On Applicative Similarity, Sequentiality, and Full Abstraction

Raphaëlle Crubillé, Ugo Dal Lago, Davide Sangiorgi, Valeria Vignudelli

On Applicative Similarity, Sequentiality, and Full Abstraction*

Raphaëlle Crubillé[1], Ugo Dal Lago[2], Davide Sangiorgi[2], and Valeria Vignudelli[2]

[1] ENS-Lyon, raphaelle.crubille@ens-lyon.fr
[2] Università di Bologna & INRIA,
{ugo.dallago,davide.sangiorgi2,valeria.vignudelli2}@unibo.it

Abstract. We study how applicative bisimilarity behaves when instantiated on a call-by-value probabilistic λ-calculus, endowed with Plotkin's parallel disjunction operator. We prove that congruence and coincidence with the corresponding context relation hold for both bisimilarity and similarity, the latter known to be impossible in sequential languages.

Keywords: probabilistic lambda calculus, bisimulation, coinduction, sequentiality

1 Introduction

The work in this paper is part of a general effort in trying to transport techniques and concepts for program correctness and verification that have been introduced and successfully applied to ordinary (first-order) concurrency (CCS, CSP, Petri Nets), following pioneering work by Bergstra, Hoare, Milner, Olderog, and others, onto formalisms with higher-order features, in which the values exchanged or manipulated may include pieces of code. Specifically, we focus on the prototypical higher-order language, the λ-calculus, enriched with a probabilistic choice, and use coinductive methods and logics to understand and characterise behavioural equivalences.

Probabilistic models are more and more pervasive. Examples of application areas in which they have proved to be useful include natural language processing [16], robotics [23], computer vision [3], and machine learning [19]. Sometimes, being able to "flip a fair coin" while computing is a *necessity* rather than an alternative, like in cryptography (where, e.g., secure public key encryption schemes are bound to be probabilistic [10]): randomness is not only a modeling tool, but a capability algorithms can exploit.

The specification of probabilistic models and algorithms can be made easier by the design of programming languages. And indeed, various probabilistic programming languages have been introduced in the last years, from abstract ones [12, 22, 18] to more concrete ones [20, 11]. A common scheme consists in

* The authors are partially supported by the ANR project 12IS02001 PACE.

endowing deterministic languages with one or more primitives for probabilistic choice, like binary probabilistic choice or primitives for distributions. Many of them, as a matter of fact, are designed around the λ-calculus or one of its incarnations, like Scheme. This, in turn, has stimulated foundational research about probabilistic λ-calculi, and in particular about the nature of program equivalence in a probabilistic setting. This has already started to produce some interesting results in the realm of denotational semantics, where adequacy and full-abstraction results have recently appeared [7, 9].

Operational techniques for program equivalence, and in particular coinductive methodologies, have the advantage of not requiring a too complicated mathematical machinery. Various notions of bisimilarity have been proved adequate and, in some cases, fully abstract, for deterministic and nondeterministic computation [1, 17, 15]. A recent paper [5] generalizes Abramsky's applicative bisimulation [1] to a call-by-*name*, untyped λ-calculus endowed with binary, fair, probabilistic choice [6]. Probabilistic applicative bisimulation is shown to be a congruence, thus included in context equivalence. Completeness, however, fails, but can be recovered if call-by-*value* evaluation is considered, as shown in [4]. This can appear surprising, given that in nondeterministic λ-calculi, both when call-by-name *and* call-by-value evaluation are considered, applicative bisimilarity is a congruence, but *finer* than context equivalence [15]. But there is another, even less expected result: the aforementioned correspondence does not hold anymore if we consider applicative *simulation* and the contextual *preorder*.

The reason why this happens can be understood if one looks at the testing-based characterization of similarity and bisimilarity from the literature [8, 24]: the class of tests characterizing *bi*similarity is simple enough to allow any test to be implementable by a program context. This is impossible for tests characterizing similarity, which include not only conjunction (which can be implemented as copying) but also disjunction, an operator that seems to require the underlying language to be parallel.

In this paper, we show that, indeed, the presence of Plotkin's disjunction [21, 2] turns applicative similarity into a relation which coincides with the context preorder. This is done by checking that the proof of precongruence for applicative bisimilarity [5, 4] continues to hold, and by showing how tests involving conjunction and disjunction can be implemented by contexts. This somehow completes the picture about how applicative (bi)similarity behaves in a probabilistic scenario.

2 Programs and Their Operational Semantics

In this section, we present the syntax and operational semantics of $\Lambda_{\oplus or}$, the language on which we define applicative bisimulation. $\Lambda_{\oplus or}$ is a λ-calculus endowed with probabilistic choice and parallel disjunction operators.

The terms of $\Lambda_{\oplus or}$ are built up from variables, using the usual constructs of λ-calculus, binary choice and parallel disjunction. In the following, $Var = \{x, y, \ldots\}$ is a countable set of variables

Definition 1. *The terms of $\Lambda_{\oplus or}$ are expressions generated by the following grammar:*

$$M, N, L \ ::= \ x \ \Big| \ \lambda x.M \ \Big| \ M \oplus N \ \Big| \ M\, N \ \Big| \ [M \parallel N] \rightarrowtail L$$

where $x \in Var$.

In what follows, we consider terms of $\Lambda_{\oplus or}$ as α-equivalence classes of syntax trees. We let $FV(M)$ denote the set of free variables of the term M. A term M is closed if $FV(M) = \emptyset$. Given a set $\overline{x}$ of variables, $\Lambda_{\oplus or}(\overline{x})$ is the set of terms M such that $FV(M) \subseteq \overline{x}$. We write $\Lambda_{\oplus or}$ for $\Lambda_{\oplus or}(\emptyset)$. The (capture-avoiding) substitution of N for the free occurrences of x in M is denoted by $M[N/x]$.

The constructs of the λ-calculus have their usual meanings. The construct $M \oplus N$ is a binary choice operator, to be interpreted probabilistically, as in $\Lambda_{\oplus}$ [6]. The construct $[M \parallel N] \rightarrowtail L$ corresponds to the so-called parallel disjunction operator: if the evaluation of M or N terminates, then the behaviour of $[M \parallel N] \rightarrowtail L$ is the same as the behaviour of L, otherwise this term does not terminate. Since we are in a probabilistic calculus, this means that $[M \parallel N] \rightarrowtail L$ converges to L with a probability that is equal to the probability that either M or N converge. (This formulation of parallel disjunction is equivalent to the binary one, without the third term.)

Example 1. Relevant examples of terms are $\Omega = (\lambda x.xx)\,(\lambda x.xx)$, and $I = \lambda x.x$: the first one always diverges, while the second always converges (to itself). In between, one can find terms such as $I \oplus \Omega$, and $I \oplus (I \oplus \Omega)$, converging with probability one half and three quarters, respectively.

2.1 Operational Semantics

Because of the probabilistic nature of choice in $\Lambda_{\oplus or}$, a program doesn't evaluate to a value, but to a probability distribution on values. Therefore, we need the following notions to define an evaluation relation.

Definition 2. Values *are terms of the form $V \ ::= \ \lambda x.M$. We will call $\mathcal{V}_{\oplus or}$ the set of values. A* value distribution *is a function $\mathscr{D} : \mathcal{V}_{\oplus or} \to [0,1]$, such that $\sum_{V \in \mathcal{V}_{\oplus or}} \mathscr{D}(V) \leq 1$. Given a value distribution $\mathscr{D}$, we let $\mathsf{S}(\mathscr{D})$ denote the set of those values V such that $\mathscr{D}(V) > 0$. Given a set X of values, $\mathscr{D}(X)$ is the sum of the probabilities assigned to every element of X, i.e., $\mathscr{D}(X) = \sum_{V \in X} \mathscr{D}(V)$. Moreover, we define $\sum \mathscr{D} = \sum_V \mathscr{D}(V)$, which corresponds to the total weight of the distribution $\mathscr{D}$. A value distribution $\mathscr{D}$ is finite whenever $\mathsf{S}(\mathscr{D})$ has finite cardinality. If V is a value, we write $\{V^1\}$ for the value distribution $\mathscr{D}$ such that $\mathscr{D}(W) = 1$ if $W = V$ and $\mathscr{D}(V) = 0$ otherwise. We'll note $\mathscr{D} \leq \mathscr{E}$ for the pointwise preorder on value distributions.*

