Compositional Reasoning for Markov Decision Processes
(Extended Abstract)

Yuxin Deng[1,2] ⋆ and Matthew Hennessy[3] ⋆⋆

[1] Dept. Comp. Sci. & Eng. and MOE-Microsoft Key Lab for Intell. Comp. & Syst., Shanghai Jiao Tong University, China
[2] State Key Lab of Comp. Sci., Inst. of Software, Chinese Academy of Sciences
[3] Trinity College Dublin, Ireland

Abstract. Markov decision processes (MDPs) have long been used to model qualitative aspects of systems in the presence of uncertainty. However, much of the literature on MDPs takes a monolithic approach, by modelling a system as a particular MDP; properties of the system are then inferred by analysis of that particular MDP. In this paper we develop compositional methods for reasoning about the qualitative behaviour of MDPs. We consider a class of labelled MDPs called weighted MDPs from a process algebraic point of view. For these we define a coinductive simulation-based behavioural preorder which is compositional in the sense that it is preserved by structural operators for constructing MDPs from components.

For finitary convergent processes, which are finite-state and finitely branching systems without divergence, we provide two characterisations of the behavioural preorder. The first uses a novel qualitative probabilistic logic, while the second is in terms of a novel form of testing, in which benefits are accrued during the execution of tests.

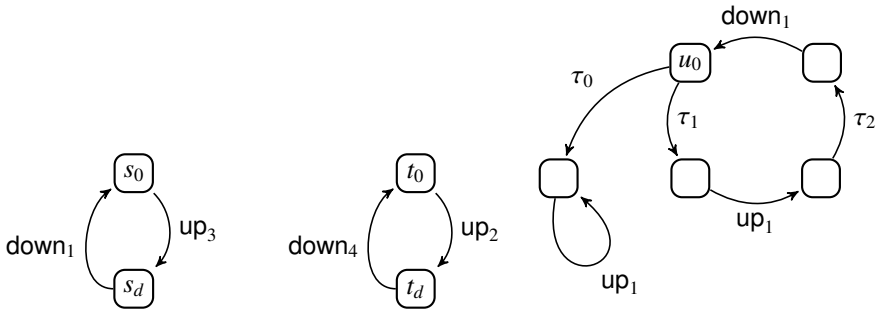
1 Introduction

Markov decision processes (MDPs) have long been used to model qualitative aspects of systems in the presence of uncertainty [12, 13, 1]. A comprehensive account of analysis techniques may be found in [12], while [13] provides a good account of *model-checking*.

However, much of the literature on MDPs takes a monolithic view of systems; essentially a system is modelled using a particular MDP, and properties of the system are then inferred by analysis of that MDP. In this paper, instead, we would like to develop compositional methods for reasoning about qualitative behaviour of Markov decision processes. This involves defining an appropriate method for comparing the behaviour MDPs which is susceptible to compositional analysis; the behaviour of a composite system should be determined by that of its components.

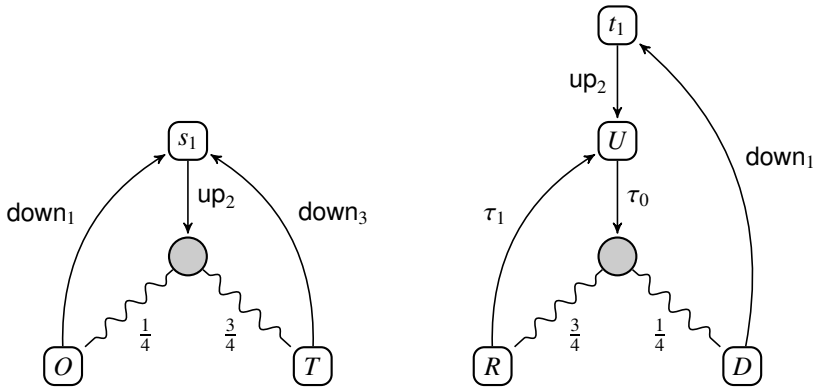
Our starting point is the idea of one system being able to *simulate* another. For example consider the following three systems:

⋆ Supported by the National Natural Science Foundation of China (61033002, 61011140074).
⋆⋆ Supported by SFI project SFI 06 IN.1 1898.



The first, a two-state machine, continually performs an up action, which accrues a benefit of 3 units, followed by a down action, which accrues a benefit of 1. The second machine performs the same actions but with benefits 2 and 4 respectively. In some sense t_0 is an improvement on s_0; intuitively t_0 can simulate the behaviour of s_0 but in so doing accrue more benefits; this is true even if one of its actions up is less beneficial than the corresponding action of s_0. The same is true for the machine u_0; it can also simulate the behaviour of s_0, with more benefit, although in this case some internal weighted actions, denoted by τ, participate in the simulation and add to the accumulation of benefits. In our terminology we will write $s_0 \sqsubseteq_{sim} t_0$, $s_0 \sqsubseteq_{sim} u_0$. However, we will have $t_0 \not\sqsubseteq_{sim} u_0$ because although u_0 can simulate the behaviour of t_0 it accumulates less benefit.

Similar informal reasoning can also be applied to probabilistic systems. Consider the following systems:

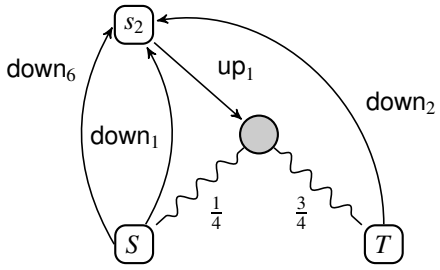


The first, from state s_1, can perform the up action with benefit 2 and a quarter of the time it ends up in a state in which down can be performed with benefit only 1. But for the remaining three-quarters it ends up in a state in which down can be performed for the larger benefit 3. The circular darkended node represents a distribution of states, with its outgoing edges describing the associated probabilities. Again intuitively we can see that s_1 is an improvement on s_0 because it can simulate s_0 and on average accrue slightly more benefits; in our theory we will have $s_0 \sqsubseteq_{sim} s_1$.

The mixture of probabilistic behaviour and internal actions introduces complications. Consider the system t_1 above which after performing an up probabilistically decides internally whether to perform a down action for benefit 1, or branch back to make

another probabilistic choice. However, each time it reverts back it accumulates a non-zero benefit via the internal weighted action τ_1, albeit with diminishing probability. Nevertheless, it will turn out that $s_0 \sqsubseteq_{sim} t_1$ and indeed $s_1 \sqsubseteq_{sim} t_1$.

Systems exhibiting both probabilistic and nondeterministic behaviour require more complicated analysis. Consider the following system:



After performing the action up it finds itself either in a state in which the action down will accrue the benefit 2, or 25% of the time there will be a nondeterministic choice between it accruing either 1 or 6. In the literature there are numerous mechanisms, such as policies, schedulers, adversaries, etc. [12, 14, 13] for resolving such choices. Here one can see if this choice systematically leads to the lower benefit 1 then s_2 will not simulate s_0 as it does not accrue sufficient benefits. This is a pessimistic outlook; an optimistic outlook means that the best choices are systematically made. If this is assumed then we will have $s_0 \sqsubseteq_{sim} s_2$; in s_2 one execution of up followed by down will yield on average the benefit $1 + (\frac{3}{4} \cdot 2 + \frac{1}{4} \cdot 6) = 4$.

The main contribution of the paper is a coinductively defined behavioural preorder $\sqsubseteq_{sim}$ between MDPs based on simulations which validate the examples discussed informally above. We confine our attention to the optimistic approach to the resolution of nondeterministic choices, although in a later paper we hope to investigate the pessimistic approach. We also show that this preorder is compositional in the sense that it is preserved by structural operators for constructing MDPs from components. The main operator is one for composing two MDPs in parallel. In $P \mid Q$ the two MPDs P and Q remain independent, execute in parallel and may communicate by synchronising on complementary actions; these internal synchronisations accrue the combined benefits of the associated complementary actions.

