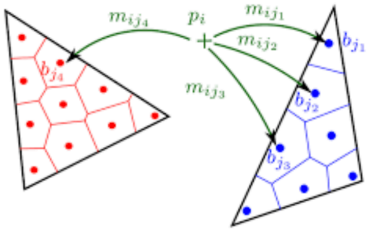
AGENCE NATIONALE DE LA RECHERCHE

ANR

REPAS

RELIABLE AND PRIVACY-AWARE SOFTWARE SYSTEMS



Deliverable D2.a (1)

Metric Reasoning About λ-Terms: The General Case (Long Version)

1 Introduction

Probability theory offers computer science models which enable system abstraction (at the price of introducing uncertainty), but which can also be seen as a *a way to compute*, like in randomized computation. Domains in which probabilistic models play a key role include machine learning [28], robotics [35], and linguistics [25]. In cryptography, on the other hand, having access to a source of uniform randomness is essential to achieve security, e.g., in the public key setting [20]. This has stimulated the development of concrete and abstract programming languages, which most often are extensions of their deterministic siblings. Among the many ways probabilistic choice can be captured in programming, the simplest one consists in endowing the language of programs with an operator modelling the flipping of a fair coin. This renders program evaluation a probabilistic process, and under mild assumptions the language becomes universal for probabilistic computation. Particularly fruitful in this sense has been the line of work on the functional paradigm, both at a theoretical [22, 30, 27] and at a more practical level [21].

We are still far, however, from a satisfactory understanding of higher-order probabilistic computation. As an example, little is known about how much of the classic, beautiful, theory underlying the λ-calculus [1] can be lifted to probabilistic λ-calculi, although the latter have been known from forty years now [31]. Until the beginning of this decade, indeed, most investigations were directed towards domain theory, which has been proved to be much more involved in presence of probabilistic choice than in a deterministic scenario [23]. In the last ten years, however, some promising results have appeared. As an example, both quantitative semantics and applicative bisimilarity have been shown to coincide with context equivalence for certain kinds of probabilistic λ-calculi [14, 5]. This not only provides us with new proof methodologies for program equivalence, but also sheds new light on the very nature of probabilistic higher-order computation. As an example, recent results tell us that program equivalence in presence of probabilistic choice lies somehow in between determinism and non-determinism [5].

But are equivalences the most proper way to compare terms? Actually, this really depends on what the underlying observable is. If observables are boolean, then equivalences (and preorders) are indeed natural choices: two programs are dubbed equivalent if they give rise to the same observable (of which there are just two!) in any context. If, on the other hand, the observable is an element of a metric space, which happens for example when we observe (the probability of) convergence in a probabilistic setting, one may wonder whether replacing equivalences with metrics makes sense. This is a question that has recently been given a positive answer in the *affine* setting [7], i.e., in a λ-calculus in which copying is simply not available. More specifically, a notion of context distance has been shown to model differences between terms satisfactorily, and has also been shown to be characterized by notions of trace metrics, and to be approximated from below by behavioural metrics.

Affine λ-calculi are very poor in terms of the computations they are able to model. Measuring the distance between terms in presence of copying, however, is bound to be problematic. On the one hand, allowing contexts to copy their argument has the potential risk of *trivialising* the underlying metric. On the other hand, finding handier characterizations of the obtained notion of metric in the style of behavioural or trace metrics is inherently hard. A more thorough discussion

on these issues can be found in Section 2 below.

In this paper, we attack the problem of analyzing the distance between λ-terms in its full generality. More specifically, the contributions of this paper are fourfold:

- First of all, we define a linear probabilistic λ-calculus, called $\Lambda^{!,\|}_{\oplus}$, in which copying and a nonstandard construct, namely Plotkin's parallel disjunction, are both available. A very liberal type system prevents deadlocks, but nevertheless leaves the expressive power of the calculus very high. This choice has been motivated by our will to put ourselves in the most general setting, so as to be able to talk about different fragments. The calculus is endowed with a notion of context distance, in Morris' style. This is covered in Section 3 below.
- We study trivialization of the obtained notion(s) of metric for different fragments of $\Lambda^{!,\|}_{\oplus}$, showing that both parallel disjunction and strong normalization give us precisely the kind of discriminating power we need to arbitrarily amplify distances, while in the most natural fragment, namely $\Lambda^{!}_{\oplus}$, trivialization does *not* hold. This is the subject of Section 4.
- In Section 5, we prove that context distance can be characterized as a coinductively defined distance on a labelled Markov chain of *tuples*. The way (tuples of) terms interact with their environment makes proofs of soundness laborious and different from their affine counterparts from [7]. An up-to-context notion of bisimulation is proved to be sound, and to be quite useful when evaluating the distance between concrete programs.
- Finally, we show that the results from Section 5 can be lifted back to ordinary probabilistic λ-calculi from the literature [10, 5]. Both when call-by-name evaluation and call-by-value are considered, our framework can be naturally adapted, and helps in facilitating concrete proofs. This is in Section 6.

2 Metrics and Trivialisation, Informally

The easiest way to render the pure λ-calculus a universal probabilistic computation model [10] consists in endowing it with a binary construct $\oplus$ for probabilistic choice. The term $M \oplus N$ evolves as either M or N, each with probability $\frac{1}{2}$. The obtained calculus can be given meaning by an operational semantics which puts terms in correspondence with *distributions of* values. The natural notion of observation, at least in an untyped setting like the one we will consider in this paper, is thus the *probability of convergence* of the observed term M, which will be denoted as $\sum_{[\![M]\!]}$. One could then define a notion of *context equivalence* following Morris' pattern, and stipulate that two terms M and N should be equivalent whenever they terminate with *exactly* the same probability when put in *any* context:

$$M \equiv N \;\Leftrightarrow\; \forall C. \sum\nolimits_{[\![C[M]]\!]} = \sum\nolimits_{[\![C[N]]\!]}.$$

The anatomy of the obtained notion of equivalence has been recently studied extensively, the by-products of this study being powerful techniques for it in the style of bisimilarity and logical relations [24, 5, 3].

As observed by various authors (see, e.g., [26] for a nice account), probabilistic programs and processes are naturally compared by *metrics* rather than *equivalences*: the latter do not give any quantitative information about *how different* two non-equivalent programs are. Given that the underlying notion of observation is inherently quantitative, on the other hand, generalizing context equivalence to a *pseudometric* turns out to be relatively simple:

$$\delta(M,N) = \sup_C \left| \sum\nolimits_{[\![C[M]]\!]} - \sum\nolimits_{[\![C[N]]\!]} \right|.$$

Observe that the obtained notion of context *distance* between two terms is a real number between 0 and 1, which is minimal precisely when the considered terms are context equivalent. It is the least discriminating pseudometric which is non-expansive and adequate, and as such it provides some quite precise information about how far the two argument programs are, observationally. A similar notion has recently been studied by the authors [7], but only in a purely affine setting.

Let us now consider two prototypical examples of non-equivalent terms, namely $I = \lambda x.x$ (the identity) and Ω (the always-divergent term). The context distance $\delta^c(I, \Omega)$ between them is maximal: when applied, e.g., to the trivial context $[\cdot]$, they converge with probability 1 and 0, respectively. A term which is conceptually "in the middle" of them is $M = I \oplus \Omega$. Indeed, in a purely affine λ-calculus, $\delta^c(I, M) = \delta^c(M, \Omega) = \frac{1}{2}$.

If we render the three terms duplicable (by putting them in the scope of a !-operator), however, the situation becomes much more complicated. Consider the terms $!I$ and $!(I \oplus \Omega)$. One can easily define a family of contexts $\{C_n\}_{n\in\mathbb{N}}$ such that the probability of convergence of $C_n[!I]$ and $C_n[!(I \oplus \Omega)]$ tend to 1 and 0 (respectively) when n tends to infinity. It suffices to take C_n as $(\lambda!x.\ \underbrace{x \ldots x}_{n \text{ times}})[\cdot]$. Allowing contexts to have the capability to duplicate their argument seems to mean that they can arbitrarily *amplify* distances. Indeed, the argument above also works when $(I \oplus \Omega)$ is replaced by any term which behaves as Ω with probability ε and as I with probability $1 - \varepsilon$, provided of course $\varepsilon > 0$. But how about $!\Omega$ and $!(I \oplus \Omega)$? Are they at maximal distance, i.e. is it that $\delta^c(!\Omega, !(I \oplus \Omega)) = 1$? Apparently, this is *not* the case. The previously defined contexts C_n cannot amplify the "linear" distance between the two terms above, namely $\frac{1}{2}$, up to 1. But what is the distance between $!\Omega$ and $!(I \oplus \Omega)$, then? Evaluating it is hard, since you need to consider all contexts, which do not have a nice structure. In Section 5, we will introduce a different, better behaved, notion of distance, this way being able to prove that, indeed, $\delta^c(!\Omega, !(I \oplus \Omega)) = \frac{1}{2}$.