We first define an *approximation* semantics, which attributes *finite* probability distributions to terms, and only later define the actual semantics, which is the least upper bound of all distributions obtained through the approximation semantics. Big-step semantics is given by means of a binary relation $\Downarrow$ between

$$\frac{}{M \Downarrow \emptyset}\, b_e \qquad \frac{}{V \Downarrow \{V^1\}}\, b_v \qquad \frac{M \Downarrow \mathscr{D} \qquad N \Downarrow \mathscr{E}}{M \oplus N \Downarrow \frac{1}{2}\mathscr{D} + \frac{1}{2}\mathscr{E}}\, b_s$$

$$\frac{M \Downarrow \mathscr{K} \qquad N \Downarrow \mathscr{F} \qquad \{P[V/x] \Downarrow \mathscr{E}_{P,V}\}_{\lambda x.P \in S(\mathscr{K}),\, V \in S(\mathscr{F})}}{MN \Downarrow \sum_{V \in S(\mathscr{F})} \mathscr{F}(V) \cdot \left(\sum_{\lambda x.P \in S(\mathscr{K})} \mathscr{K}(\lambda x.P) \cdot \mathscr{E}_{P,V}\right)}\, b_a$$

$$\frac{M \Downarrow \mathscr{D} \qquad N \Downarrow \mathscr{E} \qquad L \Downarrow \mathscr{F}}{[M \parallel N] \rightarrowtail L \Downarrow (\sum \mathscr{D} + \sum \mathscr{E} - (\sum \mathscr{D} \cdot \sum \mathscr{E})) \cdot \mathscr{F}}\, b_{or}$$

Fig. 1. Evaluation

closed terms and value distributions, which is defined by the set of rules from Figure 1. This evaluation relation is the natural extension to $\Lambda_{\oplus or}$ of the evaluation relation given in [6] for the untyped probabilistic λ-calculus. Since the calculus has a call-by-value evaluation strategy, function arguments are evaluated before being passed to functions.

Lemma 1. *For every term M, if $M \Downarrow \mathscr{D}$, and $M \Downarrow \mathscr{E}$, then there exists a distribution $\mathscr{F}$ such that $M \Downarrow \mathscr{F}$ with $\mathscr{D} \leq \mathscr{F}$, and $\mathscr{E} \leq \mathscr{F}$.*

Proof. The proof is by induction on the structure of derivations for $M \Downarrow \mathscr{D}$. We only consider two cases, since the others are the same as in [6]:

- If the derivation for $M \Downarrow \mathscr{D}$ is: $\frac{}{M \Downarrow \emptyset}\, b_e$: Then it is enough to take $\mathscr{F} = \mathscr{E}$, and since $\emptyset \leq \mathscr{E}$ and $\mathscr{E} \leq \mathscr{E}$, the result holds.
- If the derivation for $M \Downarrow \mathscr{D}$ is of the form:

$$\frac{P \Downarrow \mathscr{G} \qquad N \Downarrow \mathscr{H} \qquad L \Downarrow \mathscr{I}}{M = [P \parallel N] \rightarrowtail L \Downarrow \mathscr{D} = (\sum \mathscr{G} + \sum \mathscr{H} - (\sum \mathscr{G} \cdot \sum \mathscr{H})) \cdot \mathscr{I}}\, b_{or}$$

  Since $M = [P \parallel N] \rightarrowtail L$, there are only two possible structures for the derivation of $M \Downarrow \mathscr{E}$: either $\mathscr{E} = \emptyset$ and the result holds by $\mathscr{F} = \mathscr{D}$, or the structure of $M \Downarrow \mathscr{E}$ is the following:

$$\frac{P \Downarrow \mathscr{G}_2 \qquad N \Downarrow \mathscr{H}_2 \qquad L \Downarrow \mathscr{I}_2}{M = [P \parallel N] \rightarrowtail L \Downarrow \mathscr{E} = (\sum \mathscr{G}_2 + \sum \mathscr{H}_2 - (\sum \mathscr{G}_2 \cdot \sum \mathscr{H}_2)) \cdot \mathscr{I}_2}\, b_{or}$$

  By applying the induction hypothesis, we obtain that there exist $\mathscr{J}, \mathscr{K}, \mathscr{L}$ value distributions such that $P \Downarrow \mathscr{J}$, $N \Downarrow \mathscr{K}$, $L \Downarrow \mathscr{L}$, and, moreover, $\mathscr{G}, \mathscr{G}_2 \leq \mathscr{J}$, $\mathscr{H}, \mathscr{H}_2 \leq \mathscr{K}$, and $\mathscr{I}, \mathscr{I}_2 \leq \mathscr{L}$. We define $\mathscr{F} = (\sum \mathscr{J} + \sum \mathscr{K} - (\sum \mathscr{J} \cdot \sum \mathscr{K})) \cdot \mathscr{L}$, and we have that $M \Downarrow \mathscr{F}$. We must show that $\mathscr{D} \leq \mathscr{F}$ and $\mathscr{E} \leq \mathscr{F}$. Let $f : [0,1] \times [0,1] \to [0,1]$ be the function defined by $f(x,y) = x + y - x \cdot y$. The result follows from the fact that f is an increasing function, which holds since its two partial derivatives are positive. □

Definition 3. *For any closed term M, we define the* big-steps semantics $[\![M]\!]$ *of M as* $\sup_{M \Downarrow \mathscr{D}} \mathscr{D}$.

Since distributions form an ω-complete partial order, and for every M the set of those distributions $\mathscr{D}$ such that $M \Downarrow \mathscr{D}$ is a countable directed set (by Lemma 1), this definition is well-posed, and associates a unique value distribution to every term.

2.2 The Contextual Preorder

The general idea of the contextual preorder is the following: a term M is smaller than a term N if the probability of convergence of any program L where M occurs is less than or equal to the probability of convergence of the program obtained by replacing M by N in L. The notion of context allows us to formalize this idea.

Definition 4. *A* context C *of* $\Lambda_{\oplus or}$ *is a syntax tree with a unique hole:*

$$C ::= [\cdot] \mid \lambda x.C \mid CM \mid MC \mid C \oplus M \mid M \oplus C$$
$$\mid [C \parallel M] \rightarrowtail N \mid [M \parallel C] \rightarrowtail N \mid [M \parallel N] \rightarrowtail C\,.$$

We let $\mathscr{C}$ *denote the set of all contexts.*

Definition 5. *Terms* $M, N \in \Lambda_{\oplus or}(\overline{x})$ *are put in relation by the* contextual preorder *(*$M \leq N$*) if for every context* C *of* $\Lambda_{\oplus or}$ *such that* $C[M]$ *and* $C[N]$ *are closed terms, it holds that* $\sum[\![C[M]]\!] \leq \sum[\![C[N]]\!]$. M, N *are* contextually equivalent *(*$M = N$*) if* $M \leq N$*, and* $N \leq M$*.*

Note that the contextual preorder is directly defined on open terms, by requiring contexts to bind the free variables of terms. It is easy to verify that the contextual preorder is indeed a preorder, and analogously for equivalence.

Example 2. To see how things differ when we consider the contextual preorder in $\Lambda_{\oplus}$ and in $\Lambda_{\oplus or}$, consider the following terms of $\Lambda_{\oplus}$:

$$M = \lambda y.(\Omega \oplus I) \qquad N = (\lambda y.\Omega) \oplus (\lambda y.I).$$

where Ω and I are defined as in Example 1. We let $\leq_{\oplus}$ and $=_{\oplus}$ respectively denote the contextual preorder and equivalence for the language $\Lambda_{\oplus}$, i.e., the relations restricted to terms and contexts without the parallel disjunction construct. In [4] it is proved that $M \leq_{\oplus} N$. The converse does not hold, since if we take the $\Lambda_{\oplus}$ context

$$C = (\lambda x.(xI)(xI))[\cdot]$$

we have that in $C[M]$ the term $\lambda y.(\Omega \oplus I)$ is copied with probability one, while in $C[N]$ both term $\lambda y.\Omega$ and term $\lambda y.I$ are copied with probability one half. Hence, $C[M]$ converges with probability one quarter (i.e., the probability that $\Omega \oplus I$ converges two times in a row) while $C[N]$ has probability one half of diverging (i.e., one half times the probability that Ω diverges two times in a row) and one half of converging (i.e., one half times the probability that I converges two

times in a row). In $\Lambda_{\oplus or}$ we still have that $N \not\lesssim M$, since the contexts of $\Lambda_{\oplus}$ are contexts of $\Lambda_{\oplus or}$ as well, but we also have that $M \not\lesssim N$. Consider the context

$$C = (\lambda x.\, [(xI) \parallel (xI)] \rightarrowtail I\,)[\cdot]$$

If we put term M in context C then $\lambda y.(\Omega \oplus I)$ is copied, which has probability one half of converging when applied to I. Hence, by summing the probabilities of convergence of the two copies of $(\lambda y.(\Omega \oplus I))I$ and subtracting the probability that they both converge, we obtain that $[\![C[M]]\!] = \frac{3}{4} \cdot \{I^1\}$. Term $C[N]$ only converges with probability one half, since with one half probability we have the parallel disjunction of two terms that never converge and with one half probability we have the parallel disjunction of two terms that always converge. Hence, both in $\Lambda_{\oplus}$ and in $\Lambda_{\oplus or}$ terms M, N are not contextually equivalent, but it is only in $\Lambda_{\oplus or}$ that neither M is below N nor N is below M in the contextual preorder. We will see in the following section that this corresponds to what happens when we consider the simulation preorder.