For finitary convergent MDPs, which are finite-state and finitely branching systems without divergence, we also provide two characterisations for the behavioural preorder $\sqsubseteq_{sim}$. The first is in terms of a *qualitative probabilistic logic* $\mathcal{L}_Q$. In addition to the standard logical connectives $\wedge$, $\vee$ and both maximal and minimal fixpoints this contains a novel qualitative *possibility* modality $\langle\alpha\rangle_w(\phi_1 \,{}_p\!\oplus\, \phi_2)$, where p is some probability between 0 and 1. Intuitively this is satisfied by an MDP which can accrue at least the benefit w by performing the action α, and subsequently satisfy the probabilistic assertion $\phi_1 \,{}_p\!\oplus\, \phi_2$. It turns out that the simulation preorder is completely determined by the logic $\mathcal{L}_Q$. Further evidence of the compatibility between the logic and the simulation relation is the fact that every system P has a *characteristic formula* $\phi(P)$ in the logic which captures its behaviour; informally system Q can simulate P if and only if it satisfies the characteristic formula $\phi(P)$.

Our second characterisation is in terms of a novel form of testing called *benefits* testing. Intuitively a system P can be tested by running it in parallel with another testing system T, and seeing the possible accrued benefits. In the presence of nondeterminism the execution of the combined system $(T \mid P)$ will result in a non-empty set of benefits, **Benefits**$(T \mid P)$. Then systems P and Q can be compared by comparing the associated benefit sets **Benefits**$(T \mid P)$ and **Benefits**$(T \mid Q)$ where T ranges over some collection of possible tests. We show that the simulation preorder $\sqsubseteq_{sim}$ is also determined in this manner by a suitable collection of tests T.

The rest of this paper is organised as follows. In Section 2 we introduce the model of weighted MDPs, the notation of hyper-derivations and some important properties. Then we define a behavioural preorder based on amortised weighted simulations, which turns out to be a precongruence in a CCS-like process calculus for MDPs. Next, we provide logical and testing characterisations of the behavioural preorder over finitary convergent processes. In Section 3 we present a qualitative probabilistic logic whose formulae completely determine the behavioural preorder. We also show that characteristic formulae can be constructed for any state in such an MDP. In Section 4 we propose a testing framework where our behavioural preorder is sound and complete for may testing preorder. Finally, we conclude in Section 5.

Due to lack of space, we omit all detailed proofs: they are reported in [4].

2 Simulations for weighted Markov decision processes

There is considerable variation in the literature in the formal definition of a (labelled) Markov decision process [13, 12]. For the purpose of this paper we use Definition 1.

We first fix some notation. A (discrete) probability subdistribution over a countable set S is a function $\Delta : S \to [0,1]$ with $\sum_{s\in S} \Delta(s) \leq 1$; the support of such an Δ is the set $\lceil \Delta \rceil = \{\, s \in S \mid \Delta(s) > 0 \,\}$. A subdistribution is a (total, or full) distribution if $\sum_{s\in S} \Delta(s) = 1$. The point distribution $\overline{s}$ assigns probability 1 to s and 0 to all other elements of S , so that $\lceil \overline{s} \rceil = s$. We use $\mathcal{D}_{sub}(S)$ to denote the set of subdistributions over S , and $\mathcal{D}(S)$ its subset of full distributions.

Let $\{\Delta_k \mid k \in K\}$ be a set of subdistributions, possibly infinite. Then $\sum_{k\in K} \Delta_k$ is the real-valued function in $S \to \mathbb{R}$ defined by $(\sum_{k\in K} \Delta_k)(s) := \sum_{k\in K} \Delta_k(s)$. This is a partial operation on subdistributions because for some state s the sum of $\Delta_k(s)$ might exceed 1. If the index set is finite, say $\{1..n\}$, we often write $\Delta_1 + \ldots + \Delta_n$. For p a real number from $[0,1]$ we use $p \cdot \Delta$ to denote the subdistribution given by $(p \cdot \Delta)(s) := p \cdot \Delta(s)$. Finally we use ε to denote the everywhere-zero subdistribution that thus has empty support. These operations on subdistributions do not readily adapt themselves to distributions; yet if $\sum_{k\in K} p_k = 1$ for some collection of $p_k \geq 0$, and the Δ_k are distributions, then so is $\sum_{k\in K} p_k \cdot \Delta_k$.

Definition 1 (Weighted Markov decision process). *A* weighted Markov decision process *or wMDP is a 4-tuple* $\langle S, \mathsf{A}, W, \longrightarrow \rangle$ *where* S *is a set of states,* A *a set of actions,* W *a set of weights, and* $\longrightarrow \subseteq S \times \mathsf{A} \times W \times \mathcal{D}(S)$*. We normally write* $s \xrightarrow{\alpha}_w \Delta$ *to mean* $(s, \alpha, w, \Delta) \in \longrightarrow$*. In this paper we set* W *to be* $\mathbb{R}_{\geq 0}$*, the set of non-negative real numbers, and we assume* A *has the structure* $\mathsf{Act}_\tau = \mathsf{Act} \cup \{\tau\}$ *where, for the purpose of communication, each* a *in* Act *has an inverse* $\overline{a}$ *satisfying* $\overline{\overline{a}} = a$*.* □

A wMDP is

- *finite-state* if S is a finite set;
- *finitely branching* if for each state s, the set $\{(\alpha, w, \Delta) \mid s \xrightarrow{\alpha}_w \Delta\}$ is finite;
- *finitary* if it is both finite-state and finitely branching.

In the Introduction we have used a straightforward graphical representation for wMDPs; a state s is represented by a node $\boxed{s}$ while darkened circular nodes are used for distributions, and arrows between nodes and distributions are annotated with their weights. Often a point distribution is represented by the unique state in its support; see the first series of examples with initial states s_0, t_0 and u_0.

2.1 Hyper-derivations

As we have seen in the Introduction, when reasoning informally that t_1 can simulate s_0, the limiting behaviour of internal computations must be taken into account. We formalise this by extending the approach originally given in [5].

In a wMDP actions are only performed by states, in that actions are given by relations from states to distributions. But formally, systems or processes in general correspond to distributions over states, so in order to define what it means for a process to perform an action, we need to *lift* these relations so that they also apply to distributions. In fact we will find it convenient to lift them to subdistributions.

Definition 2. *Let $\mathcal{R} \subseteq S \times (\mathbb{R}_{\geq 0} \times \mathcal{D}_{sub}(S))$ be a relation from states to pairs of weights and subdistributions. Then $\overline{\mathcal{R}} \subseteq \mathcal{D}_{sub}(S) \times (\mathbb{R}_{\geq 0} \times \mathcal{D}_{sub}(S))$ is the smallest relation that satisfies:*

(i) $s \,\mathcal{R}\, \langle r, \Theta \rangle$ *implies* $s \,\overline{\mathcal{R}}\, \langle r, \Theta \rangle$, *and*

(ii) (Linearity) $\Delta_i \,\overline{\mathcal{R}}\, \langle r_i, \Theta_i \rangle$ *for* $i \in I$ *implies* $(\sum_{i\in I} p_i \cdot \Delta_i) \;\overline{\mathcal{R}}\; (\sum_{i\in I} p_i \cdot \langle r_i, \Theta_i \rangle)$ *for any* $p_i \in [0,1]$ *(*$i \in I$*) with* $\sum_{i\in I} p_i \leq 1$. □

An application of Definition 2 to the arrow relation $\xrightarrow{\alpha}$ in a wMDP gives a relation $\overline{(\xrightarrow{\alpha})} \subseteq \mathcal{D}(S) \times (W \times \mathcal{D}(S))$; for convenience we also denote elements of this relation as $\Delta \xrightarrow{\alpha}_w \Theta$. Thus, as source of a relation $\xrightarrow{\alpha}$ we now also allow distributions, and even subdistributions.

