' \approx_{\text{PA}} \mathcal{A}$  and  $\mathcal{A}'$  is  $\subseteq_D$ -minimal.

Note the use of  $\approx_{\text{PA}}$  in this lemma, instead of  $\approx_{\text{MA}}$  or  $\approx_{\delta}$ . In Section 8.7, we have established that on LTS,  $\approx_{\text{PA}}$  and  $\approx_{\text{LTS}}$  coincide. Hence, obviously this result immediately carries over to considerations for  $\approx_{\text{LTS}}$ . As we will discuss in Section 11.3.2, even for our considerations with respect to  $\approx_{\delta}$ , this lemma will suffice, as we will only apply the corresponding reduction after another reduction that will guarantee that  $\approx_{\delta}$  and  $\approx_{\text{PA}}$  also coincide on the reduced automaton.

**Definition 11.11 (Convex Transitive Reduction).** Let  $\mathcal{A}$  be a MA. We write  $\mathcal{A} \xrightarrow{T} \mathcal{A}'$  if  $\mathcal{A}' \subseteq_D \mathcal{A}$  and  $\mathcal{A}'$  is  $\subseteq_D$ -minimal.  $\triangleleft$

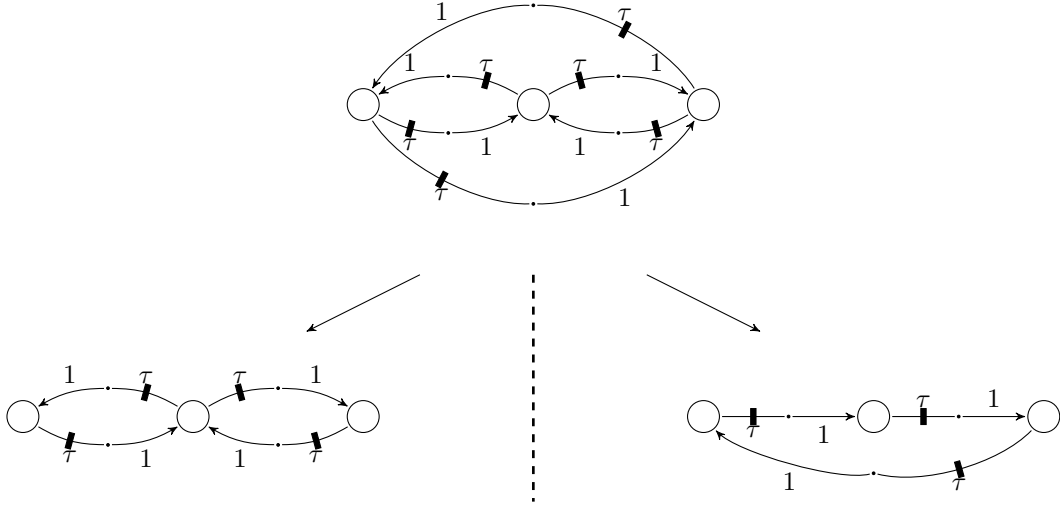
**Definition 11.12 (Stability-Preserving Convex Transitive Reduction).**  $\mathcal{A} \xrightarrow{T}_{\downarrow} \mathcal{A}''$  if and only if  $\mathcal{A} \xrightarrow{T} \mathcal{A}'$  and

- $S'' = S'$
- $\bar{s}'' = \bar{s}'$
- $\multimap'' = \multimap'$
- $\multimap'' = D' \cup T$  where  $(s, \tau, \delta(s)) \in T$  if and only if  $s \downarrow$  in  $\mathcal{A}'$ , but  $s \xrightarrow{\tau} \delta(s)$  in  $\mathcal{A}$ , but not  $s \xrightarrow{\tau} \mu$  for any  $\mu \neq \delta(s)$ .

 $\triangleleft$ 

As a special case, this reduction can be applied to non-probabilistic transition systems (LTS), where it then, in essence, coincides with transitive reduction of directed graphs [AGU72]. For this it is irrelevant that this reduction allows to combine transitions, as long as we work on a quotient reduced system, because in that system bisimilar states have been collapsed into a single representative. Thus, a Dirac transition to a single state can only be matched by a Dirac transition to precisely that state. In the LTS setting,  $\xrightarrow{T}$  preserves  $\approx_{\text{LTS}}$ , and in fact is a necessary step to arrive at the transition minimal quotient. Notably, different to all other reduction relations we have considered so far, this relation is non-deterministic, i.e., non-functional. This is the case already in a non-probabilistic setting.

**Example 11.2.** Consider the automaton on top of Figure 11.3. Depending on which transition we remove first, we obtain different minimal automata with respect to transitive reduction. The lower two automata are both minimal reductions of this automata, but with different number of transitions. This insight is by far not new. The automata are taken from [AGU72], with slight modification.  $\triangleleft$



**Figure 11.3.:** Different Ways of Reducing a Complete Graph with Three Nodes

Furthermore, the simplistic reduction strategy of removing one transition after the other, as long as the transitive closure of the resulting automaton is not changed, is not guaranteed to yield an automaton, that is *minimum* with respect to the number of edges, already in the non-probabilistic case.

**Example 11.3.** It is easy to see that the automaton on the left of Figure 11.3 cannot be further reduced by removing any transition. However, as the automaton on the right shows, we can find an alternative set of transitions which preserves reachability, but contains one transition less in total.  $\triangleleft$

This is especially interesting, as we aim not only on minimal representations, but on unique and minimum representations if they exist. As we will show in Section 11.3, this reduction is, in fact, unique up to isomorphism<sup>1</sup> (structural identity), *if* applied to a quotient reduced automaton. The overall result will therefore be unique up to isomorphism.

This may be surprising at first sight, given the insights from Example 11.3. The explanation why this phenomenon disappears in our case is that we do not apply reduction to arbitrary automata, but to the quotient with respect to  $\approx_{\text{LTS}}$  and  $\approx_{\text{PA}}$ . This ensures that there are no states on cycles of  $\tau$ -transitions, as all such states are bisimilar, and thus cannot exist in the quotient. Thus, in every cycle there must be at least one transition labelled by an observable transition. Thus, in structures, in which two cycles are interweaved, as in Example 11.3, two such transition must exist. For an example, say the transition from state ① to ② and the transition from state ② to ③ are both labelled with  $a$ . Now recall that the kind of transitivity that we are interested in always extends only to transitions labelled by  $\tau$ , and at most one transition labelled by an observable action. However, it does not include sequences of multiple observable transitions. Thus, it is no longer the case that state ② can reach itself by a weak, i.e. transitive, transition labelled by  $a$  in the automaton on the right hand side, as this would imply using the transition from ① to ②, which, is again labelled by  $a$ .

<sup>1</sup>see Definition 11.15 on page 194 for a formal definition

### 11.2.6. Rescaling

A particular fine point of weak probabilistic bisimilarities [BH97] is related to internal transitions that induce a nonzero chance of residing inside the class. If looking at the quotient, this concerns any internal transition  $(s, \tau, \mu)$  that contains the source state  $s$  with nontrivial probability, i.e.,  $0 < \mu(s) < 1$ . For those transitions, we can renormalize the probability of all other states in the support set by  $1 - \mu(s)$  without breaking weak bisimilarity. In other words, such  $\mu$  can be replaced by the rescaled distribution  $\mu \ominus s$ .

**Definition 11.13 (Rescaling).** Let  $\mathcal{A} = (S, \bar{s}, Act, \dashv\dashv, \dashv\dashv)$  be a MA. We write  $\mathcal{A} \xrightarrow{L} \mathcal{A}'$  if  $\mathcal{A}' = (S', \bar{s}', Act, \dashv\dashv', \dashv\dashv')$  such that

1.  $S = S'$ ,  $\bar{s} = \bar{s}'$  and  $\dashv\dashv = \dashv\dashv'$ , but
2. for each  $(s, a, \mu') \in \dashv\dashv'$ , either  $a \neq \tau$  and  $(s, a, \mu') \in \dashv\dashv$ , or  $a = \tau$  and there exists  $(s, \tau, \mu) \in \dashv\dashv$  such that  $\mu(s) < 1$  and  $\mu' = \mu \ominus s$ .

&lt;

As it will turn out, this reduction is the final step to obtain minimality with respect to  $\approx_{PA}^{[D]}$  if applied a posteriori to the other two reductions,  $\approx_{PA}$  and  $\xrightarrow{T}$ . Figure 11.1(c) depicts the result of applying this sequence of reductions on the MA in Figure 11.1(a). Figure 11.1(b) shows the automaton after it has been subjected to quotient reduction only.

### 11.2.7. Redundant State Elimination

It is folklore knowledge, that the quotient of an automaton with respect to a bisimulation equivalence is minimum with respect to the number of states. We also make repeated use of this fact in this chapter. However, this all is only true as long as the equivalence relation is defined over states, and not over distributions. The quotient with respect to  $\approx_\delta$ , the state equivalence within weak distribution bisimilarity, does not satisfy this property in general: in the quotient, states may exist that are still removable without affecting the behaviour of the automaton.

**Example 11.4.** The states of the automaton on the right of Figure 11.4 are all distinguishable with respect to  $\approx_\delta$  (assuming all non-round states of the same shape to represent a single state, respectively). For instance, states  $t$  and  $u$  can reach  $\blacksquare$  and  $\blacklozenge$  with different probabilities. While the total probability of reaching  $\blacklozenge$  for  $t$  is  $\frac{1}{3}$ , for  $u$  it is  $\frac{2}{3}$ . The fact that all states are distinguishable with respect to  $\approx_\delta$  implies that the automaton is already isomorphic to its quotient automaton. However, states  $t$  and  $u$  are still removable without changing the overall behaviour of the automaton: the automaton on the left is equivalent to the automaton on the right (or more precisely, its initial states are) with respect to  $\approx_\delta$ , but pruned of states  $t$  and  $u$ .

&lt;

The reason behind this phenomenon is that  $\approx$  has been designed to fuse distributions along deterministic sequences of internal transitions, as long as their overall probabilistic behaviour is not changed, even though the intermediate states along the fused sequence may be behaviourally distinct. In Section 8.4, we have called these intermediate states *redundant*, as their probabilistic behaviour is fully preserved by one of their internal successor *distributions* (not states!), and

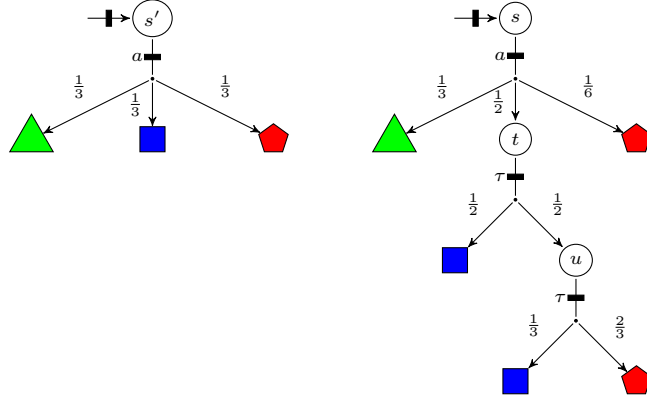


Figure 11.4.: Two Equivalent Automata with Different State Sets

they are thus behaviourally redundant. Therefore, redundant states can always be removed without affecting behaviour, as long as their incoming transitions are forwarded accordingly. Since for each distribution, there is potentially more than one behaviourally equivalent successor distribution that is not redundant, in our reduction, we need to make some choice among them. Therefore, the reduction relation we define will be non-deterministic, depending on which of these distributions we choose. For a given possibly redundant distribution  $\gamma$ , each distribution  $\mu$  satisfying

$$\gamma \xRightarrow{P} \mu$$

is a suitable candidate, when  $P$  denotes the set of preserving transitions with respect to  $\approx_\delta$ .

**Definition 11.14 (Redundant State Reduction).** Let  $\mathcal{A} = (S, \bar{s}, Act, \dashv\!\!\!\rightarrow, \dashv\!\!\!\rightarrow')$  be a weakly image-finite MA. Let  $P$  be the set of preserving transitions with respect to  $\approx_\delta$ . Whenever we write  $\mathcal{A} \xrightarrow{R} \mathcal{A}'$  then  $\mathcal{A}' = (S', \bar{s}', Act, \dashv\!\!\!\rightarrow', \dashv\!\!\!\rightarrow')$  is some Markov automaton satisfying

1.  $S' = \{s \in S \mid s \text{ is not redundant}\}$
2.  $\bar{s}' = \begin{cases} \bar{s} & \text{if } \bar{s} \text{ is not redundant} \\ t' & \text{otherwise, for some fresh } t' \end{cases}$
3.  $\dashv\!\!\!\rightarrow$  and  $\dashv\!\!\!\rightarrow'$  are some sets of minimal size containing
  - a transition  $\bar{s}' \dashv\!\!\!\rightarrow \mu$  for some  $\mu$  satisfying  $\mu_0 \xRightarrow{P} \mu$ , if and only if  $\bar{s}' = t'$ , i.e.  $\bar{s}$  is redundant.
  - for every weak hypertransition  $s \xRightarrow{a} \mu'$  of  $\mathcal{A}$ , a transition  $s \dashv\!\!\!\rightarrow' \mu$  for some  $\mu$  satisfying  $\mu' \xRightarrow{P} \mu$ .
  - for every transition  $s \xrightarrow{\chi(\lambda)} \mu'$  of  $\mathcal{A}$ , a transition  $s \xrightarrow{\chi(\lambda)'} \mu$  for some  $\mu$  satisfying  $\mu' \xRightarrow{P} \mu$ .

◁

Note that in the last line of this definition we use the special notation  $\xrightarrow{\chi(\lambda)}$  to implicitly define the transition relation  $\rightarrow$ . As we have discussed in Chapter 5, it is actually not possible to uniquely derive  $\rightarrow$  from this notation. Therefore, we have given a standard interpretation in Chapter 5 (see Remark 5.1). While this structurally may change the branching structure of  $\rightarrow$  in a way that is not obvious from the intention of this reduction, – it may for example replace two transitions  $s \xrightarrow{\lambda} t$  by one transition  $s \xrightarrow{2\lambda} t$  – it has no semantic consequences for our purpose of minimizing the automaton while preserving  $\approx$ . In fact, in the worst case we anticipate the reductions that will be done by  $\xrightarrow{\Sigma}$  in a later step during the minimization process anyway.

The fact that this reduction relation is non-deterministic in general is not relevant in our intended practical application, where we only apply this reduction to automata that are minimal with respect to  $\approx_\delta$  already.

**Lemma 11.3.** If  $\mathcal{A} = (S, \bar{s}, Act, \rightarrow, \rightarrow)$  is weakly image-finite, and a quotient automaton with respect to  $\approx_\delta$ , i.e.  $\mathcal{A} \xrightarrow{\approx_\delta} \mathcal{A}$ , then the redundant state reduction is deterministic, i.e.  $\mathcal{A} \xrightarrow{R} \mathcal{A}_1$  and  $\mathcal{A} \xrightarrow{R} \mathcal{A}_2$  implies  $\mathcal{A}_1 = \mathcal{A}_2$ .

*Proof.* Follows immediately from Lemma D.16, and the fact that  $\approx_s = \approx_\delta$  (Theorem 8.14).  $\square$

This lemma shows that the reduction becomes deterministic in the practically relevant case. Surprisingly, however, the reduction may increase the number of transitions, as a transition is inserted in the reduction whenever a weak transition has existed in the input automaton. Fortunately, this increase can be undone by an subsequent application of  $\xrightarrow{C}$ . When we discuss the algorithmic complexity of the reduction, we will see that no increase in the number of transitions is necessary with a clever implementation.

### 11.2.8. Properties of Reductions

We summarize preservation and computability properties of the reduction relations.

**Lemma 11.4** (*Preservation of Bisimilarities*).

1.  $\mathcal{A} \xrightarrow{\sim} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \mathbf{MA}$  and  $\asymp \in \mathbb{B}$ .
2.  $\mathcal{A} \xrightarrow{C} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \mathbf{MA}$  and  $\asymp \in \{\sim_{PA}, \sim_{MA}\}$ .
3.  $\mathcal{A} \xrightarrow{T} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}'$  and  $\asymp \in \{\approx_{LTS}, \approx_{PA}\}$ .
4.  $\mathcal{A} \xrightarrow{T}_\downarrow \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}'$  and  $\asymp \in \{\approx_{MC}, \approx_{MA}, \approx_\delta\}$ .
5.  $\mathcal{A} \xrightarrow{L} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \mathbf{MA}$  and  $\asymp \in \{\approx_{PA}, \approx_{MA}, \approx_\delta\}$ .
6.  $\mathcal{A} \xrightarrow{R} \mathcal{A}'$  implies  $\mathcal{A} \approx_\delta \mathcal{A}'$  for weakly image-finite  $\mathbf{MA}$ .
7.  $\mathcal{A} \xrightarrow{M} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \mathbf{MA}$  and  $\asymp \in \{\sim_{MC}, \approx_{MC}, \approx_{MA}, \approx_\delta\}$ .

8.  $\mathcal{A} \xrightarrow{\Sigma} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \{\sim_{\text{IMC}}, \approx_{\text{IMC}}, \approx_{\text{MA}}, \approx_{\delta}\}$ .

The proof of this lemma can be found in Appendix F.1.

**Lemma 11.5 (Computability of Reductions).** For every finite MA  $\mathcal{A}$ , a MA  $\mathcal{A}'$  with  $\mathcal{A} \rightsquigarrow \mathcal{A}'$  can be computed in

- *polynomial time* if  $\rightsquigarrow = \widetilde{\rightsquigarrow}$  and  $\asymp \in \{\sim_{\text{LTS}}, \approx_{\text{LTS}}, \sim_{\text{PA}}, \approx_{\text{PA}}, \sim_{\text{MA}}, \approx_{\text{MA}}\}$
- *polynomial time* if  $\rightsquigarrow \in \{\rightsquigarrow^C, \rightsquigarrow^T, \rightsquigarrow^{\downarrow}, \rightsquigarrow^L, \rightsquigarrow^M, \rightsquigarrow^{\Sigma}\}$ ,
- *exponential time* if  $\rightsquigarrow = \widetilde{\rightsquigarrow}_{\delta} \circ \rightsquigarrow^R$ .

The proof of this lemma can be found in Appendix F.2.

In the last item, we again make use of the fact that  $\approx_{\delta}$  and  $\approx_s$  coincide for finite automata. Both the reduction  $\widetilde{\rightsquigarrow}_{\delta}$  and  $\rightsquigarrow^R$  have exponential complexity. However, computing  $\rightsquigarrow^R$  comes only with polynomial cost once  $\widetilde{\rightsquigarrow}_{\delta}$  has been computed, as identifying redundant states is a side-product of the decision algorithm, Algorithm 10.1, which is already used to compute the quotient. As we will use these reductions only in this subsequent manner, we do not consider their complexity in isolation.

The results suggest that constructing a minimal representative of an equivalence class of automata with respect to all LTS, IMC and PA bisimilarities is possible in polynomial time. Only weak distribution bisimilarity has resisted a polynomial time treatment so far. In the following section, we will discuss these insights in greater detail.

## 11.3. Normal Forms

We are now concerned with minimality and uniqueness properties induced by the reduction operations with respect to the metrics discussed. To discuss uniqueness, it is convenient to introduce normal forms as means to canonically represent automata in such a way that two automata are equivalent if and only if they have identical normal forms. Or better, if and only if the normal forms are identical up to isomorphism (structural identity).

*Remark 11.1.* Throughout this section, we assume all MA to be finite.

**Definition 11.15 (Isomorphism).** Two MA  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv\!\!\rightarrow, \dashv\!\!\rightarrow')$  and  $\mathcal{A}' = (S', \bar{s}', \text{Act}, \dashv\!\!\rightarrow', \dashv\!\!\rightarrow')$  are called *isomorphic*, denoted by  $\mathcal{A} =_{\text{iso}} \mathcal{A}'$ , if their uniform representations satisfy  $\text{Act}^X = \text{Act}^{X'}$  and if there is a bijective mapping  $b: S \rightarrow S'$  such that  $b(\bar{s}) = \bar{s}'$  and  $(s, a, \mu) \in \dashv\!\!\rightarrow$  if and only if  $(b(s), a, b(\mu)) \in \dashv\!\!\rightarrow'$ .  $\triangleleft$

**Definition 11.16 (Normal Form).** Given an equivalence relation  $\asymp$  over MA, we call  $NF: \text{MA} \rightarrow \text{MA}$  a *normal form*, if it satisfies for every MA  $\mathcal{A}$

- $NF(\mathcal{A}) \asymp \mathcal{A}$ , and
- for every MA  $\mathcal{A}'$  it holds that  $\mathcal{A} \asymp \mathcal{A}'$  if and only if  $NF(\mathcal{A}) =_{\text{iso}} NF(\mathcal{A}')$ .

$\triangleleft$

It is natural to strive for normal forms that are distinguished in a certain sense. Not surprisingly, we will strive for normal forms that are distinguished as being the smallest possible representation of the behaviour they represent.

It will turn out that for LTS, PA, and IMC bisimilarities we can find a normal form that is optimal with respect to all three metrics, i.e. number of states, number of transitions and transition fanout. In contrast, for weak distribution bisimilarity, a similar result is impossible. Due to this fundamental difference, we will work out the results for MA and the other models separately.

### 11.3.1. LTS, PA, IMC Bisimilarities

In the following, when we do not explicitly restrict the equivalences over which the symbol  $\asymp$  ranges, we assume the set  $\{\sim_{\text{LTS}}, \approx_{\text{LTS}}, \sim_{\text{PA}}, \approx_{\text{PA}}, \sim_{\text{IMC}}, \approx_{\text{IMC}}\}$ . We call a total and functional subset of a binary relation  $r \subseteq \text{MA} \times \text{MA}$  a *function in  $r$* . Note that every function in  $r$  is a mapping  $\text{MA} \rightarrow \text{MA}$ .

**Definition 11.17 (Normal Form Instances).**

- Let  $NF_{\sim_{\text{LTS}}} = \sim_{\text{LTS}}^{\text{TS}}$ .
- Let  $NF_{\approx_{\text{LTS}}}$  be an arbitrary function in  $\approx_{\text{LTS}}^{\text{TS}} \circ \sim_{\text{LTS}}^{\text{TS}}$ .
- Let  $NF_{\sim_{\text{PA}}} = \sim_{\text{PA}}^{\text{PA}} \circ \sim_{\text{PA}}^{\text{C}}$ .
- Let  $NF_{\approx_{\text{PA}}}$  be an arbitrary function in  $\approx_{\text{PA}}^{\text{PA}} \circ \sim_{\text{PA}}^{\text{TS}} \circ \sim_{\text{PA}}^{\text{L}}$ ,
- Let  $NF_{\sim_{\text{IMC}}} = \sim_{\text{IMC}}^{\text{MC}} \circ \sim_{\text{IMC}}^{\text{M}} \circ \sim_{\text{IMC}}^{\Sigma}$ .
- Let  $NF_{\approx_{\text{IMC}}}$  be an arbitrary function in  $\approx_{\text{IMC}}^{\text{MC}} \circ \sim_{\text{IMC}}^{\text{TS}} \circ \sim_{\text{IMC}}^{\text{L}} \circ \sim_{\text{IMC}}^{\text{M}} \circ \sim_{\text{IMC}}^{\Sigma}$ .

◁

The next theorem is the main result of this section. It states that the automata we obtain by the reductions are normal forms and furthermore, that they are optimally reduced with respect to the three metrics. We want to emphasize that they are not only minimal with respect to these metrics, but that even no automaton exists that is smaller with respect to *any* of the three metrics. Thus, our reductions are optimal in all respects.

**Theorem 11.1.** Let  $\asymp \in \mathbb{B}$ .

1. **Minimality:**  $NF_{\asymp}(\mathcal{A})$  is  $\preceq^{|\mathcal{S}|}$ ,  $\preceq^{|\mathcal{D}|}$ , and  $\preceq^{|\mathcal{D}|}$ -minimal for every  $\mathcal{A} \in \text{MA}$ .
2. **Uniqueness of minimals:** If  $\mathcal{A}$  and  $\mathcal{A}'$  are  $\preceq^{|\mathcal{S}|}$ ,  $\preceq^{|\mathcal{D}|}$ , and  $\preceq^{|\mathcal{D}|}$ -minimal automata, and  $\mathcal{A} \asymp \mathcal{A}'$ , then also  $\mathcal{A} =_{\text{iso}} \mathcal{A}'$ ,
3. **Normal forms:**  $NF_{\asymp}$  is uniquely defined up to  $=_{\text{iso}}$ , and is a normal form.

It is straightforward to verify that all normal forms  $NF_{\asymp}$  above are indeed mappings. Furthermore, by Lemma 11.4, it follows that in each of the cases  $NF_{\asymp}(\mathcal{A}) \asymp \mathcal{A}$ .

The remainder of this section is devoted to the proof of Theorem 11.1. We begin with a lemma that we use often.

**Lemma 11.6 (Preservation of Minimality).** Let  $\preceq \in \{\preceq^{[S]}, \preceq^{[D]}, \preceq^{[D]}, \subseteq_D\}$ . If  $\mathcal{A} =_{iso} \mathcal{A}'$  and  $\mathcal{A}$  is  $\preceq$ -minimal, then  $\mathcal{A}'$  is  $\preceq$ -minimal, too.

For each normal form, the proof will refer to the following crucial, but already folklore insight, that the quotient automaton is minimal with respect to the number of states. Note that this is not true for  $\approx_\delta$ , which we will consider in the next section.

**Lemma 11.7 (State Minimality of Quotient Automata).** For every  $\mathcal{A} \in \text{MA}$ , the automaton  $\mathcal{A}'$  with  $\mathcal{A} \rightsquigarrow \mathcal{A}'$  is  $\preceq^{[S]}$ -minimal.

Next, we show that  $\preceq^{[S]}$  and  $\preceq^{[D]}$ -minimality can be achieved at the same time in one automaton. For bisimilarities on LTS, this is enough to conclude also  $\preceq^{[D]}$ -minimality, as this always coincides with  $\preceq^{[D]}$ -minimality here (as all transition have the form  $(s, a, \delta(t))$ ).

**Lemma 11.8 (Compatibility of  $\preceq^{[S]}$  and  $\preceq^{[D]}$ -minimality).** For every MA  $\mathcal{A}$  there exists a MA  $\mathcal{A}'$  with  $\mathcal{A}' \asymp \mathcal{A}$ , which is  $\preceq^{[S]}$  and  $\preceq^{[D]}$ -minimal.

*Proof.* By Lemma 11.1, there exists a MA  $\mathcal{A}$  that is  $\preceq^{[D]}$ -minimal. Consider  $[\mathcal{A}]_\asymp$ . From Definition 11.4 it is clear that for every transition of  $[\mathcal{A}]_\asymp$  there exists a transition in  $\mathcal{A}$ . Thus,  $[\mathcal{A}]_\asymp \preceq^{[D]} \mathcal{A}$ , and hence,  $[\mathcal{A}]_\asymp$  must also be  $\preceq^{[D]}$ -minimal. Furthermore, by Lemma 11.7,  $[\mathcal{A}]_\asymp$  must also be  $\preceq^{[S]}$ -minimal, and finally with Lemma 11.4  $\mathcal{A} \asymp \mathcal{A}'$  follows.  $\square$

## Strong Bisimilarities

**Lemma 11.9 (Canonicity of Normal Form).** Let  $\asymp \in \{\sim_{\text{LTS}}, \sim_{\text{PA}}, \sim_{\text{MC}}\}$ ,  $\mathcal{A} \in \text{MA}$ , and  $\mathcal{A}' = \text{NF}_\asymp(\mathcal{A})$ . For every  $\preceq^{[S]}$  and  $\preceq^{[D]}$ -minimal MA  $\mathcal{A}_m$  with  $\mathcal{A}_m \asymp \mathcal{A}$ , also  $\mathcal{A}_m =_{iso} \mathcal{A}'$ .

*Proof.* We skip the proof for  $\asymp = \sim_{\text{LTS}}$  and proceed with the more complicated case of  $\asymp = \sim_{\text{PA}}$ . Let  $\mathcal{A}_m$  be a  $\preceq^{[S]}$  and  $\preceq^{[D]}$ -minimal automaton. Recall that  $\text{NF}_{\sim_{\text{PA}}} = \sim_{\sim_{\text{PA}}}^{\text{PA}} \circ \overset{C}{\sim}$ . As applying  $\sim_{\sim_{\text{PA}}}^{\text{PA}}$  to  $\mathcal{A}$  leads to a  $\preceq_{\text{PA}}^{[S]}$ -minimal automaton according to Lemma 11.7, and  $\overset{C}{\sim}$  obviously does not alter the number of states,  $\text{NF}_{\sim_{\text{PA}}}(\mathcal{A}) = \mathcal{A}'$  is  $\preceq_{\text{PA}}^{[S]}$ -minimal, and thus  $|S_m| = |S'|$ , as  $\mathcal{A}_m$  is  $\preceq_{\text{PA}}^{[S]}$ -minimal by assumption.

Since  $\mathcal{A}' \sim_{\text{PA}} \mathcal{A}$  and  $\mathcal{A} \sim_{\text{PA}} \mathcal{A}_m$ , we have  $\mathcal{A}' \sim_{\text{PA}} \mathcal{A}_m$ . We will now argue that  $b = \sim_{\text{PA}} \cap (S' \times S_m)$  is in fact a suitable mapping to establish  $\mathcal{A}' =_{iso} \mathcal{A}_m$ . We start by showing that  $b$  is functional, injective and surjective. Assume  $b$  is not injective. Then there must exist states  $s_1, s_2 \in S'$  and  $t \in S_m$ , such that  $b(s_1) = t$  and  $b(s_2) = t$ . But this implies  $s_1 \sim_{\text{PA}} t$  and  $s_2 \sim_{\text{PA}} t$ . By transitivity, this implies  $s_1 \sim_{\text{PA}} s_2$ , contradicting Lemma 11.7. Functionality can be shown similarly. We skip the details. If  $b$  is not surjective, this would immediately contradict the assumption that  $\mathcal{A}_m$  is  $\preceq_{\text{PA}}^{[S]}$ -minimal, since then any state  $t \in \mathcal{A}_m$  for which no  $s \in S'$  exists, such that  $b(s) = t$  could be removed without violating  $\mathcal{A}' \sim_{\text{PA}} \mathcal{A}_m$ .

Most of the other conditions that have to be checked to show that  $b$  is an isomorphism are straightforward, except for the condition

$$(s, a, \mu) \in \multimap \quad \text{if and only if} \quad (b(s), a, b(\mu)) \in \multimap'. \quad (\star)$$

The set of combined transitions any state  $s$  of  $\mathcal{A}'$  can perform, must equal the set of combined transitions that  $b(s)$  can do as  $s \sim_{\text{PA}} b(s)$ . By reduction  $\overset{C}{\sim}$ , the set of transitions leaving  $s$



must be minimal, according to Lemma 2.5, and must also be unique. As the transitions of  $b(s)$  are minimal by assumption, the uniqueness of the minimal set of generators guarantees Condition  $(\star)$ .

For  $\sim_{\text{LTS}}$ , recall that  $NF_{\sim_{\text{MC}}} = \sim_{\text{MC}}^{\sim} \circ \overset{M}{\sim} \circ \overset{\Sigma}{\sim}$ . Note that the reductions  $\overset{M}{\sim}$  and  $\overset{\Sigma}{\sim}$  do not alter the number of states. Hence, with the same arguments as for  $\sim_{\text{PA}}$  we can establish that  $b = \sim_{\text{MC}} \cap (S' \times S_m)$  is a suitable mapping to establish  $\mathcal{A}' =_{\text{iso}} \mathcal{A}_m$ , except for the conditions that

$$\begin{aligned} (s, a, \mu) \in \dashv\!\rightarrow & \quad \text{if and only if} & \quad (b(s), a, b(\mu)) \in \dashv\!\rightarrow' \text{ and} \\ (s, \lambda, t) \in \dashv\!\rightarrow & \quad \text{if and only if} & \quad (b(s), \lambda, b(\mu)) \in \dashv\!\rightarrow'. \end{aligned}$$

The definition of  $\sim_{\text{MC}}$  guarantees the coincidence of transitions from  $\dashv\!\rightarrow$  analogously to the case of  $\sim_{\text{PA}}$ . Concerning  $\dashv\!\rightarrow$ , we first note that from the definition of bisimulation, we can only infer that the

1. the total exit rate of a state agrees,
2. that the state agree with respect to stability, and
3. that the reached distribution agree.

This is expressed by the condition  $s \xrightarrow{\chi(r)} \mu$  in the bisimulation. However, this notation does neither guarantee that no transition relation exist at unstable states, nor that the state-wise delay transitions coincide. For  $\mathcal{A}_m$ , it is guaranteed that no such transition exists at unstable state, and furthermore, that for each state  $t \in \text{Supp}(\mu)$ , there is exactly one transition in  $\dashv\!\rightarrow'$  with  $(s, \mu(s) \cdot r, t)$ . For  $NF_{\sim_{\text{MC}}}$  this is, however, guaranteed by the two reductions  $\overset{M}{\sim}$  and  $\overset{\Sigma}{\sim}$ , respectively.  $\square$

For  $\sim_{\text{LTS}}$ ,  $\sim_{\text{PA}}$  and  $\sim_{\text{MC}}$  Theorem 11.1 now follows almost immediately by Lemma 11.8, Lemma 11.9 and Lemma 11.4, with the following respective further observations: for  $\sim_{\text{LTS}}$ , we, in addition, need the observation that  $\mathcal{A}$  is  $\preceq^{|\mathcal{D}|}$ -minimal if and only if it is  $\preceq^{|\mathcal{D}|}$ -minimal, as we remarked before Lemma 11.8. The same holds for  $\sim_{\text{MC}}$ , when we restrict our attention to transitions from  $\dashv\!\rightarrow$ . For transitions in  $\dashv\!\rightarrow$ , reduction  $\overset{\Sigma}{\sim}$  guarantees a minimal fanout. For  $\sim_{\text{PA}}$ , the same observation holds, but follows from the uniqueness of the minimal set of generators (Lemma 2.5).

## Weak Bisimilarities

The following two lemmas are the weak counterparts to Lemma 11.9.

**Lemma 11.10.** Let  $\mathcal{A}$  be a LTS and  $\mathcal{A}' = NF_{\sim_{\text{LTS}}}(\mathcal{A})$ . Let  $\mathcal{A}_m$  be a  $\preceq_{\text{LTS}}^{|\mathcal{S}|}$  and  $\preceq_{\text{LTS}}^{|\mathcal{D}|}$ -minimal LTS satisfying  $\mathcal{A}_m \approx_{\text{LTS}} \mathcal{A}$ . Then  $\mathcal{A}' =_{\text{iso}} \mathcal{A}_m$ .

We skip the proof of this lemma, as it is similar to, but simpler than the proof of the according lemma for probabilistic automata, Lemma 11.11. Theorem 11.1 can then be proven in complete analogy to the proof for  $\sim_{\text{LTS}}$ .

We now turn to the proof of Theorem 11.1 and the necessary lemmas for PA and weak probabilistic bisimilarity. It is the most involved case in this respect. It is instructive to note that in

the following lemma, we need to apply the reduction  $\xrightarrow{L}$  to arrive at an uniqueness result. Only applying  $\xrightarrow{\approx_{PA}^{[S]}}$  followed by  $\xrightarrow{T}$  will still lead to  $\xrightarrow{\approx_{PA}^{[S]}}$  and  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal automata, but they will not agree up to  $=_{iso}$ , in full generality. Relative to Lemma 11.10 and 11.9, the following lemma is slightly more general.

**Lemma 11.11.** Let  $\mathcal{A}$  be a  $\xrightarrow{\approx_{PA}^{[S]}}$ -minimal MA,  $\mathcal{A} \xrightarrow{T} \circ \xrightarrow{L} \mathcal{A}'$ , and  $\mathcal{A}'_m$  be a  $\xrightarrow{\approx_{PA}^{[S]}}$  and  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal MA satisfying  $\mathcal{A}'_m \approx_{PA} \mathcal{A}$ . Now let  $\mathcal{A}'_m \xrightarrow{L} \mathcal{A}_m$  for some  $\mathcal{A}_m$ . Then  $\mathcal{A}' =_{iso} \mathcal{A}_m$ .

The proof of this lemma can be found in Appendix F.3.

**Corollary 11.1.** Let  $\mathcal{A}$  be a  $\xrightarrow{\approx_{PA}^{[S]}}$ -minimal MA.

$\mathcal{A}$  is  $\subseteq_D$ -minimal if and only if it is  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal.

*Proof.* Let  $\mathcal{A}$  be  $\xrightarrow{\approx_{PA}^{[S]}}$ -minimal. For the first direction of the *if and only if*, note first that by Lemma 11.8, a MA  $\mathcal{A}'_m$  must exist, which is minimal with respect to  $\xrightarrow{\approx_{PA}^{[D]}}$  and  $\xrightarrow{\approx_{PA}^{[S]}}$ . Let  $\mathcal{A}'_m \xrightarrow{L} \mathcal{A}_m$ . Clearly,  $\mathcal{A}_m$  must be  $\xrightarrow{\approx_{PA}^{[S]}}$  and  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal, too. As by assumption,  $\mathcal{A}$  is  $\subseteq_D$ -minimal,  $\mathcal{A} \xrightarrow{T} \mathcal{A}$ . Let  $\mathcal{A}'$  satisfy  $\mathcal{A} \xrightarrow{L} \mathcal{A}'$ . We combine the two reductions and see that  $\mathcal{A} \xrightarrow{T} \circ \xrightarrow{L} \mathcal{A}'$ . This allows us to apply Lemma 11.11 to obtain  $\mathcal{A}' =_{iso} \mathcal{A}_m$ . As  $\mathcal{A}' =_{iso} \mathcal{A}_m$  implies that both have the same number of transitions, also  $\mathcal{A}'$  must be  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal. If we can now show that also  $\mathcal{A}$  and  $\mathcal{A}'$  have the same number of transitions, we are done. Assume the contrary to arrive at a contradiction. As  $\mathcal{A} \xrightarrow{L} \mathcal{A}'$ , this is only possible if there are two transitions  $(s, \tau, \mu)$  and  $(s, \tau, \gamma)$  in  $\mathcal{A}$  such that  $\mu \ominus s = \gamma \ominus s$ . But then, one of them could have been removed without changing the combined weak transitions  $s$  can perform, contradicting the assumption that  $\mathcal{A}$  is  $\subseteq_D$ -minimal.

For the other direction, assume  $\mathcal{A}$  is in addition  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal. As removing transitions from  $\mathcal{A}$  would lead to an automaton that is smaller with respect to  $\xrightarrow{\approx_{PA}^{[D]}}$ , it must be the case that any such automaton  $\mathcal{A}'$  does not satisfy  $\mathcal{A}' \approx_{PA} \mathcal{A}$ , otherwise contradicting the assumption that  $\mathcal{A}$  was  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal. But then it immediately follows that  $\mathcal{A}$  is also  $\subseteq_D$ -minimal.  $\square$

**Lemma 11.12.** If  $\mathcal{A}$  is  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal, then there also exists  $\mathcal{A}'$ , such that  $\mathcal{A} \approx_{PA} \mathcal{A}'$  and  $\mathcal{A}'$  is  $\xrightarrow{\approx_{PA}^{[S]}}$ ,  $\xrightarrow{\approx_{PA}^{[D]}}$ , and  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal.

*Proof.* We first show that for every  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal automaton  $\mathcal{A}$  there is one that is also  $\xrightarrow{\approx_{PA}^{[S]}}$ -minimal. As candidate, we take the unique automaton  $\mathcal{A}'$  such that  $\mathcal{A} \xrightarrow{\approx_{PA}} \mathcal{A}'$ . From Definition 11.4 and 11.5 it is clear that the transitions of  $\mathcal{A}'$  can be surjectively mapped to transitions of  $\mathcal{A}$ , such that every transition of  $\mathcal{A}'$  is smaller or equal with respect to  $\|\cdot\|$  than its image transition in  $\mathcal{A}$ . Thus, minimality with respect to  $\xrightarrow{\approx_{PA}^{[D]}}$  is preserved.

Now we show that any  $\mathcal{A}''$ , which satisfies  $\mathcal{A}' \xrightarrow{T} \mathcal{A}''$  is in addition  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal. Clearly, the numbers of states of  $\mathcal{A}'$  and  $\mathcal{A}''$  are the same. Furthermore, the transitions of  $\mathcal{A}''$  form a subset of the transitions of  $\mathcal{A}'$ . Thus, as  $\mathcal{A}'$  is  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal, also  $\mathcal{A}''$  must be  $\xrightarrow{\approx_{PA}^{[D]}}$ -minimal. By Definition 11.11,  $\mathcal{A}''$  is minimal with respect to  $\subseteq_D$ , and thus, by Corollary 11.1, also with respect to  $\xrightarrow{\approx_{PA}^{[D]}}$ .  $\square$

**Corollary 11.2.** *For every MA  $\mathcal{A}$  there exists a MA  $\mathcal{A}'$  with  $\mathcal{A}' \approx_{\text{PA}} \mathcal{A}$ , which is  $\approx_{\text{PA}}^{|S|}$ ,  $\approx_{\text{PA}}^{|D|}$  and  $\approx_{\text{PA}}^{|D|}$ -minimal.*

*Proof.* Follows immediately from Lemma 11.1 and Lemma 11.12.  $\square$

**Lemma 11.13 (Canonicity of Normal Form).** Let  $\mathcal{A}' = NF_{\approx_{\text{PA}}}(\mathcal{A})$ . Let  $\mathcal{A}_m$  be a  $\approx_{\text{PA}}^{|S|}$ ,  $\approx_{\text{PA}}^{|D|}$ , and  $\approx_{\text{PA}}^{|D|}$ -minimal automaton satisfying  $\mathcal{A}_m \approx_{\text{PA}} \mathcal{A}$ . Then  $\mathcal{A}' =_{\text{iso}} \mathcal{A}_m$ .

*Proof.* By Corollary 11.2 we know that  $\mathcal{A}_m$  exists such that  $\mathcal{A}_m \approx_{\text{PA}} \mathcal{A}$  and  $\mathcal{A}_m$  is  $\approx_{\text{PA}}^{|S|}$ ,  $\approx_{\text{PA}}^{|D|}$  and  $\approx_{\text{PA}}^{|D|}$ -minimal. Furthermore, as  $\mathcal{A}_m$  is  $\approx_{\text{PA}}^{|D|}$ -minimal, it must hold  $\mathcal{A}_m \xrightarrow{L} \mathcal{A}_m$ . Finally, as  $\mathcal{A}' = NF_{\approx_{\text{PA}}}(\mathcal{A})$ , there must exist  $\mathcal{A}''$  such that  $\mathcal{A} \xrightarrow{\approx_{\text{PA}}} \mathcal{A}''$  and  $\mathcal{A}'' \xrightarrow{T} \mathcal{A}'$ , and by the Definition of  $\approx_{\text{PA}}$  and Lemma 11.7,  $\mathcal{A}''$  is  $\approx_{\text{PA}}^{|S|}$ -minimal. Thus, we may apply Lemma 11.11 to obtain our result.  $\square$

Theorem 11.1 now follows for  $\approx_{\text{PA}}$  with Corollary 11.2 and Lemma 11.13.

We now finally turn to the respective proof of this lemma for IMC and weak IMC bisimilarity. IMC and LTS only differ in the presence of timed transitions and preservation of stability in the bisimilarity. Thus, most results follow completely analogously. We will state the crucial lemma for this proof, which corresponds to Lemma 11.10 and 11.11 for LTS and PA, respectively. Similar as in the PA case, uniqueness of the normal form does not immediately follow by the two initial reduction steps, quotient reduction and transitive reduction. Instead, further reductions are needed. For PA, this reduction was a kind of normalization of the distributions internal transitions  $\xrightarrow{\tau}$  lead to. The corresponding reduction is  $\xrightarrow{L}$ . In IMC, immediate transitions  $\xrightarrow{\rightarrow}$  do not lead to distributions, but to single states, exactly as for LTS. Therefore, no further reduction is needed here. However, IMC are in addition equipped with timed transitions  $\xrightarrow{\rightarrow}$ . For them, two more reduction steps are necessary to arrive at a minimal and unique system:  $\xrightarrow{M} \circ \xrightarrow{\Sigma}$ .

**Lemma 11.14.** Let  $\mathcal{A}$  be an IMC and  $\mathcal{A}' = NF_{\approx_{\text{IMC}}}(\mathcal{A})$ . Let  $\mathcal{A}'_m$  be a  $\approx_{\text{IMC}}^{|S|}$  and  $\approx_{\text{IMC}}^{|D|}$ -minimal IMC satisfying  $\mathcal{A}_m \approx_{\text{IMC}} \mathcal{A}$ . Let  $\mathcal{A}'_m \xrightarrow{M} \circ \xrightarrow{\Sigma} \mathcal{A}_m$ . Then  $\mathcal{A}' =_{\text{iso}} \mathcal{A}_m$ .

*Proof Sketch.* Recall that  $NF_{\approx_{\text{IMC}}}$  can be any function in  $\approx_{\text{IMC}} \circ \xrightarrow{T}_{\downarrow} \circ \xrightarrow{L} \circ \xrightarrow{M} \circ \xrightarrow{\Sigma}$ . Similarly to the proof of Lemma 11.11, we can establish that the first two reductions  $\approx_{\text{IMC}} \circ \xrightarrow{T}_{\downarrow}$  lead to a state minimal system. The only, but important difference is that we must use  $\xrightarrow{T}_{\downarrow}$  instead of  $\xrightarrow{T}$  in order to preserve stability in the correct way. Finally, reductions  $\xrightarrow{\Sigma}$  and  $\xrightarrow{M}$  ensure that the transitions in  $\xrightarrow{\rightarrow}$  becomes minimal in number and structurally unique with respect to  $=_{\text{iso}}$ . The arguments here are the same as for  $\sim_{\text{IMC}}$ .  $\square$

### 11.3.2. MA Bisimilarities

Both strong bisimilarity and naïve weak probabilistic bisimilarity are straightforward combinations of the respective bisimilarities for PA and IMC. Their normal forms can be obtained by simply joining the reductions that lead to normal forms for the respective bisimilarities of PA and IMC. While formally, we cannot immediately conclude this from the previous result, the

proofs do not need any new ideas. We therefore skip the details here and only state a summarizing theorem.

**Definition 11.18 (Normal Form Instances for MA).**

$$\text{Let } NF_{\sim_{\text{MA}}} = \sim_{\text{MA}} \circ \overset{M}{\rightsquigarrow} \circ \overset{\Sigma}{\rightsquigarrow} \circ \overset{C}{\rightsquigarrow}.$$

and

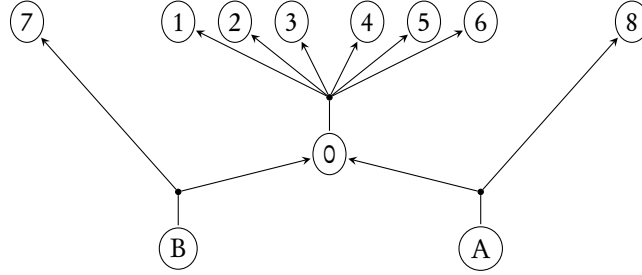
$$\text{Let } NF_{\approx_{\text{MA}}} = \approx_{\text{MA}} \circ \overset{T}{\rightsquigarrow}_{\downarrow} \circ \overset{L}{\rightsquigarrow} \circ \overset{M}{\rightsquigarrow} \circ \overset{\Sigma}{\rightsquigarrow}.$$

◁

**Theorem 11.2.** Let  $\asymp \in \{\sim_{\text{MA}}, \approx_{\text{MA}}\}$ .

1. **Minimality:**  $NF_{\asymp}(\mathcal{A})$  is  $\preceq^{[S]}$ ,  $\preceq^{[D]}$ , and  $\preceq^{[D]}$ -minimal for every  $\mathcal{A} \in \text{MA}$ .
2. **Uniqueness of minimals:** If  $\mathcal{A}$  and  $\mathcal{A}'$  are  $\preceq^{[S]}$ ,  $\preceq^{[D]}$ , and  $\preceq^{[D]}$ -minimal automata, and  $\mathcal{A} \asymp \mathcal{A}'$ , then also  $\mathcal{A} =_{\text{iso}} \mathcal{A}'$ .
3. **Normal forms:**  $NF_{\asymp}$  is uniquely defined up to  $=_{\text{iso}}$ , and is a normal form.

**Weak Distribution Bisimilarity** As weak distribution bisimilarity is the weakest bisimilarity of all we have considered, it clearly bears the potential to minimize a Markov automaton, or a probabilistic automaton more than the other bisimilarities when we focus on  $\preceq^{[S]}$  and  $\preceq^{[D]}$ . As we will see, minimizing with respect to these two metrics can come at the cost of an increase in  $\preceq^{[D]}$ . Conversely, introducing additional states and transitions can decrease the overall fanout of the automaton.



**Figure 11.5.:** Adding States and Transitions Reduces Fanout

**Example 11.5.** By weak distribution bisimilarity one may remove state 0 in the left automaton in Figure 11.5, because it is behaviourally equivalent to its unique successor distribution. On the right-hand side, we see the result of this reduction. Most notably, while we have reduced the number states by 1 and the number of transitions by 1, we have increased the total fanout by 4! While this increase seems not too high, we see its full range by considering instead a parametric example where the number of states in the support of the outgoing transition of state 0 is  $n$

(currently  $n = 6$ ). We then obtain an increase of  $n - 2$ , while still the decrease in both the number of states and transitions stays 1. We see that in the worst case, we have an *increase* that is *linear* in the number of states versus a *constant decrease*.

The example also illustrates that finding an optimally reduced automaton with respect to  $\preceq^{[D]}$  is a non-trivial task. Note that in the automaton on the right, all probabilities of the outgoing transitions of state  $B$  and  $A$  are different. Still, the probabilities for states 1 to 6 only differ in a constant factor. If we do not intend to reduce the number of the total fanout, we need to identify within *all* transitions subdistributions that are identical up to some constant factor, and thus in some sense shared between several transitions. Then, we can replace them by a new state, like state 0 in our example, that basically represents the shared subdistribution. However, identifying the optimal shared subdistributions seems a computational complex task, as taking the largest shared subdistribution (in terms of states its support contains) may be suboptimal, as well as taking a shared subdistribution that is shared by the largest number of transitions.  $\triangleleft$

Finding an automaton that is optimal with respect to  $\preceq^{[D]}$  seems highly non-trivial. We leave the problem to future research.

In the following, we will present the necessary steps to minimize an automaton with respect to  $\preceq^{|S|}$  and  $\preceq^{|D|}$ . In addition, we will show that this automaton can be further transformed into a normal form by the reduction we are already familiar with. However, this normal form will not be minimal with respect to all three metrics, and thus is not canonical in the way the normal forms for the other bisimilarities are.

The first step is to show that two Markov automata are weak distribution bisimilar if and only if they are naïve weak probabilistic bisimilar after reduction with respect to  $\approx_\delta^R$  and  $\approx^R$ .

**Theorem 11.3.** Let  $\mathcal{A}$  and  $\mathcal{A}'$  be two MA, and let  $\mathcal{A} \approx_\delta^R \circ \approx^R \hat{\mathcal{A}}$  and  $\mathcal{A}' \approx_\delta^R \circ \approx^R \hat{\mathcal{A}}'$ . Then

$$\mathcal{A} \approx_\delta \mathcal{A}' \text{ if and only if } \hat{\mathcal{A}} \approx_{\text{MA}} \hat{\mathcal{A}}'$$

*Proof.* It suffices to prove that

$$\hat{\mathcal{A}} \approx_\delta \hat{\mathcal{A}}' \text{ if and only if } \hat{\mathcal{A}} \approx_{\text{MA}} \hat{\mathcal{A}}',$$

as by Lemma 11.4  $\mathcal{A} \approx_\delta \hat{\mathcal{A}}$  and  $\mathcal{A}' \approx_\delta \hat{\mathcal{A}}'$  and transitivity of  $\approx_\delta$ .

We first derive an important property. Recall from Definition 11.14 and 8.15 that if  $\mathcal{A} \approx^R \hat{\mathcal{A}}$ , then all states  $s$  in  $\hat{\mathcal{A}}$  satisfy: whenever  $s \xrightarrow{P} \mu$  then  $\delta(s) \mathcal{L}(\approx_\delta) \mu$ . But this implies that for every state  $t \in \text{Supp}(\mu)$  also  $s \approx_\delta t$  must hold. Now we recall that the automata  $\hat{\mathcal{A}}$  and  $\hat{\mathcal{A}}'$  that we consider are also reduced with respect to  $\approx_\delta^R$ . But this immediately implies that  $s \approx_\delta t$  implies in fact  $s = t$ . Thus, whenever  $s \xrightarrow{\tau} \mu$  then either  $\mu = \delta(s)$  or  $\delta(s) \not\approx_\delta \mu$ . In other words, whenever  $s \xRightarrow{\tau \mid P}_c \mu$  then either  $\mu = \delta(s)$  or *not*  $s \xRightarrow{\tau \mid P}_c \mu$ . Let us denote this fact by  $(\star)$ .

For the proof of our claim, we use the semi-weak distribution bisimulation characterization of naïve weak probabilistic bisimulation, Definition 8.11, which is validated by Theorem 8.12. With this characterization it becomes clear that semi-weak distribution bisimulation and naïve weak probabilistic bisimulation coincide in the first and the third condition. Only the second condition, the splitting condition, differs. We recall that for semi-weak distribution bisimulation it states that whenever  $\mu \mathcal{R} \gamma$  then for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there must exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma = \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ . In turn,

for weak distribution bisimulation, it states that then for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there must exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \Longrightarrow_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R} \gamma_i$  for  $i \in \{1, 2\}$ . The precise difference lies in the way  $\gamma$  is allowed to answer the splitting:

$$\gamma = \gamma_1 \oplus_p \gamma_2 \text{ versus } \gamma \Longrightarrow_c \gamma_1 \oplus_p \gamma_2.$$

It is obvious that the left-hand side implies the right hand-side. We now treat the other direction.