3 Applicative Simulation

In this section we introduce the notions of probabilistic applicative simulation and bisimulation for $\Lambda_{\oplus or}$. Then we define probabilistic simulation and bisimulation on labelled Markov chains (LMCs, which also appear as Reactive Probabilistic Labelled Transition Systems in the literature). Bisimilarity on this class of structures was defined in [14]. We show how to define a labelled Markov chain representing terms of $\Lambda_{\oplus or}$ and their evaluation. Two states in the labelled Markov chain corresponding to terms M, N are in the simulation preorder (respectively, bisimilar) if and only if terms M, N are in the applicative simulation preorder (respectively: applicative bisimilar). Recall that, given a relation $\mathcal{R} \subseteq X \times Y$ and a set $Z \subseteq X$, $\mathcal{R}(Z) = \{y | \exists x \in Z \text{ such that } x\mathcal{R}y\}$.

Definition 6. *A relation $\mathcal{R} \subseteq \Lambda_{\oplus or} \times \Lambda_{\oplus or}$ is a probabilistic applicative simulation if $M\mathcal{R}N$ implies:*

- *for all $X \subseteq \mathcal{V}_{\oplus or}$, $[\![M]\!](X) \leq [\![N]\!](\mathcal{R}(X))$*
- *if $M = \lambda x.L$ and $N = \lambda x.P$ then $L[V/x]\mathcal{R}P[V/x]$ for all $V \in \mathcal{V}_{\oplus or}$.*

A relation $\mathcal{R}$ is a probabilistic applicative bisimulation if both $\mathcal{R}$ and $\mathcal{R}^{-1}$ are probabilistic applicative simulations. We say that M is simulated by N ($M \precsim_a N$) if there exists a probabilistic applicative simulation $\mathcal{R}$ such that $M\mathcal{R}N$. Terms M, N are bisimilar ($M \sim_a N$) if there exists a probabilistic applicative bisimulation $\mathcal{R}$ such that $M\mathcal{R}N$.

Definition 7. *A labelled Markov chain (LMC) is a triple $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$, where $\mathcal{S}$ is a countable set of* states*, $\mathcal{L}$ is a set of* labels*, and $\mathcal{P}$ is a* transition probability matrix*, i.e., a function $\mathcal{P} : \mathcal{S} \times \mathcal{L} \times \mathcal{S} \to \mathbb{R}$ such that for every state $s \in \mathcal{S}$ and for every label $l \in \mathcal{L}$, $\sum_{u \in \mathcal{S}} \mathcal{P}(s, l, u) \leq 1$.*

Definition 8. *Let $(\mathcal{S}, \mathcal{L}, \mathcal{P})$ be a labelled Markov chain. A* probabilistic simulation *is a relation $\mathcal{R}$ on $\mathcal{S}$ such that $(s, t) \in \mathcal{R}$ implies that for every $X \subseteq \mathcal{S}$ and*

for every $l \in \mathcal{L}$, $\mathcal{P}(s,l,X) \leq \mathcal{P}(t,l,\mathcal{R}(X))$. *A* probabilistic bisimulation *is a relation* $\mathcal{R}$ *on* $\mathcal{S}$ *such that both* $\mathcal{R}$ *and* $\mathcal{R}^{-1}$ *are probabilistic simulation relations. We say that* s *is simulated by* t ($s \precsim t$) *if there exists a probabilistic simulation* $\mathcal{R}$ *such that* $s\mathcal{R}t$. *States* s,t *are bisimilar* ($s \sim t$) *if there exists a probabilistic bisimulation* $\mathcal{R}$ *such that* $s\mathcal{R}t$.

Labelled Markov chains allow for external nondeterminism (every state can reach different probability distributions, depending on the chosen label) but they do not allow for internal nondeterminism (given a state and a label there is only one associated probability distribution). This is the reason why bisimilarity coincides with simulation equivalence on labelled Markov chains, i.e., $\sim = \precsim \cap \precsim^{-1}$.

Lemma 2. *For any labelled Markov chain* $(\mathcal{S},\mathcal{L},\mathcal{P})$*:*

1. *relations* $\precsim$ *and* $\sim$ *are the largest simulation and the largest bisimulation on* $\mathcal{S}$*, respectively;*
2. *relation* $\precsim$ *is a preorder and relation* $\sim$ *is an equivalence.*

Proof. Let us examine the two points separately:

1. Simulations and bisimulations are closed under union, hence the results follows.
2. The identity relation is a simulation, hence $\precsim$ is reflexive. Given two simulation relations $\mathcal{R}_1, \mathcal{R}_2$, relation $\mathcal{R}_1;\mathcal{R}_2 = \{(s,t) | s\mathcal{R}_1 u \mathcal{R}_2 t \text{ for some } u\}$ is a simulation. Hence, $\precsim$ is transitive as well. By definition, relation $\sim$ is symmetric, which implies that it is an equivalence. □

We will now define a labelled Markov chain that has among its states all terms of $\Lambda_{\oplus or}$ and that models the evaluation of these terms.

Definition 9. *The labelled Markov chain* $\mathcal{M}_{\oplus or} = (\mathcal{S}_{\oplus or}, \mathcal{L}_{\oplus or}, \mathcal{P}_{\oplus or})$ *is given by:*

- *A set of states* $\mathcal{S}_{\oplus or} = \{\Lambda_{\oplus or}\} \uplus \{\hat{\mathcal{V}}_{\oplus or}\}$*, where terms and values are taken modulo* α*-equivalence and* $\hat{\mathcal{V}}_{\oplus or} = \{\hat{V} | V \in \mathcal{V}_{\oplus or}\}$ *is a set containing copies of the values in* $\Lambda_{\oplus or}$ *decorated with* $\hat{}$*. We call these values* distinguished values*.*
- *A set of labels* $\mathcal{L}_{\oplus or} = \mathcal{V}_{\oplus or} \uplus \{eval\}$*, where, again, terms are taken modulo* α*-equivalence.*
- *A transition probability matrix* $\mathcal{P}_{\oplus or}$ *such that:*
  - *for every* $M \in \Lambda_{\oplus or}$ *and for every* $\hat{V} \in \hat{\mathcal{V}}_{\oplus or}$, $\mathcal{P}_{\oplus or}(M, eval, \hat{V}) = [\![M]\!](V)$ *and* $\mathcal{P}_{\oplus or}(M, eval, M') = 0$ *for all* $M' \in \Lambda_{\oplus or}$*.*
  - *for every* $\lambda\hat{x}.M \in \hat{\mathcal{V}}_{\oplus or}$ *and for every* $V \in \mathcal{V}_{\oplus or}$, $\mathcal{P}_{\oplus or}(\lambda\hat{x}.M, V, M[V/x]) = 1$ *and* $\mathcal{P}_{\oplus or}(\lambda\hat{x}.M, V, M') = 0$ *for all* $M' \in \Lambda_{\oplus or}$ *such that* $M' \neq M[V/x]$*.*

Please observe that if $V \in \mathcal{V}_{\oplus or}$, then both V and $\hat{V}$ are states of the Markov chain $\mathcal{M}_{\oplus or}$. A similar labelled Markov chain is defined in [13] for a call-by-name untyped probabilistic λ-calculus $\Lambda_{\oplus}$, and for a call-by-value typed probabilistic version of PCF in [4]. Actions in $\mathcal{V}_{\oplus or}$ and action *eval* respectively represent the application of a term to a value and the evaluation of a term.

Following [8], given a state and an action we allow the sum of the probabilities of reaching other states in the labelled Markov chain to be smaller than 1,

modelling divergence this way. The definition of simulation implies that whenever M is simulated by N we have that $\sum[\![M]\!] \leq \sum[\![N]\!]$. Analogously, if M is bisimilar to N, then $\sum[\![M]\!] = \sum[\![N]\!]$.

An applicative simulation $\mathcal{R}$ on terms of $\Lambda_{\oplus or}$ can be easily seen as a simulation relation $\mathcal{R}'$ on states of $\mathcal{M}_{\oplus or}$, obtained by adding to relation $\mathcal{R}$ the pairs $\{(\hat{V}, \hat{W}) | V \mathcal{R} W\}$. Analogously, a simulation relation on $\mathcal{M}_{\oplus or}$ corresponds to an applicative simulation for $\Lambda_{\oplus or}$.

Theorem 1. *On terms of* $\Lambda_{\oplus or}$, $\precsim_a = \precsim$ *and* $\sim_a = \sim$.