All this hints at even more difficult examples, like the one in which $M_\varepsilon = !(\Omega \oplus^\varepsilon I)$, where $\oplus^\varepsilon$ is the natural generalization of $\oplus$ to a possibly unfair coin flip, and one is interested in evaluating $\delta^c(M_\varepsilon, M_\mu)$. In that case, we can easily see that the "linear" distance between them is $|\varepsilon - \mu|$. In some cases, it is possible to amplify it: the most natural way is again to consider the contexts C_n defined above. Indeed, we see that the probabilities of convergence of $C_n[M]$ and $C_n[N]$ are ε^n and μ^n, respectively. It follows that $\delta^c(M_\varepsilon, M_\mu) \geq \sup_{n\in\mathbb{N}} |\varepsilon^n - \mu^n|$. For some ε and μ (for example if $\varepsilon + \mu > 1$), the context distance can be greater than $|\varepsilon - \mu|$. But there is no easy way to know *how far* amplification can lead us. The terms M_ε and M_μ will be running examples in the course of this paper. Despite their simplicity, evaluating the distance between them is quite challenging.

We are also going to consider the case where contexts can evaluate terms *in parallel*, converging if and only if at least one of the copies converges. This behaviour is not expressible in the usual λ-calculus, but is captured by well-known constructs and in particular by Plotkin's parallel disjunction [29]. In Section 4 below, we prove that all this is not accidental: the presence of parallel disjunction turns a non-trivialising metric into a trivialising one. The proof of it, by the way, relies on building certain amplifying contexts which are then shown to be universal using tools from functional analysis.

3 A Linear Probabilistic λ-Calculus

In this section, we present the syntax and operational semantics of our language $\Lambda^{!,\|}_\oplus$, on which we will later define metrics. $\Lambda^{!,\|}_\oplus$ is a probabilistic and linear λ-calculus, designed not only to allow copying, but to have a better control on it. It is based on a probabilistic variation of the calculus defined in [34], whose main feature is to never reduce inside exponential boxes. As we will see in Section 6, the calculus is capable of encoding both call-by-value and call-by-name fully-fledged probabilistic λ-calculi. We add a parallel disjunction construct to the calculus, being inspired by Plotkin's parallel disjunction [29]. Noticeably, it has been recently shown [8] that adding parallel disjunction to a (non-linear) λ-calculus increases the expressive power of contexts to the point of enabling coincidence between the contextual preorder and applicative similarity. The choice of studying a very general calculus is motivated by our desire to be as general as possible. This being said, many of our results hold only in *absence* of parallel disjunction.

Definition 1 *We assume a countable set of variables $\mathscr{X}$. The set of* terms *of $\Lambda^{!,\|}_\oplus$ (denoted $\mathscr{T}$) is defined by the following grammar:*

$$M \in \mathscr{T} ::= x \ \Big|\ MM \ \Big|\ \lambda x.M \ \Big|\ \lambda!x.M \ \Big|\ !M \ \Big|\ M \oplus M \ \Big|\ ([M \parallel M] \rightarrowtail M)\ ,$$

where $x \in \mathscr{X}$. The fragment of $\Lambda^{!,\|}_{\oplus}$ without the $([\cdot \parallel \cdot] \rightarrowtail \cdot)$ construct will be indicated as $\Lambda^{!}_{\oplus}$. Values *are those terms derived from the following grammar:*

$$V \in \mathscr{V} ::= \lambda x.M \;\Big|\; \lambda !x.M \;\Big|\; !M.$$

As already mentioned, $M \oplus N$ can evolve to either M or N, each with probability $\frac{1}{2}$. The term $!M$ is a duplicable version of M, often called an *(exponential) box*. We have two distinct abstraction operators: $\lambda x.M$ is a *linear* abstraction, while the *non-linear* abstraction $\lambda !x.M$ requires exponential boxes as arguments. The term $([M \parallel N] \rightarrowtail L)$ behaves as L if either M or N converges. Please observe that both abstractions and boxes are values—our notion of reduction is *weak* and *surface* [34].

We are now going to define an operational semantics for the closed terms of $\Lambda^{!,\|}_{\oplus}$ in a way similar to the one developed for a (non-linear) λ-calculus in [10]. We need to first define a family of *approximation semantics*, and then to take the semantics of a term as the least upper bound of all its approximations. The approximation semantics relation is denoted $M \Rightarrow \mathcal{D}$, where M is a closed term of $\Lambda^{!,\|}_{\oplus}$, and $\mathcal{D}$ is a *(sub)distribution on values* with finite support, i.e., a function from $\mathscr{V}$ to $\mathbb{R}_{[0,1]}$ which sums to a real number $\sum_{\mathcal{D}} \leq 1$. For any distribution $\mathcal{D}$ on a set X, we call *support of* $\mathcal{D}$, and we note $\mathsf{S}(\mathcal{D})$, the set $\{x \in X \mid \mathcal{D}(x) > 0\}$. We say that $\mathcal{D}$ is *finite* if $\mathsf{S}(\mathcal{D})$ is a finite set.

The rules deriving the approximation semantics relation are given in Figure 1, and are based on the notion of an *evaluation context*, which is an expression generated from the following grammar:

$$E ::= [\cdot] \;\Big|\; EV \;\Big|\; ME \;\Big|\; ([M \parallel E] \rightarrowtail N) \;\Big|\; ([E \parallel M] \rightarrowtail N)\ .$$

As usual, $E[M]$ stands for the term obtained by filling the sole occurrence of $[\cdot]$ in E with M. In Figure 1 and elsewhere in this paper, we indicate the distribution assigning probability p_i to V_i for every $i \in \{1, \ldots, n\}$ as $\{V_1^{p_1}, \ldots, V_n^{p_n}\}$. Similarly for the expression $\{V_i^{p_i}\}_{i \in I}$, where I is any countable index set. Observe how we first define a one-step reduction relation $\cdot \to \cdot$ between closed terms and *sequences* of terms, only later extending it to a small-step reduction relation $\cdot \Rightarrow \cdot$ between closed terms and *distributions* on values. A reduction step can be a linear or

$$\frac{}{M \oplus N \hookrightarrow M, N} \qquad \frac{}{(\lambda x.M)V \hookrightarrow M\{x/V\}} \qquad \frac{}{(\lambda !x.M)!N \hookrightarrow M\{x/N\}}$$

$$\frac{M \hookrightarrow N_1, \ldots, N_n}{E[M] \to E[N_1], \ldots E[N_n]}$$

$$\frac{}{([V \parallel M] \rightarrowtail N) \hookrightarrow N} \qquad \frac{}{([M \parallel V] \rightarrowtail N) \hookrightarrow N}$$

$$\frac{}{V \Rightarrow \{V^1\}} \qquad \frac{}{M \Rightarrow \emptyset} \qquad \frac{M \to N_1, \ldots, N_n \qquad (N_i \Rightarrow \mathcal{D}_i)_{1 \leq i \leq n}}{M \Rightarrow \sum_{1 \leq i \leq n} \frac{1}{n} \cdot \mathcal{D}_i}$$

Figure 1: Approximation Semantics for $\Lambda^{!}_{\oplus}$

non-linear β-reduction, or a probabilistic choice. Moreover, there can be more than one active redex in any closed term M, due to the presence of parallel disjunction. For any term M, the set of sub-distributions $\mathcal{D}$ such that $M \Rightarrow \mathcal{D}$ is a countable directed set. Since the set of sub-distributions (with potentially infinite support) is an ω-complete partial order, we can define the *semantics* of a term M as $[\![M]\!] = \sup\{\mathcal{D} \mid M \Rightarrow \mathcal{D}\}$. We could also define alternatively a big-step semantics, again in the same way as that of the probabilistic λ-calculus considered in [10].

Not all irreducible terms are values in $\Lambda^{!,\|}_{\oplus}$, e.g. $(\lambda !x.x)(\lambda x.x)$. We thus need a *type-system* which guarantees the absence of deadlocks. Since we want to be as general as possible, we consider recursive types as formulated in [2], which are expressive enough to type the image of the embeddings we will study in Section 6. The grammar of *types* is the following:

$$\sigma \in \mathscr{A} ::= \alpha \;\Big|\; \mu\alpha.\sigma \multimap \sigma \;\Big|\; \mu\alpha.!\sigma \;\Big|\; \sigma \multimap \sigma \;\Big|\; !\sigma$$

Types are defined up to the equality $=^{\mathscr{A}}$, defined in Figure 2. $\sigma[\alpha \to \tau]$ stands for the type obtained by substituting all free occurrences of α by τ in σ. An *environment* is a set of expressions in the

$$\frac{}{\mu\alpha.\sigma \multimap \tau =^{\mathscr{A}} \sigma[\alpha \to (\mu\alpha.\sigma \multimap \tau)] \multimap \tau[\alpha \to (\mu\alpha.\sigma \multimap \tau)]} \qquad \frac{\sigma =^{\mathscr{A}} \gamma[\alpha \to \sigma] \qquad \tau =^{\mathscr{A}} \gamma[\alpha \to \tau]}{\sigma =^{\mathscr{A}} \tau}$$

$$\frac{}{\mu\alpha.!\sigma =^{\mathscr{A}} !(\sigma[\alpha \to \mu\alpha.!\sigma])}$$