Definition 3 (Hyper-derivations). *A hyper-derivation consists of a collection of subdistributions $\Delta, \Delta_k^{\rightarrow}, \Delta_k^{\times}$, for $k \geq 0$, with the following properties:*

$$\begin{aligned}
\Delta &= \Delta_0^{\rightarrow} + \Delta_0^{\times} \\
\Delta_0^{\rightarrow} &\xrightarrow{\tau}_{w_0} \Delta_1^{\rightarrow} + \Delta_1^{\times} \\
&\vdots \\
\Delta_k^{\rightarrow} &\xrightarrow{\tau}_{w_k} \Delta_{k+1}^{\rightarrow} + \Delta_{k+1}^{\times} \\
&\vdots \\
\Delta' &= \sum_{k=0}^{\infty} \Delta_k^{\times}
\end{aligned} \tag{1}$$

Then we call $\Delta' = \sum_{k=0}^{\infty} \Delta_k^{\times}$ *a* hyper-derivative *of* Δ*, and write* $\Delta \stackrel{\tau}{\Longrightarrow}_w \Delta'$*, where* $w = \sum_{k\geq 0} w_k$*. Note that in general* $w \in \mathbb{R}_{\geq 0} \cup \{\infty\}$*; that is there is no guarantee that the sum* $\sum_{k\geq 0} w_k$ *has a finite limit.* □

The reader is referred to [5] for a detailed discussion of the concept of hyper-derivation.

Example 1. Consider the wMDP with initial state t_1 discussed in the Introduction. Then we have the following hyper-derivation:

$$\begin{array}{rcl}
\overline{U} & = & \overline{U} + \varepsilon \\
\overline{U} & \stackrel{\tau}{\longrightarrow}_0 & \frac{3}{4}\cdot\overline{R} + \frac{1}{4}\cdot\overline{D} \\
\frac{3}{4}\cdot\overline{R} & \stackrel{\tau}{\longrightarrow}_{\frac{3}{4}} & \frac{3}{4}\cdot\overline{U} + \varepsilon \\
\frac{3}{4}\cdot\overline{U} & \stackrel{\tau}{\longrightarrow}_0 & (\frac{3}{4})^2\cdot\overline{R} + (\frac{3}{4})\frac{1}{4}\cdot\overline{D} \\
(\frac{3}{4})^2\cdot\overline{R} & \stackrel{\tau}{\longrightarrow}_{(\frac{3}{4})^2} & (\frac{3}{4})^2\cdot\overline{U} + \varepsilon \\
 & \vdots & \\
(\frac{3}{4})^k\cdot\overline{U} & \stackrel{\tau}{\longrightarrow}_0 & (\frac{3}{4})^{(k+1)}\cdot\overline{R} + (\frac{3}{4})^k\frac{1}{4}\cdot\overline{D} \\
(\frac{3}{4})^{(k+1)}\cdot\overline{R} & \stackrel{\tau}{\longrightarrow}_{(\frac{3}{4})^{(k+1)}} & (\frac{3}{4})^{(k+1)}\cdot\overline{U} + \varepsilon \\
 & \vdots &
\end{array}$$

That is, $\overline{U} \stackrel{\tau}{\Longrightarrow}_w \sum_{k\geq 0} (\frac{3}{4})^k(\frac{1}{4}\cdot\overline{D})$ where $w = \sum_{k\geq 1}(\frac{3}{4})^k$. However this weight evaluates to 3, while the sum of the sub-distributions is the full point distribution $\overline{D}$. In other words $\overline{U} \stackrel{\tau}{\Longrightarrow}_3 \overline{D}$. □

Hyper-derivations satisfy the transitivity property: if $\Delta \stackrel{\tau}{\Longrightarrow}_{w_1} \Delta_1$ and $\Delta_1 \stackrel{\tau}{\Longrightarrow}_{w_2} \Delta_2$ then $\Delta \stackrel{\tau}{\Longrightarrow}_{w_1+w_2} \Delta_2$. The generation of a hyper-derivative is in general highly nondeterministic. In (1) of Definition 3 the calculation of $\Delta_{k+1}^{\rightarrow}$ and $\Delta_{k+1}^{\times}$ from $\Delta_k^{\rightarrow}$ involves making nondeterministic choices. But these choices can be governed by *policies*.

Definition 4 (Static policies). *A* static policy *for a wMDP is a partial function* $\mathsf{pp} : S \rightharpoonup \mathbb{R}_{\geq 0} \times \mathcal{D}(S)$ *such that if* $\mathsf{pp}(s) = \langle w, \Delta \rangle$ *then* $s \stackrel{\tau}{\longrightarrow}_w \Delta$. □

Let us write $\Delta \stackrel{\tau}{\Longrightarrow}_{\mathsf{pp},w} \Delta'$ to mean that the hyper-derivative Δ' is generated with weight w from Δ using the policy pp. Formally this means that in (1) of Definition 3 the weights w_k and subdistributions $\Delta_k^{\rightarrow}$, $\Delta_k^{\times}$ are calculated as follows:

- $s \in \lceil \Delta_k^{\times} \rceil$ if and only if $\mathsf{pp}(s)$ is undefined
- $s \in \lceil \Delta_k^{\rightarrow} \rceil$ if and only if $\mathsf{pp}(s)$ is defined
- $\langle w_{k+1}, \Delta_{k+1}^{\rightarrow} + \Delta_{k+1}^{\times} \rangle = \sum_{s\in\lceil\Delta_k^{\rightarrow}\rceil} \Delta_k^{\rightarrow}(s) \cdot \mathsf{pp}(s)$ for all $k \geq 0$.

Theorem 1 (Finite generability). *Let* $\mathsf{pp}_1, ..., \mathsf{pp}_n$ *(*$n \geq 1$*) be all the static policies in a finitary wMDP. Suppose* $\Delta \stackrel{\tau}{\Longrightarrow}_{\mathsf{pp}_i,w_i} \Delta_i'$ *and* $w_i < \infty$ *for all* $1 \leq i \leq n$*. If* $\Delta \stackrel{\tau}{\Longrightarrow}_w \Delta'$ *then there are probabilities* p_i *for all* $1 \leq i \leq n$ *with* $\sum_{i=1}^n p_i = 1$ *such that* $\langle w, \Delta' \rangle = \sum_{i=1}^n p_i \cdot \langle w_i, \Delta_i' \rangle$. □

In later developments it will be important to rule out the possibility of hyper-derivatives generating an infinite weight.

Definition 5. *A wMDP is* convergent *if* $\overline{s} \stackrel{\tau}{\Longrightarrow}_w \varepsilon$ *for no state* s *and weight* w*; in other words there is no divergent internal computation from any state. A wMDP is* bounded *if it is finitary and whenever* $\Delta \stackrel{\tau}{\Longrightarrow}_w \Delta'$ *then* $w \in \mathbb{R}_{\geq 0}$. □

A simple source of unboundedness is divergence. Consider the trivial wMDP consisting of one state s and one arrow $s \stackrel{\tau}{\longrightarrow}_1 \overline{s}$. This is not a bounded wMDP because $\overline{s} \stackrel{\tau}{\Longrightarrow}_\infty \overline{s}$. In fact for finitary wMDPs, this is the only source of unboundedness:

Theorem 2. *Every finitary convergent wMDP is also bounded.* □

The proof of the above theorem relies on Theorem 1. Another important consequence of Theorem 1 is the following.