In what follows, we will mainly use the definitions of simulation and bisimulation for the labelled Markov chain $\mathcal{M}_{\oplus or}$. By Lemma 2, $\precsim$ coincides with the simulation preorder defined in [4], which requires simulations to be preorders themselves. For instance, I and II are (applicative) bisimilar since $\mathcal{R} =$

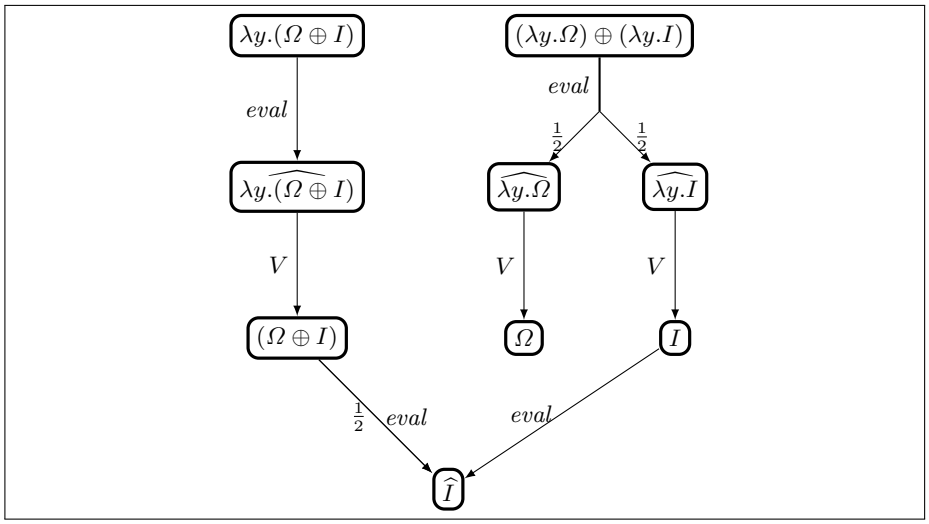

Fig. 2. LMC for M, N.

$\{(I, (II))\} \cup \mathcal{ID} \cup \{(\hat{V}, \hat{V}) | V \in \mathcal{V}_{\oplus or}\}$, where $\mathcal{ID}$ is the identity relation on $\Lambda_{\oplus or}$, is a bisimulation on $\mathcal{M}_{\oplus or}$. Consider now the terms M and N defined in Example 2 and represented in Figure 2 as states in $\mathcal{M}_{\oplus or}$. Term M is not simulated by N: if a simulation $\mathcal{R}$ relates them, then it must also relate term $(\Omega \oplus I)$ to both term Ω and term I. However, $(\Omega \oplus I)$ can perform *eval* and reach I with probability one half, while Ω has zero probability of becoming a value, which means that $\mathcal{R}$ cannot be a simulation relation. In the other direction, we have that N cannot be simulated by M either. If $\mathcal{R}$ is simulation such that $N \mathcal{R} M$ then it must relate term I to term $(\Omega \oplus I)$, but the former has probability one of convergence and the latter has probability one half of convergence.

4 The Simulation Preorder is a Precongruence

The extension $\precsim_\circ$ of the applicative simulation preorder to open terms is defined by considering all closing substitutions, i.e., for all $M, N \in \Lambda_{\oplus or}(x_1, \ldots, x_n)$, we have $M \precsim_\circ N$ if

$$M[V_1, \ldots, V_n/x_1, \ldots, x_n] \precsim_\circ N[V_1, \ldots, V_n/x_1, \ldots, x_n], \text{ for all } V_1, \ldots, V_n \in \mathcal{V}_{\oplus or}.$$

Here we show that $\precsim_\circ$ is a precongruence, i.e., closed with respect to the operators of $\Lambda_{\oplus or}$.

It is here convenient to work with generalizations of relations called $\Lambda_{\oplus or}$-relations, i.e. sets of triples in the form $(\overline{x}, M, N)$, where $M, N \in \Lambda_{\oplus or}(\overline{x})$. Given a relation $\mathcal{R}$ on open terms, if $M \mathcal{R} N$ and $M, N \in \Lambda_{\oplus or}(\overline{x})$ then the triple $(\overline{x}, M, N)$ is in the corresponding $\Lambda_{\oplus or}$-relation. We denote this by $\overline{x} \vdash M \mathcal{R} N$. We extend the usual notions of symmetry, reflexivity and transitivity to $\Lambda_{\oplus or}$-relations as expected.

Definition 10. *A $\Lambda_{\oplus or}$-relation $\mathcal{R}$ is compatible if and only if the following conditions hold:*
(Com1) $\forall \overline{x}, \forall x \in \overline{x},\ \overline{x} \vdash x \,\mathcal{R}\, x$;
(Com2) $\forall \overline{x}, \forall x \notin \overline{x}, \forall M, N,\ \overline{x} \cup \{x\} \vdash M \,\mathcal{R}\, N \implies \overline{x} \vdash \lambda x.M \,\mathcal{R}\, \lambda x.N$;
(Com3) $\forall \overline{x}, \forall M, N, P, Q,\ \overline{x} \vdash M \,\mathcal{R}\, N \wedge \overline{x} \vdash P \,\mathcal{R}\, Q \implies \overline{x} \vdash M P \,\mathcal{R}\, N Q$;
(Com4) $\forall \overline{x}, \forall M, N, P, Q,\ \overline{x} \vdash M \,\mathcal{R}\, N \wedge \overline{x} \vdash P \,\mathcal{R}\, Q \implies \overline{x} \vdash M \oplus P \,\mathcal{R}\, N \oplus Q$;
(Com5) $\forall \overline{x}, \forall M, N, P, Q, T,\ \overline{x} \vdash M \,\mathcal{R}\, N \wedge \overline{x} \vdash P \,\mathcal{R}\, Q \implies \overline{x} \vdash [M \parallel P] \rightarrowtail T \,\mathcal{R}\, [N \parallel Q] \rightarrowtail T$;

It follows from these properties that a compatible relation is reflexive, since this holds by *(Com1)* on variables, and it is preserved by the other operators by *(Com2)-(Com5)*:

Proposition 1. *If a relation is compatible, then it is reflexive.*

4.1 Howe's Method

The main idea of Howe's method consists in defining an auxiliary relation $\precsim_\circ^H$ such that it is easy to see that it is compatible, and then prove that $\precsim_\circ = \precsim_\circ^H$.

Definition 11. *Let $\mathcal{R}$ be a relation. We define inductively the relation $\mathcal{R}^H$ by the rules in Figure 3.*

We are now going to show that if the relation $\mathcal{R}$ we start from satisfies minimal requirements, namely that it is reflexive and transitive, then $\mathcal{R}^H$ is guaranteed to be compatible and to contain $\mathcal{R}$. This is a direct consequence of the following results, whose proofs are standard inductions:

- Let $\mathcal{R}$ be a reflexive relation. Then $\mathcal{R}^H$ is compatible.
- Let $\mathcal{R}$ be transitive. Then:

$$\left(\overline{x} \vdash M \,\mathcal{R}^H\, N\right) \wedge \left(\overline{x} \vdash N \,\mathcal{R}\, L\right) \Rightarrow \left(\overline{x} \vdash M \,\mathcal{R}^H\, L\right) \tag{1}$$

$$\frac{\overline{x} \cup \{x\} \vdash x \,\mathcal{R}\, M}{\overline{x} \cup \{x\} \vdash x \,\mathcal{R}^H\, M} \qquad \frac{\overline{x} \cup \{x\} \vdash M \,\mathcal{R}^H\, N \qquad \overline{x} \vdash \lambda x.N \,\mathcal{R}\, L}{\overline{x} \vdash \lambda x.M \,\mathcal{R}^H\, L}$$

$$\frac{\overline{x} \vdash M \,\mathcal{R}^H\, N \qquad \overline{x} \vdash L \,\mathcal{R}^H\, P \qquad \overline{x} \vdash NP \,\mathcal{R}\, R}{\overline{x} \vdash ML \,\mathcal{R}^H\, R}$$

$$\frac{\overline{x} \vdash M \,\mathcal{R}^H\, N \qquad \overline{x} \vdash L \,\mathcal{R}^H\, P \qquad \overline{x} \vdash N \oplus P \,\mathcal{R}\, R}{\overline{x} \vdash M \oplus L \,\mathcal{R}^H\, R}$$

$$\frac{\overline{x} \vdash M \,\mathcal{R}^H\, N \qquad \overline{x} \vdash L \,\mathcal{R}^H\, P \qquad \overline{x} \vdash [N \parallel P] \rightarrowtail T \,\mathcal{R}\, R}{\overline{x} \vdash [M \parallel L] \rightarrowtail T \,\mathcal{R}^H\, R}$$

Fig. 3. Howe's Construction

- If $\mathcal{R}$ is reflexive, then $\overline{x} \vdash M \,\mathcal{R}\, N$ implies $\overline{x} \vdash M \,\mathcal{R}^H\, N$.