Figure 2: Equality of Types

form $x : \sigma$ or $!x :!\sigma$ in which any variable occurs at most once. Environments are often indicated with metavariables like $!\Gamma$, which stands for an environment in which all variables occur as $!x$, or Δ in which, on the contrary, *only* variables can occur, i.e. Δ is of the form $x_1 : \sigma_1, \ldots, x_n : \sigma_n$. *Typing judgments* are thus of the form $!\Gamma, \Delta \vdash M : \sigma$. The typing system is given in Figure 3. The role of this type system is *not* to guarantee termination, but rather to guarantee a form of type soundness:

Lemma 1 *If $\vdash M : \sigma$ and $M \Rightarrow \mathcal{D}$, then $\vdash V : \sigma$ for every V in the support of $\mathcal{D}$. Moreover, if $\vdash M : \sigma$ and M is irreducible (i.e. $M \not\to N$ for every N), then M is value.*

Example 1 *The term $I = \lambda x.x$ can be typed as $\vdash I : \sigma \multimap \sigma$ for every $\sigma \in \mathscr{A}$. We define $\Omega_!$ to be the term $(\lambda!x.x!x)(!(\lambda!x.x!x))$, which is the counterpart in our linear calculus of the prototypical diverging term of the λ-calculus, namely $\Omega = (\lambda x.xx)(\lambda x.xx)$. We can type this divergent term with any possible type: indeed, if we take $\tau ::= \mu\alpha.!\alpha \multimap \sigma$, then $\tau =^{\mathscr{A}} !\tau \multimap \sigma$ and $\vdash \lambda!x.x!x : \sigma$. Using that, we can see that $\vdash \Omega_! : \sigma$ for every type σ. We will see in Section 6 that, more generally, there are several ways to turn any pure λ-term M into a $\Lambda^!_\oplus$ term in such a way as to obtain meaningful typing and semantics: $\Lambda^!_\oplus$ is actually at least as powerful as the usual untyped probabilistic λ-calculus [10].*

Termination could in principle be guaranteed if one considers *strictly positive* types, as we will do in Section 4.1 below. Let $\mathbb{D}$ be the set of dyadic numbers (i.e. those rational numbers in the form $\frac{n}{2^m}$ (with $n, m \in \mathbb{N}$ and $n \leq 2^m$). It is easy to derive, for every $\varepsilon \in \mathbb{D}$, a new binary operator on terms $\cdot \oplus^\varepsilon \cdot$ such that $[\![M \oplus^\varepsilon N]\!] = (1-\varepsilon)[\![M]\!] + \varepsilon[\![N]\!]$ for every closed M, N. It can be defined,

$$\frac{}{!\Gamma, !x :!\sigma \vdash x : \sigma} \qquad \frac{}{!\Gamma, x : \sigma \vdash x : \sigma} \qquad \frac{!\Gamma, x : \sigma, \Delta \vdash M : \tau}{!\Gamma, \Delta \vdash \lambda x.M : \sigma \multimap \tau}$$

$$\frac{!x : \sigma, !\Gamma, \Delta \vdash M : \tau}{!\Gamma, \Delta \vdash \lambda!x.M : \sigma \multimap \tau} \qquad \frac{!\Gamma, \Delta \vdash M : \sigma \multimap \tau \qquad !\Gamma, \Theta \vdash N : \sigma}{!\Gamma, \Delta, \Theta \vdash MN : \tau}$$

$$\frac{!\Gamma \vdash M : \sigma}{!\Gamma \vdash !M :!\sigma} \qquad \frac{!\Gamma, \Delta \vdash M : \sigma \qquad !\Gamma, \Delta \vdash N : \sigma}{!\Gamma, \Delta \vdash M \oplus N : \sigma}$$

$$\frac{!\Gamma, \Delta \vdash M : \sigma \qquad !\Gamma, \Theta \vdash N : \sigma \qquad !\Gamma, \Xi \vdash L : \tau}{!\Gamma, \Delta, \Theta, \Xi \vdash ([M \parallel N] \rightarrowtail L) : \tau}$$

Figure 3: Typing Rules

e.g., as follows by induction on m:

$$M \oplus^0 N = M$$
$$M \oplus^1 N = N$$
$$M \oplus^{\frac{n}{2^m}} N = \begin{cases} M \oplus (M \oplus^{\frac{n}{2^{m-1}}} N) & \text{if } n \leq 2^{m-1} \\ N \oplus (M \oplus^{\frac{n-2^{m-1}}{2^{m-1}}} N) & \text{if } n > 2^{m-1} \end{cases}$$

Example 2 *We define here a family of terms that we use as a running example. We consider terms of the form $M_\varepsilon = !(\Omega_! \oplus^\varepsilon I)$, for $\varepsilon \in \mathbb{D}$. It holds that $\vdash M_\varepsilon : !(\sigma \multimap \sigma)$ for every σ. M_ε corresponds to a duplicable term each copy of which behaves as I with probability ε, and does not terminate with probability $1 - \varepsilon$.*

3.1 Some Useful Terminology and Notation

In this paper, we will make heavy use of sequences of terms and types. It is thus convenient to introduce some terminology and notation about them.

A finite (ordered) sequence whose elements are $e_1, \ldots, e_n$ will be indicated as $\boldsymbol{e} = [e_1, \ldots, e_n]$, and called an *$n$-sequence*. Metavariables for sequences are boldface variations of the metavariables for their elements. Whenever $E = \{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}$ and $i_1 < \ldots < i_m$, the sub-sequence $[e_{i_1}, \ldots, e_{i_m}]$ of an n-sequence $\boldsymbol{e}$ will be indicated as $\boldsymbol{e}_E$. If the above holds, E will be called an *n-set*. If $\boldsymbol{e}$ is an n-sequence, and φ is a permutation on $\{1, \ldots, n\}$, we note e_φ the n-sequence $[e_{\varphi(1)}, \ldots, e_{\varphi(n)}]$. We can turn an n-sequence into a $(n+1)$-sequence by adding an element at the end: this is the role of the semicolor operator. We denote by $[e^n]$ the n-sequence where all components are equal to e.

Whenever this does not cause ambiguity, notations like the ones above will be used in conjunction with syntactic constructions. For example, if $\boldsymbol{\sigma}$ is an n-sequence of types, then $!\boldsymbol{\sigma}$ stands for the sequence $[!\sigma_1, \ldots, !\sigma_n]$. As another example, if $\boldsymbol{\sigma}$ is an n-sequence of types and E is an n-set, then $\boldsymbol{x}_E : \boldsymbol{\sigma}_E$ stands for the environment assigning type σ_i to x_i for every $i \in E$. As a final example, if $\boldsymbol{M}$ is an n-sequence of terms and $\boldsymbol{\sigma}$ is an n-sequence of types, $\vdash \boldsymbol{M} : \boldsymbol{\sigma}$ holds iff $\vdash M_i : \sigma_i$ is provable for every $i \in \{1, \ldots, n\}$.

3.2 Context Distance

A *context of type σ for terms of type τ* is a term C which can be typed as $\mathit{hole} : \tau \vdash C : \sigma$, where hole is a distinguished variable. $\mathcal{C}^\tau_\sigma$ collects all such terms. If $C \in \mathcal{C}^\tau_\sigma$ and M is a closed term of type τ, then the closed term $C\{\mathit{hole}/M\}$ has type σ and is often indicated as $C[M]$.

The *context distance* [7] is the natural quantitative refinement of context equivalence. Intuitively, it corresponds to the maximum separation that contexts can induce between two terms. Following [7], we take as observable the probability of convergence: for any term M, we define its *observable* $\mathrm{Obs}(M)$ as $\sum_{[\![M]\!]}$. Then, for any terms M, N such that $\vdash M : \sigma$ and $\vdash N : \sigma$, we define:

$$\delta^c_{\sigma,!,\parallel}(M, N) = \sup_{C \in \mathcal{C}^\tau_\sigma} |\mathrm{Obs}(C[M]) - \mathrm{Obs}(C[N])|.$$

Please observe that this distance is a pseudometric, and that moreover we can recover context equivalence by considering its *kernel*, that is the set of pairs of terms which are at distance 0. The binary operator $\delta^c_{\sigma,!}$ is defined similarly, but referring to terms (and contexts) from $\Lambda^!_\oplus$.

Example 3 *What can we say about $\delta^c_{\sigma,!,\parallel}(M_\varepsilon, M_\mu)$? Not much apparently, since* all *contexts should be considered. Even if we put ourselves in the fragment $\Lambda^!_\oplus$, the best we can do is to conclude that $\delta^c_{\sigma,!}(M, N) \geq \sup_{n \in \mathbb{N}} |\varepsilon^n - \mu^n|$, as explained in Section 2.*

4 On Trivialisation

As we have already mentioned, there can well be classes of terms such that the context distance collapses to context equivalence, due to the copying abilities of the language. The question of trivialisation can in fact be seen as a question about the expressive power of contexts: given two duplicable terms, how much can a context amplify the observable differences between their behaviours?

More precisely, we would like to identify *trivialising* fragments of $\Lambda^{!,\|}_{\oplus}$, that is to say fragments such that for any pair of duplicable terms, their context distance (with respect to the fragment) is either 0 or 1. This is not the case in $\Lambda^{!}_{\oplus}$ (see Example 9 below).