Corollary 1. *In a bounded wMDP, for every* Δ *the set* $\{ \langle w, \Delta' \rangle \mid \Delta \stackrel{\tau}{\Longrightarrow}_w \Delta' \}$ *is compact, in the standard Euclidean topology.* □

2.2 (Amortised weighted) simulations

Here we assume some wMDP $\langle S, \mathsf{Act}_\tau, \mathbb{R}_{\geq 0}, \longrightarrow \rangle$. Our simulation relation is parametrised on an initial investment $r \in \mathbb{R}_{\geq 0}$ and relates states to distributions, rather than states to states. It also uses *weak* arrow relations, defined using hyper-derivations: we write $\Delta \stackrel{a}{\Longrightarrow}_w \Theta$ whenever $\Delta \stackrel{\tau}{\Longrightarrow}_{w_1} \Delta' \stackrel{a}{\longrightarrow}_{w_2} \Theta' \stackrel{\tau}{\Longrightarrow}_{w_3} \Theta$, where $w = w_1 + w_2 + w_3$.

Definition 6. *Given a relation* $\mathcal{R} \subseteq S \times (\mathbb{R}_{\geq 0} \times \mathcal{D}(S))$*, let* $\mathcal{S}(\mathcal{R}) \subseteq S \times (\mathbb{R}_{\geq 0} \times \mathcal{D}(S))$ *be the relation defined by letting* $s \, \mathcal{S}(\mathcal{R}) \, \langle r, \Theta \rangle$ *whenever*

$s \stackrel{\alpha}{\longrightarrow}_v \Delta$ *implies the existence of some* w *and* Θ' *with* $\Theta \stackrel{\alpha}{\Longrightarrow}_w \Theta'$ *and* $\Delta \, \overline{\mathcal{R}} \, \langle r+w-v, \Theta' \rangle$.

We say $\mathcal{R}$ *is an (amortised weighted) simulation if* $\mathcal{R} \subseteq \mathcal{S}(\mathcal{R})$*. The operator* $\mathcal{S}(-)$ *is (pointwise) monotonic and so it has a maximal fixpoint, which is also a simulation, and which we denote by* $\lhd$*. We often write* $s \lhd_r \Theta$ *for* $s \lhd \langle r, \Theta \rangle$ *and use* $\Delta \sqsubseteq_{sim} \Theta$ *to mean that there is some initial investment* r *such that* $\Delta \, \overline{\lhd}_r \, \Theta$. □

The basic idea here is that $s \lhd_r \Theta$ intuitively means that Θ can simulate the actions of s but with *more benefit*, or at least not less benefit. The parameter r should be viewed as compensation which Θ has accumulated and can be used in local comparisons between the benefits of individual actions. Thus when we simulate $s \stackrel{\alpha}{\longrightarrow}_v \Delta$ with $\Theta \stackrel{\alpha}{\Longrightarrow}_w \Theta'$ there are two possibilities:

(i) $w > v$; here the accumulated compensation is increased from r to $r + (w - v)$. In subsequent rounds this extra compensation may be used to successfully simulate a heavier action with a lighter one.

(ii) $w \leq v$; here the compensation is decreased from r to $r - (v - w)$.

Finally it is important that $r \geq 0$, and remains greater than or equal to zero, or otherwise the presence of weights would have no effect. Thus in case (ii) if $(v - w) > r$ then the attempted simulation is not successful.

We now show that with this formal definition of the relation $\sqsubseteq_{sim}$ the various statements asserted in the Introduction are true:

$$\text{(L-ACT)}\quad \alpha_w.(\oplus_{i\in I} p_i \cdot P_i) \xrightarrow{\alpha}_w \mathcal{D}ist(\{(p_i, P_i) \mid i \in I\})$$

$$\text{(L-ALT)}\quad \frac{P_1 \xrightarrow{\alpha}_w \Delta}{P_1 + P_2 \xrightarrow{\alpha}_w \Delta}$$

$$\text{(L-COMM)}\quad \frac{P_1 \xrightarrow{a}_{w_1} \Delta_1,\ \ P_2 \xrightarrow{\overline{a}}_{w_2} \Delta_2}{P_1 \mid P_2 \xrightarrow{\tau}_w \Delta_1 \mid \Delta_2}\ w = w_1 + w_2$$

$$\text{(L-PAR)}\quad \frac{P_1 \xrightarrow{\alpha}_w \Delta}{P_1 \mid P_2 \xrightarrow{\alpha}_w \Delta \mid \overline{P_2}}$$

$$\text{(L-HIDE)}\quad \frac{P \xrightarrow{\alpha}_w \Delta}{P\backslash a \xrightarrow{\alpha}_w \Delta\backslash a}\ \alpha \neq a, \overline{a}$$

$$\text{(L-DEF)}\quad \frac{P_A \xrightarrow{\alpha}_w \Delta}{A \xrightarrow{\alpha}_w \Delta}\ A \Leftarrow P_A$$

Fig. 1. Weighted actions for CCMDP

Example 2. Consider the first two systems, s_0 and t_0, viewed as wMDPs. Then the relation $\mathcal{R}$ given by $\mathcal{R} = \{(s_0, \langle r, \overline{t_0}\rangle) \mid r \geq 1\} \cup \{(s_d, \langle r, \overline{t_d}\rangle) \mid r \geq 0\}$ is a simulation. Thus $s_0 \lhd_r \overline{t_0}$ for any $r \geq 1$. As pointed out in [11] this example shows the need for the parametrisation with respect to initial investments r; Because of the weights associated with the action up an initial investment of at least one is required in order for $\overline{t_0}$ to be able to match s_0.

We also have $s_0 \lhd_r \overline{s_1}$ for any $r \geq 1$ because of the following simulation:

$$\mathcal{R} = \{(s_0, \langle r, \overline{s_1}\rangle) \mid r \geq 1\} \cup \{(s_d, \langle r, \Delta\rangle) \mid r \geq 0\}$$

where Δ is the distribution $\frac{1}{4} \cdot \overline{O} + \frac{3}{4} \cdot \overline{T}$. Note that this is indeed a simulation because $\Delta \xrightarrow{\text{down}}_{2.5} \overline{s_1}$. Incidently this example shows the necessity of relating states to distributions, rather than states; no individual state accessible from s_1 can simulate s_d.

Similarly $s_1 \lhd_r \overline{t_1}$ for every $r \geq 0$ because of the simulation:

$$\mathcal{R} = \{(s_1, \langle r, \overline{t_1}\rangle) \mid r \geq 0\} \cup \{(O, \langle r, \overline{U}\rangle) \mid r \geq 0\} \cup \{(T, \langle r, \overline{U}\rangle) \mid r \geq 0\}$$

This relies on the fact that $\overline{U} \overset{\text{down}}{\Longrightarrow}_4 \overline{t_1}$, which follows by transitivity, since we have already seen in Example 1 that $\overline{U} \overset{\tau}{\Longrightarrow}_3 \overline{D}$.

Finally $s_0 \lhd_2 s_2$ because of the following simulation:

$$\mathcal{R} = \{(s_0, \langle r, \overline{s_2}\rangle) \mid r \geq 2\} \cup \{(s_d, \langle r, \overline{\Delta}\rangle) \mid r \geq 0\}$$

where Δ is the distribution $\frac{1}{4} \cdot \overline{S} + \frac{3}{4} \cdot \overline{T}$. Note that $\Delta \overset{\text{down}}{\Longrightarrow}_3 \overline{s_2}$ although it is also possible for it to do the down action for much less benefit. □

The simulation relations $\lhd_r$ are defined coinductively. But in bounded wMDPS they can also be characterised inductively.