We can now apply Howe's construction to $\precsim_\circ$, since it is clearly reflexive and transitive. The properties above then tell us that $\precsim_\circ^H$ is compatible and that $\precsim_\circ \subseteq \precsim_\circ^H$. What we are left with, then, is proving that $\precsim_\circ^H$ is also a simulation.[3]

Lemma 3. *$\precsim_\circ^H$ is value-substitutive: for all terms M, N and values V, W such that $x \vdash M \precsim_\circ^H N$ and $\emptyset \vdash V \precsim_\circ^H W$, it holds that $\emptyset \vdash M[V/x] \precsim_\circ^H N[W/x]$*

Proof. By induction on the derivation of $x \vdash M \precsim_\circ^H N$.

We also need an auxiliary, technical, lemma about probability assignments:

Definition 12. *$\mathbb{P} = \left(\{p_i\}_{1 \le i \le n}, \{r_I\}_{I \subseteq \{1,\ldots,n\}}\right)$ is said to be a* probability assignment *if for every $I \subseteq \{1,..,n\}$, it holds that $\sum_{i \in I} p_i \le \sum_{J \cap I \neq \emptyset} r_J$.*

Lemma 4 (Disentangling Sets). *Let $P = \left(\{p_i\}_{1 \le i \le n}, \{r_I\}_{I \subseteq \{1,\ldots,n\}}\right)$ be a probability assignment. Then for every non-empty $I \subseteq \{1, \ldots, n\}$, and for every $k \in I$, there is an $s_{k,I} \in [0,1]$ satisfying the following conditions:*

- *for every I, it holds that $\sum_{k \in I} s_{k,I} \le 1$;*
- *for every $k \in 1, \ldots, n$, it holds that $p_k \le \sum_{\{I \mid k \in I\}} s_{k,I} \cdot r_I$.*

The proof is an application of the Max-Flow Min-Cut Theorem, see e.g., [5, 4].

Given a set of set of open terms X, let $\lambda x.X = \{\lambda x.M | M \in X\}$.

Lemma 5 (Key Lemma). *For all terms M, N, if $\emptyset \vdash M \precsim_\circ^H N$, then for every $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$ it holds that $[\![M]\!]\,(\lambda x.X) \le [\![N]\!]\left(\precsim_\circ \left(\lambda x.\, \precsim_\circ^H (X)\right)\right)$.*

Proof. We show that the inequality holds for every approximation of the semantics of M, which implies the result since the semantics is the supremum of the

[3] In the proof of congruence for the probabilistic call-by-value λ-calculus presented in [4], the transitive closure of $\precsim_\circ^H$ is considered, since the definition of simulation required the relation to be preorder, which implies that the transitivity of $\precsim_\circ^H$ is needed. Since we relaxed the definition of simulation, this is not anymore necessary.

approximations. In particular, we prove by induction on the structure of the derivation of $M \Downarrow \mathscr{D}$ that, for any M, N, if $M \Downarrow \mathscr{D}$ and $\emptyset \vdash M \precsim^H_\circ N$, then for every $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$ it holds that $\mathscr{D}(\lambda x.X) \leq [\![N]\!]\left(\precsim_\circ \left(\lambda x. \precsim^H_\circ (X)\right)\right)$. We consider separately every possible rule which can be applied at the bottom of the derivation:

- If the rule is $\dfrac{}{M \Downarrow \emptyset}\, bv$ then $\mathscr{D} = \emptyset$, and for all set of values $\lambda x.X$, $\mathscr{D}(\lambda x.X) = 0$, and it concludes the proof.
- If M is a value $V = \lambda x.L$ and the last rule of the derivation is $\dfrac{}{V \Downarrow \{V^1\}}\, b_v$ then $\mathscr{D} = \{V^1\}$ is the Dirac distribution for V and, by the definition of Howe's lifting, $\left(\emptyset \vdash \lambda x.L \precsim^H_\circ N\right)$ was derived by the following rule:
$$\frac{x \vdash L \precsim^H_\circ P \qquad \emptyset \vdash \lambda x.P \precsim_\circ N}{\emptyset \vdash \lambda x.L \precsim^H_\circ N}$$
It follows from the definition of simulation and from $(\emptyset \vdash \lambda x.P \precsim_\circ N)$ that $1 = [\![N]\!](\precsim_\circ \{\lambda x.P\})$. Let $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$. If $\lambda x.L \notin \lambda x.X$ then $\mathscr{D}(\lambda x.X) = 0$ and the thesis holds. Otherwise, $\mathscr{D}(\lambda x.X) = \mathscr{D}(\lambda x.L) = 1 = [\![N]\!](\precsim_\circ \{\lambda x.P\})$. It follows from $L \precsim^H_\circ P$ and from $\lambda x.L \in \lambda x.X$ that $\lambda x.P \in \lambda x.(\precsim^H_\circ X)$; hence, $[\![N]\!](\precsim_\circ \{\lambda x.P\}) \leq [\![N]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))$.
- If the derivation of $M \Downarrow \mathscr{D}$ is of the following form:
$$\frac{M_1 \Downarrow \mathscr{K} \qquad M_2 \Downarrow \mathscr{F} \qquad \{P[V/x] \Downarrow \mathscr{E}_{P,V}\}_{\lambda x.P \in \mathsf{S}(\mathscr{K}), V \in \mathsf{S}(\mathscr{F})}}{M_1 M_2 \Downarrow \sum_{V \in \mathsf{S}(\mathscr{F})} \mathscr{F}(V) \left(\sum_{\lambda x.P \in \mathsf{S}(\mathscr{K})} \mathscr{K}(\lambda x.P).\mathscr{E}_{P,V}\right)}$$
Then $M = M_1 M_2$ and we have that the last rule used in the derivation of $\emptyset \vdash M \precsim^H_\circ N$ is:
$$\frac{\emptyset \vdash M_1 \precsim^H_\circ M_1' \qquad \emptyset \vdash M_2 \precsim^H_\circ M_2' \qquad \emptyset \vdash M_1' M_2' \precsim_\circ N}{\emptyset \vdash M_1 M_2 \precsim^H_\circ N}$$
Let $\mathsf{S}(\mathscr{K}) = \{\lambda x.P_1, \ldots, \lambda x.P_n\}$ and $K_i = \precsim_\circ \{\lambda x.L \,|\, x \vdash P_i \precsim^H_\circ L\}$ and, symmetrically, $\mathsf{S}(\mathscr{F}) = \{V_1, \ldots, V_l\}$ and $X_k = \precsim_\circ \{\lambda x.L \,|\, V_k = \lambda x.M'$ and $x \vdash M' \precsim^H_\circ L\}$. Then by the inductive hypothesis on $M_1 \Downarrow \mathscr{K}$ and $M_2 \Downarrow \mathscr{F}$ we have that $\mathscr{K}\left(\bigcup_{i \in I} \{\lambda x.P_i\}\right) \leq [\![M_1']\!]\left(\bigcup_{i \in I} K_i\right)$ for every $I \subseteq \{1, .., n\}$ and $\mathscr{F}(\bigcup_{k \in I} \{V_k\}) \leq [\![M_2']\!]\left(\bigcup_{k \in I} X_k\right)$ for every $I \subseteq \{1, .., l\}$.
Lemma 4 allows us to derive that for all $U \in \bigcup_{1 \leq i \leq n} K_i$ there exist probability values $r_1^U, \ldots, r_n^U$ and for all $W \in \bigcup_{1 \leq k \leq l} X_k$ there exist probability values $s_1^W, .., s_l^W$ such that:
$$[\![M_1']\!](U) \geq \sum_{1 \leq i \leq n} r_i^U \quad [\![M_2']\!](W) \geq \sum_{1 \leq k \leq l} s_k^W \quad \forall U \in \bigcup_{1 \leq i \leq n} K_i, W \in \bigcup_{1 \leq k \leq l} X_k$$
$$\mathscr{K}(\lambda x.P_i) \leq \sum_{U \in K_i} r_i^U \quad \mathscr{F}(V_k) \leq \sum_{W \in X_k} s_k^W \qquad \forall\, 1 \leq i \leq n, 1 \leq k \leq l$$
Hence, for every value $Z \in \mathcal{V}_{\oplus or}$, we have that:
$$\begin{aligned}\mathscr{D}(Z) &= \sum_{1 \leq k \leq l} \mathscr{F}(V_k) \cdot \sum_{1 \leq i \leq n} \mathscr{K}(\lambda x.P_i) \cdot \mathscr{E}_{P_i,V_k}(Z) \\ &\leq \sum_{1 \leq k \leq l} \sum_{W \in X_k} s_k^W \cdot \sum_{1 \leq i \leq n} \sum_{U \in K_i} r_i^U \cdot \mathscr{E}_{P_i,V_k}(Z)\end{aligned}$$