We will now argue that in our current context, both conditions must coincide. As  $\mu = \mu_1 \oplus_p \mu_2$  and thus also trivially  $\mu \approx_\delta \mu_1 \oplus_p \mu_2$ , by Remark 8.6,  $\gamma_1 \oplus_p \gamma_2 \approx \mu_1 \oplus_p \mu_2 \approx \mu$  must hold if  $\mu \approx \gamma$ . But then, by transitivity of  $\approx$ , it follows that  $\gamma \approx \gamma_1 \oplus_p \gamma_2$ . By  $(\star)$  and Lemma 8.2, we can conclude that actually  $\gamma_1 \oplus_p \gamma_2$  must equal  $\gamma$ . As all automata we currently consider are quotient automata with respect to  $\approx_\delta$ , they are clearly also *mec*-contracted. Thus, from  $\gamma \Longrightarrow_c \gamma_1 \oplus_p \gamma_2$  and  $\gamma \approx \gamma_1 \oplus_p \gamma_2$  and Lemma D.15, we can conclude that also  $\gamma \xrightarrow{\tau \downarrow P} \gamma_1 \oplus_p \gamma_2$ . However, this is an immediate contradiction to  $(\star)$ , except if  $\gamma = \gamma_1 \oplus_p \gamma_2$ . But then this implies that we have found a suitable splitting of  $\gamma$  directly. And this is all that is needed to satisfy the left-hand side above.  $\square$

The interesting insight of this theorem is that the difference between weak distribution bisimilarity and naïve weak probabilistic bisimilarity lies mainly in the ability of the former to eliminate more states, namely those, which we called redundant states. Once redundant states are removed, every reduction that is valid with respect to naïve weak probabilistic bisimilarity is also valid with respect to weak distribution bisimilarity. This theorem hence allows us to reduce the problem of minimizing an MA with respect to weak distribution bisimilarity to the problem of minimizing it with respect to naïve weak probabilistic bisimilarity.

We thus define the reduction for weak distribution bisimilarity as

$$NF_{\approx_\delta} := \widetilde{\approx_\delta} \circ \widetilde{R} \circ NF_{\approx_{\text{MA}}}.$$

We could actually omit the first reduction of  $NF_{\approx_{\text{MA}}}$ , i.e.  $\widetilde{\approx_{\text{MA}}}$ , as this will not change the state space at all, since  $\approx_{\text{MA}} \subseteq \approx$ . It is straightforward to derive from here that  $NF_{\approx_\delta}$  satisfies the following properties.

**Theorem 11.4.**

1. **Minimality:**  $NF_{\approx_\delta}(\mathcal{A})$  is  $\preceq^{|S|}$  and  $\preceq^{|D|}$ -minimal for every  $\mathcal{A} \in \text{MA}$ .
2. **Uniqueness of minimals:** If  $\mathcal{A}$  and  $\mathcal{A}'$  are  $\preceq^{|S|}$  and  $\preceq^{|D|}$ -minimal automata, and  $\mathcal{A} \approx_\delta \mathcal{A}'$ , then also  $\mathcal{A} =_{\text{iso}} \mathcal{A}'$ ,
3. **Normal forms:**  $NF_{\approx_\delta}$  is uniquely defined up to  $=_{\text{iso}}$ , and is a normal form.

We emphasize that different to similar theorems we have seen before, minimality with respect to  $\preceq^{|D|}$  does not hold for this normal form. We have already discussed this issue in the beginning of this section. This problem shows that in general a reduction with respect to one measure may lead to an increase in other measures. So, it may be necessary to develop different minimization strategies than only state space reduction by quotienting in the future, depending on which measure has the greatest influences on the run-time complexity of consecutive analyses applied on the automaton

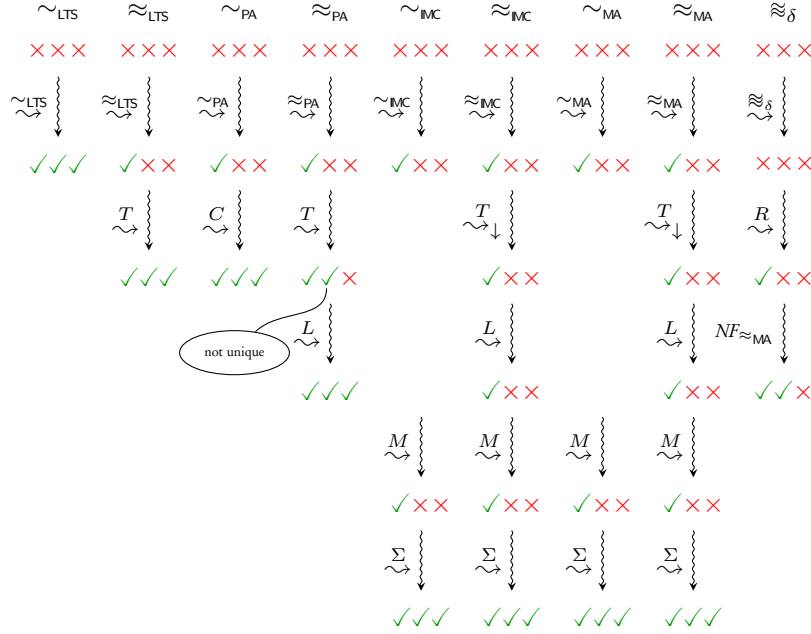


Figure 11.6.: Algorithmic steps in minimal quotient computation.

## 11.4. Summary and Discussion

This chapter has answered the question how to compute the minimal, canonical representation of labelled transition system, probabilistic automata and interactive Markov chains with respect to strong and weak bisimilarity, together with polynomial time minimization algorithms in the finite case. Canonical forms have also appeared in axiomatic treatments of probabilistic calculi [Den05], but are obtained by adding transitions via saturation, so without aiming for minimality.

For Markov automata, we have established identical results for strong bisimilarity and naïve weak probabilistic bisimilarity. For weak distribution bisimilarity, we have shown, however, that no canonical normal form exists that is optimal in size with respect to all three measures, as a reduction with respect to state space and transitions goes along with an increase in the fanout. Still, a reduction with respect to weak distribution bisimilarity may still be beneficial, as the decrease in the size of the state space and number of transitions compared to naïve weak probabilistic bisimilarity (or weak probabilistic bisimilarity) is significant. The reduction algorithm has currently exponential run-time complexity, which has inherited from the decision algorithm for weak distribution bisimilarity (see Chapter 10). As this complexity is only an upper bound, but not a lower bound, a polynomial algorithm may still be available in future.

The algorithms we developed can be exploited in an effective compositional minimization strategy for MA and its submodels, because all considered bisimilarities are congruence relations for the standard process algebraic operators. With this, we see a rich spectrum of potential applications. While quotient reduction is a well-known and widely exploited technique to reduce computation effort, the other reduction techniques presented here are novel or – at least – have not undergone a systematic review with respect to their minimization potential. In this regard

we would like to remark that the usage of reduction  $\stackrel{T}{\sim}$  must have already been considered in the context of tools exploiting weak bisimilarity [FM91; CL11], we are not aware of publications mentioning this point.

Figure 11.6 summarizes —from left to right— what steps are needed to perform the minimization in labelled transition systems, probabilistic automata, interactive Markov chains and Markov automata. The triplets indicate minimality (✓) or non-minimality (✗) with respect to  $|S|$ , then  $|D|$ , then  $\|D\|$ . For example, ✓✓✗ indicates that state and transition numbers are minimal, but transition fanout size can be non-minimal.



## 12. A Semantics for Every *GSPN*

Generalized Stochastic Petri Nets (*GSPN*) [Chi+93; Mar+94; Bal07] constitute a formalism to model concurrent computing systems involving stochastically governed timed behaviour. *GSPN* are based on Petri nets, and are in wide-spread use as a modelling formalism in different engineering and scientific communities. From Petri Nets they inherit the underlying bipartite graph structure, partitioned into *places* and *transitions*, but extend the formalism by distinguishing between *timed transitions* and *immediate transitions*. The latter can fire immediately and in zero time upon activation. The firing time of a timed transition is governed by a *rate*, which serves as a parameter of a negative exponential distribution.

While *GSPN* owe their modelling power to the combination of these two types of transitions, they also lead to a intricate semantic problem, that leaves certain definable nets –so called *confused* nets– semantically undefined. In this chapter, we will show that this semantic gap is not compulsive, and can be amended by a novel semantics for *GSPN* based on Markov automata.

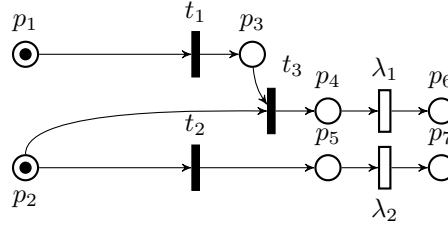
**Outline and Contributions.** We first discuss the phenomenon of confused *GSPN* in greater detail, highlight its semantic intricacies and motivate why the current *status quo* is unsatisfactory. Before we present our solution to the problem, we first recall the definition of *GSPN* in Section 12.2. Then, in Section 12.3 we present the MA semantics for *GSPN* based on the marking graph, which we then refine into the bisimulation semantics in Section 12.4. The provided semantics is *complete* in the sense that it gives a meaning to *every* *GSPN*. The semantics is furthermore *conservative* with respect to the well-established existing semantics of *well-defined* nets. More precisely, we show that for well-defined *GSPN*, our semantics is weak bisimulation equivalent to the classical *CTMC* semantics. This entails that measures of interest, such as steady-state and transient probabilities are identical.

This chapter is an extended version of [Eis+13a].

### 12.1. Confusion – A Question of Semantic Perspective

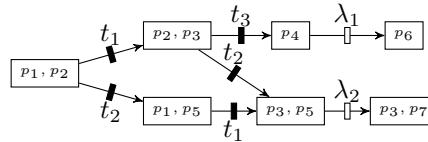
The precise semantics of a *GSPN* may conceptionally be considered as consisting of two stages. The first stage is an abstract, high-level semantics describing *when* which transitions may fire, and *with what probability*. Speaking figuratively in terms of a token game, this semantics determines how tokens can be moved from place to place by the firing of transitions. The second stage is a lower-level mathematical description of the underlying stochastic process, typically a continuous time Markov chain (*CTMC*, for short), which is derived to represent the intended stochastic behaviour captured in the first stage. This Markov chain is then subject to the analysis of steady-state or transient probabilities of markings, or more advanced analysis such as stochastic model checking.

The modelling power of *GSPN* is particularly owed to the presence of immediate transitions [Chi+93]. Unfortunately, this characteristic strength of the formalism may lead to semantically intricate situations [Bau+02; Chi+93; Chi+93; CZ96; DS99; Kat12; EP03]. One of the most prominent cases is *confusion* [Mar+94; Bal07]. In confused nets, the firing order of two concurrently enabled, non-conflicting immediate transitions determines whether two subse-



**Figure 12.1.:** Confused *GSPN*, see [Mar+94, Fig. 21]. Timed transitions are depicted as non-solid bars and immediate transitions, as solid.

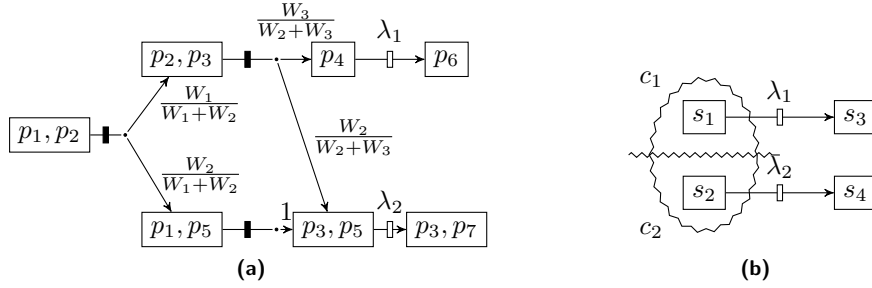
quent transitions are in conflict or not. The net in Figure 12.1 is confused, since transitions  $t_1$  and  $t_2$  are not in direct conflict, but firing transition  $t_1$  first leads to a direct conflict between  $t_2$  and  $t_3$ , which does not occur if  $t_2$  fires first instead. Confusion is not a problem of the high-level (token game) semantics of a net (stage one), as it is entirely clear which transition may fire, and how tokens are moved in either case. It is rather a problem of the underlying stochastic process (stage two) that ought to be defined by this net. Recall that the transitions  $t_1$  through  $t_3$  are all immediate, and thus happen without elapse of time. Thus, their firing is basically transparent to a continuous time evolution. Places  $p_4$  and  $p_5$  enable two distinct timed transitions with rate  $\lambda_1$  and  $\lambda_2$  respectively, cf. Figure 12.1. Now, depending on how the confusion between the transitions (and potentially the direct conflict between  $t_2$  and  $t_3$ ) is resolved, the underlying stochastic behaviour *either* corresponds to an exponential delay with rate  $\lambda_1$ , *or* to a delay with rate  $\lambda_2$ . Which of the two delays happens is not determined by the net structure, and as such is *non-deterministic*.



**Figure 12.2.:** Non-deterministic behaviour of the confused *GSPN* of Figure 12.1

Figure 12.2 shows a graphical representation of this phenomenon as a marking graph. States correspond to markings of the net in Figure 12.1, and there is an obvious graphical correspondence with respect to the representation of the firing of timed or immediate transitions by similarly shaped edges. In state  $\{p_2, p_3\}$  the direct conflict between  $t_2$  and  $t_3$  in the net yields a non-deterministic choice.

As the resulting process is not a *CTMC*, workarounds have been developed. To resolve (or: avoid) non-determinism, *priorities* and *weights* have been introduced [Mar+87]. Intuitively, weights are assigned to immediate transitions at the net level so as to induce a probabilistic



**Figure 12.3.:** (a) Probabilistic behaviour of *weighted* confused GSPN in Figure 12.1; (b) the resulting CTMC

choice instead of a non-deterministic choice between (equally-prioritized) immediate transitions. Ignoring priorities, whenever more than one immediate transition is enabled, the probability of selecting a certain enabled immediate transition is determined by its weight relative to the sum of the weights of *all* –including those that are independent– enabled transitions.

For example, for the marking depicted in Figure 12.1, transition  $t_1$  is selected with probability  $\frac{W_1}{W_1+W_2}$  where  $W_i$  is the weight of transition  $t_i$ . In this way, we obtain an unambiguous stochastic process for this GSPN, cf. Figure 12.3a. Now, the unlabelled edges have multiple endpoints and denote probability distributions over markings. We can consider this as a semi-Markov process, which has both zero-time delay and exponentially distributed time delay edges, as worked out, for instance by Balbo [Bal07]. In order to derive a CTMC from this process, sequences of zero-time delay edges are fused into probability distributions over states. For our example net, we obtain the CTMC in Figure 12.3b with initial distribution  $\mu_0$  with  $\mu_0(s_1) = c_1$  and  $\mu_0(s_2) = c_2$  where

$$c_1 = \frac{W_1}{W_1+W_2} \cdot \frac{W_3}{W_2+W_3} \text{ and } c_2 = \frac{W_2}{W_1+W_2} + \frac{W_1}{W_1+W_2} \cdot \frac{W_2}{W_2+W_3}.$$

These quantities correspond to the reachability probability of marking  $\{p_4\}$  and  $\{p_3, p_5\}$ , respectively, from the initial marking. Unfortunately, this approach has a drawback, related to the *dependence* and *independence* of transitions, an important concept in Petri net theory. In our example net of Figure 12.1, the transitions  $t_1$  and  $t_2$  are *independent*. Their firings happen independent from each other, as the two transitions share no places. Transitions  $t_2$  and  $t_3$ , in contrast, are *dependent*, as the firing of one of them influences the firing of the other (by disabling it) via the shared input place  $p_2$ . However, the expected independence between  $t_1$  and  $t_2$  is not reflected in our GSPN above after introducing weights. Instead, the probability to reach marking  $p_4$  (and marking  $p_5$ ) under the condition that transition  $t_2$  has fired will differ from the corresponding probability under the condition that  $t_1$  has fired. A further conceptual drawback from a modelling perspective, is that when a new immediate transition is inserted between  $t_1$  and  $t_3$ , then this changes these probabilities. This is irritating, since we only refine one immediate transition into a sequence of two immediate transitions. Since immediate transitions do not take time, this procedure should not result in a change of the underlying stochastic model. However, it does. We can also consider this phenomenon as a problem of locality. A local change of the net has unexpected global consequences with respect to the probabilities.

To remedy this defect, several approaches to define the stochastic process at the net level have been proposed. At the core of these approaches, immediate transitions are usually partitioned according to their conflict behaviour, based on a structural analysis of the net. The standard

approach is to partition them into *extended conflict sets* (shortly, *ECSs*) [Mar+87], which is a generalization of structural conflicts in the presence of priorities (which are not treated here). Intuitively, two transitions are in structural conflict in a marking, if both are enabled in this marking, and firing any of them will disable the other. Inside an *ECS*, weights are used to decide immediate transition firings, while no choice is resolved probabilistically across *ECSs*. For confusion-free nets, the *ECS* does provide a way of resolving conflicts probabilistically with a localized interpretation of weights. Unfortunately, for confused nets, this solution approach suffers from the same problem as our initial approach: The *ECSs* for the net in Figure 12.1 are given by the partition  $\{\{t_1\}, \{t_2, t_3\}\}$ . As transitions  $t_2$  and  $t_3$  are in the same *ECS*, the decision which to fire will be resolved probabilistically according to their weights. Transitions  $t_1$  and  $t_2$ , in contrast, are in different *ECS*. Thus, the decision will still need to be resolved non-deterministically, given that they may be enabled at the same moment. Inserting immediate transition  $t_4$  between  $t_1$  and  $t_3$  as mentioned above will lead to the *ECSs*  $\{\{t_1\}, \{t_4\}, \{t_2, t_3\}\}$ . Thus, still only the decision between transitions  $t_2$  and  $t_3$  is resolved probabilistically and not influenced by  $t_4$ . So, since some decisions are forced to be non-deterministic, this approach does in general not yield a mathematically well-defined stochastic process. Moreover, it is easy to see that in our example, any partition of immediate transitions will suffer from one of the semantic problems discussed.

In summary, certain nets lead to undesirable semantic problems. Due to this fact, several researchers have identified certain classes of nets as *not well-defined* (aka. *ill-defined*) [Mar+94; CZ96; DS99]. Such nets are excluded both semantically and from an analysis point of view. Several different definitions have appeared in the literature. However, ill-defined nets, with confused nets being a prominent example, are not *bad* nets *per se*. As Balbo states [Bal00]: “*this underspecification [in confused nets] could be due either to a precise modelling choice [...] or to a modelling error*”. We firmly believe that the modeller should have full freedom of modelling choices, and that such choices should not be treated as errors *by definition*.

## 12.2. Generalized Stochastic Petri Nets

This section introduces *GSPN*, where, for the sake of simplicity, we do not consider transition priorities. For a set  $X$ , we use  $\Sigma(X)$  to denote the set of all partitions of  $X$ . For a set of places  $P$ , a *marking*  $m$  is a multi-set over  $P$  of the form  $m : P \rightarrow \mathbb{N}$ . We let  $M$  denote the set of all markings over  $P$ , and use  $m, m_0$  etc to denote its elements.

**Definition 12.1 (Generalized stochastic Petri net).** A *generalized stochastic Petri net*  $G$  (*GSPN*) is a tuple  $(P, T, I, O, H, m_0, W, \mathcal{D})$  where:

- $P$  is a finite set of *places*,
- $T = T_i \cup T_t$  is a finite set of transitions ( $P \cap T = \emptyset$ ) partitioned into the sets  $T_t$  and  $T_i$  of *timed* and *immediate* transitions,
- $I, O, H : T \rightarrow M$  defines the transitions' *input* places, *output* places, *inhibitor* places<sup>1</sup>,
- $m_0 \in M$  is the initial marking,
- $W : T \rightarrow \mathbb{R}_{>0}$  defines the transitions' *weights*, and

<sup>1</sup>If transition  $t$  has no inhibitor places, we let  $H(t) = \infty$ .

- $\mathcal{D} : M \rightarrow \Sigma(T)$  is a marking-dependent *partition* satisfying the condition that  $T_t \in \mathcal{D}(m)$  for all markings  $m \in M$ .

◁

The above definition agrees, except for the last component  $\mathcal{D}$ , with the classical *GSPN* definition in the literature [Mar+91; Mar+94; Bal07]. We use the marking-dependent partition function  $\mathcal{D}$  as a generalization of the extended conflict set mentioned before. It serves to express for which immediate transitions choices are resolved probabilistically, and for which non-deterministically. This information is usually not provided in the net definition. Instead the (marking independent) *ECS* are derived based on a structural analysis of the net at hand. The reason why we include this information in an explicit form in the definition is mainly ought to formal reasons. However, it also enables (but does not enforce) a view where the choices between immediate transitions are resolved as a consequence of a conscious modelling decision, possibly decoupled from the net structure. The constraint  $T_t \in \mathcal{D}(m)$  is due to the fact that all enabled timed transitions are always weighted against each other in a race. At the expense of slightly more complicated definitions in the following, we could eliminate this technicality and let  $\mathcal{D} : M \rightarrow \Sigma(T_i)$ .

The *input*, *output* and *inhibition* functions assign to each transition a mapping  $P \rightarrow \mathbb{N}$ , specifying the corresponding cardinalities. A transition has *concession* if sufficiently many tokens are available in all its input places, while the corresponding inhibitor places do not contain sufficiently many tokens for an inhibitor arc to become effective. Firing a transition yields a (possibly) new marking, which is obtained by removing one or more tokens from each input place and adding tokens to the transition's output places. Immediate transitions execute immediately upon becoming enabled, whereas timed transitions are delayed by an exponentially distributed duration which is uniquely specified by a *transition rate* (i.e., a positive real number defined by the weights).

For notational convenience, we write cascaded function application with indexed notation of the first parameter. For example, we write  $I_t$ ,  $O_t$  and  $H_t$  for  $I(t)$ ,  $O(t)$  and  $H(t)$ , respectively. The semantics of a *GSPN* is defined by its *marking graph*, which is obtained by playing the “token game”. Immediate transitions are fired with priority over timed transitions [Mar+91; Chi+93; Mar+94]. Accordingly, if both timed and immediate transitions have concession in a marking, only the immediate transitions become enabled. Let  $G$  be a *GSPN* with marking  $m \in M$ .

**Definition 12.2 (Concession and enabled transitions).**

1. The set of transitions with *concession* in marking  $m$  is defined by:

$$\text{conc}(m) = \{t \in T \mid \forall p \in P. m(p) \geq I_t(p) \wedge m(p) < H_t(p)\}.$$

2. The set of *enabled transitions* in marking  $m$  is defined by:  $\text{en}_m = \text{conc}(m) \cap T_i$  if  $\text{conc}(m) \cap T_i \neq \emptyset$ , and  $\text{en}_m = \text{conc}(m)$  otherwise.

◁

A marking  $m$  is *vanishing* whenever an immediate transition is enabled in  $m$ , otherwise it is *tangible*. Given the priority of immediate transitions over timed ones, the sojourn time in vanishing markings is zero. In a vanishing marking, none of the timed transitions which have concession is enabled. In a *tangible* marking  $m$ , only timed transitions can be enabled. The residence time in tangible marking  $m$  is determined by a negative exponential distribution with rate  $\sum_{t \in \text{en}_m} W(t)$ . The effect of executing a transition is formalized in the classical way:

**Definition 12.3 (Transition execution).** Let the *transition execution relation*  $[\cdot] \subseteq M \times T \times M$  be such that for all markings  $m, m' \in M$  and transitions  $t \in T$  it holds:

$$m[t]m' \iff t \in en_m \wedge \forall p \in P. m'(p) = m(p) - I_t(p) + O_t(p).$$

◁

We now recall the notion of *marking graph*, obtained from reachable markings:

**Definition 12.4 (Reachable marking graph).** The *marking graph* of the GSPN  $G$  is the labelled digraph  $MG(G) = (RS, E)$ , where

- $RS$  is the smallest set of reachable markings satisfying:  $m_0 \in RS$ , and  $m \in RS \wedge m[t]m'$  implies  $m' \in RS$ .
- The edge between  $m$  and  $m'$  is labelled by the transition  $t$  such that  $m[t]m'$ .

◁

This graph describes how a net may evolve in terms of its markings. However, it fails to faithfully represent the stochastic aspects of the net. This is made more precise below.

Recall the idea that we consider certain immediate transitions probabilistically dependent on some other transitions (mainly when they are in conflict), while we consider them independent from others. Traditionally, these relations are captured by extended conflict sets (ECSs [Mar+87]). Here, we consider a generalization of this concept in the form of an arbitrary immediate transitions partition  $\mathcal{D}_m$ . For each marking  $m$ , the partition  $\mathcal{D}_m$  determines a way of resolving conflicts between immediate transitions. Each set  $C \in \mathcal{D}_m$  consists of transitions whose conflicts are resolved probabilistically in  $m$ . On the other hand, transitions of different sets are considered to behave in an independent manner, i.e., we make a non-deterministic selection if several of them are enabled in  $m$ . Our semantics will be general enough that we may allow the latter even if there is a *structural* conflict between these transitions. Let us make this precise.

Assume that some transitions in the set  $C \in \mathcal{D}_m$  are enabled and  $C$  is chosen to be fired. Under this condition, the probability that a specific transition fires is given as the normalized weight of the enabled transitions in  $C$ . Precisely,  $\mathbf{P}_C\{t \mid m\} = 0$  if  $t \notin C \cap en_m$ , and otherwise:

$$\mathbf{P}_C\{t \mid m\} = \frac{W(t)}{W_C(m)} \quad \text{where} \quad W_C(m) = \sum_{t \in C \cap en_m} W(t). \quad (12.1)$$

If  $m$  is a vanishing marking,  $W_C(m)$  denotes the cumulative weight of all enabled (i.e., immediate) transitions in  $C$ . In this case the probability  $\mathbf{P}_C\{t \mid m\}$  of taking the immediate transition  $t$  in  $m$  is determined by the weight assignment  $W$ . Note that  $\mathbf{P}_C\{t \mid m\}$  is 0 if  $t$  is neither enabled nor an element from  $C$ . The case that  $m$  is tangible is similar. Then only timed transitions are enabled, and recall that the set of timed transitions  $T_t$  is an element in  $\mathcal{D}_m$ . Thus,  $C = T_t$ . Accordingly,

$$W_C(m) = \sum_{t \in en_m} W(t)$$

is the exit rate from the tangible marking  $m$ . In this case,  $\mathbf{P}_C\{t \mid m\}$  is the probability of taking the transition  $t$  if the tangible marking  $m$  is left.

In both cases, several distinct transition firings may lead from  $m$  to the same marking  $m'$ . These need to be accumulated. With some overload of notation we define

$$\mathbf{P}_C(m, m') = \sum_{m \{t\} m'} \mathbf{P}_C\{t \mid m\}.$$

## 12.3. Markov Automata Semantics for GSPN

Our aim is to provide a semantics to every GSPN. In particular, this includes nets in which multiple immediate transitions are enabled in a marking, nets with cycles of immediate transitions, as well as confused nets. Obviously, stochastic processes such as CTMC do not suffice for this purpose, as they cannot express non-determinism. Markov automata, however, permit to represent the concepts above, including a formulation in terms of a semi-Markov process with zero-timed delay and exponentially distributed time delays [Bal07], while in addition supporting non-determinism between transition firings in vanishing markings. The attentive reader may have already realized that Figure 12.2 and Figure 12.3a are in fact graphical representations of MA.

### 12.3.1. Preliminary Concepts

For the technical development of this chapter, it is convenient to assume that Markov automata start their execution not from an initial *state*, as defined in Chapter 7, but from an initial *distribution* over states. In this way, as for CTMC, MA start their execution in a specific state with a certain probability. In the following, we use  $\mu_0$  to denote the initial distribution of a MA.

These adaptations have only little consequences on the definitions and notations we used throughout the thesis. We only need to generalize our definition bisimulation between two MA in the obvious way: two MA  $\mathcal{A}$  and  $\mathcal{A}'$  are weak distribution bisimilar if their initial distributions satisfy  $\mu_0 \approx \mu_0'$  in the disjoint union of the two automata.

Note that for a clearer distinction with respect to GSPN transitions, we will call transitions of Markov automata MA-transition in the following.

**Notations and Nomenclature.** We will routinely use the uniform representation of an MA in the following without explicit mentioning. Hence, we make use of the notation  $\longrightarrow$  to uniformly denote immediate and timed transitions. As before, the differentiation between the two underlying kinds of transitions happens by the concrete choice of the action taken from  $Act^X$ .

The nomenclature of GSPN deviates at some points for phenomena also known within automata-based models. In the context of this chapter, it is convenient to adopt it also for MA. Our notion of *stability* is replaced by a notion of *tangibility* in GSPN. Accordingly, we call a state  $s \in S$  *tangible* if no immediate MA-transition originates from  $s$ . A *probability distribution* over states is called *tangible* if all states in its support set are tangible. We recall from Section 7.3 that a CTMC can be considered a special case of MA: A CTMC is a MA with  $\dashrightarrow = \emptyset$ . Note that for this chapter, CTMC are equipped with initial distributions instead of initial states, too, as we adapted the definition of MA accordingly.

### 12.3.2. Basic semantics of GSPN

We are now in the position to define the semantics of every GSPN—including the *non* well-defined ones—by means of a MA. The intuition is rather simple. Basically the semantics of a GSPN corresponds to its reachable marking graph, cf. Definition 12.4. States correspond to markings, taking an immediate MA-transition in the MA is the counterpart to firing an immediate transition in the net, and likewise for timed MA-transitions and timed transitions. The marking graph can therefore directly be interpreted as a Markov automaton.

**Definition 12.5 (Basic MA semantics for GSPN).** The MA semantics of the GSPN  $G = (P, T, I, O, H, m_0, W, \mathcal{D})$  is the MA  $A_G = \mathcal{A} = (S, \mu_0, Act, \xrightarrow{x}, \dashrightarrow)$ , where

- $S = RS$  is the reachable set of markings in the marking graph,
- $\mu_0 = \delta(m_0)$ ,
- for every  $m \in RS$ , and each equivalence class  $C \in \mathcal{D}_m$ ,
  1. there is an MA-transition  $m \xrightarrow{x} \mu$  if and only if  $m$  is a tangible marking,  $r = W_C(m)$  and  $\mu(m') = \mathbf{P}_C(m, m')$  for all  $m' \in RS$ ,
  2. there is an MA-transition  $m \dashrightarrow \mu$  if and only if  $m$  is a vanishing marking and  $\mu(m') = \mathbf{P}_C(m, m')$  for all  $m' \in RS$ .

◁

So, the basic MA semantics is the marking graph of a GSPN. Every marking of the GSPN that is reachable by a sequence of (net) transitions from the initial marking corresponds to a state in the MA. As discussed before, in marking  $m$  of the net all enabled timed transitions  $t$  induce an exponentially distributed stochastic delay with a rate  $r$  that is the sum of all weights of enabled transitions. In this case, the probability to reach a marking  $m'$ , say, by MA-transition  $t$  is given as the MA-transition's relative weight. This is reflected in Clause 1 of the above MA semantics. If no timed transition is enabled in marking  $m$ , then no timed MA-transition originates from state  $m$ .

In contrast, the enabled immediate transitions in a marking need to be represented by more than one immediate MA-transition in the MA. Recall that each equivalence class  $C \in \mathcal{D}_m$  corresponds to an ECS in GSPN terminology. For every such set  $C$ , the enabled transitions in  $C$  fire with a probability that is equal to their weight in relation to the sum of the weights of all enabled transitions in  $C$ . However, transitions that are in different sets in  $\mathcal{D}_m$  are entirely independent. More precisely, transitions from different sets in  $\mathcal{D}_m$  compete in a non-deterministic way. This is reflected in Clause 2 of the above definition. The non-deterministic choice between transitions across different sets of  $\mathcal{D}_m$  is represented by introducing an immediate MA-transition for *every* set in the partition  $\mathcal{D}_m$ . The probabilistic decision among transitions *within a single* set, in turn, is reflected by the distribution over markings the corresponding immediate MA-transition leads to.

### 12.3.3. Well-defined GSPN

The aim of this section is to formalize and generalize well-defined GSPN in terms of our new semantics. Central for this purpose is the concept of (hyper) weak MA-transitions.



It is crucial to recall in the following that any MA-transition itself is a weak MA-transition, and that from any state  $s$ , there is always a weak MA-transition  $s \Longrightarrow \delta(s)$ , even if  $s$  is tangible.

**Definition 12.6 (Well-defined GSPN).** Let  $G = (P, T, I, O, H, m_0, W, \mathcal{D})$  be a GSPN with MA semantics  $A_G$ . We say  $G$  is *well-defined*, if for every state  $m \in RS$ , and every pair  $(\mu, \mu')$  of distributions over *tangible states* it holds:  $m \Longrightarrow \mu$  and  $m \Longrightarrow \mu'$  implies  $\mu = \mu'$ .  $\triangleleft$

Intuitively, this expresses that from every marking a *unique* marking over *tangible* states is reachable. Different to [EP03], we are only interested in the probability to reach a marking, and whether it is uniquely specified, but not in the sequences of MA-transitions leading to tangible markings. Phrased differently, we are only interested in tangible state to tangible state probabilities [CZ96; Bal07].

It is not surprising that a well-defined GSPN induces a unique CTMC: states will correspond to those tangible markings, MA-transition  $\xrightarrow{t}$  is obtained by extending the weak MA-transition until tangible states are reached. The uniqueness is guaranteed by the definition of well-defined GSPN. We have not given a explicit definition of CTMC in this thesis, as it can be considered as a special case of a Markov automaton (see Section 7.3). For notational convenience, we consider a CTMC as the tuple  $(S, \mu_0, \rightarrow)$  in the following, focusing only on the components of an MA which are not trivial or empty when it represents a CTMC.

**Definition 12.7 (CTMC induced by a well-defined GSPN).** The well-defined GSPN  $G$  induces the CTMC  $C_G = (S, \mu_0, \rightarrow)$ , where

- $S$  is the set of reachable tangible markings of  $G$ ,
- $m \xrightarrow{t} \mu$  if and only if  $\mu$  is the unique distribution over tangible markings such that a distribution  $\mu'$  exists with  $m \xrightarrow{t} \mu'$  and  $\mu' \Longrightarrow \mu$  in the basic MA semantics of  $G$ ,
- $\mu_0$  is the unique distribution over tangible markings such that  $m_0 \Longrightarrow \mu_0$ .

$\triangleleft$

**Lemma 12.1.** The induced CTMC of a well-defined GSPN is unique (up to isomorphism).

## 12.4. Bisimulation Semantics

The basic MA semantics we have introduced already has several advantages. It is complete, i.e. it provides semantics for every net, and it is amenable to several analysis techniques that are being established (see Chapter 13 for further details). Nevertheless, we want to address more desirable properties the current proposal does not have:

1. the semantics should be conservative with respect to the existing standard semantics for well-defined nets,
2. immediate MA-transitions should be disregarded as much as possible, and exponential delays should be only distinguished up to lumpability. This ensures that the actual formal semantics agrees with the intuitive behaviour of a net and semantic redundancies are avoided as much as possible. For instance, the introduction of a new immediate transition between

$t_1$  and  $t_3$  in Figure 12.1, which should be independent of every other concurrently enabled transition, should not affect the underlying semantics.

We now will implement the above requirements by defining the semantics of a *GSPN* as its basic MA semantics modulo weak distribution bisimilarity. This semantics will exactly represent the behavioural kernel of the *GSPN*.

**Definition 12.8 (Bisimulation Semantics for GSPN).** The *bisimulation semantics* of the *GSPN*  $G = (P, T, I, O, H, m_0, W, D)$  is the set of all Markov automata that are weak distribution bisimilar to its basic MA semantics  $A_G$ .  $\triangleleft$

For *GSPN*, whose basic MA semantics is *finite*, we can use  $NF_{\approx_s}(A_G)$  as a canonical representative (see Chapter 11). Still, our general definition can be applied to an arbitrary *GSPN*  $G$ .

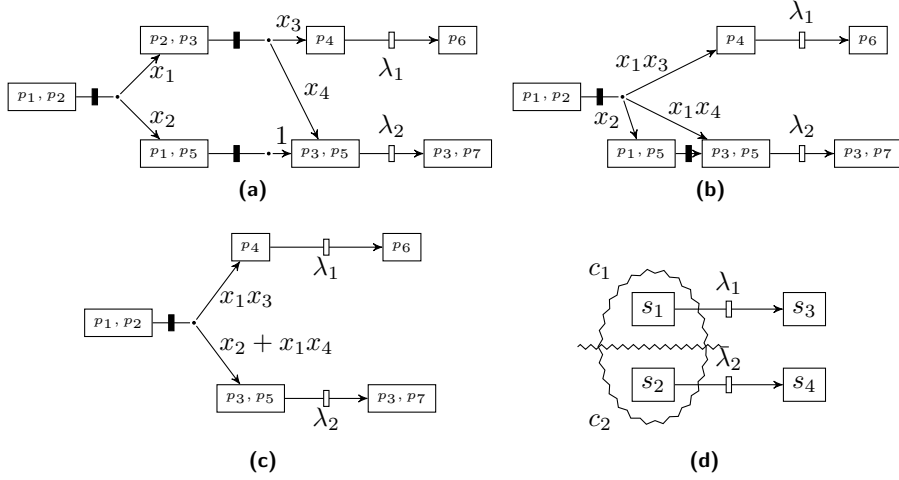
### 12.4.1. Revisiting Well-definition

To illustrate why we consider weak MA bisimilarity a semantic equivalence especially well-suited for *GSPN* semantics, let us recall the standard procedure applied to derive a *CTMC* from the basic MA semantics underlying a well-defined *GSPN*. We illustrate this process with the MA from Figure 12.3a as an example. For convenience, we repeat it in Figure 12.4a below. This figure shows the basic MA semantics of the *GSPN* in Figure 12.1 in the case that every immediate MA-transition is weighted, and choices among immediate MA-transitions are always resolved probabilistically. For a shorter notation, we now denote MA-transition probabilities by  $x_1, x_2$  and so on. When we want to transform this MA into a *CTMC*, we successively remove every immediate MA-transition by replacing a state with an outgoing immediate MA-transition by the distribution that this immediate MA-transition leads to. The result of this replacement is shown in Figure 12.4b and Figure 12.4c. Finally, when no such states remain, we obtain the *CTMC* in Figure 12.4d, where  $c_1 = x_1x_3$  and  $c_2 = x_2 + x_1x_4$ . The effect of this iterative process of fusing transitions can also be formulated via matrix operations [Mar+94].

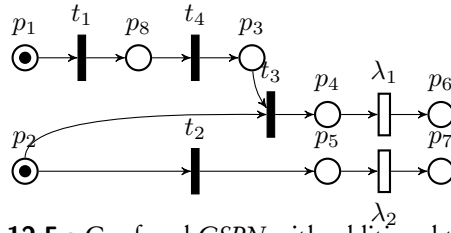
In this example, this procedure leads to a unique result, as every state has at most one outgoing immediate MA-transition. In general, it leads to unique results whenever the net is well-defined. For nets with non-determinism, however, this approach does not lead to mathematically well-defined results. To illustrate this, consider the net in Figure 12.5. Assume now that we do not resolve every choice of immediate transitions probabilistically, but only the conflict between  $t_2$  and  $t_3$ . Hence let  $\mathcal{D}_m = \{\{t_2, t_3\}, \{t_1\}, \{t_4\}\}$ . Note that these are exactly the *ECSs* of the net. We then obtain the non-deterministic basic MA semantics in Figure 12.6a. Applying the fusing procedure as before is clearly not possible, since already in the initial state of the MA, the marking  $\{p_1, p_2\}$ , we have two outgoing immediate MA-transitions, which will finally lead to two different distributions over tangible markings.

Although it is thus not possible to fully remove immediate MA-transitions here – as they are a necessary semantic component to express non-deterministic choice – we want to remove immediate MA-transitions whenever they can be fused. In our example, this would lead to the MA in Figure 12.6b. Only in the first state two immediate MA-transitions remain. They fully capture the non-deterministic behaviour of this *GSPN*.

Weak MA bisimilarity has been designed to exactly perform the task of removing immediate MA-transitions by fusion when the result is uniquely defined. In fact, the MA in Figure 12.6b is the (state- and transition-wise) minimal MA that is weakly bisimilar to the MA in Figure 12.6a.



**Figure 12.4.:** From the MA semantics (a) a CTMC is obtained (d) by step-wise fusing immediate MA-transitions in (b) and (c).

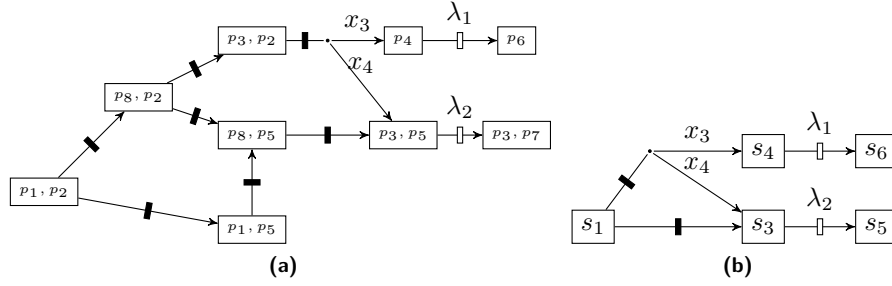


**Figure 12.5.:** Confused GSPN with additional transition

Speaking more generally, weak bisimilarity gives us a powerful means to conservatively generalize the notion of tangible and vanishing markings. Formally, a tangible marking has been defined as a marking that has no outgoing immediate transitions. Markings that are not tangible are called vanishing. More intuitively speaking, as the words *tangible* and *vanishing* suggest, vanishing markings are semantically insignificant, while tangible markings constitute the semantic essence of a net's behaviour. Now, in the context of non-deterministic behaviour, besides of those states without immediate transitions, also those states with a non-deterministic choice between immediate transitions are semantically *tangible* in the literal sense (as long as the choice makes a behaviour difference in the end).

To make this precise, we will define the notion of *significant* markings as a conservative extension of tangible markings, and show that for well-defined nets, they coincide with tangible markings and vice versa.

**Definition 12.9 (Significant marking).** Given a GSPN  $G$  and its basic MA semantics  $A_G$ , we call a marking  $m$  *insignificant* if it is vanishing *and* – in  $A_G$  –  $m$  is a state that has at least one outgoing immediate MA-transition  $m \rightarrow \mu$  such that  $\mu \approx \delta(m)$ . Otherwise we call marking  $m$  *significant*.  $\triangleleft$



**Figure 12.6.:** A basic MA (a) with non-determinism and the smallest MA weakly bisimilar to it (b). In (b), state  $s_1$  subsumes markings  $\{p_1, p_2\}$  and  $\{p_8, p_2\}$  from (a). All other markings with immediate behaviour are removed as a result of fusing them.

Whereas every tangible marking is also significant, not every vanishing marking is insignificant. Only those vanishing markings are also insignificant, which have an immediate successor distribution that is semantically equivalent to the marking itself, and could thus fully replace the marking without affecting the behaviour of the net. Only in *well-defined GSPN* significant and tangible, and vanishing and insignificant coincide respectively, as stated in the following proposition.

**Proposition 3** (Preservation). *If  $G$  is a well-defined GSPN, then a marking  $m$  of  $G$  is tangible if and only if it is significant.*

Furthermore, the CTMC associated with a well defined GSPN enjoys the following strong relation to the original net in terms of the MA semantics.

**Proposition 4.** *The basic MA semantics  $A_G$  of a well-defined GSPN  $G$  is weakly bisimilar to the CTMC  $C_G$  induced by  $G$ .*

Before we present the proof of this proposition, an auxiliary notation and a claim is needed.

For the notation, we remark that due to the uniqueness of the distribution  $\mu$  over *tangible* states that is reachable from each distribution  $\gamma$  over markings of a well-defined GSPN, we can conveniently use the relation  $\Longrightarrow$  from Section 8.4 and write  $\gamma \Longrightarrow \mu$  to express that  $\mu$  is the unique distribution over tangible states such that  $\gamma \Longrightarrow \mu$ .

The following insight is crucial for the understanding of the proof that follows: The second clause of the definition of the CTMC induced by a well-defined GSPN (Definition 12.7) can be rewritten as follows: in the CTMC  $m \xrightarrow{x} \mu$  if and only if  $m \xrightarrow{x} \mu' \Longrightarrow \mu$  in the basic MA semantics of  $G$ .

**Claim 12.1.** *Let  $G$  be a well-defined GSPN. Then for every distribution  $\gamma$  and  $\gamma'$  over states of the automaton  $A_G$ , denoting the basic MA semantics of  $G$ , it holds that  $\gamma \Longrightarrow \gamma'$  implies  $\gamma \Longrightarrow \mu$  if and only if  $\gamma' \Longrightarrow \mu$ .*

This follows immediately from the uniqueness of  $\mu$  we have just recalled.

*Proof of Proposition 4.* In order to prove  $A_G \approx C_G$ , we will provide a bisimulation  $\mathcal{R}$  and show that the pair of initial distributions of  $A_G$  and  $C_G$  is contained in  $\mathcal{R}$ .

Recall that the state space of  $A_G$  is the set  $RS$  of all reachable markings. Let  $S_t$  be the state space of  $C_G$ , which is by definition of  $C_G$  also the set of all reachable tangible markings of  $A_G$ . Let  $\mathcal{R}$  be the symmetric closure of the relation  $\{(\gamma, \mu) \in \text{Dist}(RS) \times \text{Dist}(S_t) \mid \gamma \Longrightarrow \mu\}$ . We denote by  $\mathcal{R}^{-1}$  the symmetric complement of  $\mathcal{R}$ .

The pair of initial distributions of  $A_G$  and  $C_G$  is contained in  $\mathcal{R}$ , which follows immediately from the definition of the initial distribution  $\mu_0$  of  $C_G$ .

We will now check that every pair of  $\mathcal{R}$  satisfies the bisimulation conditions. First, consider an arbitrary pair  $(\mu, \gamma) \in \mathcal{R}^{-1}$ . As  $\gamma \Longrightarrow \mu$ , also  $\gamma \Longrightarrow_c \mu$ . In addition  $(\mu, \mu) \in \mathcal{R}$  for every distribution  $\mu$  over states of  $C_G$ . This holds because, first,  $\mu$  is tangible and thus trivially  $\mu \Longrightarrow \mu$ , and, second,  $S_t \subseteq RS$ . Therefore, it is straightforward to see that conditions (b) and (c) of weak distribution bisimulation (Definition 8.4) are trivially met. For Condition (a), assume  $\mu \xrightarrow{\chi(r)} \mu'$ . By definition of  $C_G$  the distribution  $\mu'$  is tangible again. Now, by our choice of  $\mathcal{R}$ ,  $\gamma \Longrightarrow \mu$  and thus  $\gamma \Longrightarrow_c \mu$ . Even though  $\mu'$  is a distribution over states of both  $C_G$  and  $A_G$ , it is not immediately the case that  $\mu \xrightarrow{\chi(r)} \mu'$  also in  $A_G$ , if  $\mu \xrightarrow{\chi(r)} \mu'$  in  $C_G$ . However, by the definition of  $C_G$ , there must exist  $\mu''$  and  $\mu \xrightarrow{\tau} \mu'' \Longrightarrow \mu'$ . Hence, together, we infer that  $\gamma \Longrightarrow \mu \xrightarrow{\chi(r)} \mu'' \Longrightarrow \mu'$  which implies  $\gamma \xrightarrow{\chi(r)} \mu'$ . Again with  $(\mu', \mu') \in \mathcal{R}$  we are done.

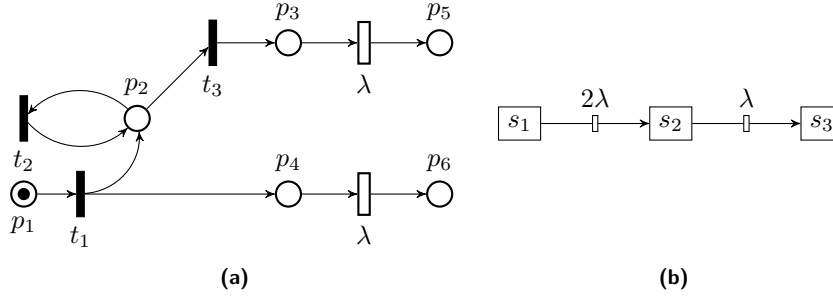
Second, we consider an arbitrary pair  $(\gamma, \mu) \in \mathcal{R} \cap \text{Dist}(RS) \times \text{Dist}(S_t)$ . Following Condition (a) of Definition 8.4, let  $\gamma \xrightarrow{a} \gamma'$ . As  $A_G$  results from a *GSPN*,  $a$  must either be  $\tau$  or  $\chi(r)$ . The second case implies that  $\gamma$  is tangible, and hence  $\gamma = \mu$ , which renders this case trivial. In the second case, let  $\gamma \xrightarrow{\tau} \gamma'$ . By Claim 12.1, this implies that  $\gamma' \Longrightarrow \mu$  as already  $\gamma \Longrightarrow \mu$ . Hence, we obtain  $(\gamma', \mu) \in \mathcal{R}$  immediately. For Condition (b), let  $\gamma_1 \oplus_p \gamma_2$  be an arbitrary splitting of  $\gamma$ . Now  $\gamma \Longrightarrow \mu$  as  $\gamma \Longrightarrow \mu^2$ . By the definition of  $\Longrightarrow$ , it is easy to derive that there must exist  $\mu_1$  and  $\mu_2$  such that  $\mu = \mu_1 \oplus_p \mu_2$  and  $\gamma_1 \Longrightarrow \mu_1$  and  $\gamma_2 \Longrightarrow \mu_2$ . As  $\mu$  is tangible, so are  $\mu_1$  and  $\mu_2$ . Thus, also  $\gamma_1 \Longrightarrow \mu_1$  and  $\gamma_2 \Longrightarrow \mu_2$ . Together, this implies  $(\gamma_i, \mu_i) \in \mathcal{R}$  for  $i \in \{1, 2\}$ . Finally, the premise of Condition (c),  $\gamma \downarrow$ , is only true if  $\gamma$  is tangible, and thus  $\gamma = \mu$  must hold. Then trivially  $\mu \downarrow$ .  $\square$

Proposition 4 provides us with a kind of correctness criterion for the setup we presented. The MA weak bisimulation semantics indeed conservatively extends the classical semantics. Furthermore, many traditionally ill-defined and confused nets can still be related to a CTMC modulo weakly bisimilarity. This is linked to the fact that weak bisimilarity embodies the notion of lumpability, apart from immediate transition fusing.

### 12.4.2. Timeless Traps

Cycles of immediate transitions are an intricate problem in classical *GSPN* theory, their circular firing is often called a timeless trap [Bau+02], see Figure 12.7a for an example. *GSPN* with timeless traps are traditionally excluded from the analysis, basically because the firing precedence of immediate over timed transitions makes the system diverge on the cycle without letting time progress. This is an awkward phenomenon, related to Zeno computations. In our MA reformulation, timeless traps are represented as cycles in the MA, and as such do not pose specific semantic problems. Furthermore, weak bisimilarity is sensitive to cycles of immediate transitions, but only to those that cannot be escaped by firing an alternative immediate transition. This is due to a built-in fairness notion in the weak bisimulation semantics, (rooted in the inclusion of a

<sup>2</sup>Strictly, from  $\Longrightarrow$  we can only defer that  $\gamma \Longrightarrow_c \mu$ . But as  $\mu$  is unique in our context, we also obtain  $\gamma \Longrightarrow \mu$ .



**Figure 12.7.:** A timeless trap that can be escaped by an immediate transition firing (a), and the smallest MA weakly bisimilar to its semantics (b). In (b), state  $s_1$  subsumes markings  $\{p_1\}$ ,  $\{p_2, p_4\}$ , and  $\{p_3, p_4\}$ . State  $s_2$  subsumes markings  $\{p_3, p_6\}$ , and  $\{p_4, p_5\}$ , while state  $s_3$  represents marking  $\{p_5, p_6\}$ .

tangibility check inside the definition of the abbreviation  $\xrightarrow{a}$ ). As a consequence, if a timeless trap can be left by firing a (finite sequence of) immediate transitions leading to a tangible marking, this is equivalent to a single immediate transition firing. This implies that the net in Figure 12.7a is in fact weak bisimilar to the small chain-structured 3-state CTMC in Figure 12.7b. And thus the net is analysable via the classical *CTMC* machinery. This example shows that the combination of lumping and fusing of immediate transitions as supported by weak bisimulation can have powerful effects. Variations to the definition of  $\xrightarrow{a}$  can induce more liberal notions of weak bisimilarity, including the option to escape timeless traps unconditionally [LDH05]. That option is not supported by our current setup, since it has originally been designed to support strong compositionality properties (see the discussions in Section 8.2.2 and Section 8.2.4). Since compositionality is not a first-class concern in the Petri net world, this avenue seems worthwhile to be investigated further.

## 12.5. Summary and Discussion

This chapter has presented a semantics of *GSPN* in terms of Markov automata. We have shown that for well-defined *GSPN*, our semantics is weak distribution bisimilar to the *CTMC* semantics existing in the literature [Bal07; Chi+93; Chi+93; Mar+94]. Since weak distribution bisimilarity coincides with lumping equivalence on *CTMC* (cf. Section 8.7 on page 152), this result establishes “backward compatibility” of our semantics in the sense that it is the same as the classical *GSPN* semantics, up to an equivalence that preserves all quantitative measures of interest such as transient, steady-state probabilities and CSL (without next) formulae [Bai+03b; Bai+05]. Thus, any tool based on our MA-semantics yields the same results as popular *GSPN* tools such as GreatSPN, SMART, and MARCIE.

The main contribution of this chapter is that our semantics applies to *every* *GSPN*. That is to say, our semantic framework is not restricted to well-specified or confusion-free nets. The key to treating confused nets is (not surprisingly) the use of non-determinism. We claim that our approach can also be applied to other stochastic net formalisms such as SANs [MMS85; SM].