Definition 7. *For every $k \geq 0$ we define the relation $\lhd^k \subseteq S \times (\mathbb{R}_{\geq 0} \times \mathcal{D}(S))$ as follows:*

(i) $\lhd^0 = S \times (\mathbb{R}_{\geq 0} \times \mathcal{D}(S))$

(ii) $\lhd^{k+1} = \mathcal{S}(\lhd^k)$ *for every* $k \geq 0$.

Finally we let $\lhd^\infty$ be $\bigcap_{k=0}^{\infty} \lhd^k$. □

Theorem 3. *In a bounded wMDP, the two relations $\lhd$ and $\lhd^\infty$ coincide.* □

Corollary 1 plays a crucial role in proving the above theorem.

The simplest approach to discussing compositionality is, as in [8], to introduce a process calculus-like syntax for wMDPs. Our calculus, CCMDP, is based on CCS:

$$P ::= \alpha_w.(\oplus_{i\in I} p_i \cdot P_i) \mid P \mid P \mid P + P \mid \mathbf{0} \mid P\backslash a \mid A \tag{2}$$

The main operator is prefixing, $\alpha_w.(\oplus_{i\in I} p_i \cdot P_i)$. Here α is taken from Act_τ, w from $\mathbb{R}_{\geq 0}$, I is a non-empty finite index set and p_i are probabilities satisfying $\sum_{i\in I} p_i = 1$. We also assume a set of definitional constants, ranged over by A, and each such A has a definition associated with it, a process term P_A. We often write these definitions as $A \Leftarrow P_A$.

Intuitively, we view each process term as describing a wMDP. Formally we describe one overarching wMDP where the states are all terms P in the grammar (2) and the weighted actions $P \xrightarrow{\alpha}_w \Delta$ are those which can be derived by the rules in Figure 1; obvious symmetric counterparts to the rules (L-ALT) (L-PAR) are omitted. In rule (L-ACT) we use the obvious notation $\mathcal{D}ist(\{\,(p_i, P_i) \mid i \in I\,\})$ for constructing a distribution from the formal term $\oplus_{i\in I} p_i \cdot P_i$. In rules (L-COMM) and (L-PAR) we use an abbreviation for distributing parallel composition over a distribution: e.g. $\Delta_1 \mid \Delta_2$ is the distribution given by $(\Delta_1 \mid \Delta_2)(R) = \Delta_1(P_1) \cdot \Delta_2(P_2)$ if $R = P_1 \mid P_2$ and 0 otherwise. Similar is the hiding operator in (L-HIDE). Note that all of the wMDPs described graphically in the Introduction can be described in CCMDP. In the sequel we will not distinguish between the syntactic term P, its interpretation as a state in the wMDP defined in Figure 1, and the wMDP it induces by considering only those states accessible from it.

Theorem 4 (Compositionality). *The preorders $\overline{\lhd}_r$, for each $r \in \mathbb{R}_{\geq 0}$, are preserved by each of the operators in the language* CCMDP. □

Example 3. Let P, Q be two processes with $P \lhd_0 \overline{Q}$. Consider the following processes:

$$\begin{aligned} U &\Leftarrow \tau_0.(\tau_1.U \,{}_{\frac{3}{4}}\!\oplus \mathsf{down}_1.Q) \\ P' &\equiv \mathsf{up}_2.(\mathsf{down}_1.P \,{}_{\frac{1}{4}}\!\oplus \mathsf{down}_3.P) \\ Q' &\equiv \mathsf{up}_2.U \end{aligned}$$

By the analysis in Example 1 we know that $\overline{U} \overset{\tau}{\Longrightarrow}_3 \overline{\mathsf{down}_1.Q}$, thus $\overline{U} \overset{\mathsf{down}}{\Longrightarrow}_4 \overline{Q}$. Then it is easy to see that $\mathsf{down}_1.P \lhd_0 \overline{U}$ and $\mathsf{down}_3.P \lhd_0 \overline{U}$. It follows from the compositionality of $\overline{\lhd}_0$ that $(\mathsf{down}_1.P \,{}_{\frac{1}{4}}\!\oplus \mathsf{down}_3.P) \,\overline{\lhd}_0\, \overline{U}$ and furthermore $\overline{P'} \,\overline{\lhd}_0\, \overline{Q'}$. □

3 A qualitative probabilistic logic

Let Var be a set of variables, ranged over by X. We define the set of formulae as follows:

$$\begin{aligned} \phi ::= {} & \mathsf{tt} \mid \mathsf{ff} \mid \langle\alpha\rangle_w(\phi_1 \,{}_p\!\oplus \phi_2),\ \alpha \in \mathsf{Act}_\tau, w \in \mathbb{R}_{\geq 0}, p \in [0,1] \\ & \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid X \mid \mathsf{min}\, X.\, \phi \mid \mathsf{max}\, X.\, \phi \end{aligned}$$

The two fixpoint operators $\mathsf{min}\, X.\, -$ and $\mathsf{max}\, X.\, -$ act as binders in the standard manner; we use $\mathcal{L}_\mathcal{Q}$ to denote the set of *closed* formulae, that is containing no free variables. As a shorthand, we write $\langle\alpha\rangle_w\phi$ for $\langle\alpha\rangle_w(\phi_1 \,{}_1\!\oplus \phi')$ for any ϕ'.

Let *Con* denote the set of *configurations*, pairs $\langle r, \Delta\rangle$ where $r \in \mathbb{R}_{\geq 0}$ and $\Delta \in \mathcal{D}(S)$, with S denoting the state space of some wMDP. Intuitively this represents a system which has accumulated compensation r which it can use to satisfy formulae in the future. A formula from $\mathcal{L}_\mathcal{Q}$ determines a set of configurations, those which satisfy it; their calculation is standard, apart from the novel qualitative possibility operator. An environment ρ is a function that maps each variable in Var to a subset of *Con*. For a set $V \subseteq \mathbb{R}_{\geq 0} \times \mathcal{D}(S)$ and a variable $X \in \mathsf{Var}$, we write $\rho[X \mapsto V]$ for the environment that maps X to V and Y to $\rho(Y)$ for all $Y \neq X$. The semantics of a formula ϕ is given by the set of configurations $[\![\phi]\!]_\rho$ defined as follows:

- $[\![\mathsf{tt}]\!]_\rho = Con, \qquad [\![\mathsf{ff}]\!]_\rho = \emptyset$
- $[\![\phi_1 \wedge \phi_2]\!]_\rho = [\![\phi_1]\!]_\rho \cap [\![\phi_2]\!]_\rho, \qquad [\![\phi_1 \vee \phi_2]\!]_\rho = [\![\phi_1]\!]_\rho \cup [\![\phi_2]\!]_\rho$
- $[\![\langle\alpha\rangle_v(\phi_1 \,{}_p\!\oplus \phi_2)]\!]_\rho = \{\langle r, \Delta\rangle \mid \Delta \stackrel{\alpha}{\Longrightarrow}_w \Theta$ where $\langle (r+w-v), \Theta\rangle = \langle r_1, \Theta_1\rangle \,{}_p\!\oplus \langle r_2, \Theta_2\rangle$ and $\langle r_i, \Theta_i\rangle \in [\![\phi_i]\!]_\rho\}$
- $[\![X]\!]\rho = \rho(X)$
- $[\![\min X.\ \phi]\!]_\rho = \bigcap\{V \mid [\![\phi]\!]_{\rho[X\mapsto V]} \subseteq V\}$
- $[\![\max X.\ \phi]\!]_\rho = \bigcup\{V \mid V \subseteq [\![\phi]\!]_{\rho[X\mapsto V]}\}$

When ϕ is closed the set $[\![\phi]\!]_\rho$ is independent of the environment ρ, and in this case we use the standard notation $C \models \phi$ in place of $C \in [\![\phi]\!]$.