In fact, a sufficient condition to trivialization is to require the existence of *amplification contexts*: for every observable type σ, for every $\alpha, \beta \in [0,1]$ distinct, for every $\gamma > 0$, we want to have a context $C^{\alpha,\beta,\gamma}_{\sigma}$ such that:

$$\left.\begin{array}{c} \vdash M, N : \sigma \\ \mathrm{Obs}(M) = \alpha \\ \mathrm{Obs}(N) = \beta \end{array}\right\} \Rightarrow \left|\mathrm{Obs}(C^{\alpha,\beta,\gamma}_{\sigma}[!M]) - \mathrm{Obs}(C^{\alpha,\beta,\gamma}_{\sigma}[!N])\right| \geq 1 - \gamma.$$

Fact 1 *Any fragment of $\Lambda^{!,\|}_{\oplus}$ admitting all amplifications contexts trivializes.*

4.1 Strictly Positive Types

First, let us consider the case of the fragment $\Lambda^{!,\downarrow}_{\oplus}$ of $\Lambda^{!}_{\oplus}$ obtained by considering strictly positive types, only (in a similar way to [2]), and by dropping parallel disjunction. Every term M of $\Lambda^{!,\downarrow}_{\oplus}$ is terminating (i.e. $\sum[\![M]\!] = 1$), so we need to adapt our notion of observation: we define the type $\mathbb{B} = !\alpha \multimap !\alpha \multimap \alpha$, which can be seen as boolean type using a variant of the usual boolean encoding in λ-calculi. Our new notion of observation, defined only at type $\mathbb{B}$, is $\mathrm{Obs}(M) = \sum_{[\![M!I!\Omega_!]\!]}$, which corresponds to the probability that M evaluates to **true**. While this notion of observation uses the full power of $\Lambda^{!}_{\oplus}$, the context distance $\delta^{c}_{!,\downarrow}$ based on it only consider contexts in $\Lambda^{!,\downarrow}_{\oplus}$.

Theorem 1 *$\delta^{c}_{!,\downarrow}$ trivialises.*

The proof of Theorem 1 is based on the construction of amplification contexts. We are going to use Bernstein constructive proof of the Stone-Weierstrass theorem. Indeed, Bernstein showed that for every continuous function $f : [0,1] \to \mathbb{R}$, the following sequence of polynomials converges uniformly towards f:

$$P^{f}_{n}(x) = \sum\nolimits_{0 \leq k \leq n} f\left(\frac{k}{n}\right) \cdot B^{n}_{k}(x), \text{ where } B^{n}_{k}(x) = \binom{n}{k} \cdot x^{k} \cdot (1-x)^{n-k}.$$

Let us consider the following continuous function: we fix $f(\alpha) = 0$, $f(\beta) = 1$, and f defined elsewhere in such a way that it is continuous, that it has values in $[0,1]$, and that moreover $f(\mathbb{Q}) \subseteq \mathbb{Q}$. We can easily implement P^{f}_{n} by a context, i.e. define C such that for every M, $\mathrm{Obs}(C[M]) = P^{f}_{n}(\mathrm{Obs}(M))$. In $\Lambda^{!,\downarrow}_{\oplus}$, we can indeed copy an argument n times, then evaluate it, and then for every k between 0 and n, if the number of **true**s obtained is exactly k, return the term $\mathbf{false} \oplus^{f(\frac{k}{n})} \mathbf{true}$ (that corresponds to a term returning **true** with probability $f\left(\frac{k}{n}\right)$). Please observe that this construction works only because in $\Lambda^{!,\downarrow}_{\oplus}$ all terms converge with probability 1.

4.2 Parallel Disjunction

As we have seen, trivialization can be enforced by restricting the class of terms, but we can also take the opposite road, namely increasing the discriminating power of contexts. Indeed, consider the full language $\Lambda^{!,\|}_{\oplus}$, with the usual notion of observation.

We can first see how parallel disjunction increases the expressive power of the calculus on a simple example. Consider the following two terms: $M = !\Omega_!$ and $N = !(\Omega_! \oplus I)$. We will see later that

these two terms are the simplest example of non-trivialization in $\Lambda^{!}_{\oplus}$: indeed $\delta^{c}_{!(\tau\multimap\tau),!}(M,N)=\frac{1}{2}$, while $\delta^{c}_{!(\tau\multimap\tau),!,\parallel}(M,N)=1$. In $\Lambda^{!,\parallel}_{\oplus}$, we are able to define a family of contexts $(C_n)_{n\in\mathbb{N}}$ as follows:

$$C_n = (\lambda !x.\ ([x \parallel\ ([x \parallel \ldots] \rightarrowtail I)\,] \rightarrowtail I)\,)\,[\cdot].$$

Essentially, C_n makes n copies of its argument, and then converges towards I if *at least* one of these copies itself converges. When we apply the context C_n to M and N, we can see that the convergence probability of $C_n[M]$ is always 0 independently of n, whereas the convergence probability of $C_n[N]$ tends towards 1 when n tends to infinity.

Theorem 2 $\delta^{c}_{!,\parallel}$ *trivializes.*

The proof is based on the construction of amplification contexts $C^{\alpha,\beta,\gamma}_{\sigma}$. If $\max(\alpha,\beta)=1$, we can extend the informal argument from Section 2, by taking contexts that copy an arbitrary number of times their argument. If $\min(\alpha,\beta)=0$, we can use the same idea as in the example above, by taking contexts that do an arbitrary number of disjunctions. What remains to be done to obtain the trivialization result is treating the case in which $0<\alpha,\beta<1$. The overall idea is to somehow mix the contexts we use in the previous cases. More precisely, we define a family of contexts $(C^{m}_{n})_{n,m\in\mathbb{N}}$ as follows:

$$C^{m}_{n} = \lambda !y.\left(\bigwedge\nolimits^{n}(\bigvee\nolimits^{m}(y,\ldots,y),\ldots,\bigvee\nolimits^{m}(y,\ldots,y))\right)[\cdot]$$

where

$$\bigvee\nolimits^{n}(M_1,\ldots M_n) = \ ([M_1 \parallel\ ([M_2 \parallel \ldots] \rightarrowtail I)\,] \rightarrowtail I)\ ;$$
$$\bigwedge\nolimits^{n}(M_1\ldots M_n) = (\lambda z_1.\lambda z_2.\ldots.\lambda y.(y z_1\ldots z_n))M_1\ldots M_n.$$

The term $\bigvee^{n}(M_1,\ldots,M_n)$ behaves as a n-ary disjunction: it terminates if *at least one* of the M_i terminates. On the other hand, $\bigwedge^{n}(M_1,\ldots,M_n)$ can be seen as a n-ary conjunction: it terminates if *all* the M_i terminates. The contexts C^{α}_{n} compute m-ary conjunctions of n-ary disjunction. Now, let ι be such that $\alpha<\iota<\beta$. We need to show that for every n, we can choose $m(n,\iota)\in\mathbb{N}$ such that:

$$\lim_{n\to\infty}\mathrm{Obs}(C^{m(n,\iota)}_{n}[!M]) = \begin{cases} 1 \text{ if } \mathrm{Obs}(M)>\iota; \\ 0 \text{ if } \mathrm{Obs}(M)<\iota. \end{cases}.$$

We can express this problem purely in terms of functional analysis, by observing that $\mathrm{Obs}(C^{m}_{n}[!M]) = (1-(1-\mathrm{Obs}(M))^{m})^{n}$. We are now going to show the result by applying the dominated convergence theorem to a well-chosen sequence of functions.

We choose $m(n,\gamma)=\lceil r(n,\iota)\rceil$, with

$$r(n,\iota)=\left(\frac{1}{1-\iota}\right)^{n}.$$

Now, we see that $\mathrm{Obs}(C^{m(n,\iota)}_{n}[!M])$, is equal to $f_{n,\iota}(\mathrm{Obs}(M))$, where $f_{n,\iota}$ is the real function defined by:

$$f_{n,\iota}(x)=(1-(1-x)^{n})^{\lceil r(n,\iota)\rceil}$$

The result we are trying to show can now be expressed by way of the following lemma:

Lemma 2

$$\lim_{n\to\infty} f_{n,\iota}(x) = \begin{cases} 0 \ \textit{if}\ x<\iota \\ 1 \ \textit{if}\ x>\iota \end{cases}$$

Proof. In order to simplify the proof, we show an equivalent result on $g_{n,\alpha}(x) = f_{n,1-\alpha}(1-x)$. Lemma 2 is equivalent to say that: $\lim_{n\to\infty} g_{n,\alpha}(x) = \begin{cases} 1 \text{ if } x < \alpha, \\ 0 \text{ if } x > \alpha. \end{cases}$ We express g as an integral, since we want to use known results about inversion of integral and limits:

$$g_{n,\alpha}(x) = \begin{cases} \int_0^x g'_{n,\alpha}(t)dt + 1 & \text{if } x < \alpha \\ -\int_x^1 g'_{n,\alpha}(t)dt \text{ if } x > \alpha \end{cases}$$

We can now express our goal using g': we have to show that

$$\lim_{n\to\infty} \int_0^x g'_{n,\alpha}(t)dt = 0 \qquad \text{and} \qquad \lim_{n\to\infty} \int_x^1 g'_{n,\alpha}(t)dt = 0.$$