The semantics closes a gap in the formal treatment of *GSPN*, which is now no longer restricted to well-defined nets. This abandons the need for any check, either syntactically or semantically,

for well-definedness. This gap was particularly disturbing because several published semantics for higher-level modelling formalisms—e.g., UML, AADL, WSDL—map onto *GSPN* without ensuring the mapping to be free of confusion, thereby inducing ill-defined models. Our Markov automata semantics provides the basis to also cover the confused and ill-specified semantic fragments of these formalisms. Indeed, we were able to relax both notions by considering the Markov automata semantics modulo weak bisimulation. To proceed this way seemed like a natural way forward for quite some time to us, but to arrive there was an astonishingly difficult notational and technical endeavour.





**Part IV.**

**Conclusion**



## 13. Conclusion

This chapter first summarizes the main contributions of this thesis and discusses in how far the principles of Markov automata are now settled. Since this thesis has focussed on the foundational side, it has not put any emphasis on analysis approaches for or applications of Markov automata in the context of real systems. Instead, this chapter continues with a survey of a variety of research activities focussing on the algorithmic analysis of properties of specific Markov automata models. We conclude the chapter with a discussion of directions for further research.

### 13.1. The Foundations of Markov Automata

Markov automata intend to be a model for concurrent systems with both discrete and timed stochastic behaviour.

**The Model** We have introduced Markov automata as a mathematical model that

- (i) provides expressiveness for concurrency and discrete as well as continuous-time stochasticity,
- (ii) is amenable to automated analysis methods, and finally
- (iii) is free from unnecessary semantic intricacies that might lead to modelling pitfalls.

To arrive there, Markov automata are extending labelled transitions systems by blending the established models of probabilistic automata and interactive Markov chains. All three models can be fully recovered from Markov automata as strict submodels. We have provided a natural and fully compositional semantics for Markov automata with respect to parallel composition and abstraction.

**The Bisimulations** We initially have proposed the following requirements on a suitable notion of equality for Markov automata.

- (i) It should be a congruence relation with respect to the composition operators.
- (ii) It should preserve the non-deterministic branching structure.
- (iii) It should enable abstraction from internal details as much as possible.

We have discussed that for probabilistic systems, be it probabilistic automata or Markov automata, the known bisimilarities fail to satisfactorily address the third requirement, with respect to an intuitive notion of observability of probabilistic behaviour: We demand that rolling a four-sided die cannot be distinguished from throwing a two-sided coin two times in direct succession.

This motivation has led us to the development of a novel bisimulation relation for Markov automata that is based on distributions instead of states: relaxed bisimilarity. While it has the above mentioned property, it fails to be a congruence relation. To remedy this, we have introduced weak distribution bisimilarity, which we have proven to be the coarsest bisimulation relation contained in relaxed bisimilarity. Most notably, it still abstracts from internal details as desired. And it enjoys the congruence property.

Concerning our second demand, the preservation of the non-deterministic branching structure, a principle remark is needed. Among the known variants of weak bisimulation only branching bisimulation preserves the non-deterministic branching structure in a strict sense and thus fully preserves properties in CTL\* and similar logics. Though, without the next operator, which expresses that a certain property will hold in the next state to occur, a notion that is difficult to make use of in asynchronous concurrent systems anyhow. As weak distribution bisimulation stems from (ordinary) weak bisimulation, it does not satisfy our second demand to the fullest extend possible. Our choice to yet adapt weak bisimulation instead of branching bisimulation for Markov automata is rooted in weak bisimulation's prevailing importance in the history and theory of PA and IMC, the formalism MA are based upon. A branching variation of weak distribution bisimulation can, however, be defined as well (cf. Section 8.6.2).

We have established various further properties that give insight into the inner workings of weak distribution bisimilarity, among them a state-based characterization, which allows for a direct comparison with the well-known bisimilarities on LTS, PA and IMC. In addition, we have provided distribution-based characterizations of the latter bisimilarities. The most notable finding here has been that weak distribution bisimilarity is the bisimulation kernel of probabilistic forward simulation, the coarsest sensible known process relation on PA.

In summary, we have succeeded in establishing a notion of process equivalence that satisfies the three requirements we have formulated and that in addition comes with several other favourable properties.

**Comparative Semantics** To deepen the comparative analysis of the novel bisimilarities we have provided a simple MA process algebra and a set of axioms for weak distribution bisimilarity, that is sound and complete for the recursion-free fragment. While doing so we fixed an inaccuracy in the published axiomatic treatment of PA (see the discussion in Section 6.3). We have identified a single axiom, the elimination axiom

$$a.(\delta(\tau.\mathcal{D}') \oplus_p \mathcal{D}) = a.(\mathcal{D}' \oplus_p \mathcal{D}),$$

as the one that distinguishes weak distribution bisimilarity from the well-known weak bisimilarities for PA and IMC: If any probable successor state of an action can only exhibit an internal step as its next behaviour, this step can be eliminated.

**Decision and Minimization** We have developed a decision algorithm for weak distribution bisimilarity that, as a side-result, computes the quotient state-space based on our state-based characterization. As a draw-back, the current decision algorithm has exponential time-complexity. It is, however, not clear whether this is strictly needed.

Building on the quotienting technique, we have investigated further size-compression techniques for Markov automata and its submodels based on the three metrics *number of states*, *number of transitions* and *transition fanout*, i.e. the number of states in the support of probabilistic

transitions. For all strict submodels and its corresponding notion of bisimilarity, we have provided algorithmic reduction techniques that lead to normal forms that are minimal with respect to each of the three metrics. For Markov automata in combination with weak distribution bisimilarity, we have shown that a reduction in the number of states and in the number of transitions is only possible at the cost of an non-minimum fanout, and vice versa.

**Generalized Stochastic Petri Net Semantics** We have presented a semantics for *GSPN* based on MA and weak distribution bisimilarity that is a conservative extension of the standard semantics, but extends to arbitrary *GSPN*, including so called *confused* nets. Applying weak distribution bisimilarity recovers the original *CTMC* semantics whenever the latter exists. It allows to explain the phenomenon of *confusion* in terms of internal non-determinism.

## 13.2. System Specification and Quantitative Analysis

**Markov Automata for System Specification.** This thesis has developed the foundational aspects of Markov automata. The importance of our investigations is underlined by the fact that Markov automata are a very expressive model. This makes them applicable in a wide range of industrial and business contexts, for instance as the semantic underpinning of modelling standards and tools, such as the Architectural Analysis and Design Language (AADL) [FG12; Boz+09c; Boz+11], dynamic fault trees [BBB92; BCS10; BCS07b; BCS07a; KK15; Băc+16], or Generalized Stochastic Petri Nets (Chapter 12).

The compositionality properties established in Part II of this thesis (first published in [EHZ10a]) make it possible to design expressive process algebraic languages that can then be used as assembly languages for more high-level formalisms. Indeed, in parallel to our foundational investigations on Markov automata, the Markov Automata Process Algebra, MAPA, has been coined by Timmer [Tim13], exploiting compositionality in a very elegant and far reaching manner. The MAPA language is at the core of the MAMA and SCOOP tool family [Tim+12; Guc+13]. Generalized Stochastic Petri Nets can be transformed into MAPA by means a tool component called GEMMA [Bam12]. This tool family demonstrates convincingly that the base theory of Markov automata can be wrapped into very practical and useable modelling approaches. In fact, SCOOP applies a variety of state space reduction techniques during a fully compositional model construction, in order to make sure that the intermediate models and the finally resulting Markov automata are of tractable size.

**Markov Automata for Quantitative Analysis.** While models of systems are in itself of value as engineering artefacts when designing complex systems, their true value lies in model-based analysis. The core analysis technique for models akin to Markov automata is *stochastic model checking* [Bai+10]. In the stochastic model checking approach, properties of models are analysed based on a deep algorithmic inspection of the model. These properties are usually defined by means of a temporal logic. Especially logics like Probabilistic Computation Tree Logic (PCTL) [HJ94; BA95] or Continuous Stochastic Logic (CSL) [Bai+03a; ZN10] are used. Indeed, it is possible to combine and lift these logics to the setting of Markov automata [HH12],

and the corresponding algorithms for model checking can be lifted as well. This yields a modelling and analysis framework where system models and desirable system properties are specified on the same level of abstraction, and submitted to a Markov automata model checker for their quantitative analysis. This approach has been implemented in the IMCA [Guc+13; GTB14; Guc+14] model checker, which can be used as a backend for the MAMA toolset mentioned above. This approach has also been extended, so as to be able to deal with cost annotations at states and transitions of the Markov automaton, leading to so called Markov reward automata (MRA). In this way, it is possible to express and quantify complex properties relating resource consumption, financial investments, profit and the like.

It is worthwhile to mention that the analysis, albeit being a quantitative analysis, can not be based on quantifying *the* probability of certain events to happen. This is because *non-determinism* is a core feature of Markov (reward) automata. We stipulate that such probabilities depend on the resolution of non-determinism. Rather than considering, e.g., the probability to reach a state (i.e., a marking), it is common to instead determine minimum and maximum reachability probabilities.

The base algorithmic challenges faced when model checking of Markov (reward) automata have received dedicated attention. They can roughly be split into the following categories.

**Long-run average objectives** Long-run analyses usually treat the time horizon of the analysis to be infinite. They then look at the average behaviour in the limit. In this way, average resource consumptions, average profit increases, or average time spendings can be computed. The base algorithm [Guc+14] to solve this problem first identifies all maximal end components  $C$  of the MRA and then analyses each of the  $C$  in isolation, and finally combines these results with a stochastic shortest path analysis performed for each  $C$ . The difficult step is the isolated analysis of each of the components, for which linear programming is used. As an alternative, there is a very recent and much more efficient approach which views an MRA as a compact encoding of a potentially exponentially larger continuous-time Markov decision process (CTMDP) and exploits this insight for a drastically more efficient and scalable long-run average analysis [BWH17], based on relative value iteration and dynamic programming.

**Time bounded objectives** Time-bounded analysis relates to a class of very natural problems, imposing a finite time horizon as a bound, and studying the model behaviour up to this bound. Typical quantities of interest are time-bounded reachability probabilities. The basic approach to timed analysis for Markov automata is a generalization of the analysis for IMC [ZN10]. The problem can be cast into the ERR framework [Hat16] discussed next, and so the results discussed there apply as well.

**Expected resource bounded reward** The problem of computing expectations of reward accumulation in the presence of resource bounds is a very general problem formulation that for finite time horizons admits a uniform treatment of many important problems [Hat+15; Hat16]. Central to this analysis is the problem of computing the optimal expected resource-bounded reward (ERR), in a setting supporting transient, instantaneous and final reward collection as well as transient resource consumption. The state-of-the art in this context is marked by a stable approximation scheme with a strict error bounds to solve the problem in an efficient way [Hat16].

This brief overview demonstrates that quantitative analysis of Markov automata is a very active field of research. A variety of case studies and examples have been studied in this context

[Tim13; Hat16; BWH17], ranging from stochastic job-shop scheduling problems, to a dependability study of a fault tolerant workstation cluster and a performance analysis of file system used at Google. Furthermore model simplification and compression techniques have been developed for Markov (reward) automata [Tim13; Bra16], so as to make the Markov automaton model checking approach scale to larger and more complex problem settings.

As a whole, these activities witness that the conception of Markov automata, which is rounded off with this thesis, has captured the scientific pulse of time.

## 13.3. Open Challenges and Further Work

The findings of this thesis give way to various new challenges and further research directions.

**Diversifying Bisimilarity** While the overall setup of weak distribution bisimulation is very much driven by the requirements initially put forward, certain aspects of it still leave room for variation. One aspect concerns the way timed self-loops are treated. [Bra02] defines a variation of bisimulation for IMC that is slightly coarser than the original definition of [Her02], and treats timed self-loops differently. This change appears to be orthogonal to our considerations and we believe that also our bisimilarities can be adapted similarly.

Another aspect is the treatment of divergence, i.e. the ability to perform an infinite sequence of internal transitions. In the context of stochastic time this phenomenon is called Zenoness, it allows an infinite number of actions to occur in finite time. We have shortly discussed this phenomenon in Section 12.4.2. The way divergence is treated currently is inherited from weak IMC bisimulation. Other approaches of dealing with divergence seem possible, but is yet to be investigated. The axiomatic perspective on possible alternatives is discussed in [LDH05].

**Finer GSPN semantics** More specifically for the purpose of providing a semantics for *GSPN*, weak distribution bisimilarity is arguably too coarse in some respects and abstracts possibly important information specific to *GSPN*. For instance, we have not considered the preservation (by the notion of weak bisimulation) of more detailed marking information such as the exact token occupancy in a place. It is, however, straightforward to include this information by a simple extension of weak bisimulation that respects a certain state labelling, and this is fairly routine [Des+10; Bai+03b]. The same is true for other reward structures—except rewards attached to immediate transitions, which are more involved to handle.

**Completing Completeness** The axiomatization we have provided in Chapter 9 only addresses the recursion-free fragment of MA processes. While we conjecture that the axiomatization can be extended to recursive processes, we have failed to establish a proof of completeness. We have discussed the problems that arise in Section 9.5, together with various routes we have explored in vain.

**Polynomial-Time Decision Algorithm** The decision algorithm for weak distribution bisimilarity, which we have provided in Chapter 10, has exponential time-complexity. This is rooted in our current inability to establish – or refute – the result that both the partition refinement and the computation of the set of preserving transitions can be computed at the same time. Currently, we need to have the set of preserving transitions guessed by an oracle beforehand,

and then verify it during the partition refinement loop. In the worst-case, the (albeit systematic) guessing takes an exponential number of tries. Once a more targeted polynomial approach of computing the set of preserving transitions is found, the whole algorithm will become polynomial in its time-complexity, and this carries over to the minimization algorithm.



# Appendix



## A. Alternative Definition of Weak (Hyper)Transitions

A few proofs in this thesis, whose lemmas focus on properties of weak (hyper)transitions, are based on a different formalization as the one given in the main part of the thesis, namely probabilistic execution ([Seg95]). In [Bre13] it has been shown that the formalizations are effectively equivalent. The reason why two variants appear in this thesis is that for different proof approaches the definitions are (subjectively) differently well-suited.

**Transitions.** A transition  $tr = (s, a, \mu) \in \multimap$ , also denoted by  $s \xrightarrow{a} \mu$ , is said to *leave* from state  $s$ , to be *labelled* by  $a$ , and to *lead* to  $\mu$ , also denoted by  $\mu_{tr}$ . We denote by  $src(tr)$  the *source* state  $s$ , by  $act(tr)$  the *action*  $a$ , and by  $trg(tr)$  the *target* distribution  $\mu$ . We also say that  $s$  enables action  $a$ , that action  $a$  is enabled from  $s$ , and that  $(s, a, \mu)$  is enabled from  $s$ . Finally, we denote by  $D(s)$  the set of transitions enabled from  $s$ , i.e.,  $D(s) = \{ tr \in \multimap \mid src(tr) = s \}$ , and similarly by  $D(a)$  the set of transitions with action  $a$ , i.e.,  $D(a) = \{ tr \in \multimap \mid act(tr) = a \}$ .

**Weak Transitions.** An *execution fragment* of a PA  $\mathcal{A}$  is a finite or infinite sequence of alternating states and actions  $\alpha = s_0 a_1 s_1 a_2 s_2 \dots$  starting from a state  $s_0$ , also denoted by  $first(\alpha)$ , and, if the sequence is finite, ending with a state denoted by  $last(\alpha)$ , such that for each  $i > 0$  there exists a transition  $(s_{i-1}, a_i, \mu_i) \in \multimap$  such that  $\mu_i(s_i) > 0$ . The *length* of  $\alpha$ , denoted by  $|\alpha|$ , is the number of occurrences of actions in  $\alpha$ . If  $\alpha$  is infinite, then  $|\alpha| = \infty$ . Denote by  $frags(\mathcal{A})$  the set of execution fragments of  $\mathcal{A}$  and by  $frags^*(\mathcal{A})$  the set of finite execution fragments of  $\mathcal{A}$ . An execution fragment  $\alpha$  is a *prefix* of an execution fragment  $\alpha'$ , denoted by  $\alpha \leq \alpha'$ , if the sequence  $\alpha$  is a prefix of the sequence  $\alpha'$ . The *trace*  $trace(\alpha)$  of  $\alpha$  is the sub-sequence of external actions of  $\alpha$ ; we denote by  $\epsilon$  the empty trace. Similarly, we define  $trace(a) = a$  for  $a \in E$  and  $trace(\tau) = \epsilon$ .

A *scheduler* for a PA  $\mathcal{A}$  is a function  $\sigma: frags^*(\mathcal{A}) \rightarrow Subdist(\multimap)$  such that for each finite execution fragment  $\alpha$ ,  $\sigma(\alpha) \in Subdist(D(last(\alpha)))$ . Note that by using sub-probability distributions, it is possible that with some non-zero probability no transition is chosen after  $\alpha$ , that is, the computation stops after  $\alpha$ . A scheduler is *determinate* [CS02] if for each pair of execution fragments  $\alpha, \alpha'$ , if  $trace(\alpha) = trace(\alpha')$  and  $last(\alpha) = last(\alpha')$ , then  $\sigma(\alpha) = \sigma(\alpha')$ . A scheduler is *Dirac* if for each  $\alpha$ ,  $\sigma(\alpha)$  is a Dirac distribution. Given a scheduler  $\sigma$  and a finite execution fragment  $\alpha$ , the distribution  $\sigma(\alpha)$  describes how transitions are chosen to move on from  $last(\alpha)$ . A scheduler  $\sigma$  and a state  $s$  induce a probability distribution  $\mu_{\sigma,s}$  over execution fragments as follows. The basic measurable events are the cones of finite execution fragments, where the cone of  $\alpha$ , denoted by  $C_\alpha$ , is the set  $\{ \alpha' \in frags(\mathcal{A}) \mid \alpha \leq \alpha' \}$ . The probability  $\mu_{\sigma,s}$  of a cone  $C_\alpha$  is

recursively defined as:

$$\mu_{\sigma,s}(C_\alpha) = \begin{cases} 0 & \text{if } \alpha = t \text{ for a state } t \neq s, \\ 1 & \text{if } \alpha = s, \\ \mu_{\sigma,s}(C_{\alpha'}) \cdot \sum_{tr \in D(a)} \sigma(\alpha')(tr) \cdot \mu_{tr}(t) & \text{if } \alpha = \alpha'at. \end{cases}$$

Standard measure theoretical arguments ensure that  $\mu_{\sigma,s}$  extends uniquely to the  $\sigma$ -field generated by cones. We call the resulting measure  $\mu_{\sigma,s}$  a *probabilistic execution fragment* of  $\mathcal{A}$  and we say that it is generated by  $\sigma$  from  $s$ . Given a finite execution fragment  $\alpha$ , we define  $\mu_{\sigma,s}(\alpha)$  as  $\mu_{\sigma,s}(\alpha) = \mu_{\sigma,s}(C_\alpha) \cdot \sigma(\alpha)(\perp)$ , where  $\sigma(\alpha)(\perp)$  is the probability of terminating the computation after  $\alpha$  has occurred.

We say that there is a *weak combined transition* from  $s \in S$  to  $\mu \in \text{Dist}(S)$  labelled by  $a \in \text{Act}^X$ , denoted by  $s \xRightarrow{a}_c \mu$ , if there exists a scheduler  $\sigma$  such that the following holds for the induced probabilistic execution fragment  $\mu_{\sigma,s}$ :

1.  $\mu_{\sigma,s}(\text{frags}^*(\mathcal{A})) = 1$ ;
2. for each  $\alpha \in \text{frags}^*(\mathcal{A})$ , if  $\mu_{\sigma,s}(\alpha) > 0$  then  $\text{trace}(\alpha) = \text{trace}(a)$
3. for each state  $t$ ,  $\mu_{\sigma,s}(\{\alpha \in \text{frags}^*(\mathcal{A}) \mid \text{last}(\alpha) = t\}) = \mu(t)$ .

In this case, we say that the weak combined transition  $s \xRightarrow{a}_c \mu$  is induced by  $\sigma$ .

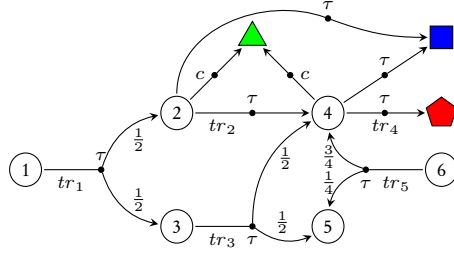
We remark that  $\text{trace}(\alpha) = \text{trace}(a)$  is equivalent to  $\text{trace}(\alpha) = \epsilon$  for  $a = \tau$  and  $\text{trace}(\alpha) = a$  for  $a \in \text{Act}^X \setminus \{\tau\}$ . Moreover, the first two conditions can be equivalently replaced by  $\mu_{\sigma,s}(\{\alpha \in \text{frags}^*(\mathcal{A}) \mid \text{trace}(\alpha) = \text{trace}(a)\}) = 1$ .

Given a set of *allowed* transitions  $\check{A} \subseteq \multimap$ , we say that there is an *allowed weak combined transition* [HT12] from  $s$  to  $\mu$  with label  $a$  respecting  $\check{A}$ , denoted by  $s \xRightarrow{a|\check{A}}_c \mu$ , if there exists a scheduler  $\sigma$  inducing  $s \xRightarrow{a}_c \mu$  such that for each  $\alpha \in \text{frags}^*(\mathcal{A})$ ,  $\text{Supp}(\sigma(\alpha)) \subseteq \check{A}$ .

Albeit the definition of weak combined transitions is somewhat intricate, this definition is just the obvious extension of weak transitions on labelled transition systems to the setting with probabilities. See [Seg06] for more details on weak combined transitions.

**Example A.1.** As an example of weak combined transition, consider the probabilistic automaton depicted in Fig. A.1 and the probability distribution  $\mu = \{(\heartsuit : \frac{3}{4}), (\circledast : \frac{1}{4})\}$ . It is immediate to verify that the weak combined transition  $\circledast \xRightarrow{\tau}_c \mu$  is induced by the Dirac determinate scheduler  $\sigma$  defined as follows:  $\sigma(\circledast) = \delta(tr_1)$ ,  $\sigma(\circledast\tau\circledast) = \delta(tr_2)$ ,  $\sigma(\circledast\tau\circledast) = \delta(tr_3)$ ,  $\sigma(\circledast\tau\circledast\tau\circledast) = \sigma(\circledast\tau\circledast\tau\circledast) = \delta(tr_4)$ , and  $\sigma(\alpha) = \delta(\perp)$  for each other finite execution fragment  $\alpha$ . If we consider all transitions but  $tr_2$  as allowed transitions  $\check{A}$ , then there is no scheduler inducing  $\circledast \xRightarrow{\tau|\check{A}}_c \mu$ . In fact, using this set of allowed transitions, the maximal probability of reaching  $\heartsuit$  from  $\circledast$  is  $\frac{1}{4}$  by the execution fragment  $\circledast\tau\circledast\tau\circledast\heartsuit$ .  $\triangleleft$

We say that there is a *weak (allowed) hyper transition* from  $\rho \in \text{Subdist}(S)$  to  $\mu \in \text{Subdist}(S)$  labelled by  $a \in \text{Act}^X$ , denoted by  $\rho \xRightarrow{a}_c \mu$  ( $\rho \xRightarrow{a|\check{A}}_c \mu$ ), if there exists a family of (allowed) weak combined transitions  $\{s \xRightarrow{a}_c \mu_s\}_{s \in \text{Supp}(\rho)}$  ( $\{s \xRightarrow{a|\check{A}}_c \mu_s\}_{s \in \text{Supp}(\rho)}$ ) such that  $\mu = \bigoplus_{s \in \text{Supp}(\rho)} \rho(s) \cdot \mu_s$ .



**Figure A.1.:** A probabilistic automaton

**Usage Summary.** The probabilistic execution approach of defining transitions only appears in the appendix and is used in the proofs of Lemma D.11, Lemma D.12, Lemma D.15, Lemma 11.4 and Definition D.3.



## B. Proofs of Chapter 5

### B.1. Lemma 5.1

Definition 5.7 and 5.8 agree.

Before we proceed with the proof of the lemma, we introduce helpful lemmas.

**Lemma B.1.** Let  $s_i, s'_i$  and  $t_i, t'_i$  for  $i = 1, 2, 3, \dots, n$  be states for an odd number  $n$ . For each  $i = 1, 2, \dots, n-1$  they satisfy  $s_i \Longrightarrow t_i$  and  $t_i \approx_{\text{MC}} t'_i$  and  $t'_i \Longrightarrow s'_{i+1}$  and  $s'_{i+1} \approx_{\text{MC}} s_{i+1}$ . And furthermore,  $s'_n \approx_{\text{MC}} s_1 = s_n$ .

Then  $s_1 \approx_{\text{MC}} t_2$ .

Note that we can generalize this lemma easily to the case where  $n$  is even by formally by duplicating the last state. As  $s \Longrightarrow s$  for every  $s$  and also  $s \approx_{\text{MC}} s$ , we can satisfy the premise of the lemma.

*Proof.* As  $s_1 \Longrightarrow t_1$ , it is obvious that whatever behaviour  $t_1$  exhibits according to Definition 5.7,  $s_1$  can mimic it weakly by first transitioning to  $t_1$  and then repeating its behaviour exactly up to  $\mathcal{I}$ , and thus also up to  $\approx_{\text{MC}}$ . If we can now show that  $t_1$  must be able to accordingly mi-mick every behaviour of  $s_1$  weakly up to  $\approx_{\text{MC}}$ . Then, it follows easily that  $\{(s_1, t_2)\} \cup \approx_{\text{MC}}$  is a weak IMC bisimulation. As  $\approx_{\text{MC}}$  is the largest weak IMC bisimulation, it must already contain  $(s_1, t_1)$ .

Hence assume that  $s_1 \xRightarrow{a} s'$  with  $a \in \text{Act}$ . Then,

- as  $s_1 \approx_{\text{MC}} s'_n$  by assumption,  $s'_n \xRightarrow{a} \hat{s}'_n$  and  $s' \approx_{\text{MC}} \hat{s}'_n$  for some  $\hat{s}'_n$ .
- Then, as  $t'_{n-1} \Longrightarrow s'_n$ , clearly also  $t'_{n-1} \Longrightarrow \hat{s}'_n$ .
- As  $t'_{n-1} \mathcal{R} t_{n-1}$ , also  $t_n \xRightarrow{a} \hat{t}_{n-1}$  for some  $\hat{t}_{n-1}$  and  $\hat{t}_{n-1} \approx_{\text{MC}} \hat{s}'_n$ . By transitivity of  $\approx_{\text{MC}}$ , also  $\hat{t}_{n-1} \approx_{\text{MC}} s'$ .

We can repeat exactly these arguments with decreasing indices until we reach  $t_1$ , which establishes the claim. The case that  $s_1 \xRightarrow{\chi(r)} \mu$  follows completely analogously, as well as the final weak stability condition.  $\square$

**Lemma B.2.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \longrightarrow, \dashv\!\!\!\rightarrow)$ . Let  $\mu_1$  and  $\mu_2$  be two distributions over  $S$ . If

$$\exists \mu'_2 : \mu_2 \Longrightarrow_c \mu'_2 \wedge \mu_1 \mathcal{L}(\approx_{\text{MC}}) \mu'_2$$

and accordingly

$$\exists \mu'_1 : \mu_1 \Longrightarrow_c \mu'_1 \wedge \mu_2 \mathcal{L}(\approx_{\text{MC}}) \mu'_1.$$

Then  $\mu_1 \mathcal{L}(\approx_{\text{MC}}) \mu_2$ .

*Proof.* Let  $[s]$  denote the equivalence class of  $s$  with respect to  $\approx_{\text{MC}}$ .

It suffices to show that  $\mu_i \mathcal{L}(\approx_{\text{MC}}) \mu'_i$  for  $i \in \{1, 2\}$ . By symmetry, it suffices to consider the problem for  $i = 1$ . Let  $\mu^1 := \mu_1$  and  $\mu^2 := \mu'_1$ .

We first note that as  $\approx_{\text{MC}}$  is a weak IMC bisimulation, it is straightforward to derive that whenever  $\mu \mathcal{L}(\approx_{\text{MC}}) \mu'$  and  $\mu \Rightarrow_c \gamma$  then also  $\mu' \Rightarrow_c \gamma'$  for some  $\gamma'$  and  $\gamma \mathcal{L}(\approx_{\text{MC}}) \gamma'$ .

Coming back to the proof, recall that we have just said that whenever  $\mu \mathcal{L}(\approx_{\text{MC}}) \mu'$  and  $\mu \Rightarrow_c \gamma$  then also  $\mu' \Rightarrow_c \gamma'$  for some  $\gamma'$  and  $\gamma \mathcal{L}(\approx_{\text{MC}}) \gamma'$ . With this we can conclude that from the premises of the lemma it follows that  $\mu^1 \Rightarrow_c \mu^2 \Rightarrow_c \mu^3 \Rightarrow_c \mu^4 \Rightarrow_c \mu^5 \dots$  with  $\mu^1 \mathcal{L}(\approx_{\text{MC}}) \mu^3 \mathcal{L}(\approx_{\text{MC}}) \mu^5 \dots$  and  $\mu^2 \mathcal{L}(\approx_{\text{MC}}) \mu^4 \mathcal{L}(\approx_{\text{MC}}) \mu^6 \dots$ . As our state space is finite, each state can thus only reach a finite number of different states. Thus, only finitely many different distributions  $\mu^k$  with  $k \in \{1, 2, 3, 4, \dots\}$  can exist. Hence, for some  $k, k' \in \mathbb{N}$  with  $k' > k$  we have  $\mu^k = \mu^{k'}$ . If  $k$  have different parity we are done. To see this, that without loss of generality,  $k$  is odd and  $k'$  is even. Then  $\mu^1 \mathcal{L}(\approx_{\text{MC}}) \mu^k = \mu^{k'} \mathcal{L}(\approx_{\text{MC}}) \mu^2$ . Now assume they have the same parity. Without loss of generality, assume  $k$  is odd and pick any even number  $k < l < k'$ . Then  $\mu^k \Rightarrow_c \mu^l$  and  $\mu^l \Rightarrow_c \mu^{k'}$  must hold and furthermore,  $\mu^k \mathcal{L}(\approx_{\text{MC}}) \mu^1$  and  $\mu^l \mathcal{L}(\approx_{\text{MC}}) \mu^2$ . If we can now show that  $\mu^k \mathcal{L}(\approx_{\text{MC}}) \mu^l$ , we are done. Thus, it suffices to show that for arbitrary distributions  $\mu$  and  $\mu'$  over states it holds that

$$\text{if } \mu \Rightarrow_c \mu' \text{ and } \mu' \Rightarrow_c \mu \text{ then } \mu \mathcal{L}(\approx_{\text{MC}}) \mu'.$$

Let  $\mathcal{C}$  and  $\mathcal{C}'$  be two *different* equivalence classes of  $S/\approx_{\text{MC}}$ . We write  $\mathcal{C} \rightsquigarrow \mathcal{C}'$  if  $\mu(\mathcal{C}) > 0$  and there is a state  $s$  in  $\mathcal{C} \cap \text{Supp}(\mu)$  such that within the transition  $\mu \Rightarrow_c \mu'$ ,  $s$  reaches a state  $t \in \mathcal{C}'$  with non-zero probability. More formally, to one of the weak hyper transitions of  $\mu$  that constitute the weak *combined* hyper transition to  $\mu'$ , the state  $s$  contributes the weak transition  $s \Rightarrow t$ . Intuitively,  $\mathcal{C} \rightsquigarrow \mathcal{C}'$  denotes that some probability mass of  $\mathcal{C}$  is converted into mass of  $\mathcal{C}'$  when  $\mu \Rightarrow_c \mu'$ . We write  $\mathcal{C} \rightsquigarrow' \mathcal{C}'$  if the above holds with the roles of  $\mu$  and  $\mu'$  exchanged.

From these definitions, we can derive that if for no single class  $\mathcal{C}$  there exists  $\mathcal{C}'$  such that  $\mathcal{C} \rightsquigarrow \mathcal{C}'$ , then  $\mu \mathcal{L}(\approx_{\text{MC}}) \mu'$ . Now assume the contrary, and let  $\mathcal{C} \rightsquigarrow \mathcal{C}'$ , and assume that  $\mathcal{C}' \not\rightsquigarrow' \mathcal{C}$  and  $\mathcal{C}' \not\rightsquigarrow \mathcal{C}$ . This means that neither in the transition  $\mu \Rightarrow_c \mu'$  nor in the transition  $\mu' \Rightarrow_c \mu$  states contained in  $\mathcal{C}'$  transition to a state of another class. Thus, it must be the case that  $\mu(\mathcal{C}') \geq \mu'(\mathcal{C}')$  and  $\mu'(\mathcal{C}') \geq \mu(\mathcal{C}')$ , as in every transition it can only be the case that  $\mathcal{C}'$  gains probability mass. As a consequence  $\mu(\mathcal{C}') = \mu'(\mathcal{C}')$ . At the same time, as  $\mathcal{C} \rightsquigarrow \mathcal{C}'$ , it must be the case that  $\mu'(\mathcal{C}') > \mu(\mathcal{C})$ . A contradiction. Thus, it cannot be the case that  $\mathcal{C}' \not\rightsquigarrow' \mathcal{C}$  and  $\mathcal{C}' \not\rightsquigarrow \mathcal{C}$  for any class  $\mathcal{C}'$  that has an incoming flow. As a consequence, there must be a infinite sequence of classes  $\mathcal{C}_i$  such that  $\mathcal{C}_i \rightsquigarrow \mathcal{C}_{i+1}$  or  $\mathcal{C}_i \rightsquigarrow' \mathcal{C}_{i+1}$ . As we have a finite state space, there must exist indices  $i, j$  with  $i > j$  and  $\mathcal{C}_i = \mathcal{C}_j$ . For a convenient notation, let  $\mathcal{D}_1 := \mathcal{C}_i$ ,  $\mathcal{D}_2 := \mathcal{C}_{i+1}$  and so on, ending with  $\mathcal{D}_n := \mathcal{C}_j$ , where  $n := j - i + 1$ . By the definition of  $\rightsquigarrow$  and  $\rightsquigarrow'$ , this implies that for every  $i \in \{1, \dots, n-1\}$  there must exist states  $s_i \in \mathcal{D}_i$ ,  $s_{i+1} \in \mathcal{D}_{i+1}$  such that  $s_i \Rightarrow s_{i+1}$ . Finally, there must also exist a state  $s_n \in \mathcal{D}_n$  with  $s_n \Rightarrow s'_n$  with  $s'_n \in \mathcal{D}_1$ . By Lemma B.1, this immediately implies that  $s_1 \approx_{\text{PA}} s_2$ , which in turn implies that  $\mathcal{D}_1 = \mathcal{D}_2$ . However, this is a contradiction to the definition of  $\rightsquigarrow$  and  $\rightsquigarrow'$ .

Hence, we arrive at a contradiction as soon as we assume that there exists a class  $\mathcal{C}$  with  $\mathcal{C} \rightsquigarrow$ . Hence,  $\mu \mathcal{L}(\approx_{\text{MC}}) \mu'$ .  $\square$

*Proof of the Lemma.* We show that  $\approx_{\text{MC}}$  satisfies the conditions of Definition 5.8, and is thus a bisimulation in that sense. Furthermore, we show that every bisimulation satisfying the conditions of Definition 5.8 also satisfies Definition 5.7.



The first condition of the two definitions only differs by the use of combined weak transitions. As every non-combined weak transition is also a combined transition, it is obvious that every weak IMC bisimulation satisfies the conditions of the lemma. Now assume that  $s \xrightarrow{a} s'$  and  $t \xRightarrow{a}_c \mu$  and  $\delta(s') \mathcal{L}(\mathcal{R}) mu$ . The last statement implies that for every  $t' \in \text{Supp}(\mu)$  it holds that  $s' \mathcal{R} t'$ . Now, in an IMC, the weak transitions of which the weak combined transition from  $t$  is composed of, must all end in Dirac distributions. Hence, for an arbitrary  $t' \in \text{Supp}(\mu)$  we can derive that  $t \xRightarrow{a} t'$  and  $s' \mathcal{R} t'$ .

It is easy to see by the laws of logic that the third clause of Definition 5.7 follows immediately from the second clause of Definition 5.8.

Only the second clause remains to be considered. First, recall that  $\approx_{\text{IMC}}$  is a weak IMC bisimulation as given by Definition 5.7 by Theorem 5.2, i.e. it satisfies

$$s \xrightarrow{\chi(r)} \mu \text{ for some } r \in \mathbb{R}_{\geq 0} \text{ implies } t \xRightarrow{\chi(r)}_c \gamma \text{ for some } \gamma \text{ and } \mu \mathcal{L}(\approx_{\text{IMC}}) \gamma.$$

We now show that it also satisfies

$$s \downarrow \text{ implies } t \Longrightarrow t' \text{ and } t' \downarrow \text{ for some } t' \text{ and}$$

$$\sum_{(s,x,s') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x = \sum_{(t's,x,t'') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x.$$

Observe that the notation  $s \xrightarrow{\chi(r)} \mu$  implies that  $s \downarrow$ , and  $t \xRightarrow{\chi(r)}_c \gamma$  implies that  $t \Longrightarrow t'$  for some  $t'$  with  $t' \downarrow$  and  $t' \xrightarrow{\chi(r)} \circ \Longrightarrow_c \gamma$ . The first observation follows immediately from the definitions. For the second observation, we use the fact that whenever  $s \Longrightarrow_c \mu$  then  $s \Longrightarrow s'$  for every  $s' \in \text{Supp}(\mu)$  in an IMC. Obviously, these observations satisfy the statements in the first line of the second clause concerning stability.

For the rest, consider that if  $t \Longrightarrow t' \downarrow$ , then also  $s \approx_{\text{IMC}} t'$ , as  $s \approx_{\text{IMC}} t$  by assumption and since  $s \downarrow$ ,  $s$  must match the transition of  $t$  directly. The distributions  $\mu_1$  and  $\mu_2$  that  $s$  and  $t$  reach respectively via  $\xrightarrow{\chi(r)}$  are unique. Thus,  $s \xrightarrow{\chi(r)} \mu_1$  and  $t \xrightarrow{\chi(r)} \mu_2$ . Furthermore, by the bisimulation condition, there must exist  $\mu'_1$  such that  $\mu_1 \Longrightarrow_c \mu'_1$  and  $\mu'_1 \mathcal{L}(\approx_{\text{IMC}}) \mu_2$ , and accordingly, there must exist  $\mu'_2$  such that  $\mu_2 \Longrightarrow_c \mu'_2$  and  $\mu_1 \mathcal{L}(\approx_{\text{IMC}}) \mu'_2$ . By Lemma B.2, this implies that already  $\mu_1 \mathcal{L}(\approx_{\text{IMC}}) \mu_2$ . However, this condition obviously implies

$$\sum_{(s,x,s') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x = \sum_{(t's,x,t'') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x$$

for every class  $\mathcal{C}$ , and thus also for a fixed class, as demanded by Definition 5.8.

It remains to show that any bisimulation  $\mathcal{R}$  according to Definition 5.8 satisfies the second condition of Definition 5.7. The only challenge is to show that if  $t \Longrightarrow t' \downarrow$  in order to satisfy

$$\sum_{(s,x,s') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x = \sum_{(t's,x,t'') \in (-\Box \cap S \times \mathbb{R}_{>0} \times \mathcal{C})} x$$

for a fixed class  $\mathcal{C}$ , then exactly the same state  $t'$  is also suited to satisfy the equation for an arbitrary other class  $\mathcal{C}'$ . To see this, it suffices to show that actually  $s \mathcal{R} t'$ , since then  $t'$  has to

be able to satisfy this equation immediately, as it cannot perform any further  $\tau$  transitions. Now  $s \mathcal{R} t'$  follows as  $s \mathcal{R} t$  and  $t \implies t'$  implies that  $s \implies s'$  and  $s' \mathcal{R} t'$ , but by assumption,  $s' \downarrow$  and thus  $s = s'$ .  $\square$

## C. Proofs of Chapter 7

### C.1. Lemma 7.5

Let  $\mathcal{A} = (S, \bar{s}, Act, \dashv, \dashv)$  be a Markov automaton. Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be sequences of distribution over  $S$ . Let furthermore

- (a)  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  for some distribution  $\gamma$ , and
- (b)  $\gamma_i \xRightarrow{a}_c \gamma'_i$ .

Then, there exists a distribution  $\gamma'$  such that

- (A)  $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ , and
- (B)  $\gamma \xRightarrow{a}_c \gamma'$ .

Additionally, if each  $\gamma'_j$  satisfies  $\gamma'_j \downarrow$  for  $j \in J$ , then also  $\gamma' \downarrow$ .

*Proof.* The following proof relies on our compactness assumption for Markov automata (cf. Chapter 7) and Lemma 7.3 (i).

Assumptions (a) and (b) provide that  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  for some distribution  $\gamma$ , and  $\gamma_i \xRightarrow{a}_c \gamma'_i$  for each  $i \in \mathbb{N}$ . As a preliminary step we show that then also  $\gamma \xRightarrow{a}_c$ . Assume the contrary, i.e. there must exist a state  $s \in \text{Supp}(\gamma)$  with  $s \not\xRightarrow{a}$ . Let  $\varepsilon = \gamma(s)$ , and let  $\gamma_j$  denote a distribution with  $d(\gamma, \gamma_j) < \varepsilon$ . Then, also  $s \in \text{Supp}(\gamma)$  must hold, because otherwise the absence of  $s$  in  $\text{Supp}(\gamma_j)$  would at least provoke a difference of  $\varepsilon$  in the metric distance between the two distributions, i.e.  $d(\gamma, \gamma_j) \geq \varepsilon$ . This is clearly a contradiction to our choice of  $\gamma_j$ .

Now, that we know that  $\gamma$  is able to perform weak  $a$ -transitions, we will show that is can simulate the  $a$ -transitions of each of the  $\gamma_i$  up to some small error. Whenever  $\gamma_i \xRightarrow{a}_c \gamma'_i$ , by Lemma 4.3, we know that

$$\gamma'_i = \bigoplus_{s \in \text{Supp}(\gamma_i)} \gamma_i(s) \mu'_s$$

for suitable distributions  $\mu'_s$ , of which each satisfies  $s \xRightarrow{a}_c \mu'_s$ . From here we construct a weak combined transition  $\gamma \xRightarrow{a}_c \gamma''$  of the distribution  $\gamma$  as follows: Let

$$\gamma'' = \bigoplus_{s \in \text{Supp}(\gamma)} \gamma(s) \gamma'_s$$

where for each state  $s \in \text{Supp}(\gamma)$  the distribution  $\gamma'_s$  is either

- $\mu'_s$  if  $s \in \text{Supp}(\gamma_i)$ , or

- some arbitrary distribution  $\rho$  satisfying  $s \xRightarrow{a}_c \rho$ .

By the construction of  $\gamma''_i$ , it follows that

$$d(\gamma'_i, \gamma''_i) \leq d(\gamma_i, \gamma). \quad (\text{C.1})$$

By Lemma 7.3 (i), there is a converging subsequence  $\langle \gamma''_i \rangle_{i \in J}$ . We then choose the distribution  $\gamma'$ , whose existence we want to prove (conclusion (B)), to be the limiting distribution of  $\langle \gamma''_i \rangle_{i \in J}$ , i.e.

$$\langle \gamma''_j \rangle_{j \in J} \longrightarrow \gamma'. \quad (\text{C.2})$$

Now, we establish that  $\gamma'$  indeed satisfies all desired properties:

1. By the compactness assumption for Markov automata and by Corollary 7.1, from  $\gamma \xRightarrow{a}_c \gamma''_j$  and Equation C.2, it immediately follows that  $\gamma \xRightarrow{a}_c \gamma'$ . This establishes conclusion (B) of the lemma.
2. As  $d(\gamma'_i, \gamma''_i) \leq d(\gamma_i, \gamma)$  (Equation C.1) and  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  (assumption (a)), also

$$d(\gamma'_i, \gamma''_i) \longrightarrow 0.$$

This and Equation C.2 together immediately imply

$$\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'. \quad (\text{C.3})$$

Hence,  $\langle \gamma'_j \rangle_{j \in J}$  is the subsequence we sought to establish conclusion (A) of the lemma.

For the additional condition, assume  $\mathcal{P}(\gamma'_j)$ . Due to Equation C.3, every state  $s \in \text{Supp}(\gamma')$  must also be a state of  $\text{Supp}(\gamma'_k)$  for some  $\gamma'_k$ . Because otherwise,  $d(\gamma', \gamma'_k) \geq \gamma'(s)$  for all  $k$ , which is a contradiction to  $\gamma'$  being the limiting distribution. As every state  $s \in \text{Supp}(\gamma'_j)$  must satisfy  $\mathcal{P}(s)$  (as  $\mathcal{P}(\gamma'_j)$ ), hence also  $\mathcal{P}(\gamma')$ .  $\square$

## C.2. Lemma 7.6

Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \xrightarrow{\quad}, \dashv\!\!\rightarrow)$  be a Markov automaton. Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be sequences of distribution over  $S$ . Let furthermore

- (i)  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$  for some distribution  $\gamma$ , and
- (ii)  $\gamma_i \xRightarrow{a}_c \xi_i$  for some  $\xi_i$ , and
- (iii)  $\gamma \xRightarrow{a}_c \gamma'$ .

Then, there exist distributions  $\gamma'_i$  such that

- $\gamma_i \xRightarrow{a}_c \gamma'_i$ , and
- $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ .

In addition: if  $\gamma \xRightarrow{a}_c \gamma'$  is a finitary transitions, also all  $\gamma_i \xRightarrow{a}_c \gamma'_i$  are finitary transitions.

*Proof.* When  $\gamma \xRightarrow{a}_c \gamma'$ , we know by Lemma 4.3 that

$$\gamma' = \bigoplus_{s \in \text{Supp}(\gamma)} \gamma(s) \mu'_s$$

for suitable distributions  $\mu'_s$ , of which each satisfies  $s \xRightarrow{a}_c \mu'_s$  for some  $\mu'_s$ . By Precondition (ii) of the lemma, we know that each distribution  $\gamma_i$  can perform a *finitary* transition with action label  $a$  to some distribution  $\xi_i$ . This implies that each state  $s \in \text{Supp}(\gamma_i)$  can perform some finitary weak transition  $s \xRightarrow{a}_c \gamma'_s (\star)$ . We claim, that each of the distributions  $\gamma_i$  can perform a weak combined transition  $\gamma_i \xRightarrow{a}_c \gamma'_i$  where

$$\gamma'_i = \bigoplus_{s \in \text{Supp}(\gamma)} \gamma_i(s) \gamma'_s$$

where for each state  $s \in \text{Supp}(\gamma)$  the distribution  $\gamma'_s$  is either

- $\mu'_s$  if  $s \in \text{Supp}(\gamma)$ , or
- $\gamma'_s$ , otherwise.

By the construction of  $\gamma'_i$ , it follows that

$$d(\gamma', \gamma'_i) \leq d(\gamma_i, \gamma).$$

As  $d(\gamma', \gamma'_i) \leq d(\gamma_i, \gamma)$  and  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \rightarrow \gamma$ , also  $d(\gamma_i, \gamma) \rightarrow 0$ , and hence

$$d(\gamma', \gamma'_i) \rightarrow 0 \text{ and therefore } \langle \gamma'_i \rangle_{i \in \mathbb{N}} \rightarrow \gamma'$$

The additional property concerning preservation of finitary transitions follows immediately from the construction and  $(\star)$ .  $\square$



## D. Proofs of Chapter 8

### D.1. Lemma 8.5

The proof of this lemma makes use of another variant of weak challenger weak distribution bisimulation, *finitary transitions* weak challenger weak distribution bisimulation. This variant is defined exactly as weak challenger weak distribution bisimulation, with the restriction, that in each condition the transition of the challenger may only be finitary.

**Lemma D.1.** In weakly image-finite MA, every *finitary transitions* weak challenger weak distribution bisimulation is contained in some weak challenger weak distribution bisimulation.

*Proof.* Let  $\mathcal{R}$  be a finitary transitions weak challenger weak distribution bisimulation. Let  $\infty(\mu)$  be the set of all distributions  $\mu'$  that are reachable by an infinitary hyper-transition of the distribution  $\mu$ . More formally, let  $\mu'$  be an element of  $\infty(\mu)$  if and only if  $\mu \xRightarrow{a}_c \mu'$  with an *infinitary transition* and arbitrary action  $a$ .

We now construct a binary relation  $\mathcal{R}' \supseteq \mathcal{R}$  that we will show to be a weak challenger weak distribution bisimulation. We let  $\mathcal{R}'$  contain the pair  $(\mu', \gamma')$  if and only if  $(\mu', \gamma') \in \mathcal{R}$  or

1. there exists a convergent sequence  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  with limiting distribution  $\mu'$ , and
2. there exists a convergent sequence  $\langle \gamma \rangle_{i \in \mathbb{N}}$  with limiting distribution  $\gamma'$ , and
3.  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in \mathbb{N}$ .

We call the infinite sequences of distributions the *witness sequences* of  $(\mu', \gamma')$ . Note that it might be the case that a single pair has several witness sequences or distributions.

It remains to show that  $\mathcal{R}'$  is a weak challenger weak distribution bisimulation. We proceed by a case distinction between the pairs in  $\mathcal{R}'$  that are contained in  $\mathcal{R}$ , and the additional pairs that are unique to  $\mathcal{R}'$ .

**Case 1** Let  $\mu \mathcal{R}' \gamma$  because  $\mu \mathcal{R} \gamma$ . Each of the three bisimulation conditions is met immediately for finitary challenger transitions, since  $\mathcal{R} \subseteq \mathcal{R}'$ . Thus, we focus our attention to infinitary transitions.

The proof is structured according to the three bisimulation condition of Definition 8.9.

**Condition (a)** We assume that  $\mu \xRightarrow{a}_c \mu'$  and  $\mu' \in \infty(\mu')$  for some action  $a$  and distribution  $\mu'$ .

**Case 1.1** First assume that  $a = \tau$ . We now construct a sequence  $\langle \mu_i \rangle_{i \in I}$  limiting distribution  $\mu'$  satisfying  $\mu \xRightarrow{\tau}_c \mu_i$  for each  $i \in I$ . Basically, we take the original transition tree inducing  $\mu \xRightarrow{\tau}_c \mu'$  and “cut” it at level  $i$ . The  $\mu_i$  are then the distributions induced by the cut transition trees. Formally, if  $\mathcal{T}$  is the transition tree of our original transition, we define the cut transition tree  $\mathcal{T}_i$  by letting

$dom(\mathcal{T}_i) = \{ \sigma \in dom(\mathcal{T}) \mid |\sigma| \leq i \}$ , for all inner nodes  $\sigma$  letting  $\mathcal{T}_i(\sigma) = \mathcal{T}(\sigma)$  and for all leave nodes  $\sigma$ , letting

$$\begin{aligned} Sta_{\mathcal{T}_i}(\sigma) &= Sta_{\mathcal{T}}(\sigma) \\ Prob_{\mathcal{T}_i}(\sigma) &= Prob_{\mathcal{T}}(\sigma) \\ Act_{\mathcal{T}_i}(\sigma) &= \perp \end{aligned}$$

Since  $\mu \mathcal{R} \gamma$ , for each  $\mu_i$  there must exist a  $\gamma_i$  with  $\gamma \Rightarrow_c \gamma_i$  and  $\mu_i \mathcal{R} \gamma_i$ . By Lemma 7.3 (ii), it follows that there is an infinite subset  $J$  of  $I$  such that  $\langle \gamma_i \rangle_{i \in J}$  has a limiting distribution  $\gamma'$ . By definition  $\mu' \mathcal{R} \gamma'$ . This suffices to establish Condition (a) for the case  $a = \tau$ .

**Case 1.2** Now assume  $a \neq \tau$ . Again, we proceed by a case distinction.

**Case 1.2.1** Assume, that there exists  $\mu'' \notin \infty(\mu)$  and  $\mu'''$  with  $\mu \Rightarrow_c \mu'' \xrightarrow{a} \mu''' \Rightarrow_c \mu'$ . In words, this means that the infinitary transition of  $\mu$  has a transition tree that is finitary up to the point where action  $a$  occurs, and becomes infinitary only afterwards. In this case, also  $\mu''' \notin \infty(\mu)$ . Then, since  $\mu \mathcal{R} \gamma$ , it immediately follows that there is  $\gamma'''$  with  $\gamma \xRightarrow{a}_c \gamma'''$  and  $\mu''' \mathcal{R} \gamma'''$ . Now, by assumption  $\mu''' \Rightarrow_c \mu'$  and  $\mu' \in \infty(\mu''')$ . It remains to show that also  $\gamma''' \Rightarrow_c \gamma'$  for some  $\gamma'$  and  $\mu' \mathcal{R}' \gamma'$ , since then we could join the transition sequence of  $\gamma \xRightarrow{a}_c \gamma''' \Rightarrow_c \gamma'$  into one transition  $\gamma \xRightarrow{a}_c \gamma'$ , which suffices to prove this case. The existence of such a  $\gamma'$  however immediately follows from the Case 1.1, since  $\mu''' \mathcal{R} \gamma'''$ .

**Case 1.2** We again rewrite  $\mu \xRightarrow{a}_c \mu'$  as  $\mu \Rightarrow_c \mu'' \xrightarrow{a} \mu''' \Rightarrow_c \mu'$ . However this time, let  $\mu'' \in \infty(\mu)$ . This means that the transition of  $\mu$  is infinitary already before action  $a$  occurs. By the already proven Case 1.1, we know that there exists  $\gamma''$  such that  $\gamma \Rightarrow_c \gamma''$  and  $\mu'' \mathcal{R}' \gamma''$ . If we can now show that the infinitary transition  $\mu'' \xRightarrow{a}_c \mu'$  can be matched by  $\gamma''$ , we are done by combining transitions again. The proof of this fact will follow in Case 2.4 below.

**Case 1.2.3** We now assume that both  $\mu''$  and  $\mu'''$  are contained in  $\infty(\mu)$ . This case is a combination of the two cases before. We skip the details here.

**Condition (b)** For Condition (b), let  $\mu \Rightarrow_c \mu^a \oplus_p \mu^b$  for some distributions  $\mu^a$  and  $\mu^b$  with  $\mu^a \oplus_p \mu^b \in \infty(\mu)$ .