The novel qualitative formula $\langle\alpha\rangle_v(\phi_1 \,{}_p\!\oplus \phi_2)$ represents the ability to do an α action with benefit at least v and then probabilistically satisfy the property $\phi_1 \,{}_p\!\oplus \phi_2$; we have $\langle r, \Delta\rangle \models \langle\alpha\rangle_v(\phi_1 \,{}_p\!\oplus \phi_2)$ whenever $\Delta \stackrel{\alpha}{\Longrightarrow}_w \Theta_1 \,{}_p\!\oplus \Theta_2$ and $\langle r_i, \Theta_i\rangle \models \phi_i$ for some r_i satisfying $(r+w-v) = p\cdot r_1 + (1-p)\cdot r_2$. Here there are two possibilities:

(i) $v > w$: here the compensation comes into play. The action may be accepted despite being too heavy but the compensation for future use is reduced from r to $r-(v-w)$; this is split into $r_1,\ r_2$ via the probability p. Note this possibility will only exist if $r-(v-w) \geq 0$.

(ii) $v \leq w$: The action is accepted and then the compensation is increased from r to $r+(w-v)$, which again is split proportionally into $r_1,\ r_2$, to satisfy ϕ_1 and ϕ_2 respectively.

Example 4. Both liveness and safety properties can be expressed in $\mathcal{L}_Q$. For example, suppose AB denote the formula $\langle a\rangle_0(\langle b\rangle_{10}\mathsf{tt} \;{}_{\frac{9}{10}}\!\oplus\ \mathsf{tt})$ and C is a configuration. Then $C \models AB$ means that C can perform an a action such that at least 90% of the time it can subsequently perform a b action with a benefit of at least 10. So the formula

$$\min X.\ \langle \mathsf{up}\rangle_0(\langle \mathsf{down}\rangle_{10} X \;{}_{\frac{9}{10}}\!\oplus\ \mathsf{tt}) \ \vee\ \langle \mathsf{up}\rangle_0 X$$

expresses the liveness property of being able to perform a sequence of up actions to arrive at a state where at least 90% of the time a down action for benefit at least 10 can be performed. On the other hand, the formula

$$\max X.\ \langle \mathsf{up}\rangle_0(\langle \mathsf{down}\rangle_{10} X \;{}_{\frac{9}{10}}\!\oplus\ \mathsf{tt}) \ \wedge\ \langle \mathsf{stay}\rangle_0 X$$

expresses the safety property of always being able to perform a stay action and at the same time to perform an up action to arrive at a state where at least 90% of the time a down action for benefit at least 10 can be performed. □

Let $\mathcal{L}_Q(r, \Delta) = \{\phi \in \mathcal{L}_Q \mid \langle r, \Delta\rangle \models \phi\}$. We have the following logical characterisation of simulations.

Theorem 5 (Logical characterisation). *In a bounded wMDP, $s \lhd_r \Theta$ if and only if $\mathcal{L}_Q(0, \overline{s}) \subseteq \mathcal{L}_Q(r, \Theta)$.* □

The proof of Theorem 5 exploits Theorem 3, which says that $\lhd$ can be approximated by a family of stratified relations $\lhd^k$ for $k \in \mathbb{N}$. So it suffices to prove that each approximant $\lhd^k$ is completely determined by the finite fragment of the qualitative probabilistic logic.

The import of Theorem 5 is that if $s \lhd_r \Theta$ does not hold then there is a formula ϕ from $\mathcal{L}_Q$ which $\overline{s}$ satisfies but Θ does not. Furthermore, it turns out that in a bounded wMDP this distinguishing formula will always be finite; that is contains no occurrence

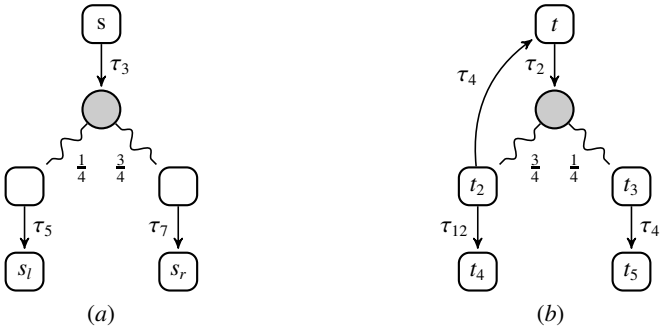


Fig. 2. Testing systems

of a fixpoint operator. For example, $\overline{s_2} \not\overline{\preceq}_0 \overline{s_0}$, where these are defined in the Introduction, because of the distinguishing formula $\langle \mathsf{up} \rangle_1 (\langle \mathsf{down} \rangle_6 \mathsf{tt} \ {}_{\frac{1}{4}}\!\oplus\ \langle \mathsf{down} \rangle_2 \mathsf{tt})$.

Our logic is expressive enough so that the whole behaviour of a state in a bounded wMDP can be captured by one formula in the logic:

Theorem 6 (Characteristic formula). *In a bounded wMDP, for every state s there is a* characteristic formula $\phi(s) \in \mathcal{L}_Q$ *such that $s \lhd_r \Theta$ if and only if $\langle r, \Theta \rangle \models \phi(s)$.* □

For example, the state s_1 in the Introduction has the following characteristic formula: $\phi(s_1) = \max X. \langle \mathsf{up} \rangle_2 (\langle \mathsf{down} \rangle_1 X \ {}_{\frac{1}{4}}\!\oplus\ \langle \mathsf{down} \rangle_3 X)$.

4 Benefits based testing

Standard theories of testing involve the idea of applying tests to processes and seeing if the result is a *success*. With the presence of weights in wMDPs we have a more elementary way of testing; we run them in parallel with other wMDPs and calculate the possible benefits which can be accrued. Then two wMDPs can be compared by examining the resulting sets of possible accrued benefits.

Consider the simple fully probabilistic wMDP in Figure 2(a), which results from running the test $T = \overline{\mathsf{up}}_1.\overline{\mathsf{down}}_4.\mathbf{0}$ in parallel with the system s_1 from the Introduction. Formally this is the sub-wMDP of the wMDP $(\overline{s_1} \mid T)$ obtained by concentrating on the internal actions τ_w; this is just the wMDP represented by $(\overline{s_1} \mid T)\backslash\mathsf{Act}$ that we denote by $\overline{s_1} \parallel T$. Every time the experiment runs we get the initial benefit 3; three-quarters of the time we also get the benefit 7 while a quarter of time we get 5. So the total benefit is $3 + \frac{3}{4} \cdot 7 + \frac{1}{4} \cdot 5 = 9.5$. In the presence of nondeterminism there will in general be a set of possible benefits, depending on the way in which the nondeterminism is resolved. Traditionally this resolution is expressed in terms of a scheduler, or adversary, which for each state decides which of its successors is chosen for execution, with the resulting set of benefits consequently depending on the choice of scheduler. Here we take a more abstract approach, following [5], and essentially allow arbitrary schedulers.