If $U = \lambda x.U' \in K_i$ then there exists S such that:

$$(2) \quad \emptyset \vdash \lambda x.S \precsim_\circ U \qquad (3) \quad x \vdash P_i \ \precsim^H_\circ \ S$$

By (2), $\emptyset \vdash S[W/x] \precsim_\circ U'[W/x]$. By (3) and by Lemma 3, for $W \in X_k$ we have that $\emptyset \vdash P_i[V_k/x] \precsim^H_\circ S[W/x]$. It follows from (1) that $\emptyset \vdash P_i[V_k/x] \precsim^H_\circ U'[W/x]$. Hence, by the induction hypothesis applied to $P_i[V_k/x]$ we have $\mathscr{E}_{P_i,V_k}(\lambda x.X) \leq [\![U'[W/x]]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))$. Therefore,

$$\begin{aligned}
&\mathscr{D}(\lambda x.X) \leq \sum_{1\leq k\leq l} \sum_{W\in X_k} s^W_k \cdot \sum_{1\leq i\leq n} \sum_{U\in K_i} r^U_i \cdot \mathscr{E}_{P_i,V_k}(\lambda x.X) \\
&\leq \sum_{W\in \bigcup\limits_{1\leq k\leq l} X_k} \ \sum_{U\in \bigcup\limits_{1\leq i\leq n} K_i} \Big(\sum_{\{k|W\in X_k\}} s^W_k \Big) \cdot \Big(\sum_{\{i|U\in K_i\}} r^U_i \Big) [\![L_{U,W}]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X)) \\
&\leq \sum_{W\in \bigcup\limits_{1\leq k\leq l} X_k} \ \sum_{U\in \bigcup\limits_{1\leq i\leq n} K_i} [\![M'_2]\!](W) \cdot [\![M'_1]\!](U) \cdot [\![L_{U,W}]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X)) \\
&\leq [\![M'_1 M'_2]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))
\end{aligned}$$

where $L_{U,W} = U'[W/x]$ for any U such that $U = \lambda x.U'$.

- If $M \Downarrow \mathscr{D}$ is derived by:

$$\frac{M_1 \Downarrow \mathscr{D}_1 \qquad M_2 \Downarrow \mathscr{D}_2}{M_1 \oplus M_2 \Downarrow \frac{1}{2}\mathscr{D}_1 + \frac{1}{2}\mathscr{D}_2}$$

then $\emptyset \vdash M \ \precsim^H_\circ \ N$ is derived by:

$$\frac{\emptyset \vdash M_1 \ \precsim^H_\circ \ N_1 \qquad \emptyset \vdash M_2 \ \precsim^H_\circ \ N_2 \qquad \emptyset \vdash N_1 \oplus N_2 \precsim_\circ N}{\emptyset \vdash M_1 \oplus M_2 \ \precsim^H_\circ \ N}$$

By the inductive hypothesis, for $i \in \{1,2\}$ we have that for any $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$,

$$\mathscr{D}_i(\lambda x.X) \leq [\![N_i]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))$$

Hence, the result follows from:

$$\frac{1}{2}\cdot\mathscr{D}_1(\lambda x.X)+\frac{1}{2}\cdot\mathscr{D}_2(\lambda x.X) \leq \frac{1}{2}\cdot[\![N_1]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))+\frac{1}{2}\cdot[\![N_2]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))$$

- If the last rule applied in the derivation of $M \Downarrow \mathscr{D}$ is of the following form:

$$\frac{M_1 \Downarrow \mathscr{D}_1 \qquad M_2 \Downarrow \mathscr{D}_2}{[M_1 \parallel M_2] \rightarrowtail T \Downarrow (\sum \mathscr{D}_1 + \sum \mathscr{D}_2 - \sum \mathscr{D}_1 \cdot \sum \mathscr{D}_2) \cdot \{T^1\}}$$

then $M = [M_1 \parallel M_2] \rightarrowtail T$ and $\emptyset \vdash M \ \precsim^H_\circ \ N$ is derived by:

$$\frac{\emptyset \vdash M_1 \ \precsim^H_\circ \ N_1 \qquad \emptyset \vdash M_2 \ \precsim^H_\circ \ N_2 \qquad \emptyset \vdash \ [N_1 \parallel N_2] \rightarrowtail T \precsim_\circ N}{\emptyset \vdash \ [M_1 \parallel M_2] \rightarrowtail T \ \precsim^H_\circ \ N}$$

By inductive hypothesis on $M_1 \Downarrow \mathscr{D}_1$ we have that for any $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$, $\mathscr{D}_1(\lambda x.X) \leq [\![N_1]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X))$. Hence, for $\lambda x.X = \mathsf{S}(\mathscr{D}_1)$ we have that:

$$\sum \mathscr{D}_1 = \mathscr{D}_1(\lambda x.X) \leq [\![N_1]\!](\precsim_\circ \lambda x.(\precsim^H_\circ X)) \leq [\![N_1]\!](\mathsf{S}([\![N_1]\!])) = \sum [\![N_1]\!]$$

and, symmetrically, by the inductive hypothesis on $M_2 \Downarrow \mathscr{D}_2$ we have $\sum \mathscr{D}_2 \leq \sum [\![N_2]\!]$. Therefore,

$$\sum \mathscr{D}_1 + \sum \mathscr{D}_2 - \sum \mathscr{D}_1 \cdot \sum \mathscr{D}_2 \leq \sum [\![N_1]\!] + \sum [\![N_2]\!] - \sum [\![N_1]\!] \cdot \sum [\![N_2]\!]$$

Let $\lambda x.X \subseteq \mathcal{V}_{\oplus or}$. If $T \notin \lambda x.X$ then $\mathscr{D} = 0$ and the result follows. Otherwise, it follows from $T = \lambda x.T' \in \precsim_\circ \lambda x.(\precsim_\circ^H \{T'\})$ (since both $\precsim_\circ$ and $\precsim_\circ^H$ are reflexive) that

$$\begin{aligned}\mathscr{D}(\lambda x.X) = \mathscr{D}(\lambda x.T') &= \sum \mathscr{D}_1 + \sum \mathscr{D}_2 - \sum \mathscr{D}_1 \cdot \sum \mathscr{D}_2 \\ &\leq \sum [\![N_1]\!] + \sum [\![N_2]\!] - \sum [\![N_1]\!] \cdot \sum [\![N_2]\!] \\ &= [\![N]\!](\lambda x.T') = [\![N]\!](\precsim_\circ \lambda x.(\precsim_\circ^H X))\end{aligned}$$

□

A consequence of the Key Lemma, then, is that relation $\precsim_\circ^H$ on closed terms is an applicative simulation, thus included in the largest one, namely $\precsim$. Hence, if M, N are open terms and $x_1, \ldots, x_n \vdash M \precsim_\circ^H N$ then it follows from Lemma 3 that for all $V_1, \ldots, V_n, W_1, \ldots, W_n$ such that $\emptyset \vdash V_i \precsim_\circ^H W_i$ we have that $\emptyset \vdash M[V_1, \ldots, V_n/x_1, \ldots, x_n] \precsim_\circ^H N[W_1, \ldots, W_n/x_1, \ldots, x_n]$, which implies (by the reflexivity of $\precsim_\circ^H$ and by $\precsim_\circ^H \subseteq \precsim_\circ$ on closed terms) that for all $V_1, \ldots, V_n$ we have that $\emptyset \vdash M[V_1, \ldots, V_n/x_1, \ldots, x_n] \precsim_\circ N[V_1, \ldots, V_n/x_1, \ldots, x_n]$, i.e., $M \precsim_\circ N$. Since $\precsim_\circ$ is itself included in $\precsim_\circ^H$, we obtain that $\precsim_\circ = \precsim_\circ^H$. Hence, it follows from the transitivity of $\precsim_\circ$ and from the fact that $\precsim_\circ^H$ is compatible that:

Theorem 2 (Congruence). *$\precsim_\circ$ is a precongruence .*

The congruence of $\precsim_\circ$ allows us to prove that it is a sound with respect to the contextual preorder.

Theorem 3 (Soundness). *If $M \precsim_\circ N$ then $M \leq N$.*

Proof. Let $M \precsim_\circ N$. Using Theorem 2, it can be easily proved by induction on C that for any context C it holds that $C[M] \precsim_\circ C[N]$. If $C[M] \precsim_\circ C[N]$ then $\sum [\![C[M]]\!] \leq \sum [\![C[M]]\!]$, which implies the result. □

5 Full Abstraction

In [24], both bisimilarity and similarity on labelled Markov chains are characterised by a language of test, refining the testing characterization of bisimilarity presented in [14]. This characterisation is used in [4] to show that the bisimilarity relation on terms is fully abstract with respect to the contextual equivalence. The language of tests used to characterize bisimulation is the following:

Definition 13. *Let $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$ be a LMC. The* test language *$\mathscr{T}_0(\mathcal{M})$ is given by the grammar $t ::= \omega \mid a \cdot t \mid \langle t, t \rangle$, where $a \in \mathcal{L}$.*

This language represents tests in the following sense: for any t in the test language $\mathscr{T}_0(\mathcal{M})$, and for any s state of $\mathcal{M}$, we can define the probability $\Pr(s,t)$ that the test t succeeds when executed on s.