Let  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  be a convergent infinite sequence of distributions with limiting distribution  $\mu'$  satisfying  $\mu \Rightarrow_c \mu_i$  for each  $i \in \mathbb{N}$ . We have already argued that such a sequence exists in Case 1.1. Now, in addition, we demand that for each  $\mu_i$  there exists a splitting  $\mu_i^a \oplus_{p_i} \mu_i^b$  such that  $\langle \mu_i^a \rangle_{i \in \mathbb{N}}$  has limiting distribution  $\mu^a$  and  $\langle \mu_i^b \rangle_{i \in \mathbb{N}}$  has limiting distribution  $\mu^b$  and  $\langle p_i \rangle_{i \in \mathbb{N}}$  has limit  $p$ . Constructing such splittings is straightforward: For every state  $s \in Supp(\mu^a) \cap Supp(\mu_i)$ , we let  $\mu_i^a(s) = \min(\mu^a(s), \mu_i(s))$ , and let  $\mu_i^b = \mu_i \ominus \mu_i^a$ . Furthermore, we choose  $p_i = p$  for each  $i \in \mathbb{N}$ .

As in the last case, since  $\mu \mathcal{R} \gamma$ , with Condition (b), for each  $\mu_i$  exists  $\gamma_i$  with  $\gamma \Rightarrow_c \gamma_i$  and  $\gamma_i$  has a splitting  $\gamma_i^a \oplus_{p_i} \gamma_i^b$  such that  $\mu_i^a \mathcal{R} \gamma_i^a$  and  $\mu_i^b \mathcal{R} \gamma_i^b$ . We thus have two infinite sequences  $\langle \gamma_i^a \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma_i^b \rangle_{i \in \mathbb{N}}$ . Applying Lemma 7.3 (ii) two times in sequence, we obtain an infinite subset  $J$  of  $\mathbb{N}$  such that both sequences  $\langle \gamma_i^a \rangle_{i \in J}$  and



$\langle \gamma_i^b \rangle_{i \in J}$  have limiting distributions  $\gamma^a$  and  $\gamma^b$ , respectively. In summary, this suffices to show  $\mu^a \mathcal{R}' \gamma^a$  and  $\mu^b \mathcal{R}' \gamma^b$ . By the compactness assumption and Corollary 7.1 furthermore  $\gamma \Rightarrow_c \gamma^a \oplus_p \gamma^b$  follows, which ends this case.

**Condition (c)** For Condition (c), assume  $\mu \Rightarrow_c \mu'$  and  $\mu' \downarrow$ . With an argument as in the Case 1.1, there must exist a converging sequence  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  with limiting distribution  $\mu'$  and  $\mu \Rightarrow_c \mu_i$  for each  $i \in \mathbb{N}$ . We split every  $\mu_i$  into  $\mu_i^\downarrow \oplus_{p_i} \mu'_i$ , where  $\mu_i^\downarrow$  has only stable states in its support, while  $\mu'_i$  only unstable states. Since  $\mu \mathcal{R} \gamma$ , and since  $\mu \Rightarrow_c \mu_i^\downarrow \oplus_{p_i} \mu'_i$  is a finitary transition, there must exist  $\gamma_i^\downarrow \oplus_{p_i} \gamma'_i$  with  $\mu \Rightarrow_c \gamma_i^\downarrow \oplus_{p_i} \gamma'_i$  and  $\mu_i^\downarrow \mathcal{R} \gamma_i^\downarrow$ . Using Condition (c), from there it follows that there must exist  $\gamma_i^{\downarrow'}$  with  $\gamma_i^\downarrow \downarrow$  and  $\gamma_i^\downarrow \Rightarrow_c \gamma_i^{\downarrow'}$ . By combining the two weak transitions, we obtain  $\gamma \Rightarrow_c \gamma_i^{\downarrow'} \oplus_{p_i} \gamma'_i$ .

Since  $\mu' \downarrow$ , the value sequence  $\langle p_i \rangle_{i \in \mathbb{N}}$  must have limit 1. Since  $\gamma_i^{\downarrow'} \downarrow$ , any limiting distribution  $\gamma'$  of any subsequence of  $\langle \gamma_i^{\downarrow'} \oplus_{p_i} \gamma'_i \rangle_{i \in \mathbb{N}}$  must also be stable, i.e.  $\gamma' \downarrow$ . The existence of such a subsequence and its limiting distribution  $\gamma'$  follows by Lemma 7.3. By the compactness property of MA and Corollary 7.1, we obtain  $\gamma \Rightarrow_c \gamma'$ , which ends the proof.

**Case 2** Let  $\mu \mathcal{R}' \gamma$  but not  $\mu \mathcal{R} \gamma$ . Then, by the definition of  $\mathcal{R}'$  there exist convergent infinite sequences  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  and  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  with limiting distribution  $\mu$  and  $\gamma$ , respectively, and  $\mu_i \mathcal{R} \gamma_i$  for each  $i \in \mathbb{N}$ .

The proof is again structured according to the three bisimulation condition of Definition 8.9.

**Condition (a)** Assume  $\mu \xrightarrow{a}_c \mu'$ . We distinguish several cases.

**Case 2.1** Assume  $a = \tau$  and  $\mu' \notin \infty(\mu)$ . Since Lemma 7.6, there exist finitary transitions  $\mu_i \Rightarrow_c \mu'_i$  for each  $i \in \mathbb{N}$  with  $\langle \mu_i \rangle_{i \in \mathbb{N}} \rightarrow \mu'$ . Since the transitions are finitary, and since  $\mu_i \mathcal{R} \gamma_i$ , there exist transitions  $\gamma_i \Rightarrow_c \gamma'_i$  for each  $i \in \mathbb{N}$ . By Lemma 7.3 (i), there exists an infinite index  $J \subseteq \mathbb{N}$  such that  $\langle \gamma'_i \rangle_{i \in J}$  has a limiting distribution  $\gamma'$ . Note that since  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  has limiting distribution  $\mu'$ , also its subsequence  $\langle \mu_i \rangle_{i \in J}$  has limiting distribution  $\mu'$ , and thus, by definition,  $\mu' \mathcal{R}' \gamma'$ . It only remains to show that  $\gamma \Rightarrow_c \gamma'$  to finish our first case. This follow immediately from Lemma 7.5.

**Case 2.2** Assume  $a = \tau$  and  $\mu' \in \infty(\mu)$ . In principle, we can proceed as in the last case. However, our construction of the transitions for each  $\mu_i$  slightly differs. Now, the transitions  $s \Rightarrow_c \mu'_s$  are not guaranteed to be finitary transition anymore, which is crucial for the further proof. We construct a finitary transition as follows. For each  $i \in \mathbb{N}$ , we take the transition tree of  $s \Rightarrow_c \mu'_s$  for each  $s$ , but cut it at a specific level, which we will determine soon. Before, let us say that this yields a finitary transition  $s \Rightarrow_c \mu'_{(s,i)}$  for each  $s \in \text{Supp}(\text{Supp}())\mu_i \cap \text{Supp}(\mu)$  and  $i \in \mathbb{N}$ . As before, the distribution  $(\bigoplus_{s \in T} \mu_i(s) \cdot \mu'_{(s,i)}) \oplus (\bar{\mu}_i)$  converges against  $\mu'$ , if we guarantee that  $d(\mu'_s, \mu'_{(s,i)})$  has limit 0. To ensure this, it suffices to cut the transition tree for each  $i$  at level  $i$ , for instance.

**Case 2.3** Assume  $a \neq \tau$  and  $\mu \xrightarrow{a}_c \mu'$ . First note that there may not exist  $\mu'_i$  such that  $\mu_i \xrightarrow{a}_c \mu'_i$  in general, as there might exist states in  $\text{Supp}(\mu_i) \setminus$

$\text{Supp}(\mu)$ . Let hence  $\mu_i^a$  be the subdistribution of  $\mu_i$  containing all states of  $\text{Supp}(\mu_i)$  that are also in  $\text{Supp}(\mu)$ . Before we continue to work with  $\mu_i^a$ , we normalize it to a full distribution. Since  $\mu \xrightarrow{a}_c$ , also for each  $i$ ,

$$\mu_i^a \xrightarrow{a}_c \xi_i \quad (\text{D.1})$$

for some distribution  $\xi_i$ . Furthermore, as  $\langle \mu_i \rangle_{i \in \mathbb{N}} \rightarrow \mu$ , it must be the case that also  $\langle \mu_i^a \rangle_{i \in \mathbb{N}} \rightarrow \mu$ . Since the transition of Equation D.1 is clearly finitary, by Lemma 7.6, there are distributions  $\mu_i'^a$  with  $\mu_i^a \xrightarrow{a}_c \mu_i'^a$  (with a finitary transition) and  $\langle \mu_i'^a \rangle_{i \in \mathbb{N}} \rightarrow \mu'$ . Since  $\mu_i \mathcal{R} \gamma_i$ , there exists distributions  $\gamma_i^a \oplus_{p_i} \gamma_i^b$  such that  $\gamma_i \xRightarrow{a}_c \gamma_i^a \oplus_{p_i} \gamma_i^b$  and  $\mu_i^a \mathcal{R} \gamma_i^a$  by Condition (b) for  $\mathcal{R}$ . By Lemma 7.3, there is an infinite index set  $J \subseteq \mathbb{N}$  such that

$$\langle \gamma_i^a \rangle_{i \in \mathbb{N}} \rightarrow \gamma^a \quad (\text{D.2})$$

for some distribution  $\gamma^a$ . Furthermore, as  $\langle p_i \rangle_{i \in \mathbb{N}} \rightarrow 1$ , also

$$\langle \gamma_i^a \oplus_{p_i} \gamma_i^b \rangle_{i \in \mathbb{N}} \rightarrow \gamma^a \quad (\text{D.3})$$

From  $\mu_i^a \mathcal{R} \gamma_i^a$  and  $\mu_i^a \xrightarrow{a}_c \mu_i'^a$  it follows that for some distribution  $\gamma_i'^a$ , also  $\gamma_i^a \xRightarrow{a}_c \gamma_i'^a$  and  $\mu_i'^a \mathcal{R} \gamma_i'^a$ . By Lemma 7.3, there is a sequence  $\langle \gamma_i'^a \rangle_{i \in K}$  for some infinite subset  $K \subseteq J$  of  $\mathbb{N}$  that has some limiting distribution  $\gamma'^a$ . By Corollary 7.1 and Equation D.2, also

$$\gamma^a \xRightarrow{a}_c \gamma'^a. \quad (\text{D.4})$$

So far, the existence of the sequences  $\langle \mu_i'^a \rangle_{i \in K}$  and  $\langle \gamma_i'^a \rangle_{i \in K}$  and the fact that  $\mu_i'^a \mathcal{R} \gamma_i'^a$  for all  $i \in K$ , suffice to establish  $\mu' \mathcal{R}' \gamma'^a$ .

It remains to show that  $\gamma \xRightarrow{a}_c \gamma'^a$ . By Lemma 7.5,  $\gamma \xRightarrow{a}_c \gamma^a$ . Together with Equation D.4 and joining the weak hypertransitions, we obtain the desired result.

**Case 2.4** Assume  $a \neq \tau$  and  $\mu \xrightarrow{a}_c \mu'$ . Then there exist  $\mu''$  and  $\mu'''$  with  $\mu \xRightarrow{a}_c \mu'' \xrightarrow{a}_c \mu''' \xRightarrow{a}_c \mu'$ . From Cases 2.1 and 2.2, we know that there exists  $\gamma''$  such that  $\gamma \xRightarrow{a}_c \gamma''$  and  $\mu'' \mathcal{R}' \gamma''$ . With Case 2.3, we derive that there exists  $\gamma'''$  such that  $\gamma'' \xRightarrow{a}_c \gamma'''$  and  $\mu''' \mathcal{R}' \gamma'''$ . Finally with Cases 2.1 and 2.2, we derive again that there exists  $\gamma'$  with  $\gamma''' \xRightarrow{a}_c \gamma'$  and  $\mu' \mathcal{R}' \gamma'$ . By joining the weak hypertransitions we finally obtain  $\gamma \xRightarrow{a}_c \gamma'$ .

**Condition (b)** The proof for Condition (b) reiterates the same ideas as the proof of Condition (a). The only difference is the initial argument, where we need to make clear that if  $\mu \xRightarrow{a}_c \mu^a \oplus_p \mu^b$  and  $\langle \mu_i \rangle_{i \in \mathbb{N}} \rightarrow \mu$ , we can then construct transitions and splittings  $\mu_i \xRightarrow{a}_c \mu_i^a \oplus_p \mu_i^b$  such that  $\langle \mu_i^a \rangle_{i \in \mathbb{N}} \rightarrow \mu^a$  and  $\langle \mu_i^b \rangle_{i \in \mathbb{N}} \rightarrow \mu^b$ . We leave the details to the reader.

**Condition (c)** Again, the general proof idea follows the same technical detail. As for Condition (b), only the initial argument is unique. Assume  $\mu \xRightarrow{a}_c \mu' \downarrow$  for some  $\mu'$ . Then as in the proof of Case 1.1, we construct transitions of the  $\mu_i$

with  $\langle \mu_i \rangle_{i \in \mathbb{N}} \longrightarrow \mu$  as follows: for each  $i$ , we cut the infinite state transition tree of  $\mu \Longrightarrow_c \mu'$  at level  $i$ . Call the distribution induced by this transition tree  $\mu'_i$ . By construction, it is obvious that  $\langle \mu'_i \rangle_{i \in \mathbb{N}} \longrightarrow \mu'$ . Note that  $\mu'_i \downarrow$  is not necessarily satisfied for any  $i \in \mathbb{N}$ . This situation is hence completely analogous to Case 2.3, and can be shown following the ideas there.

□

### D.1.1. Proof of Lemma 8.5

On weakly image-finite MA, two distribution  $\mu$  and  $\gamma$  are weak distribution bisimilar *if and only if* they are related by some weak challenger weak distribution bisimulation.

*Proof.* It is obvious that a weak challenger weak distribution bisimulation is also a weak distribution bisimulation, as we strengthen the conditions of the challenger in all of the three conditions. In Condition (a'), we allow a weak combined transition instead of a strong transition transitions. In Conditions (b') and (c'), we add the ability to the challenger to perform an additional weak combined internal hyper-transition.

For the proof of the other direction, i.e. that every weak distribution bisimulation is also a weak challenger weak distribution bisimulation, we actually show in the following that every weak distribution bisimulation is also a *finitary transitions* weak challenger weak distribution bisimulation. We can then obtain the desired result by applying Lemma D.1.

In fact, weak distribution bisimulation is not a finitary transitions weak challenger weak distribution bisimulation itself in general. However, we will show that each such bisimulation can be amended by additional pairs of distributions, such that the amended relation is a finitary transitions weak challenger weak distribution bisimulation, while remaining a weak distribution bisimulation. More formally, we show that there exists a weak distribution bisimulation  $\mathcal{R}'$  satisfying  $\mathcal{R} \subseteq \mathcal{R}'$ , that is also a finitary transitions weak challenger weak distribution bisimulation.

Concretely, we choose  $\mathcal{R}' := \mathcal{R}^\oplus$ . By Lemma 8.3,  $\mathcal{R}'$  is then also a weak distribution bisimulation. For simplicity of notation, we assume  $\mathcal{R} = \mathcal{R}^\oplus$  in the following without loss of generality.

Consider an arbitrary pair  $\mu \mathcal{R} \gamma$ . We first consider Condition (a) vs Condition (a'). We will show that if  $\mu \mathcal{R} \gamma$ , then also whenever  $\mu \xRightarrow{a}_c \mu'$  for some  $\mu'$ , then there also must exist a  $\gamma'$  such that  $\gamma \xRightarrow{a}_c \gamma'$  and  $\mu' \mathcal{R} \gamma'$  (which is exactly Condition (a')). It suffices to give the proof only for non-combined transitions, as the general case follows immediately by the fact that  $\mathcal{R} = \mathcal{R}^\oplus$  and the definition of combined transitions.

Let  $\mu \xRightarrow{a} \mu'$  be a weak hyper-transition of the challenger. By the definition of hyper-transitions, for each state  $s \in \text{Supp}(\mu)$ , there must exist a weak transitions  $s \xRightarrow{a} \mu'_s$  such that

$$\mu' = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \mu'_s.$$

As we only consider weakly image-finite MA, and finitary transitions, we may proceed by induction over the sum of the sizes of the transition trees inducing the transitions  $s \xRightarrow{a} \mu'_s$  for each  $s \in \text{Supp}(\mu)$ . We now represent the transition  $\mu \xRightarrow{a} \mu'$  as the combination of two transitions:

1. either there exists  $\mu_1$  such that  $\mu \xrightarrow{a} \mu_1$  and  $\mu_1 \Longrightarrow \mu'$ , or
2. there exists  $\mu_2 \neq \mu'$  such that  $\mu \Longrightarrow \mu_2$  and  $\mu_2 \xRightarrow{a} \mu'$ .

Note that the two cases are indeed complete, as the second decomposition, where  $\mu$  starts by performing a weak  $\tau$ -transitions, is valid as soon as at least one of the states in  $\text{Supp}(\mu)$  starts its hypertransition  $s \xRightarrow{a} \mu'_i$  with an internal transition. In the first case,  $\mu \xrightarrow{a} \mu_1$ . Clearly, then from  $\mu \mathcal{R} \gamma$  it immediately follows that there exists  $\gamma_1$  such that  $\gamma \xRightarrow{a} \gamma_1$  with  $\mu_1 \mathcal{R} \gamma_1$ . Now,  $\mu_1 \Rightarrow \mu'$  by assumption. As the total size of the transition trees inducing this hypertransition is smaller, we can apply induction. We conclude that there exists  $\gamma'$  such that  $\gamma_1 \Rightarrow_c \gamma'$  and  $\mu' \mathcal{R} \gamma'$ . All that remains to be shown is that indeed also  $\gamma \xRightarrow{a} \gamma'$ . But this holds by Lemma 4.2. In the second case, we can use the induction hypothesis twice, as both the transition from  $\mu$  to  $\mu_2$  and the transition from  $\mu_2$  to  $\mu'$  is constructed with in total smaller transition trees. Then, the proof proceeds as for the first case.

Now, for Condition (b) vs Condition (b'), we first note that by our preceding proof-step for Condition (a'), we can conclude that when  $\mu \Rightarrow_c \mu_1 \oplus_p \mu_2$ , then surely there exists  $\gamma'$  such that  $\gamma \Rightarrow_c \gamma'$  and  $\mu_1 \oplus_p \mu_2 \mathcal{R} \gamma'$ . With this, we have reduced the problem to Condition (b) of Definition 8.4, applied to  $\mu_1 \oplus_p \mu_2$  and  $\gamma'$ . From there, it follows that  $\gamma' \Rightarrow_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_1 \mathcal{R} \gamma_1$  and  $\mu_2 \mathcal{R} \gamma_2$ . Again with Lemma 4.2, we conclude that  $\gamma \Rightarrow_c \gamma_1 \oplus_p \gamma_2$  is possible immediately.

The proof of Condition (c') follows the same pattern as the last proof.  $\square$

## D.2. Lemma 8.7

Let  $\bigoplus_{i \in J} p_i \mu_i$  be a splitting of  $\mu$  with a possibly infinite index set  $J \subseteq \mathbb{N}$ .

If  $\mu \approx \gamma$  then there exist  $\gamma_i$  for each  $i \in J$  such that

- (A)  $\gamma \Rightarrow_c \bigoplus_{i \in J} p_i \gamma_i$ , and
- (B)  $\mu_i \approx \gamma_i$  for each  $i \in J$ .

*Proof.* For readability, we will use for a splitting  $\mu = \bigoplus_{i \in J} p_i \mu_i$  the notation  $(\mu_1 : p_1, \mu_2 : p_2, \dots)$ .

Furthermore, without loss of generality, we assume that the index set  $J$  is contiguous in the sense that if  $n \in J$  for any  $n \in \mathbb{N}$  then also  $n' \in J$  for each  $n' < n$ .

From the possibly infinite splitting, we first construct a (possibly infinite) sequence of finite splittings as follows: for each  $k \in \mathbb{N}$ , the finite splitting agrees with the infinite splitting in the first  $k$  components. The last component of the finite splitting is then the union of all remaining components of the infinite splitting. Let  $\mathbb{N}_k = \{i \leq k \mid i \in \mathbb{N}\}$ . We then define the  $k$ -th finite splitting as

$$(\mu_1 : p_1, \dots, \mu_k : p_k, \bigoplus_{i \in J \setminus \mathbb{N}_k} p_i \mu_i : 1 - \sum_{i \in \mathbb{N}_k} p_i).$$

By construction, this is a finite splitting of  $\mu$  with  $k + 1$  components.

By an inductive argument, we now prove that for every  $k \in \mathbb{N}$  there is a suitable splitting  $\xi^k$  according to the lemma, i.e.  $\xi^k = (\gamma_1^k : p_1, \dots, \gamma_k^k : p_k, \gamma_{k+1}^k : 1 - \sum_{i \in \mathbb{N}_k} p_i)$  with

1.  $\gamma \Rightarrow_c \xi^k$  and
2. for each  $i \in \{1, \dots, k\} : \mu_i \approx \gamma_i^k$  and also  $\bigoplus_{i \in J \setminus \mathbb{N}_k} p_i \mu_i \approx \gamma_{k+1}^k$ .

3. In addition, we will ensure that for every  $i$  the splittings  $\xi^i$  and  $\xi^{i+1}$  agree on the first  $i$  components. Formally,

$$\bigoplus_{l \leq i} p_i \gamma_l^i = \bigoplus_{l \leq i} p_i \gamma_l^{i+1}. \quad (\text{D.5})$$

Before we come to the inductive argument, we will show how the last condition helps us to establish the lemma. By construction it holds that  $d(\xi^i, \xi^{i+1}) \rightarrow 0$ , as the sum of the probability masses of the splitting components that are *identical* converges to 1 while  $i$  goes to infinity. Hence, the sequence  $\langle \xi^i \rangle_{i \in \mathbb{N}}$  has a limiting distribution  $\xi$ . As we only consider compact MA, it must furthermore hold by Corollary 7.1 that  $\gamma \Rightarrow_c \xi$ . This establishes Condition (A) of the lemma. For the Condition (B), i.e.  $\mu_i \approx \gamma_i$  for each  $i \in J$ , we assume that  $j$  is the smallest index for which the condition does not hold. By construction of the splittings  $\xi^i$ ,  $\gamma_j$  is the  $j$ -th component of the splitting  $\xi^j$  and all later splittings, i.e.  $\xi^l$  with  $l > j$  and thus also of the limiting distribution  $\xi$ . From property b), however, it follows that  $\mu_j \approx \gamma_j$ , contradicting our assumption of  $j$  being the smallest index where this property does not hold.

We now proceed with the inductive proof. The case  $k = 0$  is trivial, as the splitting has only one component. For  $k > 0$ , assume the induction hypothesis holds for the splitting

$$(\mu_1 : p_1, \dots, \mu_{k-1} : p_{k-1}, \bigoplus_{i > k-1} \mu_i : \sum_{i > k-1} p_i).$$

Hence there exist a splitting

$$\xi^{k-1} = (\gamma_1 : p_1, \dots, \gamma_{k-1} : p_{k-1}, \hat{\gamma}_k : \sum_{i > k-1} p_i)$$

for some distribution  $\hat{\gamma}_k$ , and distributions  $\gamma_i$  with  $i \in \{1, \dots, k-1\}$  with  $\gamma \Rightarrow_c \xi^{k-1}$  and  $\mu_i \approx \gamma_i$  for  $i \in \{1, \dots, k-1\}$  and

$$\bigoplus_{i > k-1} \mu_i \approx \hat{\gamma}_k. \quad (\text{D.6})$$

We can rewrite  $\bigoplus_{i > k-1} \mu_i$  as  $\mu_k \oplus_q \bigoplus_{i > k} \mu_i$  with  $q = \frac{p_k}{\sum_{i > k-1} p_i}$ . From Equation D.6 we can immediately infer that  $\hat{\gamma}_k \Rightarrow_c \gamma_k \oplus_q \hat{\gamma}_{k+1}$  for some distributions  $\gamma_k$  and  $\hat{\gamma}_{k+1}$  with  $\mu_k \approx \gamma_k$  and  $\bigoplus_{i > k} \mu_i \approx \hat{\gamma}_{k+1}$ . This allows us to construct a new splitting

$$\xi^k = (\gamma_1 : p_1, \dots, \gamma_{k-1} : p_{k-1}, \gamma_k : p_k, \hat{\gamma}_{k+1} : \sum_{i > k} p_i)$$

which satisfying the inductive proof obligations.

Furthermore,  $\xi^{k-1}$  and  $\xi^k$  are identical in the first  $k-1$  components, which is enough to satisfy the condition of Equation D.5. □

### D.3. Theorem 8.6

To arrive at the congruence result for parallel composition, we need to establish a series of helpful lemmas. The section accumulates into Lemma D.6, which establishes the congruence result. As for most results of this thesis, we assume all Markov automata considered here to be compact.

**Lemma D.2.** If  $\mu \xrightarrow{\chi(r)} \mu'$  and  $\xi \xrightarrow{\chi(s)} \xi'$  then for  $p = \frac{r}{r+s}$

$$\mu \parallel_A \xi \xrightarrow{\chi(r+s)} \mu' \parallel_A \xi \oplus_p \mu \parallel_A \xi'$$

*Proof.* Recall that the notation  $\xrightarrow{\chi(r)}$  (cf. Notation 5.2) summarizes the individual timed transitions  $\multimap$  of a process and expresses them as one single (pseudo) transition leading to the distribution over states that is induced by the timed transitions. The exit rate of  $\mu$ , i.e. the exit rate of each state in its support,  $r$ , is preserved in the form of the action label.

Also recall that  $\mu \parallel_A \xi$  is a short-hand notation for

$$\langle s \parallel_A t : \mu(s)\xi(t) \mid s \in \text{Supp}(\mu) \wedge t \in \text{Supp}(\xi) \rangle.$$

First note that

$$\mu \parallel_A \xi = \bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \mu(s) \cdot \xi(t) \cdot \delta(s \parallel_A t) \quad (\text{D.7})$$

by definition of  $\parallel_A$ . Now for each  $s \parallel_A t$  above, we can derive the following behaviour with a repeated application of **parRM** and **parLM**. The left summand is generated by applications of **parLM** and the right by **parRM**. As  $r$  is the exit rate of  $s$ , and  $s$  is the exit rate of  $t$ , the exit rate of  $s \parallel_A t$  must be  $r + s$ , as every timed transition of both  $s$  and  $t$  is fully preserved by the rules with respect to their rates.

$$s \parallel_A t \xrightarrow{\chi(r+s)} \left( \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' \parallel_A t) \right) \oplus_{\frac{r}{r+s}} \left( \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s \parallel_A t') \right)$$

Note that, as  $\multimap$  is a multirelation, several summands above may occur more than once in the expression. By the definition of lifting state-based transition relations to distribution, and Equation D.7 we infer

$$\mu \parallel_A \xi \xrightarrow{\chi(r+s)} \bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \mu(s) \cdot \xi(t) \cdot \left( \left( \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' \parallel_A t) \right) \oplus_{\frac{r}{r+s}} \left( \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s \parallel_A t') \right) \right)$$

By Lemma 2.4, we can rewrite the right-hand side to

$$\bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \left( \left( \mu(s) \cdot \xi(t) \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' \parallel_A t) \right) \oplus_{\frac{r}{r+s}} \left( \mu(s) \cdot \xi(t) \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s \parallel_A t') \right) \right)$$

Then two applications of Lemma 2.3 yield

$$\begin{aligned} & \left( \bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \left( \mu(s) \cdot \xi(t) \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' ||_A t) \right) \right) \\ & \oplus \frac{r}{r+s} \left( \bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \left( \mu(s) \cdot \xi(t) \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s ||_A t') \right) \right). \end{aligned}$$

We now swap the sums in the first summand by Lemma 2.2

$$\begin{aligned} & \left( \bigoplus_{t \in \text{Supp}(\xi)} \bigoplus_{s \in \text{Supp}(\mu)} \left( \mu(s) \cdot \xi(t) \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' ||_A t) \right) \right) \\ & \oplus \frac{r}{r+s} \left( \bigoplus_{s \in \text{Supp}(\mu)} \bigoplus_{t \in \text{Supp}(\xi)} \left( \mu(s) \cdot \xi(t) \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s ||_A t') \right) \right) \end{aligned}$$

In the first summand, we bracket  $\xi(t)$ , and in the second  $\mu(s)$

$$\begin{aligned} & \left( \bigoplus_{t \in \text{Supp}(\xi)} \xi(t) \cdot \bigoplus_{s \in \text{Supp}(\mu)} \left( \mu(s) \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s' ||_A t) \right) \right) \\ & \oplus \frac{r}{r+s} \left( \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \bigoplus_{t \in \text{Supp}(\xi)} \left( \xi(t) \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(s ||_A t') \right) \right) \end{aligned} \tag{D.8}$$

It is straightforward by the definitions of  $\xrightarrow{\lambda}$  that

$$\mu' = \bigoplus_{s \in \text{Supp}(\mu)} \left( \mu(s) \bigoplus_{s \xrightarrow{\lambda} s'} \frac{\lambda}{r} \delta(s') \right)$$

and

$$\xi' = \bigoplus_{t \in \text{Supp}(\xi)} \left( \xi(t) \bigoplus_{t \xrightarrow{\nu} t'} \frac{\nu}{s} \delta(t') \right).$$

We now use this to simplify Equation D.8 and obtain

$$\left( \bigoplus_{t \in \text{Supp}(\xi)} \xi(t) \cdot (\mu' \parallel_A \delta(t)) \right) \oplus_{\frac{r}{r+s}} \left( \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot (\delta(s) \parallel_A \xi') \right),$$

which we rewrite by the definition of  $\parallel_A$  to  $(\mu' \parallel_A \xi) \oplus_{\frac{r}{r+s}} (\mu \parallel_A \xi')$  □

**Lemma D.3.**

- 1.) a) if  $a \notin A \cup \{\chi(r) \mid r \in \mathbb{R}_{\geq 0}\}$  and  $\mu \xRightarrow{a}_c \mu'$  then

$$\mu \parallel_A \gamma \xRightarrow{a}_c \mu' \parallel_A \gamma,$$

- b) if  $a \notin A \cup \{\chi(r) \mid r \in \mathbb{R}_{\geq 0}\}$  and  $\gamma \xRightarrow{a}_c \gamma'$  then

$$\mu \parallel_A \gamma \xRightarrow{a}_c \mu \parallel_A \gamma',$$

- 2.) if  $a \in A$  and  $\mu \xRightarrow{a}_c \mu'$  and  $\gamma \xRightarrow{a}_c \gamma'$  then

$$\mu \parallel_A \gamma \xRightarrow{a}_c \mu' \parallel_A \gamma',$$

- 3.) If  $\mu \xRightarrow{\chi(r_1)}_c \mu'$  and  $\gamma \xRightarrow{\chi(r_2)}_c \gamma'$  then exist  $\mu^*$  and  $\gamma^*$  such that

$$\mu \Rightarrow_c \mu^* \text{ and } \mu^* \downarrow \text{ and } \gamma \Rightarrow_c \gamma^* \text{ and } \gamma^* \downarrow$$

and

$$\mu \parallel_A \gamma \xRightarrow{\chi(r_1+r_2)}_c (\mu' \parallel_A \gamma^*) \oplus_p (\mu^* \parallel_A \gamma')$$

with  $p = \frac{r_1}{r_1+r_2}$ .

These claims are basically liftings of the transition rules for parallel composition from states to weak hypertransitions. In Claim 3 it is surprising that it is not a straightforward generalization of Lemma D.2, but instead stable distributions  $\mu^*$  and  $\gamma^*$  need to be introduced. Stability is needed here to enable the execution of the timed transitions. A transition

$$\mu \parallel_A \gamma \xRightarrow{\chi(r_1+r_2)}_c (\mu' \parallel_A \gamma) \oplus_p (\mu \parallel_A \gamma')$$

cannot exist, if  $\gamma$  and/or  $\mu$  are unstable.

*Proof.* We prove each of these claims only for the special case that  $\mu = \delta(s)$  and  $\gamma = \delta(t)$ . It is straightforward to iterate these results on the elements of the support of arbitrary distributions  $\mu$  and  $\gamma$  to prove the full claim. Furthermore, we only consider weak transitions, and not combined weak transitions. As the latter are obtained from weak transitions by a weighted combination



with some weights  $c_i$  from the former, the result can be easily lifted to combined transitions by constructing a convex combination of the obtained distributions, using the weights  $c_i$ .

We want to remark that the weak combined transitions in this lemma are in general infinitary. Hence, a simple inductive proof is not possible. Therefore, we need to follow an approach where suitable transition trees are constructed in order to show the desired result.

For Claim 1 and 2 we only treat the case for non-combined weak hypertransition, as it is the core of our argumentation. The general case is obtained in the usual way by repeating the argument for each weak hypertransition of which the combined transition is composed of.

For Claim 1.), it suffices to proof Claim 1.a), as the other case is completely symmetric. Consider for some  $s$  and  $\xi$  a transition  $s \xRightarrow{a} \xi$ . Let this transition be induced by the transition tree  $\mathcal{T}$ . Define  $\mathcal{S}$  to be the transition tree that coincides with  $\mathcal{T}$  except that for each node  $\sigma$  :  $Sta_{\mathcal{S}}(\sigma) = Sta_{\mathcal{T}}(\sigma) \parallel_A F$ . Using the definition of parallel composition (Definition 7.9), it is straightforward to verify that  $\mathcal{S}$  is indeed a transition tree, using that every transition in the tree is either labelled by  $a \notin A \cup \{\chi(r) \mid r \in \mathbb{R}_{\geq 0}\}$  or  $\tau$ , which implies that the transitions of the left-hand-side operand of the  $\parallel_A$ -operator are completely independent of the right hand side operand. Then by construction,  $\mathcal{S}$  induces the transition  $s \parallel_A t \xRightarrow{a} \xi \parallel_A t$ .

For Claim 2.) consider for some  $s$  and  $t$  and  $\xi$  and  $\nu$  the transitions  $s \xRightarrow{a} \xi$  and  $t \xRightarrow{a} \nu$ . Let this transition be induced by the transition tree  $\mathcal{T}$  and  $\mathcal{S}$  respectively. Let  $\mathcal{T}'$  and  $\mathcal{S}'$  be the respective maximal subtrees, which are internal transition trees. Note that the for every leaf  $\sigma$  of  $\mathcal{T}'$ , the corresponding node in  $\mathcal{T}$  represents an  $a$ -transition (that has been cut off in  $\mathcal{T}'$ ); an analogous observation holds for every leaf of  $\mathcal{S}'$  and  $\mathcal{S}$ . Using the same construction as in Claim 1, it is easy to construct a transition tree  $\mathcal{R}_1$  such that the nodes of  $\mathcal{R}_1$  are identical to those of  $\mathcal{T}'$ , except that for each node  $\sigma$  :  $Sta_{\mathcal{R}_1}(\sigma) = Sta_{\mathcal{T}'}(\sigma) \parallel_A F$ . Since for each leaf  $\sigma$ ,  $Sta_{\mathcal{R}_1}(\sigma)$  is then of the form  $u \parallel_A t$  for some state  $u$ , we can use the construction from Claim 1 again. But this time we fix  $u$ , and let  $t$  perform its transitions. This gives for each leaf  $\sigma$  of  $\mathcal{R}_1$  a tree  $\mathcal{R}_\sigma$ , where each node  $\sigma'$  of  $\mathcal{R}_\sigma$  satisfies  $Sta_{\mathcal{R}_\sigma}(\sigma') = Sta_{\mathcal{R}_1}(\sigma) \parallel_A Sta_{\mathcal{S}'}(\sigma')$ , for some leaf node  $\sigma'$  of  $\mathcal{S}$ . We can now construct a new transition tree  $\mathcal{R}_2$  from  $\mathcal{R}_1$  by gluing the tree  $\mathcal{R}_\sigma$  onto every leaf node  $\sigma$ . Then by our construction, the set of leaves of  $\mathcal{R}_2$  is exactly the set of all leaves  $\sigma''$  that with some  $\sigma_{\mathcal{T}'} \in Leaf(\mathcal{T}')$  and some  $\sigma_{\mathcal{S}'} \in Leaf(\mathcal{S}')$  satisfy  $Sta_{\mathcal{R}_2}(\sigma'') = Sta_{\mathcal{T}'}(\sigma_{\mathcal{T}'}) \parallel_A Sta_{\mathcal{S}'}(\sigma_{\mathcal{S}'})$  and  $Prob_{\mathcal{R}_2}(\sigma) = Prob_{\mathcal{T}'}(\sigma_{\mathcal{T}'})Prob_{\mathcal{S}'}(\sigma_{\mathcal{S}'})$ . Thus, we have so far proven that the parallel combination of  $s$  and  $t$  can perform a transition that is an interleaving of the respective transitions of  $s$  and  $t$  up to and excluding the point where the transitions labelled by  $a$  are performed.

We now continue with the  $a$ -transitions. By construction, each leaf is now labelled by the parallel composition of two states, say  $u$  and  $v$ , that perform an  $a$ -transition in the tree  $\mathcal{T}$  and  $\mathcal{S}$ , respectively. Say  $u \xrightarrow{a} \mu_u$  and  $v \xrightarrow{a} \mu_v$  for some  $\mu_u$  and  $\mu_v$ . Therefore, following Definition 7.9, we get  $u \parallel_A v \xrightarrow{a} \mu_u \parallel_A \mu_v$ . Clearly, we can extend our transition tree at the leaves to reflect this.

Thus, so far we have constructed a transition tree that proves that the parallel combination of  $s$  and  $t$  can perform a transition that is an interleaving of the respective transitions of  $s$  and  $t$  up to and including the point where the transitions labelled by  $a$  are performed. After this point, still a series of internal transition may follow. At this point, the transition tree construction proceeds exactly as in the first part of the proof of Claim 2 for every leaf of the so far constructed tree. This finishes our proof of Claim 2. In this way, we obtain a transition tree that proves our claim.

For Claim 3, let  $s, t, \mathcal{T}, \mathcal{S}, \mathcal{T}', \mathcal{S}'$  and  $\mathcal{R}_2$  be exactly as in the proof of Claim 2. Obviously,  $\mathcal{T}'$

and  $\mathcal{S}'$  induce some distributions  $\mu_{\mathcal{T}'}$  and  $\mu_{\mathcal{S}'}$ . These distribution will be the distributions from which the distributions  $\mu^*$  and  $\gamma^*$  from the claim are constructed. In fact, they obviously satisfy  $s \Rightarrow \mu_{\mathcal{T}'}$  and  $t \Rightarrow \mu_{\mathcal{S}'}$ , respectively. And by the choice of  $\mathcal{T}'$  and  $\mathcal{S}'$  also  $\mu_{\mathcal{T}'} \downarrow$  and  $\mu_{\mathcal{S}'} \downarrow$ .

Let also  $\mathcal{R}_2$  be constructed as in Claim 2. Since  $\mu_{\mathcal{T}'} \downarrow$  and  $\mu_{\mathcal{S}'} \downarrow$ , trivially for every leaf  $\sigma$  of  $\mathcal{R}_2$  with  $\text{Sta}_{r_2}(\sigma) = G \parallel_A H$  we get  $u \downarrow$  and  $v \downarrow$ . As in Claim 2, by construction, both constituents, say  $u$  and  $v$  of the parallel state in the leaves performs an  $\chi$ -transition in the trees  $\mathcal{T}$  and  $\mathcal{S}$  respectively, say  $u \xrightarrow{\chi(r_1)} \mu_u$  and  $v \xrightarrow{\chi(r_2)} \mu_v$ .

By the way  $\xrightarrow{\chi}$  was defined and Definition 7.9, we get  $u \parallel_A H \xrightarrow{\chi r_1 + r_2} \frac{r_1}{r_1 + r_2} \mu_u \parallel_A \mu_{s'} \oplus \frac{r_2}{r_1 + r_2} \mu_{t'} \parallel_A \mu_v$ . We then continue our construction inf analogy to Claim 2.  $\square$

**Lemma D.4.**

- $(\bigoplus_{i \in J} p_i \mu_i) \parallel_A \gamma = \bigoplus_{i \in J} p_i (\mu_i \parallel_A \gamma)$ .
- $\gamma \parallel_A (\bigoplus_{i \in J} p_i \mu_i) = \bigoplus_{i \in J} p_i (\gamma \parallel_A \mu_i)$ .

*Proof.* We only prove the first statement. The second is exactly symmetric. Denote  $(\bigoplus_{i \in J} p_i \mu_i)$  by  $\mu$ . By the definitions, we derive

$$\begin{aligned} \mu \parallel_A \gamma(s \parallel_A t) &= \mu(s) \cdot \gamma(t) \\ &= (\bigoplus_{i \in J} p_i \mu_i(s)) \cdot \gamma(t) \\ &= \bigoplus_{i \in J} p_i (\mu_i(s) \cdot \gamma(t)) \\ &= \bigoplus_{i \in J} p_i ((\mu_i \parallel_A \gamma)(s \parallel_A t)) \\ &= (\bigoplus_{i \in J} p_i (\mu_i \parallel_A \gamma))(s \parallel_A t) \end{aligned}$$

$\square$

**Lemma D.5.** Let  $\mu \in \text{Dist}(S)$  satisfy  $\mu \downarrow$  and let  $\gamma \in \text{Dist}(S)$  satisfy  $\mu \approx \gamma$ . Then, whenever  $\gamma \Rightarrow_c \gamma'$ , then  $\mu \approx \gamma'$ .

*Proof.* Since  $\mu \downarrow$  there exist no outgoing internal transitions for  $\mu$ . Hence  $\mu \approx \gamma'$  follows immediately from the other premises of the lemma and Definition 8.4, Condition (a).  $\square$

We are now in the position to show that weak bisimulation is a congruence with respect to parallel composition. To arrive at the congruence result for automata composition, it suffices to show that the respective initial states are bisimilar.

**Lemma D.6.** Let  $\mathcal{A}$  and  $\mathcal{A}'$  be two MA with  $\mathcal{A} \approx_\delta \mathcal{A}'$ . Let  $s$  and  $t$  be their respective initial states. Let  $\mathcal{A}^*$  be an arbitrary MA with initial state  $u$ . Then

$$s \parallel_A u \approx_\delta t \parallel_A u.$$

*Proof.* To establish this theorem, it suffices by Lemma 8.4 to find a bisimulation-up-to-splitting  $\mathcal{R}$  containing the pair  $(\delta(s \parallel_A u), \delta(t \parallel_A u))$ . A suitable relation is

$$\mathcal{R} = \{ (\mu \parallel_A \xi, \gamma \parallel_A \nu) \mid \mu \approx \gamma \wedge \xi \approx \nu \}$$

where  $\mu$  is an arbitrary distribution over states of  $\mathcal{A}$ ,  $\gamma$  is an arbitrary distribution over states of  $\mathcal{A}'$ , and  $\xi$  and  $\nu$  are arbitrary distribution over states of  $\mathcal{A}^*$ .

We will now check the three conditions of Definition 8.8. For Condition (a), we distinguish several cases:

- Let  $\mu \parallel_A \xi \xrightarrow{a} \mu' \parallel_A \xi$  because  $\mu \xrightarrow{a} \mu'$  by rule **parL** of Definition 7.9. Then clearly,  $a \notin A$  and  $a \neq \chi(r)$ . Then, as  $\mu \approx \gamma$ , there must exist a transition  $\gamma \xrightarrow{a}_c \gamma'$  with  $\mu' \approx \gamma'$ . By Lemma D.3 1.), we can derive that  $\gamma \parallel_A \nu \xrightarrow{a}_c \gamma' \parallel_A \nu$ . Immediately,  $\mu' \parallel_A \xi \mathcal{R} \gamma' \parallel_A \nu$  follows.
- The case that  $\mu \parallel_A \xi \xrightarrow{a} \mu \parallel_A \xi'$  because  $\xi \xrightarrow{a} \xi'$  by rule **parR** of Definition 7.9 is completely symmetric.
- Let  $a \in A$  and  $\mu \parallel_A \xi \xrightarrow{a} \mu' \parallel_A \xi'$  because  $\mu \xrightarrow{a} \mu'$  and  $\xi \xrightarrow{a} \xi'$  by rule **sync** of Definition 7.9. Then, as  $\mu \approx \gamma$ , there must exist a transition  $\gamma \xrightarrow{a}_c \gamma'$  with  $\mu' \approx \gamma'$ . Accordingly, as  $\xi \approx \nu$ , there must exist a transition  $\nu \xrightarrow{a}_c \nu'$  with  $\xi' \approx \nu'$ . By Lemma D.3 2.), it then follows that also  $\gamma \parallel_A \nu \xrightarrow{a}_c \gamma' \parallel_A \nu'$ . Clearly, again  $\mu' \parallel_A \xi' \mathcal{R} \gamma' \parallel_A \nu'$ .
- The last case is the case where a stochastic timed transitions are considered. By Lemma D.2,  $\mu \parallel_A \xi \xrightarrow{\chi(r+s)} \mu' \parallel_A \xi \oplus_p \mu \parallel_A \xi'$  if  $\mu \xrightarrow{\chi(r)} \mu'$  and  $\xi \xrightarrow{\chi(s)} \xi'$ .

Continuing with our proof, as  $\mu \approx \gamma$ , there must exist a transition  $\gamma \xrightarrow{\chi(r)}_c \gamma'$  with  $\mu' \approx \gamma'$ . Accordingly, as  $\xi \approx \nu$ , there must exist a transition  $\nu \xrightarrow{\chi(s)}_c \nu'$  with  $\xi' \approx \nu'$ . By Lemma D.3 3.), it then follows from  $\gamma \xrightarrow{\chi(r)}_c \gamma'$  and  $\nu \xrightarrow{\chi(s)}_c \nu'$  that distributions  $\gamma^*$  and  $\nu^*$  exist, such that

$$\gamma \Rightarrow_c \gamma^* \text{ and } \gamma^* \downarrow \quad \text{and} \quad \nu \Rightarrow_c \nu^* \text{ and } \nu^* \downarrow$$

and

$$\gamma \parallel_A \nu \xrightarrow{\chi(r+s)}_c (\gamma' \parallel_A \nu^*) \oplus_p (\gamma^* \parallel_A \nu')$$

with  $p = \frac{r}{r+s}$ . Unfortunately, We cannot directly show that

$$(\mu' \parallel_A \xi) \oplus_p (\mu \parallel_A \xi') \mathcal{R} (\gamma' \parallel_A \nu^*) \oplus_p (\gamma^* \parallel_A \nu').$$

However, as  $\mathcal{R}$  is a bisimulation-up-to-splitting, it suffices show that

$$\mu' \parallel_A \xi \mathcal{R} \gamma' \parallel_A \nu^* \text{ and } \mu \parallel_A \xi' \mathcal{R} \gamma^* \parallel_A \nu'$$

Note that by Lemma D.5

$$\gamma \approx \gamma^* \text{ and } \nu \approx \nu^*.$$

Then, by transitivity of  $\approx$ , it is straightforward to check that the two pairs are contained in  $\mathcal{R}$ .

We now check Condition b. Therefore, assume that we split  $\mu \parallel_A \xi$  into two distribution  $\rho_1 \oplus_p \rho_2$ . As  $\mu \approx \gamma$  and  $\xi \approx \nu$ , using Lemma 8.7, we can derive that

$$\gamma \Rightarrow_c \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \gamma_s \text{ and } \nu \Rightarrow_c \bigoplus_{t \in \text{Supp}(\xi)} \xi(s) \nu_t$$

with the distributions  $\gamma_s$  and  $\nu_t$  satisfying

$$\delta(s) \approx \gamma_s \text{ and } \delta(t) \approx \nu_t \quad (\text{D.9})$$

for every  $s \in \text{Supp}(\mu)$  and  $t \in \text{Supp}(\xi)$ , respectively. By Lemma D.3,

$$\gamma \parallel_A \nu \Rightarrow_c \left( \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \gamma_s \right) \parallel_A \left( \bigoplus_{t \in \text{Supp}(\xi)} \xi(s) \nu_t \right).$$

By applying Lemma D.4 twice, it follows that

$$\begin{aligned} \left( \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \gamma_s \right) \parallel_A \left( \bigoplus_{t \in \text{Supp}(\xi)} \xi(s) \nu_t \right) \\ = \bigoplus_{s \in \text{Supp}(\mu), t \in \text{Supp}(\xi)} \mu(s) \xi(t) (\gamma_s \parallel_A \nu_t). \end{aligned}$$

Now, it is straightforward to define a suitable splitting  $\chi_1 \oplus_p \chi_2$  of this distribution that matches the splitting  $\rho_1 \oplus_p \rho_2 = \mu \parallel_A \xi$ . We simply let  $\chi_l(\gamma_s \parallel_A \nu_t) = \rho_l(s \parallel_A t)$  for  $l \in \{1, 2\}$ . Now again, we cannot directly show that  $\rho_l \mathcal{R} \chi_l$  for  $l \in \{1, 2\}$ . However, we only need to establish that  $\mathcal{R}$  is a bisimulation-up-to-splitting. This is obvious from Equation D.9. □

## D.4. Theorem 8.8

$\approx = \simeq_*$ .

*Proof.*  $\approx \subseteq \simeq_*$  follows immediately from Theorem 8.6 and 8.7. It remains to show  $\simeq_* \subseteq \approx$ . Recall that  $\simeq_*$  is defined as the coarsest observational bisimulation, that is a congruence with respect to parallel composition ( $\parallel_A$ ). Note especially that this means that  $\simeq_*$  satisfies Definition 8.3.

We now establish that  $\simeq_*$  is also a weak distribution bisimulation. Let  $\mu \simeq_* \gamma$ . Condition (c) follows immediately, as they are identical. Condition (a) of Definition 8.4 is a special case of Condition (a) of Definition 8.3 with  $p = 1$ . The most interesting case is thus to establish Condition (b) Definition 8.4. Let  $\mu = \mu_1 \oplus_p \mu_2$  be an arbitrary splitting of  $\mu$ . To show that  $\gamma$  can react accordingly, we use compositionality of  $\simeq_*$  with respect to  $\parallel_\emptyset$ . Let  $a_1$  and  $a_2$  and  $b$  be fresh actions that neither occur in  $\mu$  nor in  $\gamma$ . Consider the process

$$T = b.0 + \tau.a_1.0 + \tau.a_2.0.$$

Then, as  $\simeq_*$  is a congruence with respect to  $\parallel_\emptyset$ ,

$$\mu \parallel_\emptyset T \simeq_* \gamma \parallel_\emptyset T.$$

Then  $\mu \parallel_{\emptyset} T \Longrightarrow_c \mu_1 \parallel_{\emptyset} a_1.0 \oplus_p \mu_2 \parallel_{\emptyset} a_2.0$ . To see this, recall that  $\mu \parallel_{\emptyset} T$  is the distribution  $\langle (s \parallel_{\emptyset} T : \mu(s)) \mid s \in \text{Supp}(\mu) \rangle$ . To justify the weak hyper-transition, it is enough to provide a suitable combined transition for each  $s \in \text{Supp}(\mu \parallel_{\emptyset} T)$  that justifies this transition. We choose

$$s \parallel_{\emptyset} T \Longrightarrow_c s \parallel_{\emptyset} a_1.0 \oplus_{q_s} s \parallel_{\emptyset} a_2.0,$$

where

$$q_s = \frac{\mu_1(s)}{\mu_1(s) + \mu_2(s)}.$$

In turn, each of these combined transitions is justified by the two transitions  $s \parallel_{\emptyset} T \xrightarrow{\tau} a_1.0$  and  $s \parallel_{\emptyset} T \xrightarrow{\tau} a_2.0$ . Then, as  $\mu \parallel_{\emptyset} T \simeq_{\star} \gamma \parallel_{\emptyset} T$ , there must exist a transition  $\gamma \parallel_{\emptyset} T \Longrightarrow_c \xi$  with

$$\mu_1 \parallel_{\emptyset} a_1.0 \oplus_p \mu_2 \parallel_{\emptyset} a_2.0 \simeq_{\star} \xi.$$

The states in  $\text{Supp}(\xi)$  must all have the form  $t \parallel_{\emptyset} T'$ . As  $\mu_1 \parallel_{\emptyset} a_1.0 \oplus_p \mu_2 \parallel_{\emptyset} a_2.0$  cannot execute action  $b$ , as  $b$  was fresh for  $\mu$ , also  $\xi$  may not be able to execute  $b$ . As  $b$  is also fresh for  $\gamma$ ,  $T'$  must be of form  $a_1.0$  or  $a_2.0$ . We can thus observe that

$$\xi = \bigoplus_{i \in I_1} p_i^1(\gamma_i^1 \parallel_{\emptyset} a_1.0) \oplus_p \bigoplus_{i \in I_2} p_i^2(\gamma_i^2 \parallel_{\emptyset} a_2.0).$$

with  $\gamma \Longrightarrow_c \gamma_1 \oplus_p \gamma_2$  and  $\gamma_1 = \bigoplus_{i \in I_1} p_i^1 \gamma_i^1$  and  $\gamma_2 = \bigoplus_{i \in I_2} p_i^2 \gamma_i^2$ . In summary, we can write  $\xi = \gamma_1 \parallel_{\emptyset} a_1.0 \oplus_p \gamma_2 \parallel_{\emptyset} a_2.0$ . It remains to show that  $\mu_1 \simeq_{\star} \gamma_1$  and  $\mu_2 \simeq_{\star} \gamma_2$ . To show this, we will make use of a new claim.