We first establish a bound on the $g'_{n,\alpha}$:

$$g'_{n,\alpha}(x) = -\lceil r(n, 1-\alpha) \rceil \cdot n \cdot (1-x^n)^{\lceil r(n,1-\alpha) \rceil - 1} \cdot x^{n-1}$$

And consequently:

$$| \, g'_{n,\alpha}(x) \, | \leq \frac{1}{\alpha^2} \cdot \left(\frac{x}{\alpha}\right)^{n-1} \cdot n \cdot (1-x^n)^{(\frac{1}{\alpha})^n - 1} \tag{1}$$

We have to use the following well-known inequality, that we recall here:

$$\lim_{x\to\infty} x^m \cdot e^{-x} = 0 \text{ for every } m \in \mathbb{N}. \tag{2}$$

For every $x < \alpha$, we're going to use the bounded convergence theorem, which is a well-known theorem in real analysis, on $[0, x]$ and on $[x, 1]$ for every $x > \alpha$. We recall this theorem here:

Theorem 3 (Bounded Convergence Theorem) *If k_n is a sequence of uniformly bounded real-valued measurable functions which converges pointwise on a bounded measure space (S, Σ, μ) to a function k, then the limit k is an integrable function and: $\int_S \lim_{n\to\infty} k_n = \lim_{n\to\infty} \int_S k_n$.*

- Let's first take $0 \leq x < \alpha$. Here, we take (S, Σ, μ) as the segment $[0, x]$ with the Lebesgue measure (corresponding to the usual notion of integration on real functions). We obtain using (1), that for every $t \in [0, x]$, it holds:

  $$| \, g'_{n,\alpha}(t) \, | \leq K \cdot n \cdot \left(\frac{t}{\alpha}\right)^{n-1}$$

  where K is a constant. We can already see that $g'_{n,\alpha}$ converge pointwise to the 0-function on $[0, x]$ (using (2)). We're now going to show it's bounded. Using the fact that $t \leq x$, we obtain:

  $$\begin{aligned} | \, g'_{n,\alpha}(t) \, | &\leq K \cdot n \cdot \left(\frac{x}{\alpha}\right)^{n-1} \\ &\leq K \cdot \max_{n\in\mathbb{N}} n \cdot \left(\frac{x}{\alpha}\right)^{n-1} \end{aligned}$$

  Please observe that the max is well-defined and $< \infty$ because the sequence $n \cdot \left(\frac{x}{\alpha}\right)^{n-1}$ has a finite limit (by (2)). So we have achieved to bound $g'_{n,\alpha}(t)$ by a constant depending neither on t nor on n. Consequently, we can apply the bounded convergence theorem, and we obtain :

  $$\lim_{n\to\infty} \int_0^x g'_{n,\alpha}(t)dt = 0$$

- Now we consider the case where $\alpha < x \leq 1$. Here, we take (S, Σ, μ) as the segment $[0, x]$ with the Lebesgue measure. By using 1, we can see that for every $t \in [x, 1]$:

$$\begin{aligned} \mid g'_{n,\alpha}(t) \mid &\leq \frac{1}{\alpha^2} \cdot \left(\frac{t}{\alpha}\right)^{n-1} \cdot n \cdot exp\left(\left(\left(\frac{1}{\alpha}\right)^n - 1\right) \cdot (-t^n)\right) \\ &\leq \frac{1}{\alpha^2} \cdot \left(\frac{t}{\alpha}\right)^{n-1} \cdot n \cdot exp\left(-\left(\frac{t}{\alpha}\right)^n\right) \cdot exp(t^n) \\ &\leq \frac{1}{\alpha^2} \cdot \left(\frac{1}{\alpha}\right)^{n-1} \cdot n \cdot exp\left(-\left(\frac{x}{\alpha}\right)^n\right) \cdot exp(1) \end{aligned}$$

  Using 2 (and doing several computation), we can see that for every $a, b > 1$, it holds that:

$$\lim_{m\to\infty} a^m \cdot m \cdot exp(-b^m) = 0 \tag{3}$$

  By applying this, we obtain:

$$\lim_{n\to\infty} \frac{1}{\alpha^2} \cdot \left(\frac{1}{\alpha}\right)^{n-1} \cdot n \cdot exp\left(-\left(\frac{x}{\alpha}\right)^n\right) \cdot exp(1) = 0 \tag{4}$$

  Now, please observe that 4 gives us both the pointwise limit and the uniform bound (since $g_{n,\alpha}(t)$ is bounded by a sequence that doesn't depends on t, and has a finite limit when n goes towards infinity). So we can apply the bounded convergence theorem, and we obtain that:

$$\lim_{n\to\infty} \int_x^1 g'_{n,\alpha}(t)dt = 0$$

This concludes the proof □

5 Tuples and Full Abstraction

This section is structured as follows: first, we define a labelled Markov chain (LMC) which expresses the semantics of our calculus in an interactive way, and then we use it to give a coinductively defined notion of distance on an LTS of *distributions*, which coincides with the context distance as defined in Section 3.2. We are not considering parallel disjunction here: the motivations for that should be clear from Theorem 2.

5.1 A Labelled Markov Chain over Tuples

Labelled Markov chains are the probabilistic analogues to labelled transition systems. Formally, a LMC is a triple $\mathscr{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P})$, where $\mathcal{S}$ is a countable set of states, $\mathcal{A}$ is a countable set of *labels*, and $\mathcal{P} : \mathcal{S} \times \mathcal{A} \to \text{Distr}(\mathcal{S})$ is a *transition probability matrix* (where $\text{Distr}(X)$ is the set of all distributions over X).

Following [9], the interactive behaviour of probabilistic λ-terms can be represented by an LMC whose states are the terms of the language, whose actions are values, and where doing the action V starting from a state M corresponds to applying the value V to M. This approach is bound *not* to work in presence of pairs when metrics take the place of equivalences, due to the unsoundness of projective actions. In [7], this observation lead us to introduce a new LMC whose states are *tuples* of terms, and whose actions include one *splitting* a pair: $\mathcal{P}([\langle M, N\rangle])(\text{destruct}) = \{[M, N]^1\}$. This turns out to work well in an affine setting [7]. We are going to define a LMC $\mathscr{M}^!_\oplus = (\mathcal{S}_{\mathscr{M}^!_\oplus}, \mathcal{A}_{\mathscr{M}^!_\oplus}, \mathcal{P}_{\mathscr{M}^!_\oplus})$ which is an extension of the one from [7], and which is adapted to a language with copying capabilities. The idea is to treat exponentials in the spirit of Milner's Law: $!A \multimap A \otimes !A$.

5.1.1 States

Tuples are pairs of the form $K = (\boldsymbol{M}, \boldsymbol{V})$ where $\boldsymbol{M}$ and $\boldsymbol{V}$ are a sequence of terms and values, respectively. The set of all such tuples is indicated as $\mathscr{U}$. The first component of a tuple is called its *exponential part*, while the second one is called its *linear part*. We write $\vdash (\boldsymbol{M}, \boldsymbol{V}) : (\boldsymbol{\sigma}, \boldsymbol{\tau})$ if $\vdash \boldsymbol{M} : \boldsymbol{\sigma}$ and $\vdash \boldsymbol{V} : \boldsymbol{\tau}$. We note $\mathbf{T}$ the set of pairs $A = (\boldsymbol{\sigma}, \boldsymbol{\tau})$, and we call *tuple types* the elements of $\mathbf{T}$. Moreover, we say that $(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is a (n, m) *tuple type* if $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are, respectively, an n-sequence and an m-sequence. To any term M, we associate a tuple in a natural way: we note $\dot{M}$ the tuple $([]\,, [M])$, and similarly if σ is a type, we indicate the tuple type $([]\,, [\sigma])$ as $\dot{\sigma}$. Please observe that if $\vdash M : \sigma$, then it holds that $\vdash \dot{M} : \dot{\sigma}$.

A sequence of the form $(E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \gamma)$ is said to be an *applicative typing judgment* when $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ are, respectively, an n-sequence and an m-sequence of types, E and F are respectively an n-set and an m-set, and moreover it holds that $!\boldsymbol{x}_E :!\boldsymbol{\sigma}_E, \boldsymbol{y}_F : \boldsymbol{\tau}_F \vdash M : \gamma$. Intuitively, this means that if we have a tuple $K = (\boldsymbol{N}, \boldsymbol{V})$ of type $(\boldsymbol{\sigma}, \boldsymbol{\tau})$, we can replace free variables of M by *some* of the terms from K. More precisely, we can replace variables in linear position by the V_i with $i \in F$, and variables in non linear position by N_j, with $j \in E$. We note as $M[K]$ the closed term of type γ that we obtain this way. We note $\mathscr{J}$ the set of all applicative typing judgments. We are specially interested in those judgments $(E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \gamma)$ in $\mathscr{J}$ such that for every tuple K, the resulting term $M[K]$ is a *value*: that is when either $M = y_i$ for a $i \in \mathbb{N}$, or M is of the form $\lambda z.N$, $\lambda !z.N$, or $!N$. We note $\mathscr{J}^{\mathscr{V}}$ the set of those judgments.