The full-abstraction result in [4] is based on the fact that, when we consider the particular Markov chain used to define a bisimulation relation on terms, any of these tests can actually be simulated by a context. However, the characterisation of the simulation preorder requires to add disjunctive tests:

Definition 14. *Let $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$ be a LMC. The* test language $\mathscr{T}_1(\mathcal{M})$ *is given by the grammar $t ::= \omega \mid a \cdot t \mid \langle t, t \rangle \mid t \vee t$, where $a \in \mathcal{L}$.*

We are now going to define the success probability of a test. The success probability of ω is 1 no matter what state we are starting from. The success probability of a disjunctive test corresponds to the probability that at least one of the two tests is successful.

Definition 15. *Let $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$ be a LMC. For all $s \in \mathcal{S}$, and $t \in \mathscr{T}_1(\mathcal{M})$, we define:*

$$Pr(s,\omega) = 1; \qquad Pr(s, t \vee u) = Pr(s,t) + Pr(s,u) - Pr(s,t) \cdot Pr(s,u)$$
$$Pr(s, \langle t, u \rangle) = Pr(s,t) \cdot Pr(s,u); \qquad Pr(s, a \cdot t) = \textstyle\sum_{s' \in \mathcal{S}} \mathcal{P}(s,a,s') \cdot Pr(s',t).$$

The following theorem characterises bisimilarity and the simulation preorder on labelled Markov chains by means of sets of tests.

Theorem 4 ([24]). *Let $\mathcal{M} = (\mathcal{S}, \mathcal{L}, \mathcal{P})$ be a LMC and let $s, s' \in \mathcal{S}$. Then:*
- *$s \sim s'$ if and only if for every $t \in \mathscr{T}_0(\mathcal{M})$ it holds that: $Pr(s,t) = Pr(s',t)$*
- *$s \precsim s'$ if and only if for every $t \in \mathscr{T}_1(\mathcal{M})$ it holds that $Pr(s,t) \leq Pr(s',t)$*

Example 3. Consider the two terms $M = \lambda x.(I \oplus \Omega)$ and $N = (\lambda x.I) \oplus (\lambda x.\Omega)$ from Example 2. We already know that, since they do not verify $M \precsim N$, there exists a test $t \in \mathscr{T}_1(\mathcal{M}_{\oplus or})$ whose success probability when executed on M is strictly greater that its success probability when executed on N. We can actually explicitly give such a test: let $t = eval \cdot (I \cdot eval \cdot \omega \vee I \cdot eval \cdot \omega)$ Then it holds that:

$$\Pr(\lambda x.(I \oplus \Omega), t) = \frac{3}{4}; \qquad \Pr((\lambda x.I) \oplus (\lambda x.\Omega), t) = \frac{1}{2}.$$

5.1 From Tests to Contexts

It is shown in [4] that simulation is not fully abstract for $\mathsf{PCFL}_{\oplus}$ with respect to the contextual preorder: a direct consequence is that disjunctive tests cannot be simulated by contexts. In other terms, it is not possible to write a program that has access to two sub-programs, and terminates with a probability equal to the probability that at least one of its sub-programs terminates. The proof of [4] is based on an encoding from $\mathscr{T}_0(\mathcal{M}_{\oplus})$ to the set of contexts. We are going to extend it into two encodings from $\mathscr{T}_1(\mathcal{M}_{\oplus or})$ to the set of contexts of $\Lambda_{\oplus or}$: one encoding expresses the action of tests on states of the form M, and the other

one on states of the form $\hat{V}$. The intuitive idea behind Θ^{val} and Θ^{term} is the following: if we take a test t, its success probability starting from the state M is the same as the convergence probability of the context $\Theta^{term}(t)$ filled by M, and similarly, its success probability starting from the state $\hat{V}$ is the same as the convergence probability of the context $\Theta^{term}(t)$ filled by V.

Definition 16. *Let $\Theta^{val} : \mathscr{T}_1(\mathcal{M}_{\oplus or}) \to \mathscr{C}$ and $\Theta^{term} : \mathscr{T}_1(\mathcal{M}_{\oplus or}) \to \mathscr{C}$ be defined by:*

$$\begin{aligned}
\Theta^{term}(\omega) &= \lambda x.[\cdot]; & \Theta^{val}(\omega) &= \lambda x.[\cdot];\\
\Theta^{term}(V \cdot t) &= \Omega[\cdot]; & \Theta^{val}(V \cdot t) &= \Theta^{term}(t)[([\cdot]V)];\\
\Theta^{term}(eval \cdot t) &= \lambda x.(\Theta^{val}(t)[x])[\cdot]; & \Theta^{val}(eval \cdot t) &= \Omega[\cdot];\\
\Theta^{term}(t \vee u) &= g(\Theta^{term}(t), \Theta^{term}(u)); & \Theta^{val}(t \vee u) &= g(\Theta^{val}(t), \Theta^{val}(u));\\
\Theta^{term}(\langle t, u\rangle) &= f(\Theta^{term}(t), \Theta^{term}(u)); & \Theta^{val}(\langle t, u\rangle) &= f(\Theta^{val}(t), \Theta^{val}(u));
\end{aligned}$$

where $f, g : \mathscr{C} \times \mathscr{C} \to \mathscr{C}$ are defined by:

$$\begin{aligned}
f(C, D) &= (\lambda x.(\lambda y, z.I)(C[xI])(D[xI]))(\lambda x.[\cdot]);\\
g(C, D) &= (\lambda x.(\,[C[xI] \parallel D[xI]] \rightarrowtail I\,)(\lambda x.[\cdot]).
\end{aligned}$$

The apparently complicated structure of f and g comes from the fact that we cannot construct contexts with several holes. However, since our language has copying capability, we can emulate contexts with several holes by means of contexts with only one hole. Intuitively, we could say that $g(C, D)$ would correspond to a multihole context $[C \parallel D] \rightarrowtail I$. Please observe that the encoding of the fragment of $\mathscr{T}_1(\mathcal{M}_{\oplus or})$ corresponding to $\mathscr{T}_0(\mathcal{M}_{\oplus or})$ does not use parallel disjunction, i.e., the image of $\mathscr{T}_0(\mathcal{M}_{\oplus or})$ by the encoding is a subset of $\Lambda_{\oplus}$. We can now apply this encoding to the test we defined in Example 3.

Example 4. Recall the test $t = eval \cdot (I \cdot eval \cdot \omega \vee I \cdot eval \cdot \omega)$ defined in Example 3. We can apply the embedding to this particular test:

$$\Theta^{term}(t) = (\lambda x.\,(\lambda z.\,[(\lambda y.(\lambda w.y))zII \parallel (\lambda y.(\lambda w.y))zII] \rightarrowtail I\,)\,(\lambda y.x))\,[\cdot].$$

We can see that if we consider the terms $M = \lambda x.(I \oplus \Omega)$ and $N = (\lambda x.I) \oplus (\lambda x.\Omega)$ defined in Example 2, the context $\Theta^{term}(t)$ simulates the test t with respect to M and N:

$$\Pr(M, t) = \sum [\![\Theta^{term}(t)[M]]\!]; \qquad \Pr(N, t) = \sum [\![\Theta^{term}(t)[N]]\!].$$

Theorem 5. *Let t be a test in $\mathscr{T}_1(\mathcal{M}_{\oplus or})$. Then for every M closed term and every V closed value it holds that:*

$$Pr(M, t) = \sum [\![\Theta^{term}(t)[M]]\!]; \qquad Pr(\hat{V}, t) = \sum [\![\Theta^{val}(t)[V]]\!].$$

Proof. We are going to show the thesis by induction on the structure of t.