**Claim D.1.** *The relation*

$$\mathcal{R} = \{ (\mu, \gamma) \mid \exists \mu', \gamma' \text{ and a fresh action } c : (\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star} (\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma' \} \cup \simeq_{\star}$$

*is a relaxed bisimulation and a congruence with respect to  $\parallel_{\emptyset}$ .*

To establish the congruence property, we have to show that  $\mu \parallel_{\emptyset} \rho \mathcal{R} \gamma \parallel_{\emptyset} \rho$ , when  $\mu \mathcal{R} \gamma$ . Since  $\mu \mathcal{R} \gamma$ , there exist  $\mu', \gamma'$  and a fresh action  $c$  such that

$$(\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star} (\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma'.$$

Since  $\simeq_{\star}$  is a congruence with respect to parallel composition, also

$$[(\mu \parallel_{\emptyset} c.0) \oplus_p \mu'] \parallel_{\emptyset} \rho \simeq_{\star} [(\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma'] \parallel_{\emptyset} \rho.$$

Distributing  $\rho$  between the probabilistic sum, we obtain

$$(\mu \parallel_{\emptyset} \rho \parallel_{\emptyset} c.0) \oplus_p (\mu' \parallel_{\emptyset} \rho) \simeq_{\star} (\gamma \parallel_{\emptyset} \rho \parallel_{\emptyset} c.0) \oplus_p (\gamma' \parallel_{\emptyset} \rho)$$

By definition, this implies  $(\mu \parallel_{\emptyset} \rho) \mathcal{R} (\gamma \parallel_{\emptyset} \rho)$ .

To show that  $\mathcal{R}$  is a relaxed bisimulation, we begin with Condition (a). Assume  $\mu \Longrightarrow_c \mu''$ . We need to show that then there exists  $\gamma''$  such that  $\gamma \Longrightarrow_c \gamma''$  and  $\mu \mathcal{R} \gamma''$ .

From  $\mu \mathcal{R} \gamma$  follows  $(\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star} (\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma'$  for some  $\mu', \gamma'$  and a fresh action  $c$ . Assuming  $\mu \Longrightarrow_c \mu''$  for some  $\mu''$ , we also can derive  $(\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \Longrightarrow_c (\mu'' \parallel_{\emptyset} c.0) \oplus_p \mu'$ . Hence,  $(\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma' \Longrightarrow_c (\gamma'' \parallel_{\emptyset} c.0) \oplus_p \gamma''$  for some  $\gamma''$  and  $\gamma''$  and  $(\mu'' \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star}$

$(\gamma'' \parallel_{\emptyset} c.0) \oplus_p \gamma'''$ . Furthermore,  $\gamma \Rightarrow_c \gamma'' (\star)$  and  $\gamma' \Rightarrow_c \gamma'''$ . Now  $(\star)$  is already the first part of our proof obligation. It remains to show that  $\mu' \mathcal{R} \gamma''$ . Therefore, it suffices to mention that

$$(\mu'' \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star} (\gamma'' \parallel_{\emptyset} c.0) \oplus_p \gamma'''.$$

already satisfies the defining condition of  $\mathcal{R}$ .

To establish Condition (b), we begin with an observation. When

$$(\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \xRightarrow{c|_p}_c \mu \parallel_{\emptyset} 0$$

then also

$$(\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma' \xRightarrow{c|_p}_c \rho' \text{ and } \mu \parallel_{\emptyset} 0 \simeq_{\star} \rho'$$

for some  $\rho'$ , since  $(\mu \parallel_{\emptyset} c.0) \oplus_p \mu' \simeq_{\star} (\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma'$  and  $\simeq_{\star}$  is a relaxed bisimulation. As  $c$  is fresh,  $\gamma'$  cannot have contributed to the transition. Hence,  $\mu \parallel_{\emptyset} 0 \simeq_{\star} \rho'$  implies that  $\rho' = \gamma'' \parallel_{\emptyset} 0$  and, in addition, we can also derive that  $\gamma \Rightarrow_c \gamma''$  for some  $\gamma''$ , since the splitting factor  $p$  that appears in  $(\gamma \parallel_{\emptyset} c.0) \oplus_p \gamma'$  is fully applied also in the transition  $\xRightarrow{c|_p}_c$ . If the parameter of the transition would have been a number smaller than  $p$ , it could have been the case that only a sub-distribution of  $\gamma$  actually engages in the transition. It is straightforward to finally establish that  $\mu \parallel_{\emptyset} 0 \simeq_{\star} \gamma'' \parallel_{\emptyset} 0$  implies  $\mu \simeq_{\star} \gamma''$ . In summary, we have thus shown that there exists  $\gamma''$  such that

$$\gamma \Rightarrow_c \gamma'' \tag{D.10}$$

and

$$\mu \simeq_{\star} \gamma''. \tag{D.11}$$

Now, in order to establish Condition (b), assume  $\mu \xRightarrow{a|_q}_c \mu''$  for some  $q \in [0, 1]$ . As  $\mu \simeq_{\star} \gamma''$ , also  $\gamma'' \xRightarrow{a|_q}_c \gamma'''$  and  $\mu'' \simeq_{\star} \gamma'''$  for some  $\gamma'''$ . Together with Equation D.10, we obtain  $\gamma \Rightarrow_c \gamma'' \xRightarrow{a|_q}_c \gamma'''$ , i.e.  $\gamma \xRightarrow{a|_q}_c \gamma'''$ . Finally,  $\mu'' \simeq_{\star} \gamma'''$  implies  $\mu'' \mathcal{R} \gamma'''$  by the definition of  $\mathcal{R}$ .

For Condition (c), assume  $\mu \downarrow$ . From  $\mu \simeq_{\star} \gamma'$ , the fact that  $\simeq_{\star}$  is a relaxed bisimulation and  $\gamma \Rightarrow_c \gamma'$  everything follows. □

We want to remark that many crucial insights in this proof have been strongly inspired by the proof of Lemma 4.11 in [DH12].

## D.5. Proof of Theorem 8.10

**Lemma D.7.** The union of two semi-weak bisimulations on a *finite* PA  $\mathcal{A}$  is again a semi-weak bisimulation on  $\mathcal{A}$ .

We are now ready to prove **Theorem 8.10**:

On *finite* PA, it holds that

$$\circ \approx_{\delta}^{\downarrow} = \approx_{\text{PA}}$$

*Proof.* We first show that  $\overset{\circ}{\approx}_{\delta}^{\dagger} \subseteq \approx_{\text{PA}}$ . Therefore, we prove that  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  is also a weak probabilistic bisimulation according to Definition 4.19. The crucial point in this proof is the claim, that  $\mu \overset{\circ}{\approx}_{\delta}^{\dagger} \gamma$  implies  $\forall C \in S/\overset{\circ}{\approx}_{\delta}^{\dagger} : \mu(C) = \gamma(C)$ , since then the conditions of Definition 4.19 are met almost immediately. By Lemma D.7,  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  is a semi-weak bisimulation. Then, by a simple inductive argument, it can be shown that using the Condition ((b)) in the definition of  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  (Definition 8.12) repeatedly,  $\gamma$  can simulate any (finite) splitting of  $\mu$  ( $\star$ ).<sup>1</sup> Thus, split  $\mu$  in such way that every state in its support is isolated, i.e. we split  $\mu$  into  $\bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \delta(s)$ . Then, by ( $\star$ ),  $\gamma$  can be split into a family of distribution  $\{\gamma_s\}_{s \in \text{Supp}(\mu)}$ , such that for every  $s \in \text{Supp}(\mu)$ ,

$$\delta(s) \overset{\circ}{\approx}_{\delta}^{\dagger} \gamma_s. \quad (\text{D.12})$$

We now use the symmetry of semi-weak bisimilarity, we iterate this idea with exchanged roles. We propose for each  $\gamma_s$  a splitting into  $\bigoplus_{F \in \text{Supp}(\gamma_s)} \gamma_s(F) \delta(F)$ , which now must be simulated by  $\delta(s)$  according to Condition ((b)) again, due to Equation D.12. Clearly, as  $\text{Supp}(\delta(s))$  is a singleton set, every component of the splitting must be again  $\delta(s)$ . Thus, it must hold that

$$\delta(s) \overset{\circ}{\approx}_{\delta}^{\dagger} \delta(F). \quad (\text{D.13})$$

In summary, we have split  $\mu$  and  $\gamma$  into sets of matching Dirac distributions. From the condition, that splittings must preserve the splitting factor  $p$ , the matched Dirac distributions must contribute the same relative amount of probability mass to  $\mu$  and  $\gamma$ . From here we can immediately conclude that

$$\mu \overset{\circ}{\approx}_{\delta}^{\dagger} \gamma \text{ implies } \forall C \in S/\overset{\circ}{\approx}_{\delta}^{\dagger} : \mu(C) = \gamma(C). \quad (\text{D.14})$$

We are now ready to show that  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  satisfies Definition 4.19. First, by Theorem 8.9,  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  is an equivalence relation on  $S$ . Second, let  $s \overset{\circ}{\approx}_{\delta}^{\dagger} t$  and let  $s \xrightarrow{a} \mu$  for some  $a \in \text{Act}$  and  $\mu \in \text{Dist}(S)$ . Then, by Condition (a)) of Definition 8.12, also  $t \xrightarrow{a} \gamma$  for some distribution  $\gamma$  and  $\mu \overset{\circ}{\approx}_{\delta}^{\dagger} \gamma$ . The latter implies by D.14  $\mu \mathcal{L}(\overset{\circ}{\approx}_{\delta}^{\dagger}) \gamma$ . This ends the proof that  $\overset{\circ}{\approx}_{\delta}^{\dagger}$  is a weak probabilistic bisimulation, and thus  $\overset{\circ}{\approx}_{\delta}^{\dagger} \subseteq \approx_{\text{PA}}$ .

For the other direction, we show that the relation

$$\mathcal{R} = \mathcal{L}(\approx_{\text{PA}})$$

is a semi-weak stability insensitive distribution bisimulation i.e. it satisfies Definition 8.12. From here we can conclude that whenever  $s \approx_{\text{PA}} t$ , the pair  $(\delta(s), \delta(t))$  is contained in the semi-weak stability insensitive distribution bisimulation  $\mathcal{R}$ , which implies  $s \overset{\circ}{\approx}_{\delta}^{\dagger} t$ . Hence,  $\approx_{\text{PA}} \subseteq \overset{\circ}{\approx}_{\delta}^{\dagger}$ .

To prove that  $\mathcal{R}$  is a semi-weak stability insensitive distribution bisimulation, let us consider an arbitrary pair  $(\mu, \gamma) \in \mathcal{R}$ . Assume for Condition (a),  $\mu \xrightarrow{a} \mu'$ . Then, by the definition of weak hypertransition, every  $s \in \text{Supp}(\mu)$  contributes a transition  $s \xrightarrow{a} \mu'_s$  such that  $\mu' = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu'_s$ . Since  $\mu \mathcal{R} \gamma$  implies  $\mu \mathcal{L}(\approx_{\text{PA}}) \gamma$  by our choice of  $\mathcal{R}$ , for every  $s \in \text{Supp}(\mu)$

<sup>1</sup>In this argument, the finiteness assumption is needed in order to guarantee that the repeated application of the left-hand side condition finally terminates.

satisfies  $\gamma([s]_{\approx_{\text{PA}}}) = \mu([s]_{\approx_{\text{PA}}})$ . For each  $t \in [s]_{\approx_{\text{PA}}} \cap \text{Supp}(\gamma)$ , there exists for some distribution  $\gamma'_{s,t}$  a weak combined transition

$$t \xRightarrow{a}_c \gamma'_{s,F} \text{ with } \mu'_s \mathcal{L}(\approx_{\text{PA}}) \gamma'_{s,F}. \quad (\text{D.15})$$

We now split  $\mu$  into

$$\bigoplus_{s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \cap [s]_{\approx_{\text{PA}}}} \mu(s) \cdot \frac{\gamma(F)}{\gamma([s]_{\approx_{\text{PA}}})} \cdot \delta(s), \quad (\text{D.16})$$

which is a valid splitting, since for every  $s \in \text{Supp}(\mu)$ ,

$$\sum_{F \in \text{Supp}(\gamma) \cap [s]_{\approx_{\text{PA}}}} \frac{\gamma(F)}{\gamma([s]_{\approx_{\text{PA}}})} = 1. \quad (\text{D.17})$$

We see then by Lemma 4.3 that

$$\mu' = \bigoplus_{s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \cap [s]_{\approx_{\text{PA}}}} \mu(s) \cdot \frac{\gamma(F)}{\gamma([s]_{\approx_{\text{PA}}})} \cdot \mu'_s. \quad (\text{D.18})$$

We can also split  $\gamma$  accordingly into

$$\gamma \xRightarrow{a}_c \bigoplus_{s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \cap [s]_{\approx_{\text{PA}}}} \mu(s) \cdot \frac{\gamma(F)}{\gamma([s]_{\approx_{\text{PA}}})} \cdot \delta(F). \quad (\text{D.19})$$

By D.15 and Lemma 4.3, it then follows that

$$\gamma \xRightarrow{a}_c \gamma' := \bigoplus_{s \in \text{Supp}(\mu), t \in \text{Supp}(\gamma) \cap [s]_{\approx_{\text{PA}}}} \mu(s) \cdot \frac{\gamma(F)}{\gamma([s]_{\approx_{\text{PA}}})} \cdot \gamma'_{s,F} \quad (\text{D.20})$$

Since already  $\mu'_s \mathcal{L}(\approx_{\text{PA}}) \gamma'_{s,t}$  (which we know from D.15), we immediately see that  $\mu' \mathcal{L}(\approx_{\text{PA}}) \gamma'$  by our construction.

For Condition (b), let  $\mu = \mu_1 \oplus_p \mu_2$  be an arbitrary splitting of  $\mu$ . A matching splitting of  $\gamma = \gamma_1 \oplus_p \gamma_2$  can be constructed by defining

$$\gamma_1 = \frac{1}{p} \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \mu_1(C) \frac{\gamma(F)}{\gamma(C)} \delta(F) \quad (\text{D.21})$$

and

$$\gamma_2 = \frac{1}{1-p} \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \mu_2(C) \frac{\gamma(F)}{\gamma(C)} \delta(F). \quad (\text{D.22})$$



Straightforward calculations yield

$$\begin{aligned}
\gamma_1 \oplus \gamma_2 &= \left[ \frac{1}{p} \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \mu_1(C) \frac{\gamma(F)}{\gamma(C)} \delta(F) \right] \oplus_p \left[ \frac{1}{1-p} \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \mu_2(C) \frac{\gamma(F)}{\gamma(C)} \delta(F) \right] \\
&= \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \left( \frac{1}{p} \cdot p \cdot \mu_1(C) \frac{\gamma(F)}{\gamma(C)} \delta(F) + \frac{1}{1-p} \cdot (1-p) \cdot \mu_2(C) \frac{\gamma(F)}{\gamma(C)} \delta(F) \right) \\
&= \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} (\mu_1(C) + \mu_2(C)) \frac{\gamma(F)}{\gamma(C)} \delta(F) \\
&= \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \mu(C) \cdot \frac{\gamma(F)}{\gamma(C)} \delta(F) \\
&= \bigoplus_{C \in S/\approx_{\text{PA}}, t \in C} \gamma(F) \delta(F) \quad \text{since } \gamma() = \mu(C), \text{ as by assumption } \mu \mathcal{L}(\approx_{\text{PA}}) \gamma \\
&= \gamma
\end{aligned}$$

□

## D.6. Proof of Theorem 8.13

*Notation D.1.* Let  $S$  be a set of states. If  $\mu$  and  $\gamma \in \text{Dist}(S)$ , we write  $\mu \hat{\mathcal{L}}(\mathcal{R}) \gamma$  if and only if  $\exists \gamma' \in \text{Dist}(\text{Dist}(S)) : \text{flatten}(\gamma') = \gamma \wedge \mu \hat{\mathcal{L}}(\mathcal{R}) \gamma'$ .

**Definition D.1.** Let  $(S, \bar{s}, \text{Act}, \multimap)$  be a PA. A relation  $\mathcal{R}$  over  $\text{Dist}(S)$  is a simulation-up-to-splitting, if for each pair of distributions  $(\mu, \gamma) \in \mathcal{R}$  for every  $a \in \text{Act}$

- (a)  $\mu \xrightarrow{a} \mu'$  implies  $\gamma \xRightarrow{a}_c \gamma'$  for some  $\gamma' \in \text{Dist}(S)$  and  $\mu' \mathcal{R}^\oplus \gamma'$ ,
- (b) for every  $p \in [0, 1]$  and  $\mu_1, \mu_2 \in \text{Dist}(S)$  such that  $\mu = \mu_1 \oplus_p \mu_2$  there exist  $\gamma_1, \gamma_2 \in \text{Dist}(S)$  such that  $\gamma \xRightarrow{a}_c \gamma_1 \oplus_p \gamma_2$  and  $\mu_i \mathcal{R}^\oplus \gamma_i$  for  $i \in \{1, 2\}$ .

◁

This definition is a straightforward adaption of Definition 8.8, bisimulation-up-to-splitting. It comes without the symmetry condition. Furthermore, it demands that  $a \in \text{Act}$  and has the third condition removed. Therefore, simulation-up-to-splitting is suited to establish that two PA are weak stability insensitive distribution similar. This is stated in the following lemma.

**Lemma D.8.** Whenever  $\mathcal{R}$  is a simulation-up-to-splitting then  $\mathcal{R} \subseteq \preceq^\dagger$ .

*Proof.* Completely analogous to the proof of Lemma 8.4. This holds as the proof did not depend on the symmetry condition of  $\approx$  nor on stability preservation. □

**Lemma D.9.** Let  $\mu = \bigoplus_{i \in J} p_i \mu_i$  be a splitting of  $\mu$  with a possibly infinite index set  $J \subseteq \mathbb{N}$ .

If  $\mu \preceq^\dagger \gamma$  then there exist  $\gamma_i$  for each  $i \in J$  such that

$$\gamma \xRightarrow{a}_c \bigoplus_{i \in J} p_i \gamma_i \text{ and } \mu_i \preceq^\dagger \gamma_i.$$

*Proof.* Follows completely analogously to the proof of Lemma 8.7. This holds as the proof did not depend on the symmetry condition of  $\approx$  nor on stability preservation.  $\square$

Recall that for a distribution  $\mu \in \text{Dist}(\text{Dist}(S))$ ,

$$\text{flatten}(\mu) = \bigoplus_{\gamma \in \text{Supp}(\mu)} \mu(\gamma) \cdot \gamma.$$

The flattening operations closely corresponds to what we call a splitting of a distribution, i.e. a decomposition of a distribution in weighted distributions.

*Remark D.1.* Let  $\gamma = \text{flatten}(\mu)$ . Then  $\mu = \bigoplus_{\gamma \in \text{Supp}(\mu)} \mu(\gamma) \cdot \gamma$  is a splitting of  $\gamma$ .

**Lemma D.10.** If  $\mu \hat{\mathcal{L}}(\preceq_{fwd}) \gamma$ , then there is are splittings

$$\mu = \bigoplus_{i \in I} p_i \delta(s_i) \text{ and } \gamma = \bigoplus_{i \in I} p_i \gamma_i$$

where  $s_i \in \text{Supp}(\mu)$  and

$$\delta(s_i) \preceq_{fwd} \gamma_i.$$

Note that it may well be that case that  $s_i = s_j$  for  $i \neq j$ .

*Proof.* Let  $w$  be the weight function that justifies  $\mu \hat{\mathcal{L}}(\preceq_{fwd}) \gamma$ . Let  $I$  be a index set that provides an index for each pair  $(s, \gamma') \in \text{Supp}(\mu) \times \text{Supp}(\gamma)$  with  $w(s, \gamma') > 0$ . If  $i \in I$  is the index of the pair with index  $i$ , we denote by  $s_i$  its first component and by  $\gamma_i$  its second component. Let furthermore  $p_i = w(s, \gamma')$ .

Then by the definition of the weight  $w$ ,

$$\mu = \bigoplus_{i \in I} p_i \delta(s_i) \text{ and } \gamma = \bigoplus_{i \in I} p_i \gamma_i$$

are the desired splittings.  $\square$

### D.6.1. Proof of Theorem 8.13

We are now ready to prove Theorem 8.13. Note that the following formulation is shorter than the original version of the lemma, as we make use of Notation D.1 here (and throughout the whole section).

$$\preceq^\dagger = \hat{\mathcal{L}}(\preceq_{fwd})$$

*Proof.* We first show that the relation  $\hat{\mathcal{L}}(\preceq_{fwd})$  is a simulation-up-to-splitting. Consider a pair  $\mu \hat{\mathcal{L}}(\preceq_{fwd}) \gamma$ . For the first condition of simulation-up-to-splitting, consider  $\mu \xrightarrow{a} \mu'$ . By the definition of hypertransitions, there must be a transitions  $s \xrightarrow{a} \mu'_s$  for each  $s \in \text{Supp}(\mu)$ . From Lemma D.10, we derive the existence of a splitting

$$\mu = \bigoplus_{i \in I} p_i \delta(s_i) \text{ and } \gamma = \bigoplus_{i \in I} p_i \gamma_i.$$

As  $\delta(s_i) \preceq_{fwd} \gamma_i$  for each  $i \in I$ , there must exist a transition  $\gamma_i \xrightarrow{a}_c \gamma'_i$  with  $\mu'_s \hat{\mathcal{L}}(\preceq_{fwd}) \gamma'_i$ . We can recompose the  $\gamma'_i$  to a distribution that is reachable from  $\gamma$ , i.e.

$$\gamma \xrightarrow{a}_c \bigoplus_{i \in I} p_i \gamma'_i =: \gamma'.$$

By construction, the distributions  $\mu'$  and  $\gamma'$  can be split into distributions  $\mu'_s$  and  $\gamma'_i$  with  $i \in I$ . For each  $i \in I$ , they satisfy  $\mu'_s \hat{\mathcal{L}}(\preceq_{fwd}) \gamma'_i$ . Thus,  $\mu' \hat{\mathcal{L}}(\preceq_{fwd})^\oplus \gamma'$ .

This satisfies the first condition we need to show in order to establish that  $\hat{\mathcal{L}}(\preceq_{fwd})$  is a simulation-up-to-splitting.

We now consider the second condition. Let  $\mu = \mu_a \oplus_p \mu_b$  be an arbitrary splitting of  $\mu$ . Let

$$\mu = \bigoplus_{i \in I} p_i \delta(s_i) \text{ and } \gamma = \bigoplus_{i \in I} p_i \gamma_i$$

be splittings of  $\mu$  and  $\gamma$  according to Lemma D.10. We now define a new splitting for  $\mu$  that is a refinement of both splittings for  $\mu$ . Let  $p_i^a := p_i \cdot \frac{\mu_a(s)}{\mu(s)}$  and  $p_i^b := p_i \cdot \frac{\mu_b(s)}{\mu(s)}$ . Let both  $\mu_i^a := \mu_i$  and  $\mu_i^b := \mu_i$ . Then,

$$\mu = \left( \bigoplus_{i \in I} p \cdot p_i^a \mu_i^a \right) \oplus \left( \bigoplus_{i \in I} (1-p) \cdot p_i^b \mu_i^b \right)$$

is a splitting of  $\mu$ . Note that we use operator  $\oplus$  without index to indicate that the two sums that it joins, actually form one distribution, constructed from two substochastic distributions. Note that

$$\left( \sum_{i \in I} p \cdot p_i^a \right) + \left( \sum_{i \in I} (1-p) \cdot p_i^b \right) = 1.$$

Furthermore,  $(\bigoplus_{i \in I} p_i^a \mu_i^a)$  and  $(\bigoplus_{i \in I} (1-p) \cdot p_i^b \mu_i^b)$  are obviously splittings of  $\mu_a$  and  $\mu_b$ , respectively.

We now show how to construct a splitting  $\gamma_a \oplus \gamma_b$  such that  $\mu_a \hat{\mathcal{L}}(\preceq_{fwd})^\oplus \gamma_a$  and  $\mu_b \hat{\mathcal{L}}(\preceq_{fwd})^\oplus \gamma_b$ . This then immediately satisfies the second and also last condition of simulation-up-to-splitting. We let  $\gamma_i^a := \gamma_i^b := \gamma_i$ . Then,

$$\mu_a = \bigoplus_{i \in I} p_i^a \gamma_i^a \text{ and } \mu_b = \bigoplus_{i \in I} p_i^b \gamma_i^b$$

are splittings of  $\mu_a$  and  $\mu_b$ , respectively. By construction,  $\mu_i^a \hat{\mathcal{L}}(\preceq_{fwd}) \gamma_i^a$  and  $\mu_i^b \hat{\mathcal{L}}(\preceq_{fwd}) \gamma_i^b$ . As the splittings of  $\mu_a$  and  $\gamma_a$ , and of  $\mu_b$  and  $\gamma_b$  are constructed with the same weights  $p_i^a$  and  $p_i^b$ , respectively, this suffices to state that  $\mu_a \hat{\mathcal{L}}(\preceq_{fwd}) \gamma_a$  and  $\mu_b \hat{\mathcal{L}}(\preceq_{fwd}) \gamma_b$ . This ends the first half of our proof.

We now show that the relation  $\mathcal{R} = \{ (s, \mu) \mid \delta(s) \preceq^\dagger \mu \}$  is a probabilistic forward simulation. The crucial part of the second half of our proof is the following claim.

**Claim D.2.** *If  $\mu \preceq^\dagger \gamma$ , then there also must exist a distribution  $\bar{\gamma}^*$  with  $\gamma \preceq^\dagger \text{flatten}(\bar{\gamma}^*)$  that satisfies  $\mu \hat{\mathcal{L}}(\mathcal{R}) \bar{\gamma}^*$ .*

We first proof the claim: Let  $\mu \preceq^\dagger \gamma$ . By Lemma D.9, there are distributions  $\gamma_s$ , indexed by the states  $s \in \text{Supp}(\mu)$ , such that  $\gamma \xRightarrow{c} \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \gamma_s := \bar{\gamma}$  and for each  $s \in \text{Supp}(\mu)$ :

$\delta(s) \preceq^\dagger \gamma_s$ . We let  $\bar{\gamma}^*$  be defined such that  $\bar{\gamma}^*(\gamma_s) = \mu(s)$  for  $s \in \text{Supp}(\mu)$ , and 0 for all other distributions.

Let the function  $w$  be defined as

$$w(s, \gamma_s) = \begin{cases} \mu(s) & s \in \text{Supp}(\mu) \\ 0 & \text{otherwise} \end{cases}$$

By construction, this is a weight function and establishes  $\mu \hat{\mathcal{L}}(\mathcal{R}) \bar{\gamma}^*$ .

It is now easy to show that the relation

$$\mathcal{R} = \{ (s, \gamma) \mid \delta(s) \preceq^\dagger \gamma \}$$

is a probabilistic forward simulation. To see this assume  $s \xrightarrow{a} \mu'$ . Then since  $\delta(s) \preceq^\dagger \gamma$ , there exists  $\gamma'$  such that  $\gamma \xrightarrow{a}_c \gamma'$  with  $\mu' \preceq^\dagger \gamma'$ . By our claim,  $\gamma' \xRightarrow{c} \text{flatten}(\bar{\gamma}^*)$  for a distribution  $\bar{\gamma}^*$  that satisfies  $\mu' \hat{\mathcal{L}}(\mathcal{R}) \bar{\gamma}^*$ , which established the condition of probabilistic forward simulation, except for the last detail: by compositionality of weak combined hypertransitions, we immediately can conclude that  $\gamma \xrightarrow{a}_c \text{flatten}(\bar{\gamma}^*)$ .  $\square$

## D.7. Proof of Theorem 8.14

**Proof step:**  $t \approx_\delta t'$  implies  $t \approx_s t'$  The central insights to prove this implication are that the relation  $\approx_\delta$  is a weak state bisimulation, with the set  $P := \{ (s, \tau, \mu) \in D \mid \delta(s) \approx \mu \}$  being a suitable set of preserving transitions.

In order to prove this implication, we need several auxiliary results. Note that throughout this paragraph, we assume  $\mathcal{R} = \approx_s$  when we use the short-hand notation  $\xRightarrow{P}$  for  $\xRightarrow{P, \mathcal{R}}$ .

**Definition D.2.** Given a MA  $\mathcal{A}$  and a set  $P \subseteq D$ , let  $\rho \perp_P$  denote that  $\rho \xRightarrow{\tau \upharpoonright P}_c \rho'$  implies  $\rho' = \rho$ .  $\triangleleft$

Thus,  $\rho \perp_P$  indicates that  $\rho$  cannot perform any (non-trivial) weak transition using only transitions in  $P$ . In *mec*-contracted MA, this statement is equivalent to the second condition of the special transition predicate  $\xRightarrow{P, \mathcal{R}}$  for  $\mathcal{R} = \approx_s$ . All of the following arguments will be based on *mec*-contracted MA, and strongly rely on the predicate  $\perp_P$ . However, by Lemma 8.9 we know that  $s =_{\text{mec}} t$  implies  $s \approx_\delta t$ . Therefore,  $\rho \xRightarrow{P, \mathcal{R}} \rho$  in an arbitrary MA  $\mathcal{A}$  if  $\rho \perp_P$  in the *mec*-contracted version of  $\mathcal{A}$ .

**Lemma D.11.** Given a *mec*-contracted weakly image-finite MA  $\mathcal{A}$  and a set  $P \subseteq D$ , for every distribution  $\rho$  there exists a distribution  $\rho'$  with  $\rho \xRightarrow{\tau \upharpoonright P}_c \rho'$  and  $\rho' \perp_P$ .

*Proof.* Let  $M$  be the set of distributions reachable from  $\rho$  with Dirac determinate scheduler with  $a = \tau$ , only using transitions from  $P$ . Using [CS02, Prop. 1, 2], this is a finite set. Furthermore, every distribution  $\rho'$  reachable from  $\rho$  with a weak allowed combined transition, i.e.,  $\rho \xRightarrow{\tau \upharpoonright P}_c \rho'$  is a convex combination of the elements of  $M$ . Consider the finite directed graph  $(M, \{ (\mu, \mu') \mid \mu \xRightarrow{\tau \upharpoonright P}_c \mu' \})$ , and for all  $\mu \in M$  and  $\mu' \in \text{Dist}(S)$ .  $\mu \xRightarrow{\tau \upharpoonright P}_c \mu'$  in the MA implies that  $\mu'$  is a convex combination of distributions in  $M$ , reachable from node  $\mu$  inside the graph. Reachability arguments inside the graph thus immediately translate to reachability

arguments in the MA. As this graph is finite, it must contain a bottom strongly connected component. Within the proof, we will see that those bottom strongly connected components reduce to one single distribution without behaviour (we will refer to this statement as  $\star$  in the following). We will now show that any bottom strongly connected component  $\mu$  can be chosen as  $\rho' \perp_P$  as demanded in the lemma. By the construction of the graph, for all distributions  $\gamma, \gamma' \in M$ , we know  $\gamma \xrightarrow{\tau}_c \gamma'$  in the graph implies  $\gamma \xrightarrow{\tau \perp_P}_c \gamma'$  in the MA, and furthermore, that every distribution  $\gamma \in M$  is reachable from  $\rho$ . Thus,  $\rho \xrightarrow{\tau \perp_P}_c \mu$ . Now we prove that  $\mu$  satisfies the maximality condition, i.e.,  $\mu \perp_P$ .

In order to show that all distribution  $\mu'$  in the maximal strongly connected component of  $\mu$  satisfy  $\mu = \mu'$ , i.e. not only all mecs on states are contracted, but also on distributions. We assume that there exists  $\mu' \neq \mu$  in the mec. As every distribution  $\gamma$  such that  $\mu \xrightarrow{\tau \perp_P}_c \gamma$  must be a convex combination of the distributions in this strongly connected component, as  $M$  represents such transitions in the MA faithfully up to convex combinations, and furthermore, the equality  $=$  is preserved by convex combinations of distributions. For simpler notation, we now construct a new MA  $\mathcal{A}'$ , whose set of states is  $\text{Supp}(\mu) \cup \text{Supp}(\mu')$ , and whose transition relation is derived from certain weak combined transitions  $\mu$  and  $\mu'$  can perform. In order to define the transition relation of  $\mathcal{A}'$ , consider some arbitrary (but fixed) weak combined hyper transition  $\mu \xrightarrow{\tau \perp_P}_c \mu'$ . We split  $\mu'$  according to Lemma 4.3, and thus obtain for each  $s \in \text{Supp}(\mu)$  a distribution  $\mu_s$  with  $\text{Supp}(\mu_s) \subseteq \text{Supp}(\mu')$  and  $s \xrightarrow{\tau \perp_P}_c s'$ . We let for each  $s$  the transition  $(s, \tau, \mu_s)$  be a transition of  $\mathcal{A}'$ . With a complete symmetric construction we add a transition for the states in  $\text{Supp}(\mu')$ . Note that by this construction, every state  $s$  of  $\mathcal{A}'$  has *exactly* one outgoing transition (probably  $s \xrightarrow{\tau} \delta(s)$ ).

$\mathcal{A}'$  now has the property that whenever  $\gamma \xrightarrow{\tau} \gamma'$  for some distributions  $\gamma$  and  $\gamma'$  in  $\mathcal{A}'$ , then  $\gamma \xrightarrow{\tau \perp_P}_c \gamma'$  in the original MA. Thus, if we can prove in  $\mathcal{A}'$  that for two states  $s$  and  $t$  it holds that  $s \xRightarrow{=} \delta(t)$ , then this implies  $s \xrightarrow{\tau \perp_P}_c \delta(t)$  in the original MA, which in turn implies  $s \xRightarrow{=} \delta(t)$ . Thus,  $s \xRightarrow{=} \delta(t)$  and  $t \xRightarrow{=} \delta(s)$  in  $\mathcal{A}'$  implies  $s =_{\text{mec}} t$  in the original MA. As all mecs are contracted, it immediately follows  $s = t$  even syntactically (call this statement  $\star_1$ ).

In the following, whenever we talk about states and transitions, we consider those of  $\mathcal{A}'$ . Furthermore, we will use the notation  $\gamma \xrightarrow{\tau} \gamma'$  to mean that  $\gamma' = \bigoplus_{s \in \text{Supp}(\gamma)} \gamma(s) \cdot \gamma'_s$  where for each  $s \in \text{Supp}(\gamma)$ ,  $\gamma'_s$  results from the unique transition  $s \xrightarrow{\tau} \gamma'_s$  (note that by the construction of  $\mathcal{A}'$ , for every state  $s$  this transition exists and is unique).

For two states  $s$  and  $t$ , we write  $s \dashrightarrow_{\epsilon} t$  where  $\epsilon \in \mathbb{R}^{>0}$ , if there is a finite sequence of transitions  $s \xrightarrow{\tau} \gamma_1 \xrightarrow{\tau} \gamma_2 \xrightarrow{\tau} \dots \xrightarrow{\tau} \gamma_k$  for some  $k$  with  $\gamma_i(t) = 0$  for  $i < k$  and  $\gamma_k(t) = \epsilon > 0$ . If this sequence exists, it is unique. If the exact value of  $\epsilon$  does not matter, we write  $s \dashrightarrow t$ .

We are now ready to prove  $(\star)$ . We first show that  $s \dashrightarrow t$  implies  $t \dashrightarrow s$ . Without loss of generality, let  $\mu(s) = \kappa > 0$  and  $s \dashrightarrow_{\epsilon} t$ , but assume that  $t \dashrightarrow s$  does not hold. By definition of  $\dashrightarrow$ , there is a unique finite sequence  $s \xrightarrow{\tau} \gamma_1 \xrightarrow{\tau} \gamma_2 \dots \xrightarrow{\tau} \gamma_k = \gamma_s$  with  $\gamma_i(t) = 0$  for  $1 \leq i < k$  and  $\gamma_s(t) = \epsilon$ . In the following, we will express that  $s$  reaches  $\gamma_s$  in  $k - 1$   $\tau$ -transitions by  $s(\xrightarrow{\tau})^k \gamma_s$ . For improved readability, we will write distributions as compositions of subdistributions by  $\oplus$  in a vector notation emphasizing a splitting of (weak) transition behaviour according to Lemma 4.3.

Assume first that  $k$  is even. As every state of  $\mathcal{A}'$  has exactly one outgoing transition, and by

the property that  $\mu \xrightarrow{\tau} \mu'$  and  $\mu' \xrightarrow{\tau} \mu$ , and as  $k$  is even, clearly  $\mu(\xrightarrow{\tau})^k \mu$ . Consider

$$\mu = \begin{bmatrix} (\mu \ominus s) \\ \oplus \\ \mu(s) \cdot \delta(s) \end{bmatrix} (\xrightarrow{\tau})^k \begin{bmatrix} \mu \ominus (\mu(s) \cdot \gamma_s) \\ \oplus \\ \mu(s) \cdot \gamma_s \end{bmatrix} = \mu.$$

Then, by assumption,  $\mu(s) = \kappa$ , and by the left equality of  $\mu$  we have just arrived,  $\mu(t) \geq \mu(s) \cdot \epsilon$  follows from  $\gamma_s(t) = \epsilon$ .

Let  $\gamma_t$  be the subdistribution such that  $(\mu \ominus t)(\xrightarrow{\tau})^k \gamma_t$ . Consider

$$\mu = \begin{bmatrix} (\mu \ominus t) \\ \oplus \\ \mu(t) \cdot \delta(t) \end{bmatrix} (\xrightarrow{\tau})^k \begin{bmatrix} (\mu \ominus \gamma_t) \\ \oplus \\ \gamma_t \end{bmatrix} = \mu.$$

As  $\mu(s) = \kappa$  and  $t \not\rightarrow s$  and thus  $\gamma_t(s) = 0$ , it must be the case that  $(\mu \ominus \gamma_t)(s) = \mu(s) = \kappa$ . Now let  $\gamma'_t$  be the unique subdistribution such that  $\gamma_t(\xrightarrow{\tau})^k \gamma'_t$ . Consider

$$\mu = \begin{bmatrix} (\mu \ominus \gamma_t) \\ \oplus \\ \gamma_t \end{bmatrix} (\xrightarrow{\tau})^k \begin{bmatrix} (\mu \ominus \gamma'_t) \\ \oplus \\ \gamma'_t \end{bmatrix} = \mu.$$

Now, as  $t(\xrightarrow{\tau})^k \gamma_t$  and  $t \not\rightarrow s$  also  $\gamma_t \not\rightarrow s$  and thus,  $\gamma'_t(s) = 0$ . This and our right equality with  $\mu$  above implies that then  $(\mu \ominus \gamma'_t)(s) = \mu(s) = \kappa$ . Furthermore,  $(\mu \ominus \gamma_t)(s) = \kappa$  (as we have shown already above), and  $s(\xrightarrow{\tau})^k \gamma_s$  and  $\gamma_s(t) = \epsilon$ , also  $(\mu \ominus \gamma'_t)(t) \geq \mu(s) \cdot \epsilon$  must hold.

Now let  $\gamma_t^1 := \gamma'_t \oplus ((\mu \ominus \gamma'_t) \ominus t)$ . Note that  $|\gamma_t^1| \geq 2 \cdot \mu(s) \cdot \epsilon$ , as  $|\gamma'_t| \geq \epsilon$  and  $(\mu \ominus \gamma'_t)(t) \geq \mu(s) \cdot \epsilon$ .

Let  $\gamma_t^{1'}$  be the subdistribution such that  $\gamma_t^1(\xrightarrow{\tau})^k \gamma_t^{1'}$ . Consider

$$\mu = \begin{bmatrix} (\mu \ominus \gamma_t^1) \\ \oplus \\ \gamma_t^1 \end{bmatrix} (\xrightarrow{\tau})^k \begin{bmatrix} (\mu \ominus \gamma_t^{1'}) \\ \oplus \\ \gamma_t^{1'} \end{bmatrix} = \mu.$$

We can derive that as  $t \not\rightarrow s$ , and  $\gamma_t^1$  only consists of a convex combination of distributions reachable from  $t$ , also  $\gamma_t^1 \not\rightarrow s$  and thus  $\gamma_t^{1'}(s) = 0$ . As above, we hence conclude that  $(\mu \ominus \gamma_t^{1'})(s) = \mu(s) = \kappa$  and, also exactly as above, that  $(\mu \ominus \gamma_t^{1'})(t) = \mu(s) \cdot \epsilon$  since  $(\mu \ominus \gamma_t^1)(s) = \kappa$  and  $s(\xrightarrow{\tau})^k \gamma_s$  and  $\gamma_s(t) = \epsilon$ .

Analogously to the above, set  $\gamma_t^2 := \gamma_t^{1'} \oplus ((\mu \ominus \gamma_t^{1'}) \ominus t)$ . Again analogously, we can derive that  $|\gamma_t^2| \geq 3 \cdot \mu(s) \cdot \epsilon$ .

Proceeding with this argumentation, it is straightforward to inductively define  $\gamma_t^i$  for every  $0 < i \in \mathbb{N}$  and show that  $|\gamma_t^i| \geq (i+1) \cdot \mu(s) \cdot \epsilon$ . As  $\mu(s)$  and  $\epsilon$  are fixed numbers, this implies that for some  $i'$ ,  $|\gamma_t^{i'}| > 1$ . This is a clear contradiction to the fact that  $\gamma_t^{i'}$  is a subdistribution. Hence, our initial assumption that  $t \not\rightarrow s$  must have been false. The case for  $k$  odd a bit more technical, but is based on the same ideas.

With the just proven property, it is straightforward to establish that  $\mathcal{W} = \{(s, t) \mid s \dashrightarrow t\}$  is an equivalence relation. We now show that for every state  $s$ ,  $\mu([s]_{\mathcal{W}}) = \mu'([s]_{\mathcal{W}})$ . Recall first that  $\mu \xrightarrow{\tau} \mu'$  and  $\mu' \xrightarrow{\tau} \mu$ . Furthermore, whenever  $s \xrightarrow{\tau} \gamma$  for some  $\gamma$ , for every state

$t \in \text{Supp}(\gamma)$ , we have  $[t]_{\mathcal{W}} = [s]_{\mathcal{W}}$ . Thus, for every state  $s \in \text{Supp}(\mu)$ ,  $\mu'([s]_{\mathcal{W}}) \geq \mu([s]_{\mathcal{W}})$ , and analogously with  $\mu$  and  $\mu'$  swapped. Thus,  $\mu([s]_{\mathcal{W}}) = \mu'([s]_{\mathcal{W}})$  follows.

We will now show, that for every two states  $t, t' \in [s]_{\mathcal{W}}$ , we have  $t \Longrightarrow \delta(t')$  and  $t' \Longrightarrow \delta(t)$ . By  $(\star_1)$ , this implies  $t = t'$ . Thus,  $[s]_{\mathcal{W}}$  consists of a single state and  $\mu([s]_{\mathcal{W}}) = \mu'([s]_{\mathcal{W}})$  also implies  $\mu(s) = \mu'(s)$ , which completes the proof. Now, in order to show that for every two states  $t, t' \in [s]_{\mathcal{W}}$ , we have  $t \Longrightarrow \delta(t')$  and  $t' \Longrightarrow \delta(t)$ , it suffices to show that every state  $t \in [s]_{\mathcal{W}}$  can reach every other state  $t' \in [s]_{\mathcal{W}}$  with probability 1. For this, it suffices to construct a sequence of distributions  $\langle \rho_i \rangle_{i \in \mathbb{N}}$  with  $t \Longrightarrow \rho_i$  for every  $i \in \mathbb{N}$ , that in infinity converge against  $\delta(t')$ , i.e.  $\langle \rho_i \rangle_{i \in \mathbb{N}} \longrightarrow \delta(t')$ . Let  $\kappa = \min\{\epsilon \mid s' \in [s]_{\mathcal{W}} \wedge s' \dashrightarrow_{\epsilon} t'\}$ . The minimum  $\kappa$  exists, as  $[s]_{\mathcal{W}}$  is finite. Let  $\gamma$  be the distribution that establishes  $t \dashrightarrow_{\epsilon} t'$ . Obviously,  $\gamma(t') \geq \kappa$ . As every state  $s' \in [s]_{\mathcal{W}}$  satisfies  $s' \dashrightarrow_{\epsilon_{s'}} t'$  for some  $\epsilon_{s'} \geq \kappa$ , there must exist a distribution  $\gamma'$  such that  $(\gamma \ominus t') \Longrightarrow \gamma'$  and  $\gamma'(t') \geq \kappa$ . Let  $\gamma_1 := (\gamma \ominus t') \oplus \gamma'$ . Thus,  $\gamma_1(t') \geq \kappa + (1 - \kappa)\kappa$ . We construct  $\gamma_i$  analogously for every  $i \in \mathbb{N}$ . Clearly,  $\langle \gamma_i(t') \rangle_{i \in \mathbb{N}} \longrightarrow 1$ . Thus, the distributions must be equal. This finishes our proof.  $\square$

**Corollary D.1.** *Given a weakly image-finite MA  $\mathcal{A}$  and a set  $P \subseteq D$ , for every distribution  $\rho$  there exists a distribution  $\rho'$  with  $\rho \xrightarrow{P, \approx_s} \rho'$ .*

*Proof.* This result follows immediately from Lemma D.11 and Lemma 8.9.  $\square$

**Lemma D.12.** *Given a weakly image-finite MA  $\mathcal{A}$  and the set  $P = \{(s, \tau, \mu) \in D \mid \delta(s) \approx \mu\}$ , if  $\mu \xrightarrow{\tau \upharpoonright P}_c \mu'$ , then  $\mu \approx \mu'$ .*

*Proof.* Let  $\{s \xrightarrow{\tau \upharpoonright P}_c \mu_s\}_{s \in \text{Supp}(\mu)}$  be the family of allowed weak combined transitions justifying  $\mu \xrightarrow{\tau \upharpoonright P}_c \mu'$ , i.e.,  $\mu' = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s$ .

For each state  $s \in \text{Supp}(\mu)$ , let  $\sigma_s$  be a determinate scheduler inducing the transition  $s \xrightarrow{\tau \upharpoonright P}_c \mu_s$ ; for each  $n \in \mathbb{N}$  and each  $\alpha \in \text{frags}^*(\mathcal{A})$ , define  $\sigma_s^n(\alpha)$  as follows:

$$\sigma_s^n(\alpha) = \begin{cases} \sigma_s(\alpha) & \text{if } |\alpha| < n, \\ \delta(\perp) & \text{otherwise.} \end{cases}$$

It is worthwhile to note that for each  $0 < n \in \mathbb{N}$  and each  $\alpha \in \text{frags}^*(\mathcal{A})$ ,  $\sigma_s^{n-1}(\alpha) = \sigma_s^n(\alpha) = \sigma_s(\alpha)$  when  $|\alpha| < n - 1$ ,  $\sigma_s^{n-1}(\alpha) = \delta(\perp)$  and  $\sigma_s^n(\alpha) = \sigma_s(\alpha)$  for  $|\alpha| = n - 1$ , and  $\sigma_s^{n-1}(\alpha) = \sigma_s^n(\alpha) = \delta(\perp)$  when  $|\alpha| \leq n$ . Moreover, when  $|\alpha| = n - 1$ , we can split the probability 1 of stopping as  $\sigma_s^{n-1}(\alpha) = \delta(\perp) \cdot (\sigma_s^n(\alpha)(\perp) + \sigma_s^n(\alpha)(D(\tau)))$ , i.e., the probability of stopping according to  $\sigma_s$  plus the probability of stopping instead of performing some transition according to  $\sigma_s$ .

Let  $\mu_s^n$  be the target probability distribution of the allowed weak combined transition  $s \xrightarrow{\tau \upharpoonright P}_c \mu_s^n$  as induced by  $\sigma_s^n$ . It is immediate to see that  $\lim_{n \rightarrow \infty} \sigma_s^n = \sigma_s$  and hence  $\lim_{n \rightarrow \infty} \mu_s^n = \mu_s$ . We now prove by induction that  $\mu_s^n \approx \delta(s)$ : if  $n = 0$ , then by definition of  $\sigma_s^0$ ,  $\mu_s^0 = \delta(s)$ , hence  $\mu_s^0 \approx \delta(s)$  trivially holds. If  $n > 0$ , then consider the two distributions  $\mu_s^{n-1}$  and  $\mu_s^n$  where  $\mu_s^n$  is obtained by extending  $\mu_s^{n-1}$  with a single step according to  $\sigma_s^n$ . Let  $\nu_s^{n-1}$ ,  $\nu_s^{n\perp}$ , and  $\nu_s^n$  be three

sub-probability distributions such that for each  $t \in \text{Supp}(\mu_s^{n-1})$ ,

$$\begin{aligned}\nu_s^{n-1}(t) &= \sum_{\alpha \in F(t, <, n-1)} \mu_{\sigma_s^{n-1}, s}(\alpha), \\ \nu_s^{n\perp}(t) &= \sum_{\alpha \in F(t, =, n-1)} \mu_{\sigma_s^{n-1}, s}(C_\alpha) \cdot \sigma_s^n(\alpha)(\perp), \text{ and} \\ \nu_s^n(t) &= \sum_{\alpha \in F(t, =, n-1)} \mu_{\sigma_s^{n-1}, s}(C_\alpha) \cdot \sigma_s^n(\alpha)(D(\tau)),\end{aligned}$$

where  $t(v, \trianglelefteq, m)$  is the set  $\{\alpha \in \text{frags}^*(\mathcal{A}) \mid \text{last}(\alpha) = v, |\alpha| \leq m\}$ . It is easy to verify, by simple manipulation on probabilistic execution fragments, that  $\mu_s^{n-1} = \nu_s^{n-1} \oplus \nu_s^{n\perp} \oplus \nu_s^n = \nu_s^{n-1} \oplus \nu_s^{n\perp} \oplus \bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \delta(t)$  and that  $\mu_s^n = \nu_s^{n-1} \oplus \nu_s^{n\perp} \oplus \bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \sum_{tr \in D(\tau)} \frac{\sigma_s^n(t)(tr)}{\sigma_s^n(t)(D(\tau))} \cdot \mu_{tr}$ . Since  $\delta(t) = \bigoplus_{tr \in D(\tau)} \frac{\sigma_s^n(t)(tr)}{\sigma_s^n(t)(D(\tau))} \cdot \delta(t)$  for  $t \in \text{Supp}(\nu_s^n)$ , by the choice of  $P$  and by Corollary 8.2, it follows that  $\bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \delta(t) \approx \bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \sum_{tr \in D(\tau)} \frac{\sigma_s^n(t)(tr)}{\sigma_s^n(t)(D(\tau))} \cdot \mu_{tr}$ , thus  $\mu_s^{n-1} = \nu_s^{n-1} \oplus \nu_s^{n\perp} \oplus \bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \delta(t) \approx \nu_s^{n-1} \oplus \nu_s^{n\perp} \oplus \bigoplus_{t \in \text{Supp}(\nu_s^n)} \nu_s^n(t) \cdot \sum_{tr \in D(\tau)} \frac{\sigma_s^n(t)(tr)}{\sigma_s^n(t)(D(\tau))} \cdot \mu_{tr} = \mu_s^n$ , hence  $\delta(s) \approx \mu_s^n$  follows by induction hypothesis  $\delta(s) \approx \mu_s^{n-1}$  and transitivity of  $\approx$ .

This completes the proof that for each  $n \in \mathbb{N}$ ,  $\mu_s^n \approx \delta(s)$ , thus  $\mu_s \approx \delta(s)$  as  $\lim_{n \rightarrow \infty} \mu_s^n = \mu_s$ . Since  $\mu = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \delta(s)$ ,  $\mu' = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s$  and  $\mu_s \approx \delta(s)$  for each  $s \in \text{Supp}(\mu)$ , by Corollary 8.2, it follows that  $\mu \approx \mu'$ .  $\square$

**Definition D.3.** Given a MA  $\mathcal{A}$  and an allowed weak hyper transition  $\mu \xrightarrow{\tau \mid \bar{A}}_c \nu$ , let  $\{\sigma_s\}_{s \in \text{Supp}(\mu)}$  be the family of schedulers each one inducing the allowed weak combined transition  $s \xrightarrow{\tau \mid \bar{A}}_c \nu_s$  such that  $\nu = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \nu_s$ .

We say that  $\iota$  is an *intermediate distribution* in  $\mu \xrightarrow{\tau \mid \bar{A}}_c \nu$  if  $\iota = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \iota_s$  where the transition  $s \xrightarrow{\tau \mid \bar{A}}_c \iota_s$  is induced by a scheduler  $\sigma'_s$  such that for each finite execution fragment  $\alpha$  and transition  $tr$ ,  $\sigma'_s(\alpha)(tr) \in \{0, \sigma_s(\alpha)(tr)\}$ .  $\triangleleft$

Intuitively,  $\iota$  is any possible distribution we can reach from  $\mu$  in the weak hyper transition to  $\nu$  by stopping before reaching  $\nu$ .

**Lemma D.13.** Given a MA  $\mathcal{A}$ , an allowed weak hyper transition  $\mu \xrightarrow{\tau \mid \bar{A}}_c \nu$ , and an intermediate distribution  $\iota$ , it holds that  $\mu \xrightarrow{\tau \mid \bar{A}}_c \iota$  and  $\iota \xrightarrow{\tau \mid \bar{A}}_c \nu$ .

**Lemma D.14.** If  $s \xrightarrow{\tau} \mu$  and  $\delta(s) \not\approx \mu$ , then also for every distribution  $\gamma$  and  $0 < k \leq 1$ :

$$\delta(s) \oplus_k \gamma \not\approx \mu \oplus_k \gamma.$$

*Proof.* For a concise notation, let  $\xi = k\delta(s)$  and  $\gamma_0 = (1 - k)\gamma$  and  $\mu_0 = k\mu$ . To arrive at a contradiction, assume  $\xi \oplus \gamma_0 \approx \mu' \oplus \gamma_0$ . Then some  $\rho$  must exist such that  $\mu' \oplus \gamma_0 \Rightarrow_c \rho$  and  $\rho \approx \xi \oplus \gamma_0$  must hold. Without loss of generality, say  $\rho = \mu_0^a \oplus \mu_0^b \oplus \gamma_0^a \oplus \gamma_0^b$  where  $\xi \approx \mu_0^a \oplus \gamma_0^b$  and  $\gamma_0 \approx \mu_0^b \oplus \gamma_0^a$  and  $\gamma_0 \Rightarrow_c \gamma_0^a \oplus \gamma_0^b$  and  $\mu_0 \Rightarrow_c \mu_0^a \oplus \mu_0^b$ .