We are now in a position to define $\mathscr{M}^!_\oplus$ formally. The set of its states is indeed defined as $\mathcal{S}_{\mathscr{M}^!_\oplus} = \{(K, A) \mid K \in \mathscr{U},\, A \in \mathbf{T},\, \vdash K : A\}$.

5.1.2 Labels and Transitions

How do states in $\mathcal{S}_{\mathscr{M}^!_\oplus}$ interact with the environment? This is captured by the labels in $\mathcal{A}_{\mathscr{M}^!_\oplus}$, and the associated probability matrix. We define $\mathcal{A}_{\mathscr{M}^!_\oplus}$ as the disjoint union of $\mathcal{A}_?$, $\mathcal{A}_!$ and $\mathcal{A}_@$, where:

$$\mathcal{A}_! = \mathcal{A}_? = \{i \mid i \in \mathbb{N}\}; \qquad \mathcal{A}_@ = \{(\kappa, i) \mid i \in \mathbb{N},\, \kappa \in \mathscr{J}^{\mathscr{V}}\}.$$

In order to distinguish actions in $\mathcal{A}_!$ and $\mathcal{A}_?$, we write the action $i \in \mathbb{N}$ as $(?^i)$ if it comes from $\mathcal{A}_?$, and as $(!^i)$ if it comes from $\mathcal{A}_!$. The action $(\kappa, i) \in \mathcal{A}_@$ is often indicated as $@^i_\kappa$. The probability matrix $\mathcal{P}_{\mathscr{M}^!_\oplus}$ is defined formally in Figure 4. We give below some intuitions about it. The general idea is that $\mathscr{M}^!_\oplus$ is designed to express every possible effect that a context can have on tuples. $\mathcal{A}_?$ and $\mathcal{A}_!$ are designed to model copying capabilities, while $\mathcal{A}_@$ corresponds to applicative interactions.

Actions in $\mathcal{A}_?$ take any term of the form $!M$ from the linear part of the underlying tuple, unbox it and transfer M to the exponential part of the tuple. Please observe that this action is in fact deterministic: the resulting tuple is uniquely determined. Labels in $\mathcal{A}_!$, on the other hand, model the act of *copying* terms in the exponential part. We call its elements *Milner's actions*. More specifically, the action $(!^i)$ takes the i-th term in the exponential part of the tuple, makes a copy of it, evaluates it and adds the result to the linear part. Please observe that, contrary to $(?^i)$, this action can have a probabilistic outcome: the transferred term is evaluated.

Labels in $\mathcal{A}_@$ are analogues of the applicative actions from applicative bisimulation over terms (see, e.g. [9]). As such, they model environments passing arguments to programs. Here, we have to adapt this idea to our tuple framework: indeed, we can see the tuple as a collection of programs available to the environment, who is free to choose *with which* of the programs to interact with by passing it an argument. This argument, however, could depend on other components of the tuple, which need to be removed from it if lying in its linear part. Finally, please observe that all this should respect types. Labels in $\mathcal{A}_@$ are indeed defined as a pair of an index i corresponding to the position in the tuple of the term the environment chooses, and an applicative typing judgment, used to specify the argument. Please observe that in the definition of probability matrix for applicative actions, in Figure 4, the condition on i implies that the i-th linear component of the tuple is not used to construct the argument term.

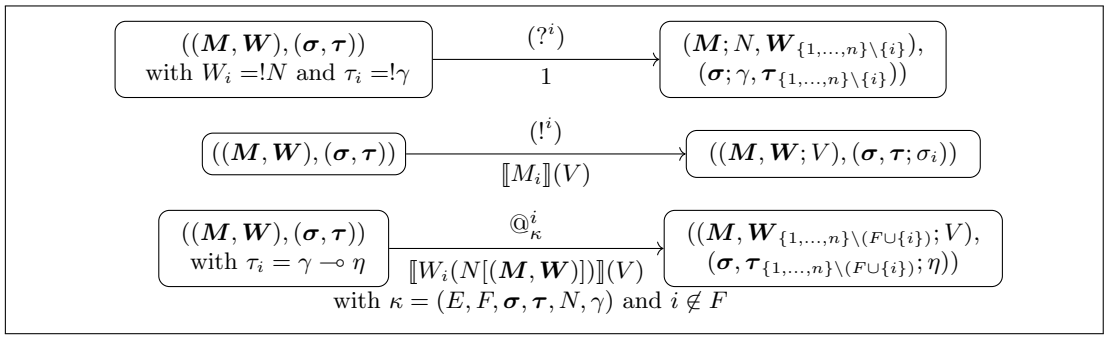


Figure 4: Definition of $\mathcal{P}_{\mathscr{M}^!_\oplus}$

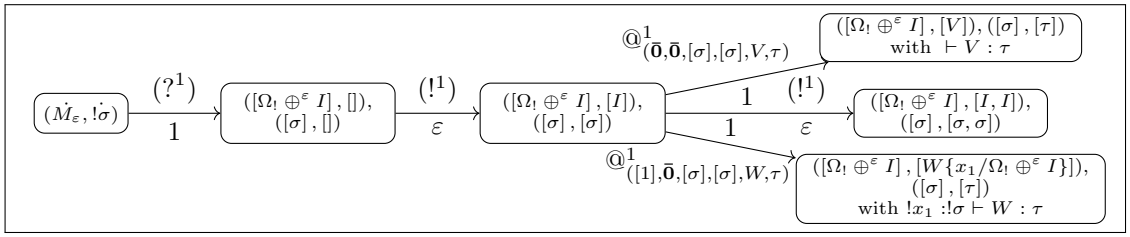


Figure 5: A Fragment of $\mathcal{P}_{\mathscr{M}^!_\oplus}$

Example 4 *We give in Figure 5 a fragment of $\mathscr{M}^!_\oplus$ illustrating our definitions on an example. Let τ be an element of $\mathscr{A}$. We consider terms of the form $M_\varepsilon = !(\Omega_! \oplus^\varepsilon I)$, for $\varepsilon \in \mathbb{D}$ and we look at* some *of the possible evolutions in $\mathscr{M}^!_\oplus$ from the associated state $(\dot{M}_\varepsilon, !(\tau \dot{\multimap} \tau)) = ([], [M_\varepsilon]), ([], [!(\tau \multimap \tau)])$. In Figure 5, we denote by σ the type $\tau \multimap \tau$.*

5.2 Distributions as States

Now that we have a LMC $\mathcal{P}_{\mathscr{M}^!_\oplus}$ modelling interaction between (tuple of) terms and their environment, we could define notions of metrics following one of the abstract definitions from the literature, e.g. by defining the *trace distance* or the *behavioural distance* between terms. This is, by the way, the approach followed in [7]. We prefer, however, to first turn $\mathcal{P}_{\mathscr{M}^!_\oplus}$ into a transition system $\mathscr{L}^!_\oplus$ whose states are *distributions* of tuples. This supports both a simple, coinductive presentation of the trace distance, but also up-to techniques, as we will see in Section 5.6 below. Both will be very convenient when evaluating the distance between concrete terms, e.g. our running example.

It turns out that the usual notion of an LTS is not sufficient for our purposes, since it lacks a way to *expose* the observables of each state, i.e., its sum. We thus adopt the following definition:

Definition 2 *A* weighted labelled transition system *(WLTS for short) is a quadruple in the form $\mathscr{L} = (\mathcal{S}, \mathcal{A}, \dot{\to}, w)$ where:*

- *$\mathcal{S}$ is a set of states and $\mathcal{A}$ is a countable set of actions,*
- *$\dot{\to}$ is a transition function, such that, for every $t \in \mathcal{S}$ and $a \in \mathcal{A}$, there exists* at most *one $s \in \mathcal{S}$, such that $t \xrightarrow{a} s$,*
- *$w : \mathcal{S} \to [0, 1]$.*

Please observe how WLTSs are *deterministic* transition systems. We define the WLTS $\mathscr{L}^!_\oplus$ by way of the equations in Figure 6.

$$\mathcal{S}_{\mathscr{L}^!_\oplus} = \bigcup_{A\in\mathbf{T}} (\mathrm{Distr}(\{K \mid \vdash K : A\}) \times \{A\})\ \mathcal{A}_{\mathscr{L}^!_\oplus} = \mathcal{A}_{\mathscr{M}^!_\oplus} \cup \{A \mid A \in \mathbf{T}\} \qquad w(\mathcal{D}, A) = \sum_{\mathcal{D}}$$
$$(\mathcal{D}, A) \xrightarrow{A} (\mathcal{D}, A) \text{ for } A \in \mathbf{T} \qquad (\mathcal{D}, A) \xrightarrow{a} \sum_{K\in\mathscr{U}} \mathcal{D}(K) \cdot \mathcal{P}_{\mathscr{M}^!_\oplus}((K, A))(a) \text{ for } a \in \mathcal{A}_{\mathscr{M}^!_\oplus}.$$

Figure 6: The WLTS $\mathscr{L}^!_\oplus = (\mathcal{S}_{\mathscr{L}^!_\oplus}, \mathcal{A}_{\mathscr{L}^!_\oplus}, \dot{\to}, w)$