Definition 8 (Extreme derivatives). *For any Δ in a wMDP we write $\Delta \stackrel{\tau}{\Longrightarrow}\!\!\!\!\rightarrow_w \Phi$ if*

- *$\Delta \stackrel{\tau}{\Longrightarrow}_w \Phi$, that is Φ is a hyper-derivative of Δ*
- *Φ is* stable*, that is $s \stackrel{\tau}{\not\mapsto}$ for every s in $\lceil \Phi \rceil$*

where $s \overset{\tau}{\not\mapsto}$ *means that* s *cannot enable any* τ*-transition. We say* Φ *is an* extreme derivative *of* Δ*, with weight* w. □

Intuitively every extreme derivation $\Delta \overset{\tau}{\Longrightarrow}_w \Phi$ represents a computation from the initial distribution Δ guided by some implicit scheduler. For example, consider the hyper-derivation of an extreme derivative:

$$\begin{array}{rcl} \Delta & = & \Delta_0^{\rightarrow} + \Delta_0^{\times} \\ \Delta_0^{\rightarrow} & \overset{\tau}{\longrightarrow}_{w_0} & \Delta_1^{\rightarrow} + \Delta_1^{\times} \\ & \vdots & \\ \Delta_k^{\rightarrow} & \overset{\tau}{\longrightarrow}_{w_k} & \Delta_{k+1}^{\rightarrow} + \Delta_{k+1}^{\times} \\ & \vdots & \\ \Phi & = & \sum_{k=0}^{\infty} \Delta_k^{\times} \end{array} \tag{3}$$

where $w = \sum_{k\geq 0} w_k$. Initially, since $\Delta_0^{\times}$ is stable, $\Delta_0^{\rightarrow}$ contains (in its support) all states which can proceed with the computation. The implicit scheduler decides for each of these states which step to take, cumulating in the first move, $\Delta_0^{\rightarrow} \overset{\tau}{\longrightarrow}_{w_0} \Delta_1^{\rightarrow} + \Delta_1^{\times}$. At an arbitrary stage, $\Delta_k^{\rightarrow}$ contains all states which can continue; the scheduler decides which step to take for each individual state and the overall result of the schedulers decision for this stage is captured in the step $\Delta_k^{\rightarrow} \overset{\tau}{\longrightarrow}_{w_k} \Delta_{k+1}^{\rightarrow} + \Delta_{k+1}^{\times}$.

Example 5. Referring to Figure 2(a) it is easy to see that $\overline{s}$ has a unique (degenerate) extreme derivative, $\overline{s_1} \overset{\tau}{\Longrightarrow}_{9.5} (\frac{1}{4}\overline{s_l} + \frac{3}{4}\overline{s_r})$, intuitively representing the unique weighted computation from $\overline{s_1}$. However, consider the wMDP in Figure 2(b), in which there is a nondeterministic choice from state t_2; here the extreme derivatives generated from $\overline{t}$, and their associated weights, will depend on the choices made during the computation by the implicit scheduler.

First suppose that the scheduler uses the static policy which maps t_2 to $\langle\, 12, \overline{t_4} \,\rangle$. Then it is easy to see that the generated extreme derivative, which is degenerate, is $\overline{t} \overset{\tau}{\Longrightarrow}_{12} (\frac{3}{4}\overline{t_4} + \frac{1}{4}\overline{t_5})$. However using the static policy which maps t_2 to $\langle\, 4, \overline{t_1} \,\rangle$ we generate, using (3), a non-degenerate extreme derivative; after some calculations this can be seen to be $\overline{t_1} \overset{\tau}{\Longrightarrow}_{24} \overline{t_5}$.

However there are many other possible implicit schedulers, for example at different times in the computations employing either of these static policies, or even choosing nondeterministically between them. But these are the only static policies and therefore we know from Theorem 1 that if $\overline{t_1} \overset{\tau}{\Longrightarrow}_w \Delta$ then w must take the form $p \cdot 12 + (1-p) \cdot 24$ for some $0 \leq p \leq 1$. That is the set of benefits which can be generated from $\overline{t_1}$ is $\{\, 24 - 12 \cdot p \mid 0 \leq p \leq 1 \,\}$. □

Definition 9 (May testing). *In a wMDP, for any* $\Delta \in \mathcal{D}(S)$*, let*

$$\textbf{Benefits}(\Delta) = \{\, w \in \mathbb{R}_{\geq 0} \mid \Delta \overset{\tau}{\Longrightarrow}_w \Phi, \textit{ for some } \Phi \in \mathcal{D}_{sub}(S) \,\}$$

Benefit sets are compared as follows:

$$B_1 \leq^r_{\mathrm{Ho}} B_2 \text{ if for every } r_1 \in B_1 \text{ there exists some } r_2 \in B_2 \text{ such that } r_1 \leq r + r_2$$

For any two distributions Δ, Θ we write $\Delta \sqsubseteq^r_{may} \Theta$ if for every finite (testing) process T, **Benefits**$(\Delta \parallel T) \leq^r_{\mathrm{Ho}}$ **Benefits**$(\Theta \parallel T)$. *We write $\Delta \sqsubseteq_{may} \Theta$ to mean that there is some $r \in \mathbb{R}_{\geq 0}$ such that $\Delta \sqsubseteq^r_{may} \Theta$.* □

This interpretation of processes is optimistic; $\Delta \sqsubseteq^r_{\mathrm{may}} \Theta$ means that, given the investment r, every possible benefit produced by Δ can in principle be improved upon by Θ.

Note that in a bounded wMDP **Benefits**(Δ) cannot contain ∞. Moreover we can show that the parallel composition of a bounded wMDP with a finite wMDP is also bounded. This means that if we confine our attention to bounded wMDPs then benefit sets will always only contain real numbers. One way of restricting to bounded wMDPS is, by Theorem 2, to only use finitary convergent wMDPs.

Our first result shows that simulations can be used as a sound proof technique for this semantics:

Theorem 7 (Soundness). *$\Delta \overline{\lhd}_r \Theta$ implies $\Delta \sqsubseteq^r_{may} \Theta$.* □

The converse is not true in general:

Example 6. Consider the two distributions $\Delta = \mathbf{0} \oplus_{\frac{1}{2}} a_1.\mathbf{0}$ and $\Theta = \tau_2.\mathbf{0} \oplus_{\frac{1}{2}} a_0.\mathbf{0}$. It is easy to see that $\Delta \overline{\not\lhd}_0 \Theta$ because there is no way to decompose Θ into $\Theta_1 \oplus_{\frac{1}{2}} \Theta_2$ for some Θ_1, Θ_2 such that $a_1.\mathbf{0} \lhd_0 \Theta_2$. However, one can show that $\Delta \sqsubseteq^0_{\mathrm{may}} \Theta$. This follows from the observations below:

(i) For all weights w and tests T, **Benefits**$(\tau_w.\mathbf{0} \parallel T) = \{v + w \mid v \in$ **Benefits**$(\mathbf{0} \parallel T)\}$.
(ii) For all weights w and tests T, **Benefits**$(a_w.\mathbf{0} \parallel T) \leq^w_{\mathrm{Ho}}$ **Benefits**$(a_0.\mathbf{0} \parallel T)$.

Both assertions can be proved by structural induction on T.

Now suppose $w \in$ **Benefits**$(\Delta \parallel T)$ for an arbitrary test T. There is some stable derivative Γ such that $\Delta \parallel T \stackrel{\tau}{\Longrightarrow}_w \Gamma$. It can be shown that there are some $w_1, w_2, \Gamma_1, \Gamma_2$ with $\mathbf{0} \parallel T \stackrel{\tau}{\Longrightarrow}_{w_1} \Gamma_1$, $a_1.\mathbf{0} \parallel T \stackrel{\tau}{\Longrightarrow}_{w_2} \Gamma_2$, $w = \frac{1}{2}w_1 + \frac{1}{2}w_2$, and $\Gamma = \frac{1}{2} \cdot \Gamma_1 + \frac{1}{2} \cdot \Gamma_2$, where both Γ_1 and Γ_2 are stable. In other words, we have $w_1 \in$ **Benefits**$(\mathbf{0} \parallel T)$ and $w_2 \in$ **Benefits**$(a_1.\mathbf{0} \parallel T)$. By (i) above, $w_1 + 2 \in$ **Benefits**$(\tau_2.\mathbf{0} \parallel T)$; by (ii) above, there exists some $w_2' \in$ **Benefits**$(a_0.\mathbf{0} \parallel T)$ with $w_2 \leq w_2' + 1$. Thus, we can infer that

$$\begin{aligned} w &= \tfrac{1}{2}w_1 + \tfrac{1}{2}w_2 \\ &< \tfrac{1}{2}(w_1 + 2) + \tfrac{1}{2}(w_2 - 1) \\ &\leq \tfrac{1}{2}(w_1 + 2) + \tfrac{1}{2}w_2' \end{aligned}$$