As  $\gamma_0 \approx \mu_0^b \oplus \gamma_0^a$  and  $\gamma_0 \Rightarrow_c \gamma_0^a \oplus \gamma_0^b$  there must exist distributions  $\mu_1^s, \mu_1^c, \gamma_1^s, \gamma_1^c$  such that  $\gamma_0^a \Rightarrow_c \gamma_1^s \oplus \gamma_1^c$  and  $\mu_0^b \Rightarrow_c \mu_1^s \oplus \mu_1^c$  and  $\gamma_0^b \approx \mu_1^s \oplus \gamma_1^s$  and  $\gamma_0^a \approx \mu_1^c \oplus \gamma_1^c$ .

By replacing equals by equals (with respect to  $\approx$ ), we can infer that  $\xi \approx \mu_0^a \oplus \mu_1^s \oplus \gamma_1^c$ .



Similarly, as  $\gamma_0^a \approx \mu_1^c \oplus \gamma_1^s$  and  $\gamma_0^a \Rightarrow_c \gamma_1^s \oplus \gamma_1^c$  there must exist distributions  $\mu_2^{cs}, \mu_2^{cc}, \gamma_2^{ss}, \gamma_2^{sc}$  such that  $\gamma_1^s \Rightarrow_c \gamma_2^{ss} \oplus \gamma_2^{sc}$  and  $\mu_1^c \Rightarrow_c \mu_2^{cs} \oplus \mu_2^{cc}$  and  $\gamma_1^s \approx \mu_2^{cc} \oplus \gamma_2^{ss}$  and  $\gamma_1^c \approx \mu_2^{cs} \oplus \gamma_2^{sc}$ . By again replacing the last component of this sum of distributions by the equivalent term we have just derived, we obtain  $\xi \approx \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \gamma_2^{sc}$ .

We can now repeat this argument completely analogously. Starting the argumentation from  $\gamma_1^s \Rightarrow_c \gamma_2^{ss} \oplus \gamma_2^{sc}$  and  $\gamma_1^s \approx \mu_2^{cc} \oplus \gamma_2^{ss}$ , we see that there must exist distributions  $\mu_3^{ccs}, \mu_3^{ccc}, \gamma_3^{ssc}, \gamma_3^{ssc}$  such that  $\gamma_2^{ss} \Rightarrow_c \gamma_3^{ssc} \oplus \gamma_3^{ssc}$  and  $\mu_2^{cc} \Rightarrow_c \mu_3^{ccs} \oplus \mu_3^{ccc}$  and  $\gamma_2^{ss} \approx \mu_3^{ccc} \oplus \gamma_3^{ssc}$  and  $\gamma_2^{sc} \approx \mu_3^{ccs} \oplus \gamma_3^{ssc}$ . By again replacing the last component of this sum of distributions by the equivalent term we have just derived, we obtain  $\xi \approx \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \gamma_3^{ssc}$ .

Repeating this pattern, we find that for every  $i \in \mathbb{N}$  we can inductively draw the following conclusion: there must exist distributions  $\mu_{i+1}^{c \dots cs}, \mu_{i+1}^{c \dots cc}, \gamma_{i+1}^{s \dots sc}, \gamma_{i+1}^{s \dots sc}$  such that  $\gamma_i^{s \dots ss} \Rightarrow_c \gamma_{i+1}^{s \dots ss} \oplus \gamma_{i+1}^{s \dots sc}$  and  $\mu_i^{c \dots c} \Rightarrow_c \mu_{i+1}^{c \dots cs} \oplus \mu_{i+1}^{c \dots cc}$  and  $\gamma_i^{s \dots ss} \approx \mu_{i+1}^{c \dots cc} \oplus \gamma_{i+1}^{s \dots ss}$  and  $\gamma_i^{s \dots sc} \approx \mu_{i+1}^{c \dots cs} \oplus \gamma_{i+1}^{s \dots sc}$ . Note that the number of  $c$  and  $s$  in the superscript always is the number stated in the subscript of the distributions. By successively repeating our replacement of equals by equals as shown above, we derive for the general case that  $\xi \approx \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \gamma_{i+1}^{s \dots sc}$ .

Now first note that for each  $i \in \mathbb{N}$ ,  $\mu_0 \Rightarrow_c \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \mu_{i+1}^{c \dots cc}$  must hold by transitivity of  $\Rightarrow_c$  and our derivations above.

As  $|\xi|$  is bound and distributions have non-negative size, the sequence

$$\langle \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \rangle_{i \in \mathbb{N}}$$

must be increasing, and thus, the sequence  $\langle \gamma_{i+1}^{s \dots sc} \rangle_{i \in \mathbb{N}}$  must be decreasing (not necessarily strictly, though). Thus, there must exist  $\lambda \in \mathbb{R}^{\geq 0}$  such that  $\langle \gamma_{i+1}^{s \dots sc} \rangle_{i \in \mathbb{N}} \rightarrow \lambda$ . First assume  $\lambda = 0$ . This implies that the difference of  $\xi$  and  $\mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \gamma_{i+1}^{s \dots sc}$  can become arbitrarily small up to  $\approx$ . More formally, when  $\rho$  is the distribution, such that

$$\langle \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \gamma_{i+1}^{s \dots sc} \rangle_{i \in \mathbb{N}} \rightarrow \rho,$$

then  $\rho \approx \xi$ . Formally, this can be shown using Corollary D.4.

As  $|\gamma_{i+1}^{c \dots cc}| = |\mu_{i+1}^{c \dots cc}|$  for all  $i \in \mathbb{N}$  and  $\langle |\gamma_{i+1}^{s \dots sc}| \rangle \rightarrow 0$ , also  $\langle |\mu_{i+1}^{c \dots cc}| \rangle \rightarrow 0$ . This implies also  $\langle \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \mu_{i+1}^{c \dots cc} \rangle_{i \in \mathbb{N}} \rightarrow \rho$ .

As for each  $i \in \mathbb{N}$ ,  $\mu_0 \Rightarrow_c \mu_0^a \oplus \mu_1^s \oplus \mu_2^{cs} \oplus \mu_3^{ccs} \oplus \dots \oplus \mu_{i+1}^{c \dots cs} \oplus \mu_{i+1}^{c \dots cc}$  holds, and by [EHZ10c, Lemma 2]) then also  $\mu_0 \Rightarrow_c \rho \approx \xi$ . Since  $\xi = k\delta(s)$  and  $\mu_0 = k\mu$ , this is a contradiction to the premise of the lemma that  $\delta(s) \not\approx \mu$ . Thus, our initial assumption that  $\xi \oplus \gamma_0 \approx \mu' \oplus \gamma_0$  must have been wrong.

Now assume  $\lambda > 0$ . Then there must exist  $0 < \lambda' < \lambda$  and  $k' \in \mathbb{N}$  such that for every  $k' > k$ ,  $|\gamma_k^{s \dots ss}| \geq \lambda'$ . As for every  $i \in \mathbb{N}$ ,  $\gamma_i^{s \dots ss} \Rightarrow_c \gamma_{i+1}^{s \dots ss} \oplus \gamma_{i+1}^{s \dots sc}$ , clearly  $|\gamma_{i+1}^{s \dots ss} \oplus \gamma_{i+1}^{s \dots sc}| = |\gamma_i^{s \dots ss}|$  and hence for  $i \geq k$  also  $|\gamma_{i+1}^{s \dots ss}| = |\gamma_i^{s \dots ss}| - |\gamma_{i+1}^{s \dots sc}| \leq |\gamma_i^{s \dots ss}| - \lambda'$ . Using these inequalities, we can bound  $|\gamma_{k+m}^{s \dots ss}|$  for an arbitrary  $m \in \mathbb{N}$  from above by  $|\gamma_k^{s \dots ss}| \leq |\gamma_{k+m}^{s \dots ss}| + m\lambda'$ . But for  $m$  large enough the right-hand side of the inequality becomes negative, which is a contradiction to the fact that distributions have non-negative size. Thus, the case  $\lambda > 0$  leads to a contradiction. Thus, again, our initial assumption that  $\xi \oplus \gamma_0 \approx \mu' \oplus \gamma_0$  must have been wrong.  $\square$

**Lemma D.15.** Given a weakly image-finite MA  $\mathcal{A}$  and the set  $P = \{(s, \tau, \mu) \in D \mid \delta(s) \approx \mu\}$ , if  $\mu \Rightarrow_c \mu'$  and  $\mu \approx \mu'$ , then also  $\mu \xRightarrow{\tau|P} \mu'$ .

*Proof.* By definition of weak hyper transition,  $\mu \xRightarrow{\tau} \mu'$  implies that there exists a family of weak combined transitions  $\{s \xRightarrow{\tau} \mu_s\}_{s \in \text{Supp}(\mu)}$  such that  $\mu = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \mu_s$ . If each

transition  $s \xrightarrow{\tau}_c \mu_s$  is already an allowed weak combined transition  $s \xrightarrow{\tau \downarrow P}_c \mu_s$ , then this implies that  $\{s \xrightarrow{\tau}_c \mu_s\}_{s \in \text{Supp}(\mu)}$  is actually the family  $\{s \xrightarrow{\tau \downarrow P}_c \mu_s\}_{s \in \text{Supp}(\mu)}$ , hence  $\mu \xrightarrow{\tau \downarrow P}_c \mu'$ , as required.

Suppose, on the contrary, that there exists some transition  $s \xrightarrow{\tau}_c \mu_s$  that is not an allowed weak combined transition  $s \xrightarrow{\tau \downarrow P}_c \mu_s$ . Let  $\sigma_s$  be the determinate scheduler inducing  $s \xrightarrow{\tau}_c \mu_s$  (cf. [HT12, Th. 2]) and define  $\sigma'_s$  as  $\sigma_s$  except for the fact that  $\sigma'_s$  chooses  $\perp$  when  $\sigma_s$  schedules a transition not in  $P$ . Formally, for each finite execution fragment  $\alpha$  and each transition  $tr$ ,

$$\sigma'_s(\alpha)(tr) = \begin{cases} \sigma_s(\alpha)(tr) & \text{if } tr \in P, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, given a transition  $tr_t = t \xrightarrow{\tau} \mu_t \notin P$ , define  $\sigma''_s$  as follows: for each finite execution fragment  $\alpha$  and each transition  $tr$ ,

$$\sigma''_s(\alpha)(tr) = \begin{cases} \sigma_s(\alpha)(tr) & \text{if } tr \in P, tr \neq tr_t, \\ \sigma_s(\alpha)(tr) & \text{if } \alpha = \alpha' \tau t, \text{last}(\alpha') \neq t, \text{ and } tr = tr_t, \text{ and} \\ 0 & \text{if } \alpha = \alpha' \tau v, v \neq t, \text{ and } \text{last}(\alpha') = t. \end{cases}$$

It is immediate to see that  $\sigma'_s$  induces an allowed weak combined transition  $s \xrightarrow{\tau \downarrow P}_c \iota_s$  for some distribution  $\iota_s$ ; and that  $\sigma''_s$  induces a weak combined transition  $s \xrightarrow{\tau}_c \iota'_s$  for the distribution  $\iota'_s$  that is  $\iota_s$  extended from state  $t$  with the transition  $tr_t$ . Since we have an allowed weak combined transition  $s \xrightarrow{\tau \downarrow P}_c \iota_s$  for each  $s \in \text{Supp}(\mu)$ , then there exists a family of allowed weak combined transitions  $\{s \xrightarrow{\tau \downarrow P}_c \iota_s\}_{s \in \text{Supp}(\mu)}$  inducing the allowed hypertransition  $\mu \xrightarrow{\tau \downarrow P}_c \iota$  where  $\iota = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \iota_s$  is an intermediate distribution and thus, by Lemma D.12, we have  $\mu \approx \iota$ . Also  $\iota' = \bigoplus_{s \in \text{Supp}(\mu)} \mu(s) \cdot \iota'_s$  is an intermediate distribution; however  $\iota' \not\approx \iota$  by Lemma D.14 where  $k = \iota(t) \cdot \sum_{s \in \text{Supp}(\mu)} \mu(s) \cdot \sigma'_s(t)(tr_t)$  and  $\gamma = (\iota \ominus t) \oplus (\iota(t) - k)\delta(t)$ .

Since  $\iota'$  is an intermediate transition, Lemma D.13 implies that  $\mu \xrightarrow{\tau}_c \iota'$  and  $\iota' \xrightarrow{\tau}_c \mu'$ , hence by Lemma 8.2 we have  $\iota' \approx \mu'$ , thus  $\iota \approx \mu \approx \mu' \approx \iota'$ , that is,  $\iota \approx \iota'$  by transitivity of  $\approx$  but this contradicts  $\iota \not\approx \iota'$ . This implies that  $tr_t \in P$ , hence  $\mu \xrightarrow{\tau \downarrow P}_c \mu'$ , as required.  $\square$

**Lemma D.16.** Let  $\mathcal{A}$  be a *mec*-contracted weakly image-finite Markov automaton and  $P = \{(s, \tau, \mu) \in D \mid \delta(s) \approx \mu\}$ . Then  $\mu \approx \gamma$  implies there exist  $\mu^*$  and  $\gamma^*$  such that  $\mu \xrightarrow{P} \mu^*$  and  $\gamma \xrightarrow{P} \gamma^*$  and  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$ . Furthermore,  $\mu^*$  and  $\gamma^*$  are unique up to  $\mathcal{L}(\approx_\delta)$ .

*Proof.* Let  $\mu \approx \gamma$  and  $P = \{(s, \tau, \mu) \in D \mid \delta(s) \approx \mu\}$ . Choose some arbitrary  $\mu^*$  and  $\gamma^*$  such that  $\mu \xrightarrow{P} \mu^*$  and  $\gamma \xrightarrow{P} \gamma^*$ . They exist by Corollary D.1. As we do not impose any further restrictions on  $\mu^*$  and  $\gamma^*$ , we will also prove their uniqueness up to  $\mathcal{L}(\approx_\delta)$  by showing that  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$  now.

It remains to establish  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$ . If already  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$ , we are done. Hence assume *not*  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$ . From our choice of  $P$ , the first condition in Definition 8.15, Corollary 8.2, and transitivity of  $\approx$ , we conclude  $\mu \approx \mu^*$ , and  $\gamma \approx \gamma^*$ . Then, from Theorem 8.15, we see that there must exist  $\mu'$  and  $\gamma'$  with  $\mu^* \xrightarrow{c} \mu'$ ,  $\gamma^* \xrightarrow{c} \gamma'$ ,  $\mu' \mathcal{L}(\approx_\delta) \gamma'$ ,  $\mu^* \approx \mu'$ , and  $\gamma^* \approx \gamma'$ . By Lemma D.15, it immediately follows that  $\mu^* \xrightarrow{\tau \downarrow P}_c \mu'$  and  $\gamma^* \xrightarrow{\tau \downarrow P}_c \gamma'$  must hold. By Condition 2 of Definition 8.15, this implies that already  $\mu^* \mathcal{L}(\approx_\delta) \mu'$  and  $\gamma^* \mathcal{L}(\approx_\delta) \gamma'$ . Thus, by transitivity of  $\approx_\delta$  and  $\mu' \mathcal{L}(\approx_\delta) \gamma'$ , we conclude  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$ .  $\square$

**Proposition 5.** *Given a mec-contracted, weakly image-finite MA  $\mathcal{A}$ , let*

$$P := \{ (s, \tau, \mu) \in D \mid \delta(s) \approx \mu \}.$$

*The pair  $(\approx_\delta, P)$  is a weak state bisimulation pair.*

*Proof.*

- We first check the necessary condition of weak state bisimulation according to Definition 8.19. Let hence  $s \approx_\delta t$ .
  1. If  $s \xRightarrow{a}_c \mu'$ , then there exists  $\mu', \gamma'$  and  $\gamma$  such that  $\mu' \xRightarrow{P} \mu$  and  $t \xRightarrow{a}_c \gamma' \xRightarrow{P} \gamma$  and  $\mu \mathcal{L}(\mathcal{R}) \gamma$ . As  $\delta(s) \approx \delta(t)$ , clearly  $t \xRightarrow{a}_c \gamma'$  holds for some  $\gamma$  and  $\mu' \approx \gamma'$ . By Lemma D.16, there exist  $\mu^*$  and  $\gamma^*$ , which are a suitable choice for  $\mu$  and  $\gamma$  respectively.
  2. Assume  $s \downarrow$  then by the definition of weak distribution bisimulation it immediately follows that  $t \xRightarrow{a}_c \gamma$  with  $\gamma \downarrow$ .
- To establish that  $P$  is indeed a set of preserving transitions, let  $(s, \tau, \mu) \in P$ , and let  $s \xRightarrow{a}_c \gamma$  for some distribution  $\gamma$ . Using Lemma 8.5, as  $\delta(s) \approx \mu$ , it must hold that there exists  $\hat{\gamma}$  such that  $\mu \xRightarrow{a}_c \hat{\gamma}$  and  $\gamma \approx \hat{\gamma}$ . Applying Lemma D.16 to  $\gamma$  and  $\hat{\gamma}$ , we see that there exist  $\gamma'$  and  $\hat{\gamma}'$  such that  $\gamma \xRightarrow{P} \gamma', \hat{\gamma} \xRightarrow{P} \hat{\gamma}'$ , and  $\gamma' \mathcal{L}(\approx_\delta) \hat{\gamma}'$ .

□

**Proof step:**  $t \approx_s t'$  implies  $t \approx_\delta t'$

Let in the following  $\mathcal{A}$  be a weakly image-finite MA and  $(\mathcal{R}, P)$  a weak state bisimulation pair. As usual, we omit  $\mathcal{R}$  in the notation  $\xRightarrow{P, \mathcal{R}}$  in the following.

**Lemma D.17.** Let  $\mathcal{R}$  be an arbitrary equivalence relation over  $S$ . Let  $\mathcal{P}$  be a predicate over  $S$  such that  $\mathcal{P}(s)$  holds if and only if there exists no distribution  $\mu$  such that  $s \xRightarrow{\tau \upharpoonright P}_c \mu$  and *not*  $\delta(s) \mathcal{L}(\mathcal{R}) \mu$ . We lift  $\mathcal{P}$  to distributions over  $S$  by the usual state-wise extension.

For an arbitrary distribution  $\mu^*$   $\mathcal{P}(\mu^*)$  holds *if and only if* whenever  $\xRightarrow{\tau \upharpoonright P}_c \mu'$ , then  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$ .

*Proof.* Let  $s \in \text{Supp}(\mu^*)$  and assume that  $\neg \mathcal{P}(s)$  and thus also  $\neg \mathcal{P}(\mu^*)$ . Clearly,  $s \xRightarrow{\tau \upharpoonright P}_c \mu$  implies that there is a transition  $\mu^* \xRightarrow{\tau \upharpoonright P}_c \mu \oplus_{\mu^*(s)} (\mu^* \ominus s) := \mu'$ . As *not*  $\mu \mathcal{L}(\mathcal{R}) \delta(s)$ , it cannot either be the case that  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$ . Hence, with  $\mu'$  we have a witness that shows that the right-hand side of the *if-and-only-if* must be false, either. For the other directions, assume that the condition on the right-hand side is violated. This means that there exists  $\mu'$  such that  $\mu^* \xRightarrow{\tau \upharpoonright P}_c \mu'$ , and *not*  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$ . Trivially, there must be at least one state in  $\text{Supp}(\mu^*)$  that, as part of this hypertransition of  $\mu^*$ , reaches a distributions  $\mu'_s$  that does not satisfy  $\delta(s) \mathcal{L}(\mathcal{R}) \mu'_s$ , as otherwise  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$  must hold. Immediately,  $\neg \mathcal{P}(\mu^*)$  follows. □

**Corollary D.2.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \xrightarrow{\quad}, \neg \rightarrow)$  be a weakly image-finite Markov automaton. Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  be a sequence of distribution over  $S$  converging against the distribution  $\gamma$ . Let  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be a sequence of distribution satisfying

$$\gamma_i \xRightarrow{\tau \upharpoonright P}_c \gamma'_i.$$

Then, there exists a distribution  $\gamma'$  such that

- $\gamma \xrightarrow{\tau \upharpoonright P}_c \gamma'$ ,
- $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ .

In addition, let  $\mathcal{P}$  be an arbitrary predicate over states that is lifted to distributions by letting  $\mathcal{P}(\gamma)$  hold if and only if  $\mathcal{P}(s)$  holds for every  $s \in \text{Supp}(\gamma)$ . Then, if each  $\gamma'_j$  satisfies  $\mathcal{P}(\gamma'_j)$  for  $j \in J$ , then also  $\mathcal{P}(\gamma')$ .

*Proof.* The transition relation  $\xrightarrow{\tau \upharpoonright P}_c$  in an automaton  $\mathcal{A}$  can be interpreted as an ordinary weak hyper transition in the automaton  $\mathcal{A}'$  that results from  $\mathcal{A}$  by removing all transitions not in  $P$  and labelled different from  $\tau$ . As we consider weakly image-finite MA only, the structure graph of  $\mathcal{A}'$  consists of possibly infinitely many *finite* connected components. Thus,  $\mathcal{A}'$  must be compact, as each of its unconnected components is by Lemma 7.2. With compactness, the proof of Lemma 7.5 carries over immediately to  $\mathcal{A}'$  and from there to  $\mathcal{A}$ .  $\square$

**Corollary D.3.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv\!\!\rightarrow, \dashv\!\!\rightarrow)$  be a weakly image-finite Markov automaton. Let  $\mathcal{R}$  be an equivalence relation over  $S$ . Let  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  be a sequence of distribution over  $S$  converging against the distribution  $\gamma$ . Let  $\langle \gamma'_i \rangle_{i \in \mathbb{N}}$  be a sequence of distribution satisfying

$$\gamma_i \xrightarrow{P, \mathcal{R}} \gamma'_i.$$

Then, there exists a distribution  $\gamma'$  such that

- $\gamma \xrightarrow{P, \mathcal{R}} \gamma'$ ,
- $\langle \gamma'_j \rangle_{j \in J} \longrightarrow \gamma'$  for some subsequence  $\langle \gamma'_j \rangle_{j \in J}$  with  $J \subseteq \mathbb{N}$ .

*Proof.* Recall that we write  $\mu \xrightarrow{P, \mathcal{R}} \mu^*$  to denote a weak combined transition from  $\mu$  to  $\mu^*$  with the following properties:

1.  $\mu \xrightarrow{\tau \upharpoonright P}_c \mu^*$ .
2. whenever  $\mu^* \xrightarrow{\tau \upharpoonright P}_c \mu'$ , then  $\mu^* \mathcal{L}(\mathcal{R}) \mu'$ .

Without the second condition, the lemma would be an immediate instance of Corollary D.2. However, Condition 2 is not obviously covered by the lemma. In Lemma D.17, we have shown that this condition can, however, be interpreted as a predicate over states, according to Corollary D.2. Thus, Corollary D.2 does indeed apply here.  $\square$

The following lemma is the core of our proof.

**Lemma D.18.** Let  $\mu \xrightarrow{\tau \upharpoonright P}_c \gamma$  for two otherwise arbitrary distributions  $\mu$  and  $\gamma$ .

Whenever  $\mu \xrightarrow{a}_c \xi'$  for some distribution  $\xi'$  then there exist distributions  $\nu, \nu'$  and  $\xi$  such that

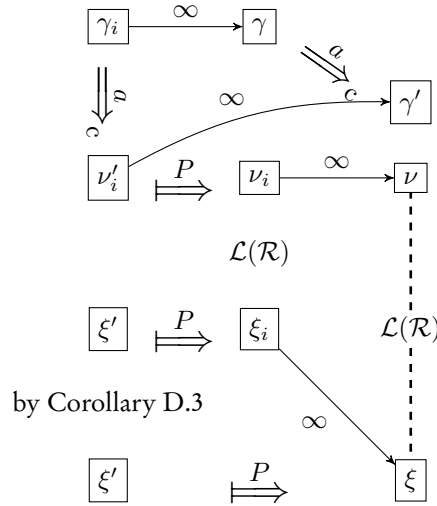
- $\gamma \xrightarrow{a}_c \nu'$ , and
- $\xi' \xrightarrow{P} \xi$  and  $\nu' \xrightarrow{P} \nu$  and  $\xi \mathcal{L}(\mathcal{R}) \nu$ .

*Proof.* We first consider the case where  $\mu \xrightarrow{\tau \mid P}_c \gamma$  is a *non-combined* and *finitary* weak hypertransition. We may thus proceed by induction on the size of transition tree inducing the transition  $\mu \xrightarrow{\tau \mid P}_c \gamma$ .

Let  $\mu \xrightarrow{\tau \mid P}_c \gamma'$  by a smaller transition tree with  $\gamma'$  chosen such that  $\gamma' \xrightarrow{\tau \mid P}_c \gamma$ . Now consider  $\mu \xrightarrow{a}_c \xi'$ . By induction, there exist  $\xi_1, \nu'_1$  and  $\nu_1$  such that  $\xi' \xrightarrow{P} \xi_1$  and  $\gamma' \xrightarrow{a}_c \nu'_1$  and  $\nu'_1 \xrightarrow{P} \nu_1$  and  $\xi \mathcal{L}(\mathcal{R}) \nu_1$ . To finish the induction step, we first remark that  $\gamma' \xrightarrow{a}_c \nu'_1$  and  $\nu'_1 \xrightarrow{P} \nu_1$  it immediately follows that  $\gamma' \xrightarrow{a}_c \nu_1$ . As  $\gamma' \xrightarrow{\tau \mid P}_c \gamma$  and by the definition of preserving transitions (Definition 8.19), it is then clear that there exist  $\nu'_1, \gamma'$  and  $\gamma''$  such that  $\nu_1 \xrightarrow{P} \nu'_1$  and  $\gamma \xrightarrow{a}_c \gamma'$  and  $\gamma' \xrightarrow{\tau} P\gamma''$  and  $\nu_1 \mathcal{L}(\mathcal{R}) \gamma''$ . But by the second clause of the definition of  $\xrightarrow{P}$ , it is clear that  $\nu_1 \mathcal{L}(\mathcal{R}) \nu'_1$ . As  $\mathcal{R}$  is an equivalence relation, this immediately implies  $\nu_1 \mathcal{L}(\mathcal{R}) \gamma''$  and as in turn  $\xi \mathcal{L}(\mathcal{R}) \nu_1$  furthermore  $\xi \mathcal{L}(\mathcal{R}) \gamma''$ .

We have shown the lemma for the case where  $\mu \xrightarrow{\tau \mid P}_c \gamma$  is a *non-combined* and *finitary* weak hypertransition. We now proceed with the argument how to generalize our result to non-combined, but *infinitary* weak hypertransitions. Assume  $\mu \xrightarrow{\tau \mid P}_c \gamma$  by an infinitary weak hypertransition. From the transition tree inducing this transition we obtain a infinite sequence of finitary weak transitions  $\mu \xrightarrow{\tau \mid P}_c \gamma_i$  whose induced distributions  $\gamma_i$  converge against  $\gamma$  by cutting the infinitary transition tree at level  $i$ .

For each of these  $\gamma_i$  we can already apply the lemma with the precondition  $\mu \xrightarrow{\tau \mid P}_c \gamma_i$ . To distinguish between the different instances of applying the lemma, we use the notation  $\eta_i$  where  $\eta \in \{\gamma, \nu', \nu, \xi\}$  to refer to the respective distribution  $\eta$  from the lemma. Note that  $\mu$  and  $\xi'$  are the same in each instance. In the following, we will use Corollary D.3 and Lemma 7.4 in order to establish the limit case, i.e.  $\mu \xrightarrow{\tau \mid P}_c \gamma$ . Figure D.1 illustrates the following lines. We apply



**Figure D.1.:** Graphical Proof Sketch

Lemma 7.5 on the  $\gamma_i$ , the  $\nu'_i$  and  $\gamma$  to derive the existence of a distribution  $\gamma'$  such that  $\gamma \xrightarrow{a}_c \gamma'$ . Then, we apply Corollary D.3 on the  $\nu'_i$ , the  $\nu_i$  and  $\gamma'$  to derive the existence of a distribution

$\nu$  with  $\gamma' \xRightarrow{P} \nu$ , which is in addition the limiting distribution of some subsequence with some index set  $J'$  of the  $\xi_i$ . Then, on  $\xi'$  and the  $\xi_i$ , (however, only applied to the subsequence with index set  $J'$ ) we apply Corollary D.3 to derive the existence of a distribution  $\xi$  with  $\xi' \xRightarrow{P} \xi$ , which in addition is the limiting distribution of the  $\xi_i$ . More correctly, it is the limiting distribution of some subsequence with some index set  $J''$ , which is a subset of  $J'$ . This allows us to apply Lemma 7.4 on  $\nu$ ,  $\xi$ , and the subsequences of the  $\nu_i$  and the  $\xi_i$  with index set  $J''$ , to derive  $\nu \mathcal{L}(\mathcal{R}) \xi$ . This establishes that the lemma also holds for the case that  $\mu \xRightarrow{\tau \downarrow P}_c \gamma$  is induced by an *infinitary* transition tree, but is not a convex weak hypertransition.

We now finally show the most general case, where  $\mu \xRightarrow{\tau \downarrow P}_c \gamma$  may also be a convex weak hyper transitions.

Let  $\gamma = \bigoplus_{i \in I} c_i \gamma_i$  for a family of weights  $c_i \in [0, 1]$  and distributions  $\gamma_i$  satisfying  $\mu \xRightarrow{\tau \downarrow P}_c \gamma_i$  with a *non-combined* weak hypertransition. Obviously, we can apply the lemma for each of the  $\gamma_i$  already by our preceding proof steps.

To distinguish between the different instances of applying the lemma, we use the notation  $\eta_i$  where  $\eta \in \{\gamma, \nu', \nu, \xi\}$  to refer to the respective distribution  $\eta$  from the lemma. Note that  $\mu$  and  $\xi'$  are the same in each instance.

Immediately, by Lemma 4.3, we can conclude that  $\gamma \xRightarrow{a}_c \bigoplus_{i \in I} c_i \nu'_i$ . As  $\xRightarrow{a}_c$  is also composed of combined transitions, we may also conclude by the same lemma that  $\nu' \xRightarrow{P} \bigoplus_{i \in I} c_i \nu_i$  and that  $\xi' \xRightarrow{P} \bigoplus_{i \in I} c_i \xi_i$ . Finally, we only need to establish that  $\bigoplus_{i \in I} c_i \xi_i \mathcal{L}(\mathcal{R}) \bigoplus_{i \in I} c_i \nu_i$ . But this follows easily from Lemma 4.1.  $\square$

**Proposition 6.**  $\mathcal{S} = \{(\mu, \gamma) \mid \exists \mu' \exists \gamma'. \mu \xRightarrow{P} \mu' \wedge \gamma \xRightarrow{P} \gamma' \wedge \mu' \mathcal{L}(\mathcal{R}) \gamma'\}$  is a weak distribution bisimulation.

*Proof.* We will now check the three conditions of weak distribution bisimulation. Let  $(\mu, \gamma)$  be an arbitrary pair in  $\mathcal{S}$ , and let  $\mu'$  and  $\gamma'$  be as above.

1. Assume  $\mu \xrightarrow{a} \xi$  for an arbitrary  $\mu'$ . By Lemma D.18, there exist  $\xi', \xi''$  and  $\xi'''$  such that  $\mu' \xRightarrow{a}_c \xi''$  and  $\xi'' \xRightarrow{P} \xi'$  and  $\xi \xRightarrow{P} \xi'''$  and  $\xi''' \mathcal{L}(\mathcal{R}) \xi'$ . Since  $\mu' \mathcal{L}(\mathcal{R}) \gamma'$ , clearly there exist  $\xi^*, \nu'$  and  $\nu'$  such that  $\xi' \xRightarrow{P} \xi^*$  and  $\gamma' \xRightarrow{a}_c \nu'$  and  $\nu' \xRightarrow{P} \nu$  and  $\xi^* \mathcal{L}(\mathcal{R}) \nu$ . By the second clause of the definition of  $\xRightarrow{P}$  and since  $\mathcal{R}$  is an equivalence relation and thus transitive, it follows that  $\xi' \mathcal{L}(\mathcal{R}) \nu$  and thus with the same argument  $\xi''' \mathcal{L}(\mathcal{R}) \nu$ . Finally, as  $\gamma \xRightarrow{P} \gamma'$  and  $\gamma' \xRightarrow{a}_c \nu'$  and  $\nu' \xRightarrow{P} \nu$  we can infer  $\gamma \xRightarrow{a}_c \nu$ . It only remains to be established that  $\xi \mathcal{L}(\mathcal{S}) \nu$ . This follows by the witnesses  $\xi'''$  and the fact that  $\nu \xRightarrow{P} \nu$  by our choice of  $\nu$ .
2. Let  $\mu = \mu_1 \oplus_p \mu_2$ . It is straightforward to establish that there is a splitting  $\mu' = \mu'_1 \oplus_p \mu'_2$  with  $\mu_i \xRightarrow{P} \mu'_i$  for  $i \in \{1, 2\}$ . By the definition of  $\mathcal{L}(\mathcal{R})$  it follows that then there is also a splitting  $\gamma' = \gamma'_1 \oplus_p \gamma'_2$ . By the definition of  $\mathcal{S}$ , as  $\gamma \xRightarrow{P} \gamma'$ , it follows that also  $\gamma \xRightarrow{a}_c \gamma'$ . These facts, together with the remark that  $\gamma'_i \xRightarrow{P} \gamma'_i$  allow to finally establish that  $\mu_i \mathcal{S} \gamma'_i$  for  $i \in \{1, 2\}$ .
3. Let  $\mu \downarrow$ . Then immediately  $\mu \mathcal{L}(\mathcal{R}) \gamma'$  follows. Then, by the definition of weak state bisimulation, it follows that  $\gamma' \xRightarrow{a}_c \gamma''$  and  $\gamma'' \downarrow$ . Clearly, as  $\gamma \xRightarrow{P} \gamma'$ , also  $\gamma \xRightarrow{a}_c \gamma''$ .

□

**Proof of Theorem 8.14**On finite MA,  $\approx_s = \approx_\delta$ .

This result is an immediate consequence of Proposition 5 and Proposition 6.

**D.8. Proof of Theorem 8.15****Necessary Lemmas****Lemma D.19.** Let  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\quad}, \dashv)$  be a Markov automaton and let  $\mathcal{R} \subseteq Dist(S) \times Dist(S)$  contain all pairs  $(\mu, \gamma)$  satisfying

1. there is a (infinite) sequence  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  with
  - a)  $\langle \mu_i \rangle_{i \in \mathbb{N}} \longrightarrow \mu$ ,
  - b) whenever  $\mu \xrightarrow{a}$  for some  $a \in Act^X$ , then  $\mu_i \xrightarrow{a}$  for all  $i \in \mathbb{N}$ ,
  - c) if  $\mu \downarrow$  then there exists  $\mu'_i$  such that  $\mu_i \Longrightarrow \mu'_i$  and  $\mu'_i \downarrow$  for all  $i \in \mathbb{N}$ ,
2. there is a (infinite) sequence  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  with
  - a)  $\langle \gamma_i \rangle_{i \in \mathbb{N}} \longrightarrow \gamma$ ,
  - b) whenever  $\gamma \xrightarrow{a}$  for some  $a \in Act^X$ , then  $\gamma_i \xrightarrow{a}$  for all  $i \in \mathbb{N}$ ,
  - c) if  $\gamma \downarrow$  then there exists  $\gamma'_i$  such that  $\gamma_i \Longrightarrow \gamma'_i$  and  $\gamma'_i \downarrow$  for all  $i \in \mathbb{N}$ ,
3. for each  $i \in \mathbb{N}$  :  $\mu_i \approx \gamma_i$

Then  $\mathcal{R}$  is a weak distribution bisimulation.*Proof.* We begin our proof with a claim.*Claim 1:* Whenever  $\mu \xrightarrow{a} \mu'$ , then for each  $i \in \mathbb{N}$ , there exists a distribution  $\mu'_i$  such that  $\mu_i \xrightarrow{a} \mu'_i$  and  $d(\mu', \mu'_i) \leq d(\mu, \mu_i)$ .*Proof.* We know by the definition of hyper transitions (Definition 4.14) that

$$\mu' = \bigoplus_{s \in Supp(\mu)} \mu(s) \mu'_s$$

for suitable distributions  $\mu'_s$  satisfying  $s \xrightarrow{a} \mu'_s$ . With the help these distributions we can construct the sought-for distribution  $\mu'_i$  as follows

$$\mu'_i = \bigoplus_{t \in Supp(\mu'_i)} \mu''_t$$

where we let  $\mu''_t = \mu'_t$  if  $t \in Supp(\mu')$ , and otherwise, we let it be an arbitrary distribution  $\rho$  satisfying  $t \xrightarrow{a} \rho$ . That  $\rho$  must exist follows from Assumption 1.b) of this lemma. By construction,  $d(\mu'_i, \mu') \leq d(\mu, \mu_i)$  holds. □

We now continue with the actual proof of this lemma. First of all, obviously,  $\mathcal{R}$  is symmetric. Now, to check Condition a.) of weak distribution bisimulation, assume  $\mu \xrightarrow{a} \mu'$ . Then, by Claim 1, for each  $i \in \mathbb{N}$ , there exists a distribution  $\mu'_i$  such that  $\mu_i \xrightarrow{a} \mu'_i$  and  $d(\mu', \mu'_i) \leq d(\mu, \mu_i)$ . Hence, the sequence  $\langle \mu'_i \rangle_{i \in \mathbb{N}} \rightarrow \mu'$ . As  $\mu_i \approx \gamma_i$ , there also exists a distribution  $\gamma'_i$  for each  $i \in \mathbb{N}$ , such that  $\gamma_i \xrightarrow{a} \gamma'_i$  and  $\mu'_i \approx \gamma'_i$ . By Lemma 7.5, there then exists a distribution  $\gamma'$  such that  $\gamma \xrightarrow{a} \gamma'$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}} \rightarrow \gamma'$ . As both  $\langle \mu'_i \rangle_{i \in \mathbb{N}} \rightarrow \mu'$  and  $\langle \gamma'_i \rangle_{i \in \mathbb{N}} \rightarrow \gamma'$ , the pair  $(\mu', \gamma')$  is contained in  $\mathcal{R}$ . This ends the proof of the first condition.

For Condition b.), let us assume that  $\mu = \mu^0 \oplus_p \mu^1$  for some arbitrary  $p \in [0, 1]$  and suitable distributions  $\mu^0, \mu^1$ . We now construct a splitting of  $\mu_i = \mu^0_i \oplus_p \mu^1_i$  such that  $\langle \mu^k_i \rangle_{i \in \mathbb{N}} \rightarrow \mu^k$  for  $k \in \{0, 1\}$ . Let for  $k \in \{0, 1\}$

$$\mu_i^k(s) = \begin{cases} \frac{\mu^k(s)}{\mu^k(s) + \mu^{1-k}(s)} \mu_i(s) & \text{if } s \in \text{Supp}(\mu) \\ \frac{1}{2} \mu_i(s) & \text{else} \end{cases}$$

Given this splitting for each  $\mu_i$ , by  $\mu_i \approx \gamma_i$ , it follows that  $\gamma_i \xrightarrow{a} \gamma'_i = \gamma_i^0 \oplus \gamma_i^1$  and  $\mu_i^k \approx \gamma_i^k$  for  $k \in \{0, 1\}$ . By Lemma 7.5, it is clear that  $\gamma \xrightarrow{a} \gamma'$  where  $\gamma'$  is the limiting distribution of a subsequence  $\langle \gamma'_i \rangle_{i \in J}$  with a suitable index set  $J \subseteq \mathbb{N}$ . It now remains to be shown that for  $k \in \{0, 1\}$

$$\langle \gamma_j^k \rangle_{j \in J} \rightarrow \gamma^k \text{ for some distribution } \gamma^k$$

and that

$$\gamma' = \gamma^0 \oplus \gamma^1.$$

This is not completely obvious, as even though the sequence converges against the distribution  $\gamma'$ , it could be possible that the splitting always happens in a non-converging way. Actually, we cannot show that the sequence converges against one single splitting. We can, however, show that there must exist a subset of indexes  $J' \subseteq J$  such that the corresponding subsequence converges for  $\langle \gamma_j^0 \rangle_{j \in J'}$  and  $\langle \gamma_j^1 \rangle_{j \in J'}$ . More precisely, we will show that there exists an infinite index set  $J' \subseteq J$  such that for every  $\varepsilon \in \mathbb{R}_{>0}$  there is a number  $N \in \mathbb{N}$  such that for every  $n, m \in J'$  with  $n, m > N$

**Condition 1**  $|\gamma_n^k(s) - \gamma_m^k(s)| \leq \varepsilon$  for every  $s \in \text{Supp}(\gamma')$  and  $k \in \{0, 1\}$ , and

**Condition 2**  $\langle \gamma_j^0(s) + \gamma_j^1(s) \rangle_{j \in J'} \rightarrow \gamma'(s)$  for every  $s \in \text{Supp}(\gamma')$ .

The first statement means that  $\langle \gamma_j^k(s) \rangle_{j \in J'}$  for both  $k = 0$  and  $k = 1$  is a Cauchy sequence over the set  $\mathbb{R}$  for every  $s \in \text{Supp}(\gamma')$ . Thus, the sequence must have a limit in  $\mathbb{R}$ . We now define  $\gamma^k(s)$  as the respective limiting distribution for each  $s \in \text{Supp}(\gamma')$  and  $k \in \{0, 1\}$ . By the second statement, it follows that  $\gamma' = \gamma^0 \oplus_p \gamma^1$ , and everything is shown.

It remains to prove the existence of the index set  $J'$ . We construct the sequence as the limit of a sequence of index set. Let  $\{s_1, s_2, \dots\} = \text{Supp}(\gamma')$  and without loss of generality let  $\gamma'(s_i) \leq \gamma'(s_{i-1})$ . Let  $J_0 = J$ . We derive  $J_l$  from  $J_{l-1}$  as follows. The sequence  $\langle \gamma_j^0(s_l) \rangle_{j \in J_{l-1}}$  is a bounded sequence of real numbers. Clearly, there must be a subsequence  $\bar{J}_{l-1}$  of the sequence  $J_{l-1}$  such that  $r_l \in \mathbb{R}$  exists with

$$\langle \gamma_j^0(s_l) \rangle_{j \in \bar{J}_{l-1}} \rightarrow r_l.$$

Without loss of generality, we may demand that  $\bar{J}_{l-1}$  agrees with  $J_{l-1}$  for the first  $l$  elements  $(\star)$ . This is possible, as any finite prefix of a sequence is irrelevant for the possible accumulation points of the sequence. We choose  $J_l := \bar{J}_{l-1}$ .



We now define

$$J' = \bigcap_{s_l \in \text{Supp}(\gamma')} J_l.$$

By  $(\star)$ , it is ensured that  $J'$  is an infinite set. Furthermore,  $J' \subseteq J_l$ . Hence, it guarantees that

$$\langle \gamma_j^0(s_l) \rangle_{j \in J'} \longrightarrow r_l.$$

We now check that  $J'$  satisfies the two conditions we have required. Condition 2 is obvious, as already

$$\langle \gamma_j^0 + \gamma_j^1 \rangle_{j \in J_0} \longrightarrow \gamma',$$

and therefore clearly also

$$\langle \gamma_j^0 + \gamma_j^1 \rangle_{j \in J'} \longrightarrow \gamma'$$

as  $J_0 \supseteq J'$ . Trivially, this implies  $\langle \gamma_j^0 + \gamma_j^1 \rangle_{j \in J'} \longrightarrow \gamma'$  for every  $s \in \text{Supp}(\gamma')$ . For Condition 2, let  $\varepsilon \in \mathbb{R}_{>0}$  be given. There exists a number  $N_0 \in J_0$ , such that for every  $n > N_0$

$$|\gamma'(s) - \gamma'_n(s)| \leq \varepsilon_1$$

for every  $s \in \text{Supp}(\gamma')$  for arbitrarily small  $\varepsilon_1 \in \mathbb{R}_{>0}$ . Furthermore, for every  $s_l \in \text{Supp}(\gamma')$  there exists a number  $N_l \in J_l$  such that for every  $n > N_l$

$$|\gamma_n^0(s_l) - \gamma_m^0(s_l)| \leq \varepsilon_2 \tag{D.23}$$

for arbitrarily small  $\varepsilon_2 \in \mathbb{R}_{>0}$ . As  $\gamma'_n = \gamma_n^0 \oplus \gamma_n^1$ , also

$$|\gamma_n^1(s_l) - \gamma_m^1(s_l)| \leq \varepsilon_1 + \varepsilon_2. \tag{D.24}$$

For any arbitrarily small  $\varepsilon_3$ , there exists a number  $o$  such that

$$\gamma'(\{\text{Supp}(\gamma') \setminus \{s_1, s_2, \dots, s_o\}\}) \leq \varepsilon_3. \tag{D.25}$$

As our last step, we now show that in order to satisfy Condition 1, we can choose  $N$  to be any  $N \in J'$  with  $N \geq \max N_0, N_1, \dots, N_o$ , and  $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \frac{1}{2}\varepsilon$ . We check the condition now for every  $s \in \text{Supp}(\gamma')$  and we distinguish two cases.

**Case 1:**  $s = s_i$  for  $i \in \{1, \dots, o\}$  As  $N \geq N_i$ , by Equation (D.23) and (D.24) everything follows.

**Case 2:**  $s = s_i$  for  $i > o$  From Equation (D.25), we can infer that  $\gamma'(s) \leq \varepsilon$ . Thus, also  $\gamma_n^k(s) \leq \varepsilon_1 + \varepsilon_3 = \varepsilon$  for every  $n > N$ . Hence,  $|\gamma_n^k(s)\gamma_m^k(s)| \leq \varepsilon$  for  $n, m > N$ , as in the worst case, in one distribution the probability of  $s$  is  $\varepsilon$  and in the other 0.

For Condition c.), let us assume  $\mu \downarrow$ . Then, for all  $\mu_i$  there exists  $\mu'_i$  with  $\mu'_i \downarrow$  and  $\mu_i \Longrightarrow_c \mu'_i$  by the assumptions of the Lemma. As  $\mu_i \approx \gamma_i$  also  $\gamma_i \Longrightarrow_c \gamma'_i$  with  $\gamma'_i \downarrow$ . Immediately, by Lemma 7.5, everything follows.  $\square$

**Corollary D.4.** Let  $\mathcal{A} = (S, \bar{s}, \text{Act}, \dashv, \dashv\dashv)$  be a Markov automaton  $\gamma$  be a distribution over  $S$ . Let  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  be a sequence of distributions over  $S$  with  $\langle \mu_i \rangle_{i \in \mathbb{N}} \longrightarrow \mu$ , satisfying in addition for every  $i \in \mathbb{N}$

$$\mu_i \approx \gamma_i.$$

Then  $\mu \approx \gamma$ .

*Proof.* We use Lemma D.19 to prove this corollary. Clearly, the sequence  $\langle \gamma_i \rangle_{i \in \mathbb{N}}$  with  $\gamma_i = \gamma$  for every  $i \in \mathbb{N}$  trivially satisfies all necessary preconditions. It remains to check the preconditions for the sequence  $\langle \mu_i \rangle_{i \in \mathbb{N}}$ . In detail, it remains to check that

1. whenever  $\mu \xrightarrow{a}$  for some  $a \in \text{Act}^X$ , then  $\mu_i \xrightarrow{a}$  for all  $i \in \mathbb{N}$ , and
2. if  $\mu \downarrow$  then there exists  $\mu'_i$  such that  $\mu_i \Rightarrow \mu'_i$  and  $\mu'_i \downarrow$  for all  $i \in \mathbb{N}$ .

For the first part, we will establish that  $\gamma \xrightarrow{a}$  if  $\mu \xrightarrow{a}$ . Then, it immediately follows from  $\mu_i \approx \gamma$  that also  $\mu_i \xrightarrow{a}$  for every  $i \in \mathbb{N}$ . Assume the contrary, i.e.  $\mu \xrightarrow{a}$  but *not*  $\gamma \xrightarrow{a}$ . Let  $\overline{\gamma^a} \oplus_p \gamma^a = \gamma$  be a splitting of  $\gamma$  with  $\gamma^a \xrightarrow{a}$  and  $\overline{\gamma^a} \not\xrightarrow{a}$ . We can now show that this statement cannot hold for any fixed  $p > 0$ . To see this, consider an arbitrary fixed  $p > 0$ . Choose  $i$  large enough such that  $d(\mu_i, \mu) < \frac{1}{2}p$ . From this we may conclude the existence of states  $X \subseteq \text{Supp}(\mu_i) \cap \text{Supp}(\mu)$  with  $\mu_i(X) \geq 1 - \frac{1}{2}p$ . From  $\mu \xrightarrow{a}$  it follows that each  $s \in X$  can perform an  $a$ -transition. Intuitively speaking, the probability for a transition, that cannot perform an  $a$ -transition is less than  $\frac{1}{2}p$ . Thus, whenever  $\xi \approx \mu_i$ , then there *cannot* be a splitting  $\xi = \overline{\xi^a} \oplus_p \xi^a$  with  $\overline{\xi^a} \not\xrightarrow{a}$  ( $\star$ ).

On the other hand, as  $\mu_i \approx \gamma$ , for  $\mu_i$  there must exist a splitting  $\mu_i \Rightarrow \overline{\mu_i^a} \oplus_p \mu_i^a$  with  $\mu_i^a \xrightarrow{a}$  and  $\overline{\mu_i^a} \not\xrightarrow{a}$ . From Lemma 8.6, we know that  $\overline{\mu_i^a} \oplus_p \mu_i^a \approx \mu_i$ . But this directly contradicts ( $\star$ ). This ends the proof of the first statement. The proof of the second statement is completely analogue.  $\square$

**Lemma D.20.** Let  $s \Rightarrow \mu$  and  $\delta(s) \not\approx \mu$ . Then for every  $0 \leq k < 1$

$$s \not\approx \delta(s) \oplus_k \mu.$$

*Proof.* Assume that on the contrary for a fixed  $k$  in fact  $s \approx \delta(s) \oplus_k \mu$  holds. Since  $s \not\approx \mu$ , but  $s \Rightarrow \mu$ , there either must be a transition  $s \xrightarrow{a} \gamma$  such that whenever  $\mu \xrightarrow{a} \gamma'$ , then  $\gamma' \not\approx \gamma$  or if  $a \neq \tau$ , then possibly directly  $\mu \not\xrightarrow{a}$ . We will refer to this fact by ( $\star$ ). The second case immediately implies  $\delta(s) \oplus_k \mu \not\xrightarrow{a}$ , which is a contradiction to our assumption. We now consider the first case that whenever  $\mu \xrightarrow{a} \gamma'$  then  $\gamma' \not\approx \gamma$ . By our assumption that  $s \approx \delta(s) \oplus_k \mu$ , there must, however, exist some  $\xi$  and some  $\gamma'$  such that  $s \xrightarrow{a} \xi$  and  $\mu \xrightarrow{a} \gamma'$  and  $\xi \oplus_k \gamma' \approx \gamma$  holds. We will refer to this fact by ( $\star_1$ ).

Again, it cannot be the case that  $\mu$  can simulate the transition  $s \xrightarrow{a} \xi$ . To see this, assume the contrary. Then, for some  $\xi' : \mu \xrightarrow{a} \xi'$  and  $\xi \approx \xi'$ . But then also  $\mu \xrightarrow{a} (\xi') \oplus_k \gamma' \approx \xi \oplus_k \gamma' \approx \gamma$ . A contradiction to ( $\star$ )!

Nevertheless, by our assumption that  $s \approx \delta(s) \oplus_k \mu$ , we must be able to repeat this argument as above, such that there must exist a  $\pi$  such that  $\mu \xrightarrow{a} \pi$  and a  $\nu$  such that  $s \xrightarrow{a} \nu$  and  $\nu \oplus_k \pi \approx \xi$ . We will refer to this fact by ( $\star_2$ ).

With the help of these results, it can be shown that not only  $\delta(s) \oplus_k \mu$  can simulate the challenging transition  $s \xrightarrow{a} \gamma$ , but also  $\delta(s) \oplus_{k^2} \mu$ . (We will show this step in the next paragraph. Before, we finish the train of thoughts for this proof.) Then, we can repeat the proofs argumentation so far for  $\delta(s) \oplus_{k^2} \mu$ , to see that everything also must hold for  $\delta(s) \oplus_{k^{2^2}} \mu$ . Repeating this argument over and over again, we see that it even must hold for  $\delta(s) \oplus_{k^{2^i}} \mu$  for every  $i \in \mathbb{N}$ .

As in our proof, we made no specific assumptions on the concrete value of  $a$  or the challenging transition itself, this means that every transition of  $s$  can be mimicked by  $\delta(s) \oplus_{k^{2^i}} \mu$  for every

$i \in \mathbb{N}$ . At the same time,  $s \implies \mu$  by the premise of the lemma, which implies that  $s$  can simulate every behaviour of  $\mu$ , we can easily establish that indeed  $\delta(s) \approx s \oplus_{k^{2^i}} \mu$  must hold. It is well-known that the countable infinite sequence  $\langle k_i \rangle_{i \in \mathbb{N}}$  with  $k_i = k^{2^i}$  has limit 0. Hence, this implies immediately that  $\langle s \oplus_{k_i} \mu \rangle_{i \in \mathbb{N}} \longrightarrow \mu$ . This is enough to use Corollary D.4 to establish  $s \approx \mu$ . A contradiction!