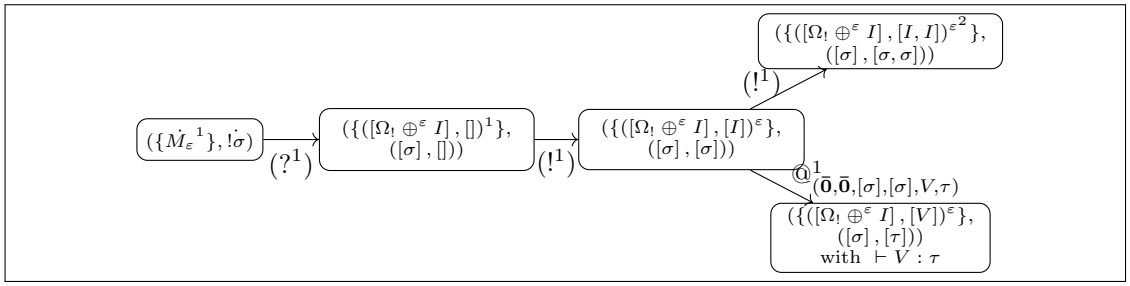


Figure 7: A Fragment of $\mathscr{L}^!_\oplus$

If $t = (\mathcal{D}, A)$ is in $\mathcal{S}_{\mathscr{L}^!_\oplus}$, we say that t is a A-state. It is easy to check that $\mathscr{L}^!_\oplus$ is nothing more than the natural way to turn $\mathcal{P}_{\mathscr{M}^!_\oplus}$ into a deterministic transition system. We illustrate this idea in Figure 7, by giving a fragment of $\mathscr{L}^!_\oplus$ corresponding to (part of) the fragment of $\mathscr{M}^!_\oplus$ given in Example 4.

5.3 A Coinductively Defined Metric

Following Desharnais et al. [13], we use a quantitative notion of bisimulation on $\mathscr{L}^!_\oplus$ to define a distance between terms. The idea is to stipulate that, for any $\varepsilon \in [0, 1]$, a relation R is an ε-bisimulation if it is, somehow, *ε-close to a bisimulation*. The distance between two states t and s is just the smallest ε such that t and s are ε-bisimilar. However, while in [13] the notion of ε-bisimulation is *local*, we want it to be more restricted by the *global* deviation we may accept considering arbitrary sequences of actions.

Definition 3 *Let $\mathscr{L} = (\mathcal{S}, \mathcal{A}, \dot{\to}, w)$ be a WLTS. Let R be a symmetric and reflexive relation on $\mathcal{S}$, and $\varepsilon \in [0, 1]$. R is a ε-bisimulation whenever the following two conditions hold:*

- *if $t\,R\,s$, and $t \xrightarrow{a} u$, then there exists v such that $s \xrightarrow{a} v$, and it holds that $u\,R\,v$.*
- *if $t\,R\,s$, then $|w(t) - w(s)| \leq \varepsilon$.*

For every $\varepsilon \in [0, 1]$, there exists a largest ε-bisimulation, that we indicate as R^ε. Please observe that it is not an equivalence relation (since it is not transitive). We can now define a metric on $\mathcal{S}$: $\delta^b_{\mathscr{L}}(t, s) = \inf\{\varepsilon \mid t\,R^\varepsilon\,s\}$.

The greatest lower bound is in fact reached as a $\delta^b_{\mathscr{L}}(t, s)$-bisimulation: that's the sens of the following lemma.

Lemma 3 *Let $t, s \in \mathcal{S}_{\mathscr{L}^!_\oplus}$. Then $\delta^b_{\mathscr{L}}(t, s) \leq \varepsilon$ if and only if $t\,R^\varepsilon\,s$.*

Proof. The right-to-left implication is given by the definition of $\delta^b_{\mathscr{L}}$. The other one comes from the fact that the relation R given by $t\,R\,s$ if $\delta^b_{\mathscr{L}}(t, s) \leq \varepsilon$ is a ε-bisimulation. □

As a corollary, it is easy to see that the infinum in Definition 3 is in fact reached.

How can we turn $\delta^b_{\mathscr{L}}$ into a metric on *terms*? The idea is to consider the distributions on *tuples* one naturally gets when evaluating the term. To every term M of type σ, we define $\hat{s}_\sigma(M) \in \mathcal{S}_{\mathscr{L}^!_\oplus}$ as $(\{\dot{V}^{[\![M]\!](V)}\}_{V\in\mathscr{V}}, \dot{\sigma})$.

Definition 4 *For every terms M and N such that $\vdash M : \sigma$ and $\vdash N : \sigma$, we set $\delta^b_{\sigma,!}(M,N) = \delta^b_{\mathscr{L}^!_\oplus}(\hat{s}_\sigma(M), \hat{s}_\sigma(N))$.*

Example 5 *Consider again the terms M_ε from Example 2. We fix a type τ, and define $\sigma = \tau \multimap \tau$. As mentioned in Example 2, it holds that $\vdash M_\varepsilon : !\sigma$. Let now ε, μ, α be in $[0,1]$, and let R be any α-bisimulation, such that $\hat{s}_{!\sigma}(M_\varepsilon)\, R\, \hat{s}_{!\sigma}(M_\mu)$. Let $\{t_i\}_{i\in\mathbb{N}}$, and $\{s_i\}_{i\in\mathbb{N}}$ be families from $\mathcal{S}_{\mathscr{L}^!_\oplus}$ such that $\hat{s}_{!\sigma}(M_\varepsilon) \xrightarrow{(?^1)} t_0 \xrightarrow{(!^1)} \ldots \xrightarrow{(!^1)} t_i \ldots$ and $\hat{s}_{!\sigma}(M_\mu) \xrightarrow{(?^1)} s_0 \xrightarrow{(!^1)} \ldots \xrightarrow{(!^1)} s_i \ldots$. Since R is an α-bisimulation, for every i, it holds that $t_i\, R\, s_i$. Looking at the definition of $\mathscr{L}^!_\oplus$, it is easy to realise that:*

$$t_i = \{([\Omega_! \oplus^\varepsilon I], [I, \ldots, I]), ([\sigma], [\sigma, \ldots, \sigma])^{\varepsilon^i}\}_{i\in\mathbb{N}};$$
$$s_i = \{([\Omega_! \oplus^\mu I], [I, \ldots, I]), ([\sigma], [\sigma, \ldots, \sigma])^{\mu^i}\}_{i\in\mathbb{N}}.$$

By the definition of a α-bisimulation, we see that this implies that $\alpha \geq |\varepsilon^i - \mu^i|$. Since this reasoning can be done for every α such that M_ε and M_μ are α-bisimilar, it means that: $\delta^b_{!\sigma,!}(M_\varepsilon, M_\mu) \geq \sup_{i\in\mathbb{N}} |\varepsilon^i - \mu^i|$. Moreover, if we consider the special case where $\varepsilon = 0$, we can actually construct a μ-bisimulation by taking

$$R = (\hat{s}_{!\sigma}(M_0), \hat{s}_{!\sigma}(M_\mu)) \cup \{(t_0, s_0)\} \cup \{((0,A),(\mathcal{D},A) \mid \sum\nolimits_{\mathcal{D}} \leq \mu\}.$$

We can easily check that R is indeed a μ-bisimulation, which tells us that $\delta^b_{!\sigma,!}(M_0, M_\mu) = \mu$.

In Section 5.5 below, we will prove that $\delta^b_{\sigma,!}$ coincides with $\delta^c_{\sigma,!}$. While, as we will see soon, this helps a lot when precisely evaluating the distance between terms, sometime a characterization of $\delta^b_{\sigma,!}$ by traces is very effective when deriving lower bounds on the distance between terms (see, e.g., Section 5.4).

5.4 A Trace Characterization

In this section, we will characterize the metric defined in the previous section in an *inductive* way, by way of traces. This will be useful when proving full abstraction. A *trace* is a (possibly empty) sequence of actions in $\mathcal{A}_{\mathscr{L}^!_\oplus}$. Formally, the set of traces is generated by the following grammar:

$$\mathsf{T} \in \mathscr{T} ::= \top \;\Big|\; a \cdot \mathsf{T} \qquad \text{where } a \in \mathcal{A}_{\mathscr{L}^!_\oplus}.$$

If $\mathsf{T} \in \mathscr{T}$, and $t \in \mathcal{S}_{\mathscr{L}^!_\oplus}$, we write $t \xrightarrow{\mathsf{T}} s$ if $\mathsf{T} = a_1, \ldots, a_n$ and there exists a sequence of states $t_0, \ldots, t_n$ where $t_0 = t$, $t_n = s$, and moreover for every $1 \leq i < n$, it holds that $t_{i-1} \xrightarrow{a_i} t_i$. A trace T is said to be *admissible* for t if there exists s such that $t \xrightarrow{\mathsf{T}} s$. The set of admissible traces for t is indicated as $\mathcal{A}(t)$. The following is a well-posed definition, because the underlying WLTS $\mathscr{L}^!_\oplus$ is by definition deterministic.