- If $t = \omega$, then for every closed term M, and every closed value V, $\Pr(M, \omega) = \Pr(\hat{V}, \omega) = 1$, and we have defined $\Theta^{term}(\omega) = \Theta^{val}(\omega) = \lambda x.[\cdot]$. Since $\Theta^{term}(\omega)[M]$ and $\Theta^{val}(\omega)[V]$ are values, the weight of their semantics is 1, and so the result holds.
- If $t = \langle u_1, u_2 \rangle$, we can directly adapt the construction proposed in [4] to the untyped case. By the inductive hypothesis, for all $1 \leq i \leq 2$ it holds that for every closed term M and every closed value V,

$$\Pr(M, u_i) = \sum [\![\Theta^{term}(u_i)[M]]\!]; \Pr(\hat{V}, u_i) = \sum [\![\Theta^{val}(u_i)[V]]\!].$$

  The overall effect of f is to copy the content of the hole into the holes of the two contexts C and D. For any closed term M, we can express the convergence probability of $f(C, D)[M]$ as a function of the convergence probability of $C[M]$ and $D[M]$:

$$\begin{aligned}\sum [\![f(C, D)[M]]\!] &= \left(\sum [\![C[(\lambda x.M)I]]\!]\right) \cdot \left(\sum [\![D[(\lambda x.M)I]]\!]\right) \\ &= \left(\sum [\![C[M]]\!]\right) \cdot \left(\sum [\![D[M]]\!]\right)\end{aligned}$$

  Please recall that we have defined:

$$\begin{aligned}\Theta^{term}(\langle u_1, u_2 \rangle) &= f(\Theta^{term}(u_1), \Theta^{term}(u_2)) \\ \Theta^{val}(\langle u_1, u_2 \rangle) &= f(\Theta^{val}(u_1), \Theta^{val}(u_2))\end{aligned}$$

  We have that, for any closed term M, and any closed value V:

$$\begin{aligned}\sum [\![\Theta^{term}(\langle u_1, u_2 \rangle)[M]]\!] &= \Pr(M, u_1) \cdot \Pr(M, u_2) = \Pr(M, \langle u_1, u_2 \rangle) \\ \sum [\![\Theta^{val}(\langle u_1, u_2 \rangle)[V]]\!] &= \Pr(\hat{V}, u_1) \cdot \Pr(\hat{V}, u_2) = \Pr(\hat{V}, \langle u_1, u_2 \rangle)\end{aligned}$$

- Now the case $t = u_1 \vee u_2$. By the inductive hypothesis, for all $1 \leq i \leq 2$ it holds that for every closed term M and every closed value V,

$$\Pr(M, u_i) = \sum [\![\Theta^{term}(u_i)[M]]\!] \qquad \Pr(\hat{V}, u_i) = \sum [\![\Theta^{val}(u_i)[V]]\!].$$

  The definition of g allows us to show:

$$\sum [\![g(C, D)[M]]\!] = \sum [\![C[M]]\!] + \sum [\![D[M]]\!] - \sum [\![C[M]]\!] \cdot \sum [\![D[M]]\!]$$

  and now it is straightforward to see that:

$$\begin{aligned}\sum [\![\Theta^{term}(u_1 \vee u_2)[M]]\!] &= \Pr(M, u_1 \vee u_2); \\ \sum [\![\Theta^{val}(u_1 \vee u_2)[V]]\!] &= \Pr(\hat{V}, u_1 \vee u_2).\end{aligned}$$

- If $t = a \cdot u$, there are two different kinds of actions:

- when $a = eval$, we first consider $\Theta^{val}(t)$: since the *eval* action is relevant only for states of $\mathcal{M}_{\oplus or}$ which are terms (and not distinguished values), we want that $\Theta^{val}(t)[V]$ always diverges. Since $\Theta^{val}(t) = \Omega[\cdot]$ and since $[\![\Omega]\!] = \emptyset$, we have that for any closed value V, $[\![\Theta^{val}(t)[V]]\!] = \emptyset$.
  Now, we consider $\Theta^{term}(t)$. By the inductive hypothesis, we know that:

$$\Pr(\hat{V}, u) = \sum [\![\Theta^{val}(u)[V]]\!].$$

  Please recall that we have defined: $\Theta^{term}(a \cdot u) = \lambda x.(\Theta^{val}(u)[x])[\cdot]$. Let be M a closed term. Then it holds that:

$$\begin{aligned}\sum [\![\Theta^{term}(a \cdot u)[M]]\!] &= \sum_V [\![M]\!](V) \cdot \sum [\![\Theta^{val}(u)[V]]\!] \\ &= \sum_V [\![M]\!](V) \cdot \Pr(\hat{V}, u) \\ &= \sum_{e \in \mathcal{S}_{\oplus or}} \mathcal{P}_{\oplus or}(M, eval, e) \cdot \Pr(e, u) = \Pr(M, u)\end{aligned}$$

- When $a = V$, with $V \in \mathcal{V}_{\oplus or}$, we consider first $\Theta^{term}(V \cdot u)$. It has been designed to be a context which diverges whatever its argument is, and so we indeed have: $\Pr(M, V \cdot u) = 0 = \sum [\![\Theta^{term}(V \cdot u)[M]]\!]$. Then we consider $\Theta^{val}(t)$. Recall that we have defined: $\Theta^{val}(V \cdot u) = \Theta^{term}(u)[[\cdot]V]$. Let $W = \lambda x.M$ be a closed value:

$$\begin{aligned}\sum [\![\Theta^{val}(V \cdot u)[W]]\!] &= \sum [\![\Theta^{term}(u)[WV]]\!] \\ &= \Pr(WV, u) \\ &= \Pr(M[x/V], u) \qquad \text{since } [\![WV]\!] = [\![M[x/V]]\!] \\ &= \Pr(W, V \cdot u).\end{aligned}$$

□

Theorem 6. *$\precsim$ is fully abstract with respect to the contextual preorder.*

Proof. We already know that $\precsim$ is sound, that is $\precsim \subseteq \leq$. Hence, what is left to show is that $\leq \subseteq \precsim$, which follows from Theorem 5. Let M and N be two closed terms such that $M \leq N$. We want to show that $M \precsim N$. The testing characterisation of simulation allows us to say that it is sufficient to show that, for every test $t \in \mathcal{T}_1(\mathcal{M}_{\oplus or})$, $\Pr(M, t) \leq \Pr(N, t)$, which in turn is a consequence of Theorem 5, since every test t of $\mathcal{T}_1(\mathcal{M}_{\oplus or})$ can be simulated by a context of $\Lambda_{\oplus or}$.

References

1. S. Abramsky. The Lazy λ-Calculus. In D. Turner, editor, *Research Topics in Functional Programming*, pages 65–117. Addison Wesley, 1990.
2. S. Abramsky and C.-H. L. Ong. Full abstraction in the lazy lambda calculus. *Inf. Comput.*, 105(2):159–267, 1993.

3. D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence,*, 25(5):564–577, 2003.
4. R. Crubillé and U. D. Lago. On probabilistic applicative bisimulation and call-by-value λ-calculi (long version). *CoRR*, abs/1401.3766, 2014.
5. U. Dal Lago, D. Sangiorgi, and M. Alberti. On coinductive equivalences for higher-order probabilistic functional programs. In *POPL*, pages 297–308, 2014.
6. U. Dal Lago and M. Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO - Theor. Inf. and Applic.*, 46(3):413–450, 2012.
7. V. Danos and R. Harmer. Probabilistic game semantics. *ACM Trans. Comput. Log.*, 3(3):359–382, 2002.
8. J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled markov processes. *Inf. Comput.*, 179(2):163–193, 2002.
9. T. Ehrhard, C. Tasson, and M. Pagani. Probabilistic coherence spaces are fully abstract for probabilistic PCF. In *POPL*, pages 309–320, 2014.
10. S. Goldwasser and S. Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.
11. N. D. Goodman. The principles and practice of probabilistic programming. In *POPL*, pages 399–402, 2013.
12. C. Jones and G. D. Plotkin. A probabilistic powerdomain of evaluations. In *LICS*, pages 186–195, 1989.
13. U. D. Lago, D. Sangiorgi, and M. Alberti. On coinductive equivalences for higher-order probabilistic functional programs (long version). *CoRR*, abs/1311.1722, 2013.
14. K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94(1):1–28, 1991.
15. S. B. Lassen. *Relational Reasoning about Functions and Nondeterminism.* PhD thesis, University of Aarhus, 1998.
16. C. D. Manning and H. Schütze. *Foundations of statistical natural language processing*, volume 999. MIT Press, 1999.
17. C.-H. L. Ong. Non-determinism in a functional setting. In *LICS*, pages 275–286, 1993.
18. S. Park, F. Pfenning, and S. Thrun. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.*, 31(1), 2008.
19. J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.
20. A. Pfeffer. IBAL: A probabilistic rational programming language. In *IJCAI*, pages 733–740. Morgan Kaufmann, 2001.
21. G. D. Plotkin. LCF considered as a programming language. *Theor. Comput. Sci.*, 5(3):223–255, 1977.
22. N. Ramsey and A. Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *POPL*, pages 154–165, 2002.
23. S. Thrun. Robotic mapping: A survey. *Exploring artificial intelligence in the new millennium*, pages 1–35, 2002.
24. F. van Breugel, M. W. Mislove, J. Ouaknine, and J. Worrell. Domain theory, testing and simulation for labelled markov processes. *Theor. Comput. Sci.*, 333(1-2):171–197, 2005.