We now prove the missing step we deferred above. We state it again for convenience. If  $\delta(s) \oplus_k \mu$  can simulate the challenging transition  $s \xrightarrow{a} \gamma$ , then also  $\delta(s) \oplus_{k^2} \mu$  can simulate it. For a better readability, in the following paragraph, we use the notation  $k\mu \oplus (1-k)\gamma$  instead of  $\mu \oplus_k \gamma$  when convenient. With straightforward transformations we obtain

$$\begin{aligned}
 k^2\delta(s) \oplus (1-k^2)\mu &= k^2\delta(s) \oplus (1-k)(k+1)\mu \\
 &= k^2\delta(s) \oplus k(1-k)\mu \oplus (1-k)\mu \\
 &= k(k\delta(s) \oplus (1-k)\mu) \oplus (1-k)\mu \\
 &= (\delta(s) \oplus_k \mu) \oplus_k \mu.
 \end{aligned}$$

From  $(\star_2)$ , we know that

$$\delta(s) \oplus_k \mu \xRightarrow{a}_c \nu \oplus_k \pi$$

and from  $(\star_1)$ , we know that  $\mu \xRightarrow{a} \gamma'$ . Hence,

$$(\delta(s) \oplus_k \mu) \oplus_k \mu \xRightarrow{a}_c (\nu \oplus_k \pi) \oplus_k \gamma'.$$

But from  $(\star_1)$  and  $(\star_2)$ , we also know that  $\nu \oplus_k \pi \approx \xi$  and  $\xi \oplus_k \gamma' \approx \gamma$ . In summary, we obtain

$$k^2\delta(s) \oplus (1-k^2)\mu \xRightarrow{a}_c (\nu \oplus_k \pi) \oplus_k \gamma' \approx \gamma.$$

□

**Corollary D.5.** *Let  $\mu_1 \approx \mu_2 \approx \delta(s)$  and  $\mu \not\approx \delta(s)$ , but  $s \xrightarrow{\tau} \mu$ . Then  $\forall 0 \leq k < 1 : \mu_1 \not\approx k\mu_2 \oplus (1-k)\mu$ .*

*Proof.* Assume the contrary. Then for some  $0 \leq k < 1 : \delta(s) \approx \mu_1 \approx k\mu_2 \oplus (1-k)\mu \approx k\delta(s) \oplus (1-k)\mu$ . A contradiction to Lemma D.20 □

## Proof of Theorem 8.15

Let  $\mathcal{A} = (S, \bar{s}, Act, \xrightarrow{\cdot}, \dashv\!\!\!\rightarrow)$  be a Markov automaton, and  $\mu, \gamma \in Dist(S)$ .

If  $\mu \approx \gamma$  then there exist  $\mu^*$  and  $\gamma^*$  such that  $\mu \xRightarrow{\cdot}_c \mu^*, \gamma \xRightarrow{\cdot}_c \gamma^*$ ,

$$\mu \approx \mu^* \mathcal{L}(\approx_\delta) \gamma^* \approx \gamma.$$

*Proof.* We first define a set of internal transition trees that we will prove to be suited to reach our sought-for distributions  $\mu'$  and  $\gamma'$ . The construction is identical for each element  $s$  in  $Supp(\mu) \cup Supp(\gamma)$ . Readers familiar with the notion of schedulers, can regard the following construction as a variant of deterministic schedulers.

Let  $\mathcal{T}$  be an internal transition tree with  $\text{Sta}_{\mathcal{T}}(\varepsilon) = s$  that is stepwise constructed as follows. Consider a leaf  $\sigma$ . Let  $\text{Sta}_{\mathcal{T}}(\sigma) = s$ . If  $s \xrightarrow{\tau} \mu$  and  $\mu \approx \delta(s)$  and  $\exists u \in \text{Supp}(\mu) : u \not\approx_{\delta} s$ , then continue  $\mathcal{T}$  with this transition. If  $s$  allows several different internal transitions, we choose for every node of  $\mathcal{T}$ , that is labelled by  $s$ , the same internal transition. In all other cases, we do not continue the tree here.

If a state is continued, following the construction above, we call it a *running state*. Otherwise, a *stopping state*.

For simplicity, we will now speak of state  $s$  in the tree  $\mathcal{T}$ , when formally we mean a node  $\sigma$  with  $\text{Sta}_{\mathcal{T}}(\sigma) = s$ . Since we continue every such node  $\sigma$  with the same internal transition, this is well justified.

In general, this transition tree has an infinite depth, and thus, it is not fully clear whether it indeed induces a transition. We will thus not directly work with this tree, but use it as a blueprint for a sequence of finite trees. This sequence is constructed by cutting the infinite tree at level  $i$  for each  $i \in \mathbb{N}$ . This sequence of trees induces a sequence of distributions  $\langle \mu_i \rangle_{i \in \mathbb{N}}$  with  $\mu \Longrightarrow \mu_i$  for each  $i \in \mathbb{N}$ , and accordingly for  $\gamma$ . In the following, we restrict our considerations to  $\mu$ , as the case for  $\gamma$  is completely analogue.

As we consider compact Markov automata, this sequence of distributions must have a subsequence that converges against a distribution  $\bar{\mu}$ , and furthermore by Corollary 7.1, also  $\mu \Longrightarrow \bar{\mu}$ . We choose  $\mu^* := \bar{\mu}$ .

We now show that every  $s \in \text{Supp}(\mu^*)$  is stopping. Assume the contrary and let  $\mu^* = \mu_1 \oplus_p \mu_2$  be a splitting of  $\mu^*$  where every state in the support of  $\mu_1$  is running, and every state in  $\mu_2$  is stopping. Whenever we talk about (weak) transitions in the following, we refer to the transitions possible in this automaton. Note that this automaton is completely deterministic. First, note that no state in  $\text{Supp}(\mu_1)$  can reach any state of  $\mu_2$ , as otherwise the probability with what they would be reached, can never reach  $\mu_1$  again, and thus,  $\mu^*$  could not be a valid limiting distribution. As  $\mu_1$  consists of running states, there must exist a *non-trivial* weak transition  $\mu_1 \xrightarrow{\tau} \circ \Longrightarrow \mu_1$ . If this was not possible, then  $\mu_1$  could not be a limiting distribution.

We now proceed with two final steps.

1. Show that there exists a state  $t \in \text{Supp}(\mu_1)$  such that  $t \xrightarrow{\tau} \circ \Longrightarrow \delta(t)$ .
2. Show that this implies that the transition that makes  $t$  running actually contradicts our definition for running transitions.

The second statement implies that  $t$  should be stopping and thus contradicts our assumption that  $t \in \text{Supp}(\mu_1)$ , as  $\mu_1$  consists only of running states. Thus, our assumption that we can split  $\mu^*$  into a running and stopping part must be wrong. Thus, all states in  $\text{Supp}(\mu^*)$  must be stopping, and we have proven our claim. We now prove the two statements

1. Recall that  $\mu_1 \xrightarrow{\tau} \circ \Longrightarrow \mu_1$ . We can thus, for every  $i \in \mathbb{N}$ , construct the weak hyper-transition

$$\mu_1 \Longrightarrow \underbrace{\mu_1 \Longrightarrow \cdots \Longrightarrow}_{i \text{ times}} \mu_1,$$

where we assume that each of the  $i$  weak hypertransitions are induced by the same transition tree. We call in the following  $i$  the size of the weak hypertransition. Let  $t$  be an arbitrary state in the support of  $\mu_1$  and let  $\xi'_i$  be the distribution that  $t$  contributes to the weak hyper-transition of size  $i \in \mathbb{N}$ , i.e. there exists a distribution  $\bar{\xi}'_i$  such that  $\mu_1 = \xi'_i \oplus_{\mu_1(F)} \bar{\xi}'_i$

and  $t \Rightarrow \xi'_i$ . Let  $\xi_i$  be a distribution that is obtained from a transition tree resembling the one that induces the transition to  $\xi'_i$ , but which is cut at every node labelled with  $t$ . Let  $\xi$  be a distribution, such that

$$\langle \xi_i \rangle_{i \in J} \longrightarrow \xi \text{ for some infinite } J \subseteq \mathbb{N}.$$

$\xi$  must exist, as  $\mathcal{A}$  is compact by Lemma 7.3. Let now  $\delta(F) \oplus_p \xi' = \xi$  be a splitting of  $\xi$  for  $p \in [0, 1]$ , such that  $\delta(t) \notin \text{Supp}(\xi')$ . The states in  $\text{Supp}(\xi')$  cannot reach  $t$  with any probability greater 0, i.e. for every  $u \in \xi'$  there is *no* weak transition  $u \Rightarrow \xi''$  with  $t \in \text{Supp}(\xi'')$ . This holds because otherwise,  $\xi$  would not be a valid limiting distribution of our sequence. This, however, means that  $t$  the larger the size  $i$  of the weak hypertransition

$$\mu_1 \Rightarrow \underbrace{\mu_1 \Rightarrow \cdots \Rightarrow}_{i \text{ times}} \mu_1,$$

the less the probability  $\mu'_i(t)$  must be. Therefore,  $\mu_1 \ominus \delta(t) \ominus \xi'$  must be able to compensate for the missing probability mass, i.e. it can reach  $t$  with probability at least  $(1 - p)\mu_1 t$ , as this is exactly the probability mass that  $t$  loses. However, this means that on the long run, also  $(1 - p)$  percent of this probability will be lost in  $\xi'$ . A contradiction.

2. Assume  $t \xrightarrow{\tau} \circ \Rightarrow \delta(t)$ . Then either  $t \xrightarrow{\tau} \delta(F)$  or there exists another distribution  $\xi$  with  $t \xrightarrow{\tau} \xi \Rightarrow \delta(F)$ . The first case is an immediate violation of our definition of running states, as  $t \approx_\delta F$ . We will now show that every state  $u \in \text{Supp}(\xi)$  can reach  $t$  with probability 1 and vice versa. Formally,

$$t \Rightarrow \delta(G) \text{ and } u \Rightarrow \delta(F).$$

If this is the case, then it is straightforward to see that  $t \approx_\delta G$ , again contradicting our assumption that the transition  $t \xrightarrow{\tau} \xi$  qualifies  $t$  as a running state. For the first direction, we consider a weak transition that is induced by a transition tree that resembles the one from the transition  $t \xrightarrow{\tau} \xi \Rightarrow \delta(t)$ , but cut it at the nodes labelled by  $u$ . Then this distribution, call it  $\xi'$  satisfies  $\xi'(u) \geq \xi(u)$  and  $\text{Supp}(\xi') = \{G, F\}$ . Then from  $\xi$ , we can continue repeat this construction. Obviously, the limiting distribution of this sequence of distributions is  $\delta(t)$ , which establishes the first direction. The other direction immediately follows from the fact that  $u \in \text{Supp}(\xi)$  and  $\xi \Rightarrow \delta(t)$ .

This ends the first part of our proof.

Let now  $\mu \approx \gamma$  and let  $\mu \Rightarrow \mu^*$  be the transition obtained by repeating the above construction for each  $s \in \mu$ . And analogously for  $\gamma \Rightarrow \gamma^*$ . Clearly,  $\mu \approx \mu^*$  and  $\gamma \approx \gamma^*$ .

We will now show that whenever  $\mu^* \Rightarrow_c \mu'$ , then

$$\text{either } \mu^* \not\approx \mu', \text{ or } \mu^* \mathcal{L}(\approx_\delta) \mu'.$$

And analogously for  $\gamma^*$ . We skip the analogous proof. Assume the contrary, i.e.  $\mu' \approx \mu^*$ , but *not*  $\mu' \mathcal{L}(\approx_\delta) \mu^*$ .

The weak combined hypertransition  $\mu^* \Rightarrow_c \mu'$  is justified by a family of non-combined weak hypertransitions and their respective weights (Definition 4.15). For each hypertransition, there must exist one transition for each  $s \in \text{Supp}(\mu^*)$  that together justify the hyper-transition by

Definition 4.14. If, as we assumed,  $(\mu' \mathcal{L}(\approx_\delta) \mu^*)$  does *not* hold, then one of these transitions – denote it by  $s \Rightarrow \mu''$  – must be induced by a transition tree  $\mathcal{T}$  that contains a node  $\sigma$  with the following properties:

- it is labelled with a state  $s'$  ( $\text{Sta}_{\mathcal{T}}(\sigma) = s'$ ), with  $s' \approx_\delta s$ ,
- $s' \xrightarrow{\tau} \mu_b$  where  $\mu_b$  is determined by the children of  $\sigma$  in  $\mathcal{T}$ , and
- $\mu_b \not\approx \delta(s')$ .

Let  $0 < c \in [0, 1]$  be the weight with what this transition contributes to the weak combined transition  $\mu^* \Rightarrow_c \mu'$ .

Without loss of generality, we assume that no such transition has occurred in the tree on a smaller tree node level. Let  $\mathcal{T}'$  be the subtree of  $\mathcal{T}$  up to and including the level of  $\sigma$  plus the children of  $\sigma$ , which represent the distribution  $\mu_b$ . Then the distribution induced by this tree  $\mathcal{T}'$  can be characterized as  $\xi_b = \mu_b \oplus_{\text{Prob}_{\mathcal{T}'(\sigma)}} \xi'$ , where  $\xi'$  is a distribution only consisting of states that are equivalent to  $s$  with respect to  $\approx_\delta$ . Then by Corollary D.5,  $\xi_b = \mu_b \oplus_{\text{Prob}_{\mathcal{T}'(\sigma)}} \xi' \not\approx \delta(s)$ .

We can rewrite  $\mu^*$  as

$$\mu^* = (\delta(s)) \oplus_{\mu^*(s)} (\mu^* \ominus s)$$

We can clearly further rewrite  $\mu^*$  to

$$(\delta(s) \oplus_c \delta(s)) \oplus_{\mu^*(s)} (\mu^* \ominus s)$$

for any  $c \in [0, 1]$ . Recall that by  $c$  we denoted the weight that the transition  $s \Rightarrow \mu''$  contributes to the weak combined transition  $\mu^* \Rightarrow_c \mu'$ . Then, we can construct a weak combined transition of  $\mu^*$  where we let all states stand idle, while only  $s$  performs a transition to  $\xi_b$ , with fraction  $c$ . In summary,

$$\begin{aligned} \mu^* \Rightarrow_c (\xi_b) & \quad \oplus_c \delta(s) \oplus_{\mu^*(s)} (\mu^* \ominus s) \\ & = (\mu_b \oplus_{\text{Prob}_{\mathcal{T}'(\sigma)}} \xi') \quad \oplus_c \delta(s) \oplus_{\mu^*(s)} (\mu^* \ominus s). \end{aligned}$$

As  $\xi \approx \delta(s)$  and  $\delta(s') \approx \delta(s)$ , by Corollary 8.2 we get

$$\mu^* \approx \underbrace{(\delta(s') \oplus_{\text{Prob}_{\mathcal{T}'(\sigma)}} \xi') \oplus_c \delta(s) \oplus_{\mu^*(s)} (\mu^* \ominus s)}_{:=\rho}.$$

By Lemma D.14, we furthermore derive

$$\rho \not\approx \underbrace{(\mu_b \oplus_{\text{Prob}_{\mathcal{T}'(\sigma)}} \xi') \oplus_c \delta(s) \oplus_{\mu^*(s)} (\mu^* \ominus s)}_{:=\rho_b}$$

By our choice of  $\rho_b$ , we know that  $\mu^* \Rightarrow_c \rho_b$  and also  $\rho_b \Rightarrow_c \mu'$ . But as now  $\mu' \approx \mu^*$ , this immediately implies by Lemma 8.2 that also  $\rho_b \approx \mu^*$ . A contradiction.

It is now easy to see that in indeed  $\mu^* \mathcal{L}(\approx_\delta) \gamma^*$  must hold. By Lemma 8.7,  $\gamma^* \Rightarrow_c \gamma'$  for some  $\gamma'$  that can be split in such a way that for each  $s \in \text{Supp}(\mu^*)$  there exists a component of the splitting, call it  $\gamma'_s$ , such that  $\delta(s) \approx \gamma'_s$ . But in fact, whenever  $\gamma^* \Rightarrow_c \gamma'$ , as detailed out above,  $\gamma^* \mathcal{L}(\approx_\delta) \gamma'$  or  $\gamma^* \not\approx \gamma'$ , which, however, would be a contradiction to the lemma. So it must

hold that already  $\gamma^*$  can be split according to the lemma. Name the components of the splitting  $\gamma_s^*$ , analogously to  $\gamma'_s$  above.

If now  $\gamma_s^*$  contains states  $t$  with  $t \not\approx_\delta s$ , then with an argument symmetric to the last, it must be possible to directly split  $s$  suitably. Since splitting  $s$  only yields  $s$  again all states in  $\text{Supp}(\gamma_s^*)$  are already equivalent to  $s$  and hence also equivalent to each other. A contradiction to our assumption.  $\square$





## E. Proof of Lemma 9.7

Throughout the chapter, we write  $E \equiv F$ , if  $E = F$  only by the axioms (COM) and (ASS). Recall that a process is in *normal form*, if it is weak saturated, weak tangible, convex reduced,  $\Delta$ -expanded, and in sum form. We will show that the following statements hold:

1. Every process can be rewritten into normal form with  $\mathcal{A}_{\text{MA}}$ , and
2. For two processes  $E$  and  $F$  in normal form,  $E \simeq_{\text{MA}} F$  implies  $E \equiv F$ .

*Proof.* We strengthen the statement as follows:

1. Every process can be rewritten into normal form with  $\mathcal{A}_{\text{MA}}$ , and
2. For two processes  $E$  and  $F$  in normal form,
  - a)  $E \simeq_{\text{MA}} F$  implies  $E \equiv F$ , and
  - b)  $E \approx F$  implies  $E \equiv F$ ,  $\tau.\delta(E) \equiv F$  or  $E \equiv \tau.\delta(F)$ .

By Lemma 9.2, we may assume that  $E$  and  $F$  are already in sum form and  $\Delta$ -expanded. The proof is by induction on the maximum of  $\mathcal{D}(E)$  and  $\mathcal{D}(F)$ . We call this value  $n$  in the following. If  $n = 0$ , both  $E$  and  $F$  must be the terminated process 0, or  $\Delta(0)$ . By the maximal progress condition, we derive that  $E \simeq_{\text{MA}} F$  implies that it cannot be the case that one of them is 0 while the other is  $\Delta(0)$ . Furthermore, by definition, 0 and  $\Delta(0)$  are already expressions in sum form and  $E \equiv F$  must hold.

Let  $n = 1$  and let without loss of generality  $\mathcal{D}(F) \leq \mathcal{D}(E)$ . Let  $E$  be of the form  $\Delta(G)$ . In fact, whenever  $E$  is in this form, then  $n = 1$  must hold. To see this, consider that since  $E$  is  $\Delta$ -expanded and in sum form,  $G$  cannot contain further  $\Delta$  expressions nor any prefix expressions, as the first is forbidden by sum form, and the later violates our assumption that  $E$  is  $\Delta$ -expanded, since then  $E$  would consist of multiple summands. So  $G$  must be 0.  $E$  is thus in normal form, which proves Statement 1. Since  $E \simeq_{\text{MA}} F$  and  $\mathcal{D}(F) \leq \mathcal{D}(E)$ ,  $F$  must be  $\Delta(0)$  as well, and we immediately obtain  $E \equiv F$ . This proves Statement 2.a.). For Statement 2.b.) very similar arguments can be applied.

Let now  $n > 2$  and let without loss of generality  $\mathcal{D}(F) \leq \mathcal{D}(E) = n$ . As we have shown above, we do not need to consider the case that  $E$  is a  $\Delta$  expression in the following.

### Proof step for Statement 1)

By Lemma 9.2, 9.4, and 9.5, we may assume that  $E$  is saturated, convex reduced,  $\Delta$ -expanded and in sum form. It thus remains to show that  $E$  can, in addition, be made weak tangible, while maintaining the (weak) saturation, convex reduction,  $\Delta$ -expanded and sum form properties.

By induction and the definition of  $\mathcal{D}$ , we can assume that all statements already hold for the successor processes of  $E$ .

Assume  $E$  is not weak tangible already. Then  $E$  cannot be tangible, and it must have at least two summands  $A$  and  $B$  with the following possible cases:

1.  $A = \Delta(G)$ ,  $E \approx A$  and  $B = \Delta(H)$
2.  $A = \Delta(G)$ ,  $E \approx A$  and  $B = a. \bigoplus_{j \in J} q_j F_j$  with  $a \in \text{Act}$
3.  $A = \tau. \bigoplus_{i \in I} p_i E_i$ ,  $E \approx \langle (E_i : p_i) \mid i \in I \rangle$  and  $B = a. \bigoplus_{j \in J} q_j F_j$  with  $a \in \text{Act}$
4.  $A = \tau. \bigoplus_{i \in I} p_i E_i$ ,  $E \approx \langle (E_i : p_i) \mid i \in I \rangle$  and  $B = \Delta(H)$

Note that as we assumed that  $E$  is in sum form already, the presence of the  $\tau$ -prefixed, or  $\Delta$  summand excludes the existence of a summand prefixed by  $\lambda \in \mathbb{R}_{>0}$ , hence we can guarantee  $a \in \text{Act}$ .

*Proofs of the cases:*

1. By the definition of sum form,  $G$  and  $H$  may not contain summands of the form  $\tau.\mathcal{D}$  or  $\Delta(L)$ , thus are unable to perform any  $\tau$  transitions. If now  $G$  and  $H$  are weak distribution bisimilar, then also  $G \simeq_{\text{MA}} H$ , since they cannot perform any  $\tau$  transitions. Thus, by induction, we can rewrite  $G$  and  $H$  to some terms  $G'$  and  $H'$  with  $G' \equiv H'$  and then clearly also  $\Delta(G') \equiv \Delta(H')$ . This allows us to eliminate one of the summands by  $(\Delta\text{-6})_{\text{MA}}$ . If there are no other summands,  $E$  is now tangible after this transformation. Otherwise, we continue with other summands and one of our four cases.

But now assume that  $G$  and  $H$  are not weak distribution bisimilar. Then, since  $\Delta(G) \approx E$ , it must hold that  $\Delta(G) \xrightarrow{\tau} \mu$  with  $\mu \approx \delta(\Delta(H))$ , since also  $E \not\xrightarrow{\tau} \delta(\Delta(H))$ . But since  $G$  cannot perform any  $\tau$  transitions, this is impossible, contradicting our assumption, that  $G$  and  $H$  are not weak distribution bisimilar.

2. With a similar argument as before, it cannot be the case that  $a = \tau$ , except if  $\bigoplus_{j \in J} q_j F_j$  is itself weak distribution bisimilar to  $E$ . But then we can treat this as an instance of the next case. So if  $a \neq \tau$ , let  $G'$  be the normalized variant of  $G$ , which exists by induction. Then there must be a summand  $a.\mathcal{D}$  of  $G'$  with  $\mathcal{D} \approx \bigoplus_{j \in J} q_j F_j$ . We then proceed as in Case 3, which we refer the reader to.
3. The proof idea now is the following: As  $E \approx \langle (E_i : p_i) \mid i \in I \rangle$ , intuitively it must be the case that the distribution  $\bigoplus_{i \in I} p_i E_i$  must contain the summand  $a. \bigoplus_{j \in J} q_j F_j$  already, such that with an application of Axiom  $(D\text{-}\tau\text{-}3)$  from right to left, we can remove the summand from  $E$ . Repeating this argument for all remaining summands (except  $\tau. \bigoplus_{i \in I} p_i E_i$  itself), it is clear that only  $\tau. \bigoplus_{i \in I} p_i E_i$  will remain solely. From here, it remains to show that all the  $E_i$  can be saturated and made tangible, as then,  $E$  itself will be weakly tangible and weakly saturated.

We now proceed to formally show why the individual summands are already contained in  $\bigoplus_{i \in I} p_i E_i$ , and as a necessary step, we will show how the  $E_i$  can be rewritten into a saturated and tangible form. As  $a. \bigoplus_{j \in J} q_j F_j$  is a summand of  $E$ , clearly  $E \xrightarrow{a} \langle (F_j : q_j) \mid j \in J \rangle$ . Call this distribution  $\mu$ . As  $E \approx \langle (E_i : p_i) \mid i \in I \rangle$  also  $\langle (E_i : p_i) \mid i \in I \rangle \xrightarrow{a}_c \gamma$  and  $\mu \approx \gamma$  must hold for some  $\gamma$ . For simplicity, first assume that  $\langle (E_i : p_i) \mid i \in I \rangle \xrightarrow{a} \gamma$  instead of  $\langle (E_i : p_i) \mid i \in I \rangle \xrightarrow{a}_c \gamma$ . Then, as the  $E_i$  are by induction weakly saturated, they must be of shape either

$$E_i \equiv E'_i + a. \bigoplus_{k \in K_i} r_k G_k \quad \text{or} \quad E_i \equiv \tau.(E'_i + a. \bigoplus_{k \in K_i} r_k G_k)$$

with

$$\langle (G_k : p_i r_k) \mid i \in I, k \in K_i \rangle = \gamma.$$

Which of the two shapes actually is attained depends on whether  $E_i$  is already saturated or only weakly saturated. For the general case, that  $\langle (E_i : p_i) \mid i \in I \rangle \xrightarrow{a}_c \gamma$ , we can use Axiom (CC) to first add a suitable  $a$ -prefixed summand to each  $E_i$  first. We will now show that actually  $\mu \mathcal{L}(\equiv) \gamma$  holds. Having shown this, we can apply Axiom (D- $\tau$ -3) to reach our goal and remove the summand  $a. \bigoplus_{j \in J} q_j F_j$  from  $E$ .

It now remains to show that  $\mu \mathcal{L}(\equiv) \gamma$ . We know that

$$\langle (F_j : q_j) \mid j \in J \rangle = \mu \approx \gamma = \langle (G_k : p_i r_k) \mid i \in I, k \in K_i \rangle.$$

We will now show that by application of Axiom (D- $\tau$ -1)<sub>MA</sub>, we can rewrite  $\langle (F_j : q_j) \mid j \in J \rangle$  and  $\langle (G_k : p_i r_k) \mid i \in I, k \in K_i \rangle$  into two new distributions  $\mu'$  and  $\gamma'$  that satisfy the stronger statement

$$\mu' \mathcal{L}(\approx_\delta) \gamma'. \quad (\star)$$

Recall that by Theorem 8.15, two distributions  $\mu'$  and  $\gamma'$  must exist with  $\mu \xRightarrow{c} \mu'$  and  $\mu \approx \mu'$  and also  $\gamma \xRightarrow{c} \gamma'$  and  $\gamma \approx \gamma'$  with  $\mu' \mathcal{L}(\approx_\delta) \gamma'$ .

We proceed with our arguments only for  $\mu$ . The case for  $\gamma$  is fully analogous. By induction hypothesis, all processes  $F_j$  in the support of  $\mu$  are in normal form, and thus in particular, also weak tangible and weak saturated.

Hence, each  $F_j$  must be either already saturated and tangible, or be of the shape

$$F_j \equiv \tau. \bigoplus_{k \in K_j} r_k H_k$$

with suitable tangible and saturated  $H_k$ . If  $F_j$  is already saturated and tangible, then  $F_j$  must be in support of  $\mu'$  itself, as it *cannot* perform any internal transition  $F_j \xrightarrow{\tau} \xi$  without violating  $F_j \approx \xi$ ; if this equality is violated, however, also the equality  $\mu \approx \mu'$  must be violated by Lemma D.14 and Lemma 8.2. In case that  $F_j$  is of shape  $F_j \equiv \tau. \bigoplus_{k \in K_j} r_k H_k$  then the only possible transition it can take is the internal transition to  $\langle (H_k : r_k) \mid k \in K_j \rangle$ . From this distribution, no further weak transitions are possible with exactly the same argument which we have just given for saturated and tangible  $F_j$ , as the  $H_k$  must be saturated and tangible themselves. Thus, we can deduce that  $\mu'$  must be of shape

$$\mu' = \bigoplus_{j \in J} q_j \mu_j$$

where either  $\mu_j = \delta(F_j)$  or  $\mu_j = \langle (H_k : r_k) \mid k \in K_j \rangle$ . We can now apply Axiom (D- $\tau$ -1)<sub>MA</sub> on every  $F_j$ , that is not tangible and saturated already, i.e. it is of the form  $F_j \equiv \tau. \bigoplus_{k \in K_j} r_k H_k$ , since the  $F_j$  are part of a prefix expression  $a. \bigoplus_{j \in J} q_j F_j$ , which allows the application of the axiom. With these transformations, we (indirectly) rewrite  $\bigoplus_{j \in J} q_j F_j$  in such a way that

- (i) every process in its support is *tangible* and *saturated*, as they then are exactly either the saturated and tangible  $F_j$  or the  $H_k$ s, which are also tangible and saturated, and

- (ii) our goal, Equation  $\star$ , is met, as after this transformation, we exactly obtain  $\mu'$  from above. If we apply the same procedure to the  $G_k$ , which are the support of  $\gamma$ , we obtain  $\gamma'$  from above.

With this, we have reduced our problem of showing that we can apply Axiom  $(D\tau\beta)$  to remove the summand  $a. \bigoplus_{j \in J} q_j F_j$  from  $E$  to the problem of showing that if  $\mu' \approx_\delta \gamma'$ , then also  $\mu' \equiv \gamma'$ . To state the idea for the remaining proof simply, assume for a second that  $\approx$  and  $\simeq_{\text{MA}}$  coincide. Then by using induction hypothesis with the second claim, we can assume that for every  $H, H' \in \text{Supp}(\mu) \cup \text{Supp}(\gamma)$  with  $H \approx_\delta H'$  also  $H \equiv H'$ . Then, from  $\mu' \approx_\delta \gamma'$  and the use of axioms  $(D\text{COM})$  and  $(D\text{ADD})$ , our claim follows.

Clearly, as in general  $\approx$  and  $\simeq_{\text{MA}}$  do not coincide for arbitrary processes, we now need to show how we can still ensure that the induction hypothesis can be applied. It suffices to show that the  $H, H' \in \text{Supp}(\mu) \cup \text{Supp}(\gamma)$  with  $H \approx_\delta H'$  can always be rewritten with the axioms to processes  $\hat{H}$  and  $\hat{H}'$  in such a way that  $\hat{H} \simeq_{\text{MA}} \hat{H}'$ .

With this result at hand, we can use Axiom  $(D\tau\beta)_{\text{MA}}$  from right to left in order to prefix  $H$  or  $H'$  as needed by  $\tau$ .

This almost ends the induction step for the first claim, as we now have ensured that  $E$  is indeed weak tangible and also weak saturated. Only convex reducedness may have been violated by the applications of  $(D\tau\beta)$  above. We then must undo these changes again with the same axiom.

4. If also  $B \approx E$ , then this is an instance of Case 2. Otherwise, there must be transition  $\bigoplus_{i \in I} p_i E_i \xrightarrow{\tau} \mu$  for some  $\mu$  with  $\mu \approx \delta(B) = \delta(\Delta(H))$ . First, since  $\mu$  is bisimilar to a Dirac-distribution, all processes  $G_i$  in the support of  $\mu$  must be weak distribution bisimilar to each other and to  $B$ . Furthermore, the  $G_i$  (and hence  $\mu$ ) can be chosen such that all  $G_i$  are  $\Delta$  expresions. Then  $\mathcal{D}(G_i)$  and  $\mathcal{D}(H)$  is smaller then  $n$ . Furthermore, from the fact that  $\Delta(G_i) \approx \Delta(H)$  we can infer that  $\Delta(G_i) \simeq_{\text{MA}} \Delta(H)$  due to the presence of  $\Delta$ , which gurantees any necessary initial  $\tau$ -transitions. Given these facts, we can apply induction to establish that we can rewrite  $H$  to some  $H'$  and the  $G_i$  to some  $G'_i$ , respectively, with  $H' \equiv G'_i$ . From here, we can now eliminate summand  $B$  by applying  $(\Delta\tau)_{\text{MA}}$  and  $(D\tau\beta)$  multiply times.

**Proof step for Statement 2.a.)** As the processes are in normal form, and thus also in sum form, it suffices to proof that every summand of  $E$  is also a summand of  $F$  and vice versa. We now show that every summand of  $E$  must also be a summand of  $F$ . We skip the completely symmetric case that every summand of  $F$  is also a summand of  $E$ .

Let  $a. \bigoplus_{i \in I} p_i E_i$  with  $a \in \text{Act}$  be a summand of  $E$ . Hence  $E \xrightarrow{a} \langle (E_i : p_i) \mid i \in I \rangle := \mu$ . As  $E \simeq_{\text{MA}} F$ , there must exist a distribution  $\gamma$  such that  $F \xrightarrow{a}_{\text{c}} \gamma$  and  $\mu \approx \gamma$ . As  $F$  is saturated, this means that also  $F \xrightarrow{a}_{\text{c}} \gamma$ . We can now use Axiom  $(\text{CC})$  to ensure the existence of a suitable summand  $a. \bigoplus_{j \in J} q_j F_j$  in  $F$  with  $\gamma = \langle (F_j : q_j) \mid j \in J \rangle$ . As the  $E_i$  and the  $E_j$  are saturated and tangible, we can apply the induction hypothesis with Statement 2.b) to show that  $a. \bigoplus_{i \in I} p_i E_i \equiv a. \bigoplus_{j \in J} q_j F_j$ .

As we can repeat this argument for every summand of both  $E$  and  $F$  accordingly, it is clear that the distributions reachable by  $a$ -prefixed summands (for each  $a \in \text{Act}$ ) of both  $E$  and  $F$  agree up to  $\equiv$ , and convex combinations. Thus, our usage of Axiom  $(\text{CC})$  before has actually not been necessary, as both  $E$  and  $F$  are convex reduced. Thus, also without the additional use of Axiom  $(\text{CC})$ , we can establish  $E \equiv F$ .

Finally, let  $\lambda. \bigoplus_{i \in I} p_i E_i$  with  $\lambda \in \mathbb{R}_{>0}$  be a summand of  $E$ . As  $E$  is in sum form, this is the only delay-prefix summand of  $E$ . For the same reason,  $E$  must be stable. Hence, if  $E \simeq_{\text{MA}} F$ , it must hold that  $F \xrightarrow{\lambda}_c \gamma$  and  $\langle (E_i : p_i) \mid i \in I \rangle \approx \gamma$ . Furthermore,  $F$  must be stable, and posses one delay-prefix summand  $\lambda. \bigoplus_{j \in J} q_j F_j$ . Then, (i) immediately  $\langle (F_j : q_j) \mid j \in J \rangle = \gamma$ , and no weak internal transition has occurred. In this case, we continue our argument as before.

Alternatively, (ii), the weak combined transition of  $F$  must be representable as a sequence of two transitions  $F \xrightarrow{\lambda} \langle (F_j : q_j) \mid j \in J \rangle \Rightarrow_c \gamma$ . As the  $F_j$  are saturated and tangible, it must hold that  $\langle (F_j : q_j) \mid j \in J \rangle \not\approx \gamma$ . As clearly  $\langle (F_j : q_j) \mid j \in J \rangle$  can mimic every behaviour of  $\gamma$ , it must be the case that  $\gamma$  cannot match every transition of  $\langle (F_j : q_j) \mid j \in J \rangle$  up to  $\approx$ . As  $\langle (E_i : p_i) \mid i \in I \rangle \approx \gamma$ , also  $\langle (E_i : p_i) \mid i \in I \rangle$  cannot mimic every behaviour of  $\langle (F_j : q_j) \mid j \in J \rangle$ . If we now repeat all these arguments with the roles of  $E$  and  $F$  exchanged, we find the existence of a distribution  $\mu$  with  $\mu \approx \langle F_j : q_j \mid j \in J \rangle$  but  $\mu \not\approx \langle E_i : p_i \mid i \in I \rangle$ . As before, we can derive that  $\bigoplus_{i \in I} p_i E_i$  must posses some behaviour  $\mu$  cannot mimic. If we now denote  $\langle F_j : q_j \mid j \in J \rangle$  in the following by  $\gamma^*$ , and  $\langle E_i : p_i \mid i \in I \rangle$  by  $\mu^*$ , we see that this means that  $\mu^*$  cannot be mimicked by  $\mu$ , which is equivalent to  $\gamma^*$ , which in turn cannot be mimicked by  $\gamma$ , which, however, is equivalent to  $\mu^*$ . But from this chain of arguments we can infer that  $\mu$  cannot mimic every behaviour of itself. A contradiction. Thus, either  $\mu = \mu^*$  or  $\gamma = \gamma^*$  must hold, as the  $E_i$  and  $F_j$  are tangible. Assume that  $\mu = \mu^*$ . But then, as  $\mu \approx \gamma^*$  and  $\mu^* \approx \gamma$ , also  $\gamma \approx \gamma^*$  must hold, which is again a contradiction to our assumption, that the  $F_j$  are tangible. Thus  $\gamma = \gamma^*$  must hold, too. In summary,  $\mu^* \approx \gamma^*$  must hold, and thus, the weak internal transition that we assumed in this case, does not exist.

Finally, assume there is a summand  $\Delta(G)$  in  $E$ . Then  $E \xrightarrow{\tau} \delta(\Delta(G))$ . Since  $E \simeq_{\text{MA}} F$ , process  $F$  can weakly simulate this behaviour. Since  $F$  is weakly saturated, an even stronger condition holds: either  $F$  has a summand  $\Delta(H)$  such that  $\delta(\Delta(H)) \approx \delta(\Delta(G))$ , or it has a summand  $\tau. \langle H_i : p_i \mid i \in I \rangle$ , such that  $\delta(H_i) \approx \delta(\Delta(G))$  for all  $i \in I$ .

In the first case, we use induction on  $G$  and  $H$  with statement 2.a.), and our claim follows, since then we have found a corresponding summand in  $F$  up to  $\equiv$ . To do so, we need to establish that indeed  $G \simeq_{\text{MA}} H$ . This can be seen by  $\Delta(G) \approx \Delta(H)$  and the fact that neither  $G$  nor  $H$  can exhibit internal transitions due to them being in sum form inside a  $\Delta$ -expression.

For the second case, we first note that whenever  $H_i \xrightarrow{\tau} \mathcal{D}$  for some  $\mathcal{D}$ , then  $\mathcal{D} \approx \delta(\Delta(G))$ . This holds since  $\delta(H_i) \approx \delta(\Delta(G))$  and since  $G$  is in sum form,  $G$  may not exhibit any  $\tau$  transitions, and thus  $\Delta(G)$  could never mimick the  $\tau$ -transition of  $H_i$  if  $\mathcal{D} \not\approx \delta(\Delta(G))$  would hold. Thus, whenever  $H_i \xrightarrow{\tau} \mathcal{D}'$ , also  $\mathcal{D}' \approx \delta(\Delta(G))$  and each  $H'_i \in \text{Supp}(\mathcal{D}')$  satisfies  $\delta(H'_i) \approx \delta(\Delta(G))$ . Since we only consider processes without recursion, there must exist  $\mathcal{D}'$  such that  $\tau.\delta(H) \xrightarrow{\tau} \mathcal{D}'$ ,  $\mathcal{D}' \approx \delta(\Delta(G))$  and each  $H'_i \in \text{Supp}(\mathcal{D}')$  satisfies  $\delta(H'_i) \approx \delta(\Delta(G))$  and, additionally,  $H'_i$  has no summand of the form  $\tau.\mathcal{D}''$  for some  $\mathcal{D}''$ . But since  $\delta(H'_i) \approx \delta(\Delta(G))$ ,  $H'_i$  cannot be stable, since  $\Delta(G)$  is not stable. Thus, it must be the case that  $H'_i \equiv \Delta(H''_i)$  for some  $H''_i$ . Hence, in summary we see that  $F \xrightarrow{\tau} \langle \Delta(H''_i) : p_i \mid i \in I \rangle$  for some index set  $I$ , process expressions  $H''_i$  and probabilities  $p_i$  and the additional property  $\delta(\Delta(H''_i)) \approx \delta(\Delta(G))$  for all  $i \in I$ . From this property we can infer that  $\delta(\Delta(H''_i)) \approx \delta(\Delta(H''_j))$  for each  $i, j \in I$ . It further follows that also  $\delta(\Delta(H''_i)) \simeq_{\text{MA}} \delta(\Delta(H''_j))$  due to the presence of  $\Delta$ , which gurantees any necessary initial  $\tau$ -transitions. Since  $\mathcal{D}(H''_i) < n$  for all  $i \in N$ , since the are contained in a  $\tau$  summand of  $F$ , we can apply induction to assume that all the  $\Delta(H''_i)$  are already in normal form and that thus  $\Delta(H''_i) \equiv \Delta(H''_j)$  for all  $i, j \in I$ . By saturatedness of  $F$ , it must also be the case

that  $F \xrightarrow{\tau} \langle \Delta(H_i'') : p_i \mid i \in I \rangle$ . Furthermore, since  $F$  is in sum form,  $\langle \Delta(H_i'') : p_i \mid i \in I \rangle$  must in fact be a Dirac distribution  $\delta(\Delta(H_i''))$ , since sum form implies that syntactically identical process expressions in the support of a distribution expression are merged. Now we have established that  $F$  must have a summand  $\tau.\Delta(\delta(H_i''))$ . Again, by saturatedness,  $\Delta(\delta(H_i''))$  itself must be a summand of  $F$  and we in fact have arrived at an instance of our first case.

**Proof step for Statement 2.b.)** Similar as in the proof of Statement 2.a.), we can show that the processes  $E$  and  $F$  have the same summands, except for the following case: without loss of generality, if  $E \xrightarrow{\tau} \mu$  and  $\delta(E) \approx \mu$ , then it may be the case that  $F$  does not exhibit an actual summand, as immediately,  $F \approx \mu \approx E$  holds. In this case, however,  $E \approx \tau.\delta(F)$  then satisfies the desired property. Clearly,  $\tau.\delta(F)$  exhibits the missing summand. All summands of  $F$ , which are now not top-level, can be reinstalled by applying  $(D\text{-}\tau\text{-}\beta)$ . This also holds for summands that are  $\Delta$  expressions, when we apply  $(\Delta\text{-}I)_{\text{MA}}$  first to convert them into a  $\tau$ -prefixed summand.  $\square$

## F. Proofs of Chapter 11

### F.1. Lemma 11.4

1.  $\mathcal{A} \xrightarrow{\sim} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \mathbb{B}$ .
2.  $\mathcal{A} \xrightarrow{C} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \{\sim_{\text{PA}}, \sim_{\text{MA}}\}$ .
3.  $\mathcal{A} \xrightarrow{T} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}'$  and  $\asymp \in \{\sim_{\text{LTS}}, \sim_{\text{PA}}\}$ .
4.  $\mathcal{A} \xrightarrow{T}_{\downarrow} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}'$  and  $\asymp \in \{\approx_{\text{MC}}, \approx_{\text{MA}}, \approx_{\delta}\}$ .
5.  $\mathcal{A} \xrightarrow{L} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \{\approx_{\text{PA}}, \approx_{\text{MA}}, \approx_{\delta}\}$ .
6.  $\mathcal{A} \xrightarrow{R} \mathcal{A}'$  implies  $\mathcal{A} \approx_{\delta} \mathcal{A}'$  for weakly image-finite MA.
7.  $\mathcal{A} \xrightarrow{M} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \{\sim_{\text{MC}}, \approx_{\text{MC}}, \approx_{\text{MA}}, \approx_{\delta}\}$ .
8.  $\mathcal{A} \xrightarrow{\Sigma} \mathcal{A}'$  implies  $\mathcal{A} \asymp \mathcal{A}'$  for each  $\mathcal{A}, \mathcal{A}' \in \text{MA}$  and  $\asymp \in \{\sim_{\text{MC}}, \approx_{\text{MC}}, \approx_{\text{MA}}, \approx_{\delta}\}$ .

*Proof.* **Proof of 1.** The result follows immediately from the definitions of the reductions.

**Proof of 2.** Follows immediately from the definitions.

**Proof of 3.** As  $S = S'$ , we can take the symmetric and transitive closure of the identity relation as a bisimulation, irrespective of which notion of bisimulation we want to establish. Then everything follows by noting that reachability with respect to weak combined transitions is fully preserved by the reduction.

Also the second condition of weak distribution bisimilarity, the splitting condition follows immediately in this way.

**Proof of 4.** Everything follows similarly to the last proof. The only interesting case we need to consider, however, is the last condition of the three bisimulations. It states that whenever  $s \mathcal{R} t$  by some bisimulation  $\mathcal{R}$ , then  $s \downarrow$  implies  $t \Rightarrow_c \mu$  and  $\mu \downarrow$ . As we only remove immediate transitions, it is guaranteed that if a state was stable in  $\mathcal{A}$ , then it must still be stable in  $\mathcal{A}'$ . Similarly, if  $s \Rightarrow_c \mu$  and  $\mu \downarrow$  in  $\mathcal{A}$ , this also holds in  $\mathcal{A}'$ , as reachability is preserved, too, by the reduction. Finally, let us assume that  $s \downarrow$  in  $\mathcal{A}'$ , but  $s$  cannot reach a stable distribution in  $\mathcal{A}$ . Then this means that all immediate transitions have been removed while preserving reachability. But this means that  $s \xrightarrow{\tau}_c \circ \Rightarrow_c \mu$  with at least one  $\tau$ -transition, must imply  $\mu = \delta(s)$ . But by the definition of  $\xrightarrow{T}_{\downarrow}$ , in this case a transition  $s \xrightarrow{\tau} \delta(s)$  exists in  $\mathcal{A}'$ .

**Proof of 5.** Since by definition of  $\xrightarrow{L}$ ,  $\mathcal{A}$  and  $\mathcal{A}'$  have the same set of states, we use  $\nu$  to refer to distributions from both  $\mathcal{A}$  and  $\mathcal{A}'$ ; however, we write  $s'$  when we mean the state  $s \in \mathcal{A}'$  in comparison to  $s \in \mathcal{A}$ .

Let  $\mathcal{I}$  be the equivalence relation on  $S \uplus S'$  whose set of classes is  $\{\{s, s'\} \mid s \in S\}$ , i.e., we relate each state  $s$  with its primed counterpart in  $\mathcal{A}'$ . We now show that  $\mathcal{I}$  is a weak probabilistic bisimulation for  $\mathcal{A}$  and  $\mathcal{A}'$ . We treat the case for weak distribution bisimilarity later. As  $\mathcal{I}$  is an equivalence relation, it also contains pairs  $s \mathcal{I} s$  or  $s' \mathcal{I} s'$ , i.e. pairs, where both states are taken from the same automaton. Since the proof is trivial here, we skip it. Let  $s \mathcal{I} s'$  and  $s \xrightarrow{a} \nu$ ; if  $a \neq \tau$ , then also  $s'$  enables exactly the same transition  $s' \xrightarrow{a} \nu$ , as the reduction only affects internal transitions, and thus  $\nu \mathcal{L}(\mathcal{I}) \nu$ . Now, consider  $a = \tau$ :  $s'$  is able to match such transition by the weak combined transition  $s' \xRightarrow{\tau}_c \nu$  as induced by the scheduler  $\sigma$  such that  $\sigma(s')(\perp) = \nu(s)$ ,  $\sigma(s')(tr) = 1 - \nu(s)$ , and  $\sigma(\alpha)(\perp) = 1$  for each finite execution fragment  $\alpha \neq s'$ , where  $tr = (s', \tau, \nu \ominus s)$ . Note that this applies also when  $\nu = \delta(s)$  as the resulting scheduler assigns  $\sigma(s')(\perp) = \nu(s) = 1$  so the induced weak combined transition is  $s' \xRightarrow{\tau}_c \delta(s')$  and  $\delta(s) \mathcal{L}(\mathcal{I}) \delta(s')$ .

Now, we exchange roles and let  $s' \xrightarrow{a} \nu$ . If  $a \neq \tau$ , the case is straightforward as before. Otherwise, by the definition of the reduction, there must exist a distribution  $\rho$  such that  $\nu = \rho \ominus s'$  or  $\nu = \rho \ominus s$ , respectively, as  $s = s'$ , and  $s \xrightarrow{\tau} \rho$ . But then, there clearly also exists a weak combined transition  $s \xRightarrow{\tau}_c \nu$ , which is obtained from a scheduler that infinitely often executes the transition  $s \xrightarrow{\tau} \rho$ , whenever it sees  $s$ , and stops for any other state. This proves that  $\mathcal{I}$  is a weak probabilistic bisimulation.

The proof for  $\approx_{\text{MA}}$  is completely analogue.

Similarly, we can establish that the straightforward lifting of  $\mathcal{I}$  to distributions over states is a weak distribution bisimulation. For the first condition, the argument is completely analogous to the above proof. For the second condition, consider an arbitrary pair  $\mu \mathcal{I} \mu'$  together with a splitting  $\mu = \mu_1 \oplus_p \mu_2$ . Recall that  $\mathcal{A}$  and  $\mathcal{A}'$  consist of exactly the same states; and we only used the primed notation for states of  $\mathcal{A}'$  to make a clearer distinction. As  $\mu$  and  $\mu'$  are identical then,  $\mu_1 \oplus_p \mu_2$  is also a splitting for  $\mu'$ , and by definition  $\mu_1 \mathcal{I} \mu_1$  and  $\mu_2 \mathcal{I} \mu_2$ . For the last case, we note that the reduction does not change the actions enabled from a state. Thus, it preserves stability. This suffices to establish the last condition of weak distribution bisimulation.

**Proof of 6.** In the reduced automaton  $\mathcal{A}'$ , the state space consists of non-redundant states of  $\mathcal{A}$  only. Furthermore, by construction of the transition relations of  $\mathcal{A}'$ , every reachable redundant distribution is associated with non-redundant distributions via the relation  $\xRightarrow{P}$ . To recall this, we repeat the definition for the relation  $\twoheadrightarrow$  for convenience:

*Remark F.1 (Reminder).*  $\twoheadrightarrow$  is the smallest relation containing for every weak hypertransition  $s \xRightarrow{a} \mu'$  of  $\mathcal{A}$ , a transition  $s \twoheadrightarrow \mu$  for some  $\mu$  satisfying  $\mu' \xRightarrow{P} \mu$ .

Our idea is to show that the symmetric closure of

$$\mathcal{R} := \{(\mu, \gamma) \in \text{Dist}(S) \times \text{Dist}(S') \mid \mu \xRightarrow{P} \circ \mathcal{L}(\approx_\delta) \gamma\} \cup \{(\delta(\bar{s}), \delta(\bar{s}'))\}$$

is a weak distribution bisimulation containing the pair of the initial states, i.e.  $(\delta(\bar{s}), \delta(\bar{s}')) \in \mathcal{R}$ . We denote by  $\mathcal{R}^{-1}$  the symmetric complement of  $\mathcal{R}$ .

The definition of  $\mathcal{R}$  has been constructed as the union of two sets. Containment of the initial states follows immediately by the definition of the second set. In the case that  $\bar{s}$  is non-redundant,  $\bar{s} = \bar{s}'$  by construction of  $\mathcal{A}'$ , adding the pair  $(\delta(\bar{s}), \delta(\bar{s}'))$  to  $\mathcal{R}$  explicitly is not really necessary, as the states are also included in the first set defining  $\mathcal{R}$ . In the case that  $\bar{s}$  is redundant,  $\bar{s}'$  is a



fresh state  $t'$  and thus  $\delta(t') \notin \text{Dist}(S')$ , which prevents the pair of initial states to be in the first set already.

In order to establish that the symmetric closure of  $\mathcal{R}$  is indeed a weak distribution bisimulation, we need to check the bisimulation condition for every pair of this relation. The pairs stemming from the second set, i.e. the initial states, needs treatment only in the case where  $\bar{s}$  has been redundant, since in the other case, the pair is already included in the first set defining  $\mathcal{R}$ , and we will hence provide its proof implicitly when we treat pairs from this set. The proof of the case that  $\bar{s}$  is redundant is not identical, but very similar to the proofs for pairs taken from the first set.

We omit it and immediately proceed with the proof for an arbitrary pair  $(\mu, \gamma)$  from the first set. Let hence  $(\mu, \gamma) \in \text{Dist}(S) \times \text{Dist}(S')$  and  $\mu \xRightarrow{P} \circ \mathcal{L}(\approx_\delta) \gamma$ .

Assume that according to Condition (a) of Definition 8.4,  $\mu \xrightarrow{a} \mu'$  in  $\mathcal{A}$ . As  $\mu \xRightarrow{P} \gamma$  (in  $\mathcal{A}$ ), it holds that  $\gamma \xRightarrow{a}_c \circ \xRightarrow{P} \gamma'$  for some  $\gamma'$  and also  $\mu' \xRightarrow{P} \mu''$  for some  $\mu''$  satisfying  $\mu'' \mathcal{L}(\approx_\delta) \gamma'$  by the fact that  $\approx_\delta$  is a weak state bisimulation on weakly image-finite MA (Theorem 8.14). By the definition of combined hypertransition, the definition  $\gamma \xRightarrow{a}_c \circ \xRightarrow{P} \gamma'$  can be split into non-combined hypertransitions  $\gamma \xRightarrow{a} \gamma'_i$  such that together with suitable weight  $p_i$  it holds that  $\gamma' = \bigoplus_i p_i \cdot \gamma'_i$ .

For each of these non-combined transitions, we know from the construction of  $\mathcal{A}'$  (cf. the remark above), that there exists a transition  $\gamma \xrightarrow{a} \gamma''_i$  in  $\mathcal{A}'$  with the property that  $\gamma'_i \xRightarrow{P} \gamma''_i$ . As  $\gamma'_i$  already resulted from an application of the relation  $\xRightarrow{P}$  and the definition of  $\xRightarrow{P}$ , we know that  $\gamma''_i \mathcal{L}(\approx_\delta) \gamma'_i$ .

Combing these transitions, we see that hence  $\gamma \xRightarrow{a}_c \bigoplus_i p_i \gamma''_i$ , and furthermore,

$$\gamma' = \bigoplus_i p_i \cdot \gamma'_i \mathcal{L}(\approx_\delta) \bigoplus_i p_i \cdot \gamma''_i$$

holds because  $\gamma''_i \mathcal{L}(\approx_\delta) \gamma'_i$  for each  $i$ . Thus, by transitivity of  $\mathcal{L}(\approx_\delta)$ , in summary, we obtain

$$\mu' \xRightarrow{P} \mu'' \text{ and } \mu'' \mathcal{L}(\approx_\delta) \bigoplus_i p_i \cdot \gamma''_i.$$

By definition, this implies

$$\mu' \mathcal{R} \bigoplus_i p_i \cdot \gamma''_i,$$

which suffices to satisfy Condition (a) of Definition 8.4.

Condition (b) can be shown by using Lemma 4.3 and the fact that  $\xRightarrow{P}$  is, after all, an ordinary weak combined hypertransition. If  $\mu_1 \oplus_p \mu_2$  is the splitting of  $\mu$ , Lemma 4.3 allows us to obtain a splitting  $\gamma_1 \oplus_p \gamma_2$  of  $\gamma$  satisfying the property that  $\mu_i \xRightarrow{P} \gamma_i$  for  $i \in \{1, 2\}$ . The last condition suffices to establish  $\mu_i \mathcal{R} \gamma_i$  as well.