It turns out that $\frac{1}{2}(w_1 + 2) + \frac{1}{2}w_2' \in$ **Benefits**$(\Theta \parallel T)$. Therefore, we have

$$\textbf{Benefits}(\Delta \parallel T) \leq^0_{\mathrm{Ho}} \textbf{Benefits}(\Theta \parallel T).$$

Since this reasoning is carried out for an arbitrary test T, it follows that $\Delta \sqsubseteq^0_{\mathrm{may}} \Theta$. □

Nevertheless we do have a testing characterisation for the unannotated simulation preorder:

Theorem 8 (Testing characterisation). *In a bounded wMDP,* $\overline{s} \sqsubseteq_{may} \Theta$ *if and only if* $\overline{s} \sqsubseteq_{sim} \Theta$.

Proof (Outline). One direction follows from Theorem 7. For the converse we carry out the proof in two steps: we first prove that $\overline{s} \sqsubseteq^{r}_{\text{may}} \Theta$ implies the existence of some compensation $r' \geq r$ with $\mathcal{L}_Q(0, \overline{s}) \subseteq \mathcal{L}_Q(r', \Theta)$, then appeal to Theorem 5. In the first step we proceed by constructing, for each formula ϕ, a *characteristic test* $T(\phi)$, such that if a process satisfies ϕ then it passes test $T(\phi)$ with some threshold benefit. □

An alternative approach to testing would be to use one special action ω in a test to report success and when applying such a test to a system to report the weighted average of the weight of each path leading to an occurrence of the success action; this we refer to as *expected benefits testing*. Here we will not give the formal definition of how these *expected benefits* are calculated, which is provided in [4], but simply give an informal argument to show that our simulation preorder is not sound with respect to it.

Example 7 (Simulation is unsound for expected benefits testing). Consider the following processes:

$$P = \tau_2.(\mathbf{0} \,{}_{\frac{1}{4}}\!\oplus a_0.\mathbf{0})$$
$$Q = \tau_1.(\tau_2.(\mathbf{0} \,{}_{\frac{1}{2}}\!\oplus a_0.\mathbf{0}) \,{}_{\frac{1}{2}}\!\oplus a_0.\mathbf{0})$$

It is easy to see that $P \lhd_0 Q$ as the transition $P \xrightarrow{\tau}_2 \mathbf{0} \,{}_{\frac{1}{4}}\!\oplus a_0.\mathbf{0}$ can be simulated by the hyper-derivative $Q \stackrel{\tau}{\Longrightarrow}_2 \mathbf{0} \,{}_{\frac{1}{4}}\!\oplus a_0.\mathbf{0}$. Now let T be the test $\bar{a}_0.\omega$. Both $P \parallel T$ and $Q \parallel T$ give rise to fully probabilistic wMDPs. The unique expected benefit resulted from $P \parallel T$ is $\frac{1}{4} \cdot 0 + \frac{3}{4} \cdot 2$, i.e. $\frac{3}{2}$. On the other hand, the unique expected benefit obtained from $Q \parallel T$ is $\frac{1}{2} \cdot 1 + \frac{1}{2}(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 3)$, i.e. $\frac{5}{4}$. As $\{\frac{3}{2}\} \not\leq^{0}_{\text{Ho}} \{\frac{5}{4}\}$, we have that P is not related to Q under expected benefits may testing; thus $\lhd$ is unsound for expected benefits testing. Note that if we consider total benefits, then $\mathbf{Benefits}(P \parallel T) = \{2\} = \mathbf{Benefits}(Q \parallel T)$. □

5 Concluding remarks

We have proposed the model of weighted Markov decision processes for compositional reasoning about the behaviour of systems with uncertainty. Amortised weighted simulation is coinductively defined to be a behavioural preorder for comparing different wMDPs. It is shown to be a precongruence relation with respect to all structural operators for constructing wMDPs from components. For finitary convergent wMDPs, we have also given logical and testing characterisations of the simulation preorder: it can be completely determined by a qualitative probabilistic logic and for each system we can find a characteristic formula to capture its behaviour; the simulation preorder also coincides with a notion of may testing preorder.

The dual of may testing is must testing. It would be interesting to investigate the must preorder given by our testing approach. We leave it as future work to provide a coinductive formulation of the preorder and study its logical characterisations.

There is a very limited literature on compositional theories of Markov decision processes particularly in the presence of weights. There is however an extensive literature on probabilistic variations of bisimulation equivalence for Markov chains; see Chapter 10 of [1] for an elementary introduction and [10] for a survey. Bisimulation equivalence has also been defined in [8] for *Interactive Markov Chains (IMCs)*, and it is shown to be compositional, in the sense of our Theorem 4: it is preserved by the operators of a process calculus interpreted as IMCs. Bisimulation and testing equivalence for Markovian process algebras are also investigated in [9, 2], but the analysis was mainly restricted to models free of nondeterminism. Recently a combination of probabilistic automata and IMCs has been studied in [7], where a notation of weak bisimulation is proposed. Since time rates are treated essentially as action names, some intuitively equivalent processes are differentiated by the weak bisimulation.

There is also an extensive literature on weighted automata [6], and probabilistic variations have also been studied [3]. However there the focus is on traditional language theoretic issues, rather than our primary concern, compositionality.

References

1. C. Baier and J.-P. Katoen. *Principles of Model Checking*. The MIT Press, 2008.
2. M. Bernardo and R. Cleaveland. A theory of testing for Markovian processes. In Proc. *CONCUR'00*, LNCS 1877, pp. 305-319. Springer, 2000.
3. K. Chatterjee, L. Doyen, and T.A. Henzinger. Probabilistic Weighted Automata. In Proc. *CONCUR'09*, LNCS 5710, pp. 244–258. Springer, 2009.
4. Y. Deng and M. Hennessy. Compositional reasoning for Markov decision processes (full version), 2010. Available at `http://basics.sjtu.edu.cn/~yuxin/temp/mdp.pdf`.
5. Y. Deng, R. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes. In Proc. *CONCUR'09*, LNCS 5710, pp. 274–288. Springer, 2009.
6. M. Droste, W. Kuich, and H. Vogler (Eds.) *Handbook of Weighted Automata*. Springer, 2009.
7. E. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In Proc. *LICS'10*, pp. 342-351. IEEE Computer Society, 2010.
8. H. Hermanns. *Interactive Markov Chains: The Quest for Quantified Quality*, LNCS 2428. Springer, 2002.
9. J. Hillston. *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.
10. B. Jonsson, K.G. Larsen, and Y. Wang. Probabilistic Extensions of Process Algebras. In *Handbook of Process Algebra*, pp. 685–710. Elsevier, 2001.
11. A. Kiehn and S. Arun-Kumar. Amortised bisimulations. In Proc. *FORTE'05*, LNCS 3731, pp. 320–334. Springer, 2005.
12. M. Puterman. *Markov Decision Processes*. Wiley, 1994.
13. J. Rutten, M. Kwiatkowska, G. Norman, and D. Parker. *Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems.* American Mathematical Society, 2004.
14. R. Segala. Modeling and verification of randomized distributed real-time systems. Technical Report MIT/LCS/TR-676, PhD thesis, MIT, Dept. of EECS, 1995.