Condition (c) is straightforward, as  $\mu \downarrow$  implies that  $\mu = \gamma$ .

We finally proceed with the proof for pairs taken from the symmetric complement  $\mathcal{R}^{-1}$ . Let  $(\gamma, \mu)$  be such a pair and let, for Condition (a),  $\gamma \xrightarrow{a} \gamma'$  in  $\mathcal{A}'$ . By the construction of  $\mathcal{A}'$ , this transition is represented in  $\mathcal{A}$  in the form of a weak combined hypertransition  $\gamma \xRightarrow{a}_c \gamma'$  (more precisely, as an instance of the relation  $\xRightarrow{a}_c \circ \xRightarrow{P}$ ). Thus,  $\gamma$  in  $\mathcal{A}$  can weakly simulate every

transition of  $\gamma$  in  $\mathcal{A}'$ . Together with the fact that  $\mu \xRightarrow{P} \gamma$ , which implies  $\mu \Longrightarrow_c \gamma$ , we obtain  $\mu \xRightarrow{a}_c \gamma'$ . Clearly  $\gamma' \mathcal{R} \gamma'$  (as by construction of  $\mathcal{A}'$ ,  $\gamma'$  is non-redundant).

Conditions (b) and (c) follow with a similarly simply argument relying on  $\mu \xRightarrow{P} \gamma$ .

**Proof of 7** The only conditions dealing with timed transitions in the definitions of the bisimulations use the predicate  $\mu \xrightarrow{\chi(\lambda)} \mu'$  in the premise of an implication (Condition a in all cases). This predicate, however, becomes logically *false* as soon as  $\mu$  is unstable i.e. there is a state in  $\text{Supp}(\mu)$  that enables an internal transition. Hence, in this case, the implication becomes trivially true immediately, without actually being concerned with the timed transitions of  $\mu$ . Thus, they are not affected by removing or adding timed transitions from unstable states. Clearly, the other conditions are not affected, too, as timed transitions are not mentioned there.

**Proof of 8** As in the last case, this follows immediately from the definition of the notation  $\mu \xrightarrow{\chi(\lambda)} \mu'$ , as there, the accumulation of rates as issued by  $\xrightarrow{\Sigma}$  already takes place implicitly.  $\square$

## F.2. Lemma 11.5

For every MA  $\mathcal{A}$ , a MA  $\mathcal{A}'$  with  $\mathcal{A} \rightsquigarrow \mathcal{A}'$  can be computed in

- *polynomial time* if  $\rightsquigarrow = \widetilde{\rightsquigarrow}$  and  $\asymp \in \{\sim_{\text{LTS}}, \approx_{\text{LTS}}, \sim_{\text{PA}}, \approx_{\text{PA}}\}$
- *polynomial time* if  $\rightsquigarrow \in \{\xrightarrow{C}, \xrightarrow{T}, \xrightarrow{T}_{\downarrow}, \xrightarrow{L}, \xrightarrow{M}, \xrightarrow{\Sigma}\}$ ,
- *exponential time* if  $\rightsquigarrow = \widetilde{\approx}_{\delta} \circ \widetilde{R}$ .

*Proof.* The polynomial time result for  $\widetilde{\rightsquigarrow}$  follows by the corresponding polynomial decision procedures [CS02; HT12; FM91; PT87; KS83; Her02] and reachability analysis. Further details can also be found in Chapter 3, 4 and 5. The exponential time result for  $\widetilde{\approx}_{\delta}$  comes from the exponential time algorithm to decide  $\approx_{\delta}$  (cf. Chapter 10). This complexity is, however, only an upper bound. The actual complexity is currently unknown. The rest follows as before.

$\xrightarrow{C}$  requires for each state and each enabled action to solve  $\mathcal{O}(|D|)$  linear programming problems (cf. [CS02, Sec. 6]) in order to find the set of generators of reachable distributions;  $\xrightarrow{L}$  can be obtained by computing for each transition  $s \xrightarrow{T} \nu$  the distribution  $\nu \ominus s$  that requires at most  $\mathcal{O}(|S|)$  operations; finally,  $\xrightarrow{T}$  and  $\xrightarrow{T}_{\downarrow}$  can be computed by iteratively removing transitions and checking, if reachability with respect to weak combined transitions has changed by operation. The check can be done in polynomial time by a convex reachability analysis from state  $s$  to distribution  $\mu$  if  $s \xrightarrow{a} \mu$  is the transition that has been removed in each of the successive elimination, using techniques from [HT12], where similar problems are described as a variation of flow problems. If it turns out that a transition cannot be removed without breaking reachability, this cannot change after removing further (reachability preserving) transitions. Thus, every transition in  $\rightarrow$  is considered exactly once for removal.  $\xrightarrow{\Sigma}$  and  $\xrightarrow{M}$  can obviously be computed by considering every state and its outgoing transitions exactly once.  $\square$

### F.3. Lemma 11.11

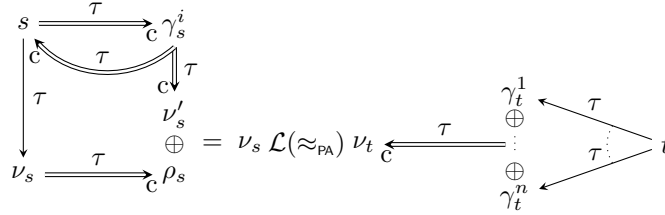
Let  $\mathcal{A}$  be a  $\lesssim_{\text{PA}}^{[S]}$ -minimal MA,  $\mathcal{A} \xrightarrow{T} \circ \xrightarrow{L} \mathcal{A}'$ , and  $\mathcal{A}'_m$  be a  $\lesssim_{\text{PA}}^{[S]}$  and  $\lesssim_{\text{PA}}^{[D]}$ -minimal MA satisfying  $\mathcal{A}'_m \approx_{\text{PA}} \mathcal{A}$ . Now let  $\mathcal{A}'_m \xrightarrow{L} \mathcal{A}_m$  for some  $\mathcal{A}_m$ . Then  $\mathcal{A}' =_{\text{iso}} \mathcal{A}_m$ .

*Proof.* Let  $\mathcal{A}_m$  and  $\mathcal{A}'$  be chosen as in the claim. We then proceed similarly as in the proof of Lemma 11.9 to show that  $b = \approx_{\text{PA}} \cap (S_m \times S')$  is a bijection. Then we will be able to establish that  $b$  is a suitable mapping to establish  $\mathcal{A}_m =_{\text{iso}} \mathcal{A}'$ .

Assume, to derive a contradiction, that  $b$  is not an isomorphism. Since  $b$  is a bijection between  $S_m$  and  $S'$  (note that all automata in this lemma must be  $\lesssim_{\text{PA}}^{[S]}$ -minimal), in order to have  $\mathcal{A}_m \neq_{\text{iso}} \mathcal{A}'$  there must exist  $s \in S_m, t \in S'$  with  $s \approx_{\text{PA}} t$  (i.e.  $b(s) = t$ ), and

- (i) either a transition  $s \xrightarrow{a} \nu_s \in \mathbf{\rightarrow}_m$  but there does not exist  $t \xrightarrow{a} \nu_t \in \mathbf{\rightarrow}'$  such that  $\nu_s \mathcal{L}(\approx_{\text{PA}}) \nu_t$ , i.e. there does not exist a transition  $t \xrightarrow{a} \nu_t \in \mathbf{\rightarrow}'$  such that  $\nu_t = b(\nu_s)$ , or
- (ii) a transition  $t \xrightarrow{a} \nu_t \in \mathbf{\rightarrow}'$  but there does not exist  $s \xrightarrow{a} \nu_s \in \mathbf{\rightarrow}_m$  such that  $\nu_s \mathcal{L}(\approx_{\text{PA}}) \nu_t$ . We proceed with the proof of (i).

Note that this cannot be caused by two transitions with  $\nu_t \neq b(\nu_s)$  but  $b(\nu_s \oplus s) = \nu_t \oplus t$ , since both automata are rescaled. However, since  $s \approx_{\text{PA}} t$ , it follows that there exists  $t \xrightarrow{a} \nu_t$  such that  $\nu_s \mathcal{L}(\approx_{\text{PA}}) \nu_t$ . Now, there are two cases: either  $a \in E$ , or  $a \in H$ . We provide the detailed proof for  $a = \tau$  whose schematic proof idea is depicted below; the case  $a \neq \tau$  is similar.



Let  $\sigma_t$  be the scheduler inducing  $t \xrightarrow{\tau} \nu_t$  and  $t \xrightarrow{\tau} \gamma_t^1, \dots, t \xrightarrow{\tau} \gamma_t^n$  be all transitions such that  $\sigma_t(t)(t \xrightarrow{\tau} \gamma_t^i) > 0$  and  $\gamma_t^i \not\mathcal{L}(\approx_{\text{PA}}) \nu_s$ , that is,  $t \xrightarrow{\tau} \gamma_t^i$  is a transition used in the first step of the weak combined transition  $t \xrightarrow{\tau} \nu_t$ ; it is immediate to see that  $(\bigoplus_{i=1}^n \gamma_t^i) \xrightarrow{\tau} \nu_t$ . Since  $s \approx_{\text{PA}} t$ , it follows that there exists  $\gamma_s^i$  for each  $1 \leq i \leq n$  such that  $s \xrightarrow{\tau} \gamma_s^i$  and  $\gamma_s^i \mathcal{L}(\approx_{\text{PA}}) \gamma_t^i$ . Furthermore,  $(\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau} \nu_s$ , as  $(\bigoplus_{i=1}^n \gamma_t^i) \xrightarrow{\tau} \nu_t$  and  $\nu_t = b(\nu_s)$ .

Now, consider a generic  $\gamma_s^j$ ; there are two cases depending on whether  $s \xrightarrow{\tau} \nu_s$  is used to reach  $\nu_s$ . If it is not used by any of the  $\gamma_s^i$ , then there exists the weak combined transition  $s \xrightarrow{\tau} (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau} \nu_s$  that does not involve  $s \xrightarrow{\tau} \nu_s$ , hence  $s \xrightarrow{\tau} \nu_s$  can be omitted. This contradicts the  $\lesssim_{\text{PA}}^{[D]}$ -minimality of  $\mathcal{A}_m$ .

So, suppose that  $s \xrightarrow{\tau} \nu_s$  is used in order to reach  $\nu_s$ . Since  $(\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau} \nu_s$ , we may split this hyper-transition into two parts according to Lemma 4.3, depending on whether  $s \xrightarrow{\tau} \nu_s$  is chosen by the scheduler with non-zero probability:  $(\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau} \nu'_s$  with weight  $c_1 \geq 0$  that does not involve  $s \xrightarrow{\tau} \nu_s$ , and  $(\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau} \delta(s)$  with weight  $c_2 > 0$  that involves  $s \xrightarrow{\tau} \nu_s$  such that  $c_1 + c_2 = 1$  and there exists  $\rho_s$  such that  $(s \xrightarrow{\tau} \nu_s)$  and  $\nu_s \xrightarrow{\tau} \rho_s$ .

and  $\nu_s = (c_1\nu'_s \oplus c_2\rho_s)$ . Note that we use  $\rho_s$  instead of  $\nu_s$  since it may be that, in order to reach distribution equivalent to  $\nu_s$ , we have to adjust probabilities by performing more steps. Now, consider the convex combination of the two weak combined transitions  $Tr_1 = s \xrightarrow{\tau}_c (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau}_c \nu'_s$  and  $Tr_2 = s \xrightarrow{\tau}_c (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau}_c \delta(s) \xrightarrow{\tau} \nu_s \xrightarrow{\tau}_c \rho_s$ , with weights  $c_1$  and  $c_2$ , respectively. Since  $(c_1\nu'_s \oplus c_2\rho_s) = \nu_s$ , we have that such convex combination corresponds to the weak transition  $s \xrightarrow{\tau}_c \nu_s$ , so we can replace the transition  $s \xrightarrow{\tau} \nu_s$  by the weak combined transition  $Tr = c_1 \cdot Tr_1 \oplus c_2 \cdot Tr_2$  with  $\nu_s = c_1\nu'_s \oplus c_2\rho_s$ . Since  $s \xrightarrow{\tau} \nu_s$  still occurs in  $Tr_2 = s \xrightarrow{\tau}_c \delta(s) \xrightarrow{\tau} \nu_s \xrightarrow{\tau}_c \rho_s$ , we can recursively replace it by the same weak combined transition  $Tr$ , hence, after  $k$  replacements, we have that  $\nu_s = c_1\nu'_s \oplus c_2c_1\nu'_s \oplus c_2^2c_1\nu'_s \oplus \dots \oplus c_2^k\rho_s = (\bigoplus_{l=0}^{k-1} c_1c_2^l\nu'_s) \oplus c_2^k\rho_s$ , that is,  $(\bigoplus_{l=0}^{k-1} (1 - c_2)c_2^l\nu'_s) \oplus c_2^k\rho_s$ . If we tend  $k$  to infinite, since  $c_2 < 1$ , we derive that  $\nu_s = \nu'_s$ , therefore there exists the weak combined transition  $s \xrightarrow{\tau}_c (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau}_c \nu_s$  that does not involve  $s \xrightarrow{\tau} \nu_s$ , hence again  $s \xrightarrow{\tau} \nu_s$  can be omitted. This contradicts the  $\approx_{\text{PA}}^{[D]}$ -minimality of  $\mathcal{A}_m$ . The proof of case (ii) is completely analogous, except that the contradictions will be derived with respect to  $\subseteq_D$ , which is a result of the fact that  $\mathcal{A}'$  has been reduced according to  $\xrightarrow{T}$ .

As final note, consider the weight  $c_2$  and suppose that  $c_2 = 1$ . Since  $s \xrightarrow{\tau}_c (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau}_c \delta(s)$  with  $(\bigoplus_{i=1}^n \gamma_s^i) \not\approx_{\text{PA}} \delta(s)$ , it follows that each state in the support of  $\bigoplus_{i=1}^n \gamma_s^i$  is actually weak bisimilar to  $s$  as the states touched in the loop  $s \xrightarrow{\tau}_c (\bigoplus_{i=1}^n \gamma_s^i) \xrightarrow{\tau}_c \delta(s)$  form a strongly connected component. But this contradicts the  $\approx_{\text{PA}}^{[S]}$ -minimality of  $\mathcal{A}_m$ .  $\square$

# Bibliography

- [AB01] S. Andova and J. C. M. Baeten. “Abstraction in Probabilistic Process Algebra”. In: *TACAS 2001*. Vol. 2031. LNCS. 2001, pp. 204–219.
- [Abr06] S. Abramsky. “What are the Fundamental Structures of Concurrency?: We still don’t know!” In: *Electronic Notes in Theoretical Computer Science* 162 (2006), pp. 37–41.
- [ABW06] S. Andova, J. C. M. Baeten, and T. A. C. Willemse. “A Complete Axiomatisation of Branching Bisimulation for Probabilistic Systems with an Application in Protocol Verification”. In: *CONCUR 2006*. Vol. 4137. LNCS. 2006, pp. 327–342.
- [AG09] S. Andova and S. Georgievska. “On Compositionality, Efficiency, and Applicability of Abstraction in Probabilistic Systems”. In: *SOFSEM 2009: Theory and Practice of Computer Science*. Springer, 2009, pp. 67–78.
- [AGT08] S. Andova, S. Georgievska, and N. Trcka. *Probabilistic Process Algebra of Communicating Processes with Abstractions*. 2008.
- [AGT12] S. Andova, S. Georgievska, and N. Trcka. “Branching Bisimulation Congruence for Probabilistic Systems”. In: *Theoretical Computer Science* 413.1 (2012), pp. 58–72.
- [AGU72] A. Aho, M. Garey, and J. Ullman. “The Transitive Reduction of a Directed Graph”. In: *SIAM Journal on Computing* 1.2 (1972), pp. 131–137.
- [AHJ01] L. de Alfaro, T. A. Henzinger, and R. Jhala. “Compositional Methods for Probabilistic Systems”. In: *CONCUR 2001*. Ed. by K. G. Larsen and M. Nielsen. Springer, 2001, pp. 351–365.
- [Alf97] L. d. Alfaro. “Formal Verification of Probabilistic Systems”. PhD thesis. Stanford University, 1997.
- [And02] S. Andova. “Probabilistic Process Algebra”. PhD thesis. Eindhoven University of Technology, 2002.
- [And99] S. Andova. “Process Algebra with Probabilistic Choice”. In: *Formal Methods for Real-Time and Probabilistic Systems*. Ed. by J.-P. Katoen. Vol. 1601. LNCS. Springer, 1999, pp. 111–129.
- [AW06] S. Andova and T. A. C. Willemse. “Branching Bisimulation for Probabilistic Systems: Characteristics and Decidability”. In: *Theoretical Computer Science* 356.3 (2006), pp. 325–355.
- [BA95] A. Bianco and L. d. Alfaro. “Model Checking of Probabilistic and Nondeterministic Systems”. In: *FSTTCS 1995*. LNCS. Springer, 1995, pp. 499–513.
- [Bäc + 16] O. Bäckström, Y. Butkova, H. Hermanns, J. Krcál, and P. Krcál. “Effective Static and Dynamic Fault Tree Analysis”. In: *SAFECOMP 2016*. LNCS. Springer, 2016.
- [Bae + 10] J. C. Baeten, T. Basten, T. Basten, and M. Reniers. *Process Algebra: Equational Theories of Communicating Processes*. Vol. 50. Cambridge University Press, 2010.

- [Bae05] J. C. Baeten. “A Brief History of Process Algebra”. In: *Theoretical Computer Science* 335.2-3 (2005), pp. 131–146.
- [Bae93] J. C. Baeten. *The Total Order Assumption*. Springer, 1993.
- [Bai+03a] C. Baier, B. R. Haverkort, H. Hermanns, and J.-P. Katoen. “Model-Checking Algorithms for Continuous-Time Markov Chains”. In: *IEEE TSE* 29.10 (2003).
- [Bai+03b] C. Baier, H. Hermanns, J.-P. Katoen, and V. Wolf. “Comparative Branching-Time Semantics for Markov Chains (Extended Abstract)”. In: *CONCUR 2003*. Vol. 2761. LNCS. 2003, pp. 492–507.
- [Bai+05] C. Baier, J.-P. Katoen, H. Hermanns, and V. Wolf. “Comparative Branching-Time Semantics for Markov Chains”. In: *Information and Computation* 200.2 (2005), pp. 149–214.
- [Bai+10] C. Baier, B. R. Haverkort, H. Hermanns, and J. Katoen. “Performance Evaluation and Model Checking Join Forces”. In: *Communications of the ACM* 53 (2010).
- [Bai96] C. Baier. “Polynomial Time Algorithms for Testing Probabilistic Bisimulation and Simulation”. In: *CAV 1996*. Ed. by R. Alur and T. Henzinger. Vol. 1102. LNCS. Springer, 1996, pp. 50–61.
- [Bai98] C. Baier. *On Algorithmic Verification Methods for Probabilistic Systems*. Habilitation Thesis. Universität Mannheim, 1998.
- [Bal00] G. Balbo. “Introduction to Stochastic Petri Nets”. In: *European Educational Forum: School on Formal Methods and Performance Analysis*. Vol. 2090. LNCS. Springer, 2000, pp. 84–155.
- [Bal07] G. Balbo. “Introduction to Generalized Stochastic Petri Nets”. In: *Formal Methods for Performance Evaluation: 7th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2007, Bertinoro, Italy, May 28-June 2, 2007, Advanced Lectures*. Ed. by M. Bernardo and J. Hillston. Vol. 4486. LNCS. Springer Berlin Heidelberg, 2007, pp. 83–131.
- [Bam12] R. Bamberg. “Non-Deterministic Generalised Stochastic Petri Nets Modelling and Analysis”. MA thesis. University of Twente, 2012.
- [Bar+11] B. Barbot, T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. “Efficient CTMC Model Checking of Linear Real-Time Objectives”. In: *TACAS 2011*. Vol. 6605 LNCS. Springer, 2011, pp. 128–142.
- [Bau+02] F. Bause, H. Beilner, M. Fischer, P. Kemper, and M. Völker. “The ProC/B Toolset for the Modelling and Analysis of Process Chains”. In: *Computer Performance Evaluation, Modelling Techniques and Tools*. Vol. 2324. LNCS. Springer, 2002, pp. 51–70.
- [BB87] T. Bolognesi and E. Brinksma. “Introduction to the ISO Specification Language LOTOS”. In: *Computer Networks* 14.1 (1987), pp. 25–59.
- [BBB92] J. Bechta Dugan, S. J. Bavuso, and M. A. Boyd. “Dynamic Fault-Tree Models for Fault-Tolerant Computer Systems”. In: *Reliability, IEEE Transactions on* 41.3 (1992), pp. 363–377.
- [BBK87] J. C. M. Baeten, J. A. Bergstra, and J. W. Klop. “Ready-Trace Semantics for Concrete Process Algebra with the Priority Operator”. In: *The Computer Journal* 30 (1987).

- [BBS95] J. Baeten, J. Bergstra, and S. Smolka. “Axiomatizing Probabilistic Processes: ACP with Generative Probabilities”. In: *Journal of Information and Computation*. Ed. by W. R. Cleaveland. Vol. 121. LNCS 2. Springer, 1995, pp. 234–255.
- [BCS07a] H. Boudali, P. Crouzen, and M. Stoelinga. “A Compositional Semantics for Dynamic Fault Trees in Terms of Interactive Markov Chains”. In: *ATVA 2007*. Vol. 4762. 642. Springer, 2007, p. 441.
- [BCS07b] H. Boudali, P. Crouzen, and M. Stoelinga. “Dynamic Fault Tree Analysis Using Input/Output Interactive Markov Chains”. In: *DSN 2007*. IEEE, 2007, pp. 708–717.
- [BCS10] H. Boudali, P. Crouzen, and M. Stoelinga. “A Rigorous, Compositional, and Extensible Framework for Dynamic Fault Tree Analysis”. In: *IEEE Transactions on Dependable and Secure Computing* 7.2 (2010), pp. 128–143.
- [BDL13] M. Bernardo, R. De Nicola, and M. Loreti. “A Uniform Framework for Modeling Nondeterministic, Probabilistic, Stochastic, or Mixed Processes and their Behavioral Equivalences”. In: *Information and Computation* (2013).
- [BEM00] C. Baier, B. Engelen, and M. Majster-Cederbaum. “Deciding Bisimilarity and Similarity for Probabilistic Processes”. In: *Journal of Computer and System Science* 60.1 (2000), pp. 187–231.
- [BG98] M. Bernardo and R. Gorrieri. “A Tutorial on EMPA: A Theory of Concurrent Processes with Nondeterminism, Priorities, Probabilities and Time”. In: *Theoretical Computer Science* 202 (1998).
- [BH97] C. Baier and H. Hermanns. “Weak Bisimulation for Fully Probabilistic Processes”. In: *CAV 1997*. Vol. 1254. LNCS. 1997, pp. 119–130.
- [BIM95] B. Bloom, S. Istrail, and A. R. Meyer. “Bisimulation Can’t Be Traced”. In: *Journal of the ACM* 42.1 (1995), pp. 232–268.
- [BK00] C. Baier and M. Kwiatkowska. “Domain Equations for Probabilistic Processes”. In: *Mathematical Structures in Comp. Sci.* 10 (2000).
- [BK08] C. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.
- [BK97] C. Baier and M. Kwiatkowska. “Domain Equations for Probabilistic Processes”. In: *EXPRESS 1997*. Elsevier, 1997.
- [BN98] F. Baader and T. Nipkow. *Term Rewriting and All That*. New York, NY, USA: Cambridge University Press, 1998.
- [Böd+09] E. Böde, M. Herbstritt, H. Hermanns, S. Johr, T. Peikenkamp, R. Pulungan, J. Rakow, R. Wimmer, and B. Becker. “Compositional Dependability Evaluation for STATEMATE”. In: *IEEE Transactions on Software Engineering* 35.2 SPEC. ISS. (2009), pp. 274–292.
- [Bou+08] H. Boudali, P. Crouzen, B. R. Haverkort, M. Kuntz, and M. Stoelinga. “Architectural Dependability Evaluation with Arcade”. In: *DSN 2008*. LNCS. Springer, 2008, pp. 512–521.
- [Boz+09a] M. Bozzano, A. Cimatti, J.-P. Katoen, V. Y. Nguyen, T. Noll, and M. Roveri. “The COMPASS Approach: Correctness, Modelling and Performability of Aerospace Systems”. In: *Computer Safety, Reliability, and Security*. Springer, 2009, pp. 173–186.

- [Boz+09b] M. Bozzano, A. Cimatti, M. Roveri, J.-P. Katoen, V. Y. Nguyen, and T. Noll. “Codesign of Dependable Systems: A Component-based Modeling Language”. In: *MEMOCODE 2009*. IEEE, 2009, pp. 121–130.
- [Boz+09c] M. Bozzano, A. Cimatti, M. Roveri, J.-P. Katoen, V. Y. Nguyen, and T. Noll. “Verification and Performance Evaluation of AADL Models”. In: *ESEC/SIGSOFT FSE*. ACM, 2009.
- [Boz+11] M. Bozzano, A. Cimatti, J.-P. Katoen, V. Y. Nguyen, T. Noll, and M. Roveri. “Safety, Dependability and Performance Analysis of Extended AADL Models.” In: *Computer Journal* 54.21171 (2011).
- [BPS01] J. A. Bergstra, A. Ponse, and S. A. Smolka. *Handbook of Process Algebra*. Elsevier, 2001.
- [Bra02] M. Bravetti. “Revisiting Interactive Markov Chains”. In: *MTCS 2002*. Vol. 68. ENTCS. 2002, pp. 65–84.
- [Bra08] M. Bravetti. *Stochastic Semantics in the Presence of Structural Congruence: Reduction Semantics for Stochastic Pi-Calculus*. Tech. rep. University of Bologna (Italy). Department of Computer Science, 2008.
- [Bra16] B. Braitling. “Spielbasierte Abstraktion von Markow-Automaten”. PhD thesis. Universität Freiburg, 2016.
- [Bre13] M. Brengel. “Probabilistic Weak Transitions”. Bachelor’s thesis. Saarland University, 2013.
- [BS00] C. Baier and M. Stoelinga. “Norm Functions for Probabilistic Bisimulations with Delays”. In: *FoSSaCS 2000*. Vol. 1784. LNCS. 2000, pp. 1–16.
- [BS01] E. Bandini and R. Segala. “Axiomatizations for Probabilistic Bisimulation”. In: *ICALP 2001*. Vol. 2076. LNCS. 2001, pp. 370–381.
- [BW90] J. C. Baeten and W. P. Weijland. *Process Algebra*. Vol. 18. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1990.
- [BWH17] Y. Butkova, R. Wimmer, and H. Hermanns. “Long-run Rewards for Markov Automata”. In: *TACAS 2017*. LNCS. to appear. 2017.
- [CC92] L. Christoff and I. Christoff. “Efficient Algorithms for Verification of Equivalences for Probabilistic Processes”. In: *CAV 1992*. Springer, 1992, pp. 310–321.
- [CH11] K. Chatterjee and M. Henzinger. “Faster and Dynamic Algorithms for Maximal End-Component Decomposition and Related Graph Problems in Probabilistic Verification”. In: *SODA 2011*. 2011, pp. 1318–1336.
- [Che+96] G. Chehaibar, H. Garavel, L. Mounier, N. Tawbi, and F. Zulian. “Specification and Verification of the PowerScale Bus Arbitration Protocol: An Industrial Experiment with LOTOS”. In: *FORTE 1996*. 1996, pp. 435–450.
- [Chi+93] G. Chiola, M. A. Marsan, G. Balbo, and G. Conte. “Generalized Stochastic Petri Nets: A Definition at the Net Level and its Implications”. In: *IEEE TSE* 19.2 (1993), pp. 89–107.
- [Chr90] I. Christoff. “Testing Equivalences and Fully Abstract Models for Probabilistic Processes”. In: *CONCUR 1990*. Springer, 1990, pp. 126–138.



- 
- [CL11] P. Crouzen and F. Lang. “Smart Reduction”. In: *FASE 2011*. Vol. 6603. 2011, pp. 111–126.
  - [Cla+07] A. Clark, S. Gilmore, J. Hillston, and M. Tribastone. “Stochastic Process Algebras”. In: *Formal Methods for Performance Evaluation*. Springer, 2007, pp. 132–179.
  - [Cos+08] N. Coste, H. Garavel, H. Hermanns, R. Hersemeule, Y. Thonnart, and M. Zidouni. “Quantitative Evaluation in Embedded System Design: Validation of Multiprocessor Multithreaded Architectures”. In: *Design, Automation and Test in Europe (YEAR)*. IEEE, 2008, pp. 88–89.
  - [Cos+09] N. Coste, H. Hermanns, E. Lantreibe, and W. Serwe. “Towards Performance Prediction of Compositional Models in Industrial GALS Designs”. In: *CAV 2009*. Vol. 5643. LNCS. Springer, 2009, pp. 204–218.
  - [CS02] S. Cattani and R. Segala. “Decision Algorithms for Probabilistic Bisimulation”. In: *CONCUR 2002*. LNCS. Springer, 2002, pp. 371–386.
  - [CSV07] L. Cheung, M. Stoelinga, and F. W. Vaandrager. “A Testing Scenario for Probabilistic Processes”. In: *Journal of the ACM* 54.6 (2007).
  - [CSZ92] R. Cleaveland, S. A. Smolka, and A. E. Zwarico. “Testing Preorders for Probabilistic Processes”. In: *ICALP 1992*. Vol. 154. LNCS. Springer, 1992, pp. 93–148.
  - [CW87] D. Coppersmith and S. Winograd. “Matrix Multiplication via Arithmetic Progressions”. In: *STOC 1987*. Vol. 9. ACM, 1987, pp. 251–280.
  - [CZ96] G. Ciardo and R. Zijal. “Well-Defined Stochastic Petri Nets”. In: *MASCOTS 1996*. 1996.
  - [CZM09] G. Chehaibar, M. Zidouni, and R. Mateescu. “Modeling Multiprocessor Cache Protocol Impact on MPI Performance”. In: *IEEE QuEST*. 2009.
  - [DD07] Y. Deng and W. Du. “Probabilistic Barbed Congruence”. In: *Electr. Notes Theor. Comput. Sci.* 190 (2007).
  - [De 99] L. De Alfaro. *The Verification of Probabilistic Systems under Memoryless Partial-Information Policies is Hard*. Tech. rep. DTIC Document, 1999, pp. 19–32.
  - [Den+07] Y. Deng, R. van Glabbeek, M. Hennessy, C. Morgan, and C. Zhang. “Remarks on Testing Probabilistic Processes”. In: *Electronic Notes in Theoretical Computer Science* 172 (2007). Computation, Meaning, and Logic: Articles dedicated to Gordon Plotkin, pp. 359–397.
  - [Den+08] Y. Deng, R. van Glabbeek, M. Hennessy, and C. Morgan. “Characterising Testing Preorders for Finite Probabilistic Processes”. In: *Logical Methods in Computer Science* 4.4 (2008), pp. 1–33.
  - [Den+09] Y. Deng, R. J. v. Glabbeek, M. Hennessy, and C. Morgan. “Testing Finitary Probabilistic Processes”. In: *CONCUR 2009*. 2009, pp. 274–288.
  - [Den05] Y. Deng. “Axiomatisations and Types for Probabilistic and Mobile Processes”. PhD thesis. École des Mines de Paris, 2005.
  - [Des+10] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. “Weak Bisimulation is Sound and Complete for PCTL<sup>\*</sup>”. In: *Inf. Comput.* 208.2 (2010), pp. 203–219.
  - [Des+99] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. “Metrics for Labeled Markov Systems”. In: *CONCUR 1999*. LNCS. 1999, pp. 258–273.

- [DH11] Y. Deng and M. Hennessy. “On the Semantics of Markov Automata”. In: *ICALP 2011*. 2011, pp. 307–318.
- [DH12] Y. Deng and M. Hennessy. “On the Semantics of Markov Automata”. In: *Information and Computation* 222 (2012), pp. 139–168.
- [DHR08] L. Doyen, T. A. Henzinger, and J.-F. Raskin. “Equivalence of Labeled Markov Chains”. In: *International Journal of Foundations of Computer Science* 19.03 (2008), pp. 549–563.
- [DP05] Y. Deng and C. Palamidessi. “Axiomatizations for Probabilistic Finite-State Behaviors”. In: *FoSSaCS 2005*. Vol. 373. LNCS 1-2. Springer Berlin Heidelberg, 2005, pp. 110–124.
- [DP07] Y. Deng and C. Palamidessi. “Axiomatizations for Probabilistic Finite-State Behaviors”. In: *Theoretical Computer Science* 373.1-2 (2007), pp. 92–114.
- [DPP05] Y. Deng, C. Palamidessi, and J. Pang. “Compositional Reasoning for Probabilistic Finite-State Behaviors”. In: *Processes, Terms and Cycles: Steps on the Road to Infinity*. Springer, 2005, pp. 309–337.
- [DS99] D. D. Deavours and W. H. Sanders. “An Efficient Well-Specified Check”. In: *Petri Nets and Performance Models, IEEE International Workshop on* (1999).
- [EHZ10a] C. Eisentraut, H. Hermanns, and L. Zhang. “On Probabilistic Automata in Continuous Time”. In: *LICS 2010*. IEEE, 2010, pp. 342–351.
- [EHZ10b] C. Eisentraut, H. Hermanns, and L. Zhang. “Concurrency and Composition in a Stochastic World”. In: *CONCUR 2010*. Vol. 6269 LNCS. LNCS. Springer, 2010, pp. 21–39.
- [EHZ10c] C. Eisentraut, H. Hermanns, and L. Zhang. *On Probabilistic Automata in Continuous Time*. Reports of SFB/TR 14 AVACS 62. ISSN: 1860-9821, <http://www.avacs.org/SFB/TR>, Aug. 2010.
- [Eis+ 13a] C. Eisentraut, H. Hermanns, J.-P. Katoen, and L. Zhang. “A Semantics for Every GSPN”. In: *Application and Theory of Petri Nets and Concurrency*. Springer, 2013, pp. 90–109.
- [Eis+ 13b] C. Eisentraut, H. Hermanns, J. Krämer, A. Turrini, and L. Zhang. “Deciding Bisimilarities on Distributions”. In: *QEST 2013*. Vol. 8054 LNCS. Springer, 2013, pp. 72–88.
- [Eis+ 13c] C. Eisentraut, H. Hermanns, J. Schuster, A. Turrini, and L. Zhang. “The Quest for Minimal Quotients for Probabilistic Automata”. In: *TACAS 2013*. LNCS. Springer, 2013, pp. 16–31.
- [Eis+ 15] C. Eisentraut, J. C. Godskesen, H. Hermanns, L. Song, and L. Zhang. “Probabilistic Bisimulation for Realistic Schedulers”. In: *FM 2015*. 2015, pp. 248–264.
- [Eis07] C. Eisentraut. “Complete Completeness for Weak Bisimulation Semantics”. MA thesis. Saarland University, 2007.
- [EP03] G. E. Teruel and M. D. Pierro. “Well-Defined Generalized Stochastic Petri Nets: A Net-Level Method to Specify Priorities”. In: *IEEE Transactions on Software Engineering* 29.11 (2003), pp. 962–973.

- 
- [Est + 12] M.-A. Esteve, J.-P. Katoen, V. Y. Nguyen, B. Postma, and Y. Yushtein. “Formal Correctness, Safety, Dependability and Performance Analysis of a Satellite”. In: *ICSE 2012*. 2012, pp. 1022–1031.
  - [Fer90] J.-C. Fernandez. “An Implementation of an Efficient Algorithm for Bisimulation Equivalence”. In: *Science of Computer Programming* 13 (1990), pp. 219–236.
  - [FG12] P. H. Feiler and D. P. Gluch. *Model-Based Engineering with AADL: An Introduction to the SAE Architecture Analysis & Design Language*. Addison-Wesley, 2012.
  - [FM91] J.-C. Fernandez and L. Mounier. “A Tool Set for Deciding Behavioral Equivalences”. In: *CONCUR 1991*. Vol. 527. Springer, 1991, pp. 23–42.
  - [FZ14] Y. Feng and L. Zhang. “When Equivalence and Bisimulation Join Forces in Probabilistic Automata”. In: *FM 2014*. Vol. 8442. 2014, pp. 247–262.
  - [GA12] S. Georgievska and S. Andova. “Probabilistic May/Must Testing: Retaining Probabilities by Restricted Schedulers”. In: *Formal Aspects of Computing* 24.4-6 (2012), pp. 727–748.
  - [GD07] S. Giro and P. R. D’Argenio. “Quantitative Model Checking Revisited: neither Decidable nor Approximable”. In: *FORMATS 2007*. LNCS. Springer, 2007, pp. 179–194.
  - [GHR92] N. Götz, U. Herzog, and M. Rettelbach. *TIPP—a Language for Timed Processes and Performance Evaluation*. Tech. rep. University of Erlangen-Nürnberg, 1992.
  - [Gla + 90] R. J. van Glabbeek, S. A. Smolka, B. Steffen, and C. M. N. Tofts. “Reactive, Generative, and Stratified Models of Probabilistic Processes”. In: *LICS 1990*. IEEE, 1990, pp. 130–141.
  - [Gla90] R. J. van Glabbeek. “The Linear Time-Branching Time Spectrum”. In: *CONCUR 1990*. Springer Berlin Heidelberg, 1990, pp. 278–297.
  - [Gla93] R. J. v. Glabbeek. “The Linear Time - Branching Time Spectrum II”. In: *CONCUR 1993*. Springer Berlin Heidelberg, 1993, pp. 66–81.
  - [Göt94] N. Götz. “Stochastische Prozessalgebren – Integration von funktionalem Entwurf und Leistungsbewertung verteilter Systeme”. PhD thesis. Universität Erlangen, 1994.
  - [GSS95] R. J. van Glabbeek, S. A. Smolka, and B. Steffen. “Reactive, Generative, and Stratified Models of Probabilistic Processes”. In: *Information and Computation* 121.1 (1995), pp. 59–80.
  - [GTB14] D. Guck, M. Timmer, and S. C. C. Blom. “Extending Markov Automata with State and Action Rewards”. In: *QAPL 2014*. Ed. by N. Bertrand and L. Bortolussi. 12238. Inria, Apr. 2014.
  - [Guc + 13] D. Guck, H. Hatefi, H. Hermanns, J.-P. Katoen, and M. Timmer. “Modelling, Reduction and Analysis of Markov Automata”. In: *Quantitative Evaluation of Systems*. Vol. 8054. LNCS. Springer, 2013, pp. 55–71.
  - [Guc + 14] D. Guck, M. Timmer, E. Ruijters, H. Hatefi, and M. Stoelinga. “Modelling and Analysis of Markov Reward Automata”. In: *ATVA 2014*. Vol. 8837. LNCS. Springer, 2014, pp. 168–184.

- [GW96] R. J. van Glabbeek and W. P. Weijland. “Branching Time and Abstraction in Bisimulation Semantics”. In: *Journal of the ACM* 43.3 (1996), pp. 555–600.
- [Hah+14] E. M. Hahn, Y. Li, S. Schewe, A. Turrini, and L. Zhang. “iscasMc: A Web-Based Probabilistic Model Checker”. In: *FM 2014*. Vol. 8442. 2014, pp. 312–317.
- [Han91] H. A. Hansson. “Time and Probability in Formal Design of Distributed Systems”. PhD thesis. Department of Computer Systems, Uppsala University, 1991.
- [Har02] J. den Hartog. “Probabilistic Extensions of Semantical Models”. PhD thesis. Uppsala University, Department of Computer Science, 2002.
- [Har99] J. I. den Hartog. “Verifying Probabilistic Programs Using a Hoare Like Logic”. In: *ASIAN 1999*. Springer, 1999.
- [Hat+15] H. Hatefi, B. Braitling, R. Wimmer, L. M. F. Fioriti, H. Hermanns, and B. Becker. “Cost vs. Time in Stochastic Games and Markov Automata”. In: *SETTA 2015*. Ed. by X. Li, Z. Liu, and W. Yi. Vol. 9409. LNCS. Springer, 2015, pp. 19–34.
- [Hat16] H. Hatefi Ardakani. “Finite Horizon Analysis of Markov Automata”. PhD thesis. Universität des Saarlandes, 2016.
- [Hav+10] B. R. Haverkort, M. Kuntz, A. Remke, S. Roolvink, and M. Stoelinga. “Evaluating Repair Strategies for a Water-Treatment Facility using Arcade”. In: *Dependable Systems and Networks*. 2010, pp. 419–424.
- [Hen12] M. Hennessy. “Exploring Probabilistic Bisimulations, Part I”. In: *Formal Asp. Comput.* 24.4-6 (2012), pp. 749–768.
- [Her02] H. Hermanns. *Interactive Markov Chains: The Quest for Quantified Quality*. Vol. 2428. LNCS. Springer, 2002.
- [HH12] H. Hatefi and H. Hermanns. “Model Checking Algorithms for Markov automata”. In: *Electronic Communications of the EASST* 53 (2012).
- [HHK02] H. Hermanns, U. Herzog, and J.-P. Katoen. “Process Algebra for Performance Evaluation”. In: *Theoretical Computer Science* 274 (2002), pp. 43–87.
- [Hil96] J. Hillston. *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.
- [HJ08] H. Hermanns and S. Johr. “May we reach it? Or must we? In what time? With what probability?” In: *MMB 2008*. VDE. 2008, pp. 125–140.
- [HJ94] H. Hansson and B. Jonsson. “A Logic for Reasoning about Time and Reliability”. In: *Formal Aspects of Computing* 6 (1994).
- [HK00] H. Hermanns and J.-P. Katoen. “Automated Compositional Markov Chain Generation for a Plain-old Telephone System”. In: *Science of Computer Programming* 36.1 (2000), pp. 97–127.
- [HK10] H. Hermanns and J.-P. Katoen. “The How and Why of Interactive Markov Chains”. In: *FMCO*. Vol. 6286. LNCS. 2010, pp. 311–337.
- [Hoa85] C. A. R. Hoare. *Communicating Sequential Processes*. Prentice Hall, 1985.
- [HT12] H. Hermanns and A. Turrini. “Deciding Probabilistic Automata Weak Bisimulation in Polynomial Time”. In: *FSTTCS 2012*. Vol. 18. 2012, pp. 435–447.

- 
- [HWZ08] H. Hermanns, B. Wachter, and L. Zhang. “Probabilistic CEGAR”. In: *CAV 2008*. Vol. 5123. LNCS. Springer, 2008, pp. 162–175.
  - [JL91] B. Jonsson and K. G. Larsen. “Specification and Refinement of Probabilistic Processes”. In: *LICS 1991*. IEEE, 1991, pp. 266–277.
  - [JS90] C.-C. Jou and S. A. Smolka. “Equivalences, Congruences, and Complete Axiomatizations for Probabilistic Processes”. In: *CONCUR 1990*. Vol. 458. LNCS. 1990, pp. 367–383.
  - [JY02] B. Jonsson and W. Yi. “Testing preorders for probabilistic processes can be characterized by simulations”. In: *Theoretical Computer Science* 282.1 (2002). Real-Time and Probabilistic Systems, pp. 33–51.
  - [Kat+07] J.-P. Katoen, T. Kemna, I. S. Zapreev, and D. N. Jansen. “Bisimulation Minimisation Mostly Speeds Up Probabilistic Model Checking”. In: *TACAS 2007*. Vol. 4424. LNCS. 2007, pp. 87–101.
  - [Kat12] J. Katoen. “GSPNs Revisited: Simple Semantics and New Analysis Algorithms”. In: *ACSD 2012*. 2012, pp. 6–11.
  - [Kat13] J.-P. Katoen. “Concurrency Meets Probability: Theory and Practice”. In: *CONCUR 2013*. LNCS. Springer, 2013.
  - [KK15] J. Král and P. Král. “Scalable Analysis of Fault Trees with Dynamic Features”. In: *DSN 2015*. IEEE, 2015, pp. 89–100.
  - [KN98] M. Kwiatkowska and G. Norman. “A Testing Equivalence for Reactive Probabilistic Processes”. In: *EXPRESS 1998*. Vol. 16(2). Electronic Notes in Theoretical Computer Science. Elsevier Science, 1998.
  - [KNP11] M. Z. Kwiatkowska, G. Norman, and D. Parker. “PRISM 4.0: Verification of Probabilistic Real-Time Systems”. In: *CAV 2011*. Vol. 6806. 2011, pp. 585–591.
  - [KS83] P. C. Kanellakis and S. A. Smolka. “CCS Expressions, Finite State Processes, and Three Problems of Equivalence”. In: *PODC 1983*. 1983, pp. 228–240.
  - [KS90] P. C. Kanellakis and S. A. Smolka. “CCS Expressions, Finite State Processes, and Three Problems of Equivalence”. In: *Information and Computation* 86.1 (1990), pp. 43–68.
  - [LDH05] M. Lohrey, P. R. D’Argenio, and H. Hermanns. “Axiomatising Divergence”. In: *Information and Computation* 203.2 (2005), pp. 115–144.
  - [LS89] K. G. Larsen and A. Skou. “Bisimulation Through Probabilistic Testing”. In: *POPL 1989*. 1989, pp. 344–352.
  - [LS91] K. G. Larsen and A. Skou. “Bisimulation through Probabilistic Testing”. In: *Information and Computation* 94.1 (1991), pp. 1–28.
  - [LS92] K. G. Larsen and A. Skou. “Compositional Verification of Probabilistic Processes”. In: *CONCUR 1992*. 1992, pp. 456–471.
  - [LSV07] N. A. Lynch, R. Segala, and F. W. Vaandrager. “Observing Branching Structure through Probabilistic Contexts”. In: *SIAM Journal on Computing* 37.4 (2007), pp. 977–1013.

- [Mar+87] M. A. Marsan, G. Balbo, G. Chiola, and G. Conte. “Generalized Stochastic Petri Nets Revisited: Random Switches and Priorities.” In: *PNPM*. IEEE, 1987, pp. 44–53.
- [Mar+91] M. A. Marsan, G. Balbo, G. Chiola, G. Conte, S. Donatelli, and G. Franceschinis. “An Introduction to Generalized Stochastic Petri Nets.” In: *Microel. \ and Rel.* 31.4 (1991), pp. 699–725.
- [Mar+94] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, 1994.
- [MCB84] M. A. Marsan, G. Conte, and G. Balbo. “A Class of Generalized Stochastic Petri Nets for the Performance Evaluation of Multiprocessor Systems”. In: *ACM Transactions on Computer Systems (TOCS)* 2.2 (1984), pp. 93–122.
- [Mil81] R. Milner. “A Modal Characterisation of Observable Machine-Behaviour”. In: *CAAP 1981*. Vol. 112. Springer, 1981, pp. 25–34.
- [Mil82] R. Milner. *A Calculus of Communicating Systems*. Springer, 1982.
- [Mil89a] R. Milner. “A Complete Axiomatisation for Observational Congruence of Finite-State Behaviours”. In: *Information and Computation* 81 (1989).
- [Mil89b] R. Milner. *Communication and Concurrency*. Prentice Hall, 1989.
- [ML12] E. Matsikoudis and E. A. Lee. “From Transitions to Executions”. In: *Coalgebraic Methods in Computer Science*. Springer, 2012, pp. 170–190.
- [MMS85] J. F. Meyer, A. Movaghar, and W. H. Sanders. “Stochastic Activity Networks: Structure, Behavior, and Application”. In: *PNPM 1085*. IEEE, 1985, pp. 106–115.
- [MS92a] R. Milner and D. Sangiorgi. “Barbed Bisimulation”. In: *ICALP 1992*. Vol. 623. LNCS. Springer, 1992, pp. 685–695.
- [MS92b] R. Milner and D. Sangiorgi. “Barbed Bisimulation”. In: *Automata, Languages and Programming*. Ed. by W. Kuich. Vol. 623. LNCS. Springer, 1992, pp. 685–695.
- [Neu10] M. R. Neuhäusser. “Model Checking Nondeterministic and Randomly Timed Systems”. PhD thesis. RWTH Aachen University / University of Twente, 2010.
- [Nic+13] R. de Nicola, D. Latella, M. Loretì, and M. Massink. “A Uniform Definition of Stochastic Process Calculi”. In: *ACM Computing Surveys (CSUR)* 46.1 (2013), pp. 1–35.
- [NV07] S. Nain and M. Y. Vardi. “Branching vs. Linear Time: Semantical Perspective”. In: *ATVA 2007*. LNCS. Springer, 2007, pp. 19–34.
- [NV09] S. Nain and M. Y. Vardi. “Trace Semantics is Fully Abstract”. In: *LICS 2009*. IEEE, 2009, pp. 59–68.
- [Par81] D. Park. “Concurrency and Automata on Infinite Sequences”. In: *Proceedings of the 5th GI-Conference on Theoretical Computer Science*. Vol. 104. LNCS. 1981, pp. 167–183.
- [Phi87] I. Phillips. “Refusal testing”. In: *Theoretical Computer Science* 50.3 (1987), pp. 241–284.
- [PLS00] A. Philippou, I. Lee, and O. Sokolsky. “Weak Bisimulation for Probabilistic Systems”. In: *CONCUR 2000*. Vol. 1877. LNCS. 2000, pp. 334–349.

- 
- [PS04] A. Parma and R. Segala. “Axiomatization of Trace Semantics for Stochastic Nondeterministic Processes”. In: *QEST 2004*. 2004, pp. 294–303.
  - [PT87] R. Paige and R. E. Tarjan. “Three Partition Refinement Algorithms”. In: *SIAM Journal on Computing* 16.6 (1987), pp. 973–989.
  - [San95] D. Sangiorgi. “Expressing Mobility in Process Algebras: First-Order and Higher-Order Paradigms”. PhD thesis. Department of Computer Science, University of Edinburgh, 1995.
  - [Seg06] R. Segala. “Probability and Nondeterminism in Operational Models of Concurrency”. In: *CONCUR 2006*. Vol. 4137. 2006, pp. 64–78.
  - [Seg95] R. Segala. “Modeling and Verification of Randomized Distributed Real-Time Systems”. PhD thesis. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1995.
  - [Seg97] R. Segala. “Compositional Verification of Randomized Distributed Algorithms”. In: *COMPOS 1997*. Vol. 1536. LNCS. Springer, 1997, pp. 515–540.
  - [SL94] R. Segala and N. A. Lynch. “Probabilistic Simulations for Probabilistic Processes”. In: *CONCUR 1994*. Vol. 836. LNCS. 1994, pp. 481–496.
  - [SL95] R. Segala and N. A. Lynch. “Probabilistic Simulations for Probabilistic Processes”. In: *Nordic Journal of Computing* 2.2 (1995), pp. 250–273.
  - [SM] W. H. Sanders and J. F. Meyer. “Stochastic Activity Networks: Formal Definitions and Concepts”. In: *FMPA 2000*. Vol. 315–343. LNCS 9975019. Springer, pp. 315–343.
  - [SS14] J. Schuster and M. Siegle. “Markov Automata: Deciding Weak Bisimulation by Means Of Non-Naively Vanishing States”. In: *Information and Computation* 237 (2014), pp. 151–173.
  - [Sto02a] M. I. A. Stoelinga. “An Introduction to Probabilistic Automata”. In: *Bulletin of the EATCS* 78 (2002).
  - [Sto02b] M. Stoelinga. “Alea Jacta Est: Verification of Probabilistic, Real-Time and Parametric Systems”. PhD thesis. University of Nijmegen, the Netherlands, 2002.
  - [SV03] M. Stoelinga and F. Vaandrager. “A Testing Scenario for Probabilistic Automata”. In: *ICALP 2003*. LNCS. Springer, 2003.
  - [SV04] A. Sokolova and E. Vink. “Probabilistic Automata: System Types, Parallel Composition and Comparison”. In: *Validation of Stochastic Systems*. Ed. by C. Baier, B. Haverkort, H. Hermanns, J.-P. Katoen, and M. Siegle. LNCS. Springer, 2004.
  - [SV99] M. Stoelinga and F. W. Vaandrager. “Root Contention in IEEE 1394”. In: *AMAST Workshop on Formal Methods for Real-Time and Probabilistic Systems*. Vol. 1601. LNCS. Springer, 1999, pp. 53–74.
  - [Tim+12] M. Timmer, J.-P. Katoen, J. van de Pol, and M. Stoelinga. “Efficient Modelling and Generation of Markov Automata”. In: *CONCUR 2012*. Ed. by M. Koutny and I. Ulidowski. LNCS CTIT Ph.D. Thesis Series No. 13-261. Springer, 2012.
  - [Tim13] M. Timmer. “Efficient Modelling, Generation and Analysis Of Markov Automata”. PhD thesis. Enschede: University of Twente, Sept. 2013.

- [TSP11] M. Timmer, M. Stoelinga, and J. van de Pol. “Confluence Reduction for Probabilistic Systems”. In: *TACAS 2011*. Ed. by P. Abdulla and K. Leino. Vol. 6605 LNCS. LNCS. Springer, 2011, pp. 311–325.
- [Van94] R. J. Van Glabbeek. “What is Branching Time Semantics and Why to Use It?” In: *Bulletin of the EATCS*. Vol. 53. 1994, pp. 191–198.
- [Var01] M. Vardi. “Branching vs. Linear Time: Final Showdown”. In: *TACAS 2001*. LNCS. Springer, 2001, pp. 1–22.
- [Wil12] V. V. Williams. “Multiplying Matrices Faster Than Coppersmith-Winograd”. In: *STOC 2012*. ACM, 2012.
- [WMB05] V. Wolf, M. Majster-Cederbaum, and C. Baier. “Trace Machines for Observing Continuous-Time Markov Chains”. In: *QAPL 2005*. 2005.
- [WZH07] B. Wachter, L. Zhang, and H. Hermanns. “Probabilistic Model Checking Modulo Theories”. In: *QEST 2007*. IEEE. 2007, pp. 129–138.
- [ZN10] L. Zhang and M. R. Neuhäusser. “Model Checking Interactive Markov Chains”. In: *TACAS 2010*. LNCS. Springer, 2010, pp. 53–68.