Definition 5 *Let be $t \in \mathcal{S}_{\mathscr{L}^!_\oplus}$, and $\mathsf{T} \in \mathcal{A}(t)$. Then there is a unique u such that $t \xrightarrow{\mathsf{T}} u$, and we define the* success probability of T starting from t, *that we note $\mathsf{PR}_\mathsf{T}(t)$, as $w(u)$.*

We can express the metric $\delta^b_{\mathscr{L}^!_\oplus}$ as the maximum separation that a trace can induce:

Proposition 1 *Let $t, s \in \mathcal{S}_{\mathscr{L}^!_\oplus}$. Then:*

$$\delta^b_{\mathscr{L}^!_\oplus}(t,s) = \begin{cases} \sup_{\mathsf{T}\in\mathscr{T}} |\mathsf{PR}_\mathsf{T}(t) - \mathsf{PR}_\mathsf{T}(s)| & \text{if } \mathcal{A}(t) = \mathcal{A}(s); \\ 1 & \text{otherwise.} \end{cases}$$

Proof. We show separately the two inequalities:

- First, if two states t and s are related by an ε-bisimulation, we can see that $|\mathsf{PR}_\mathsf{T}(t) - \mathsf{PR}_\mathsf{T}(s)| \leq \varepsilon$ (since the states obtained after having done every action in T are still ε-bisimilar).

- To show the other implication, it is sufficient to define, for every ε, a relation imposing a bound ε on the success probabilities of traces, and to show that it is a ε-bisimulation.

□

Please observe that Proposition 1 gives an inductive definition of $\delta^t_!$, while the notion of ε-bisimulation is defined coinductively.

We conclude this section by stating a corollary of Proposition 1 which will be useful in the following.

Corollary 1 *Let be M and N of type σ, such that $\hat{s}_\sigma(M)$ and $\hat{s}_\sigma(N)$ are* not *ε-bisimilar. Then there exists a trace T such that $|\mathsf{PR}_\mathsf{T}(\hat{s}_\sigma(M)) - \mathsf{PR}_\mathsf{T}(\hat{s}_\sigma(N))| \geq \varepsilon$.*

5.5 Full Abstraction

In this section, we prove that $\delta^b_{\sigma,!}$ coincides with $\delta^c_{\sigma,!}$.

5.5.1 Soundness

We first of all show that the metric $\delta^b_{\sigma,!}$ is *sound* with respect to $\delta^c_{\sigma,!}$, i.e. that $\delta^b_{\sigma,!}$ discriminates at least as much as $\delta^c_{\sigma,!}$:

Theorem 4 (Soundness) *For any terms M and N of $\Lambda^!_\oplus$, such that $\vdash M : \sigma$ and $\vdash N : \sigma$, it holds that $\delta^c_{\sigma,!}(M,N) \leq \delta^b_{\sigma,!}(M,N)$.*

The rest of this section is devoted to the proof of Theorem 4. Please remember that our definition of the tuple distance is based on the notion of ε-bisimulation. Proving the soundness theorem, thus, requires us to show that for any terms M and N of type σ such that $\hat{s}_\sigma(M)$ and $\hat{s}_\sigma(N)$ are ε-bisimilar, and for any $\tau \in \mathscr{A}, C \in \mathcal{C}^\sigma_\tau$, it holds that $| \sum_{[\![C[M]]\!]} - \sum_{[\![C[N]]\!]} | \leq \varepsilon$.

Our proof strategy is based on the fact that we can decompose every evaluation path of a term in the form $C[L]$ into *external* reduction steps (that is, steps that do *not* affect L), and *internal* reduction steps (that is, reduction steps affecting L, but which can be shown to correspond *only* to actions from $\mathscr{L}^!_\oplus$). Intuitively, if we reduce in parallel $C[M]$ and $C[N]$, we are going to have steps where only the context is modified (and the modification does not depend on whether we are considering the first program or the second), and steps where the internal part is modified, but these steps cannot induce too much of a difference between the two programs, since the two internal terms are ε-bisimilar.

We first of all need to generalize the notion of a context to one dealing with tuples rather than terms. We in particular need contexts with multiple holes having types which match those of the tuple (or, more precisely, the *typtuplone*-state) they are meant to be paired with. More formally:

Definition 6 (Tuple Contexts) Tuple contexts *are triples of the form (C, A, γ), where C is an open term, $A = (\boldsymbol{\sigma}, \boldsymbol{\tau})$ is a (n,m) tuple type, and γ is a type such that $!\boldsymbol{x}_{\{1,\ldots,n\}} :! \boldsymbol{\sigma}, \boldsymbol{y}_{\{1,\ldots,m\}} : \boldsymbol{\tau} \vdash C : \gamma$. We note $\mathscr{C}^{\mathbf{T}}$ the set of tuple contexts. A tuple context (C, A, γ) is said to be an* open value *if C is of one of the following four forms: $\lambda x.M$, $\lambda !x.M$, $!M$, y_i (where $i \in \mathbb{N}$).*

We now want to define *when* a tuple context and an A-state) can be paired together, and the operational semantics of such an object, which will be derived from that of $\Lambda^!_\oplus$ terms. This is the purpose of the following definition:

Definition 7 (Tuple Context Pairs) *We say that a pair $u = (C, t)$ is a* tuple context pair *iff $t = (A, \mathcal{D})$ is an A-state, and $\exists \gamma \in \mathscr{A}$, $(C, A, \gamma) \in \mathscr{C}^{\mathbf{T}}$. We indicate as $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ the set of tuple context pairs. Moreover, given such a $u = (C, (A, \mathcal{D}))$, we define $\boldsymbol{F}(u)$ as the (potentially infinite) distribution over $\mathscr{T}$ given by:*

$$\boldsymbol{F}(u) = \{C\{\boldsymbol{x}/\boldsymbol{M}\}\{\boldsymbol{y}/\boldsymbol{N}\}^{\mathcal{D}(\boldsymbol{M},\boldsymbol{N})}\}_{(\boldsymbol{M},\boldsymbol{N}) \in \mathcal{S}(\mathcal{D})}.$$

Giving a notion of context distance for A-states is now quite easy and natural, since we know how contexts for such objects look like. For the sake of being as general as possible, this notion of a distance is parametric on a set of tuple contexts $\mathscr{C} \subseteq \mathscr{C}^{\mathbf{T}}$.

Definition 8 *Let be $\mathscr{C} \subseteq \mathscr{C}^{\mathbf{T}}$, $A \in \mathbf{T}$, and t, s two A-states. We define:*

$$\delta^c_{\mathscr{C}}(t,s) = \sup_{(C,A,\sigma)\in\mathscr{C}} \left| \sum\nolimits_{[\![\boldsymbol{F}(C,t)]\!]} - \sum\nolimits_{[\![\boldsymbol{F}(C,s)]\!]} \right|$$

Unsurprisingly, the context distance between terms equals $\delta^c_{\mathscr{C}^{\mathbf{T}}}$ when applied to A-states obtained through $\hat{s}_\sigma(\cdot)$:

Proposition 2 *If $\vdash M, N : \sigma$, then $\delta^c_{\sigma,!}(M,N) = \delta^c_{\mathscr{C}^{\mathbf{T}}}(\hat{s}_\sigma(M), \hat{s}_\sigma(N))$.*

But why did we introduce $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$? Actually, these pairs allow for a fine analysis of how tuples behave when put in a context, which in turn is precisely what we need to prove Theorem 4. This analysis, however, is not possible without endowing $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ itself with an operational semantics, which is precisely what we are going to do in the next paragraphs.

Semantics for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$:

In this paragraph, we define, for every $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, its semantics $[\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, which is a (potentially infinite) distribution over $\{k \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U}) \mid \mathbf{F}(k) \in \mathrm{Distr}(\mathscr{V})\}$.

Two relations need to be defined. On on the one hand, we need a one-step *labelled* transition relation $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ which turn an element of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ into a distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ by perfoming an action. Intuitively, one step of reduction in $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ corresponds to *at most* one step of reduction in $\mathscr{L}^!_\oplus$. If that step exists, (i.e. if the *term* is reduced) then the label is the same, and otherwise (i.e., if only the *context* is reduced), the label is just τ. We also need a multi-step approximation semantics $\Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ between elements of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ and subdistributions over the same set. The latter is based on the former, and both are formally defined in Figure 8, where

- E is an evaluation context;
- t is an (n,m)-state from $\mathcal{S}_{\mathscr{L}^!_\oplus}$;
- h is a tuple-context pair from $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$;
- For every context C, $C_{\mathrm{remove}(E)}$ stands for the context

$$C\{y_1/y_{1-\#\{j|j\in E\wedge j<1\}}\}\ldots\{y_n/y_{n-\#\{j|j\in E\wedge j<n\}}\}$$

Definition 9 *We define a* one-step labelled transition relation for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$*, that we note $h \xrightarrow{a}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$ with $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, $a \in \mathcal{A}_{\mathscr{M}^!_\oplus} \cup \{\tau\}$, and $\mathcal{D}$ a finite distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, as specified in Figure 8.*
We call normal form *those elements in $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ that cannot be reduced by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, and we note $NF(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$ the set of normal forms.*

Observe that we can actually give a characterization of normal forms:

$$NF(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})) = \{(C,t) \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U}) \mid C \text{ is an open value }\}. \tag{5}$$

If $\mathcal{D} \in \mathrm{Distr}(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$, we note $\mathscr{V}(\mathcal{D}) = \sum_{h\in NF(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))} \mathcal{D}(h) \cdot \{h^1\}$.

Definition 10 *We define an* approximation semantics relation for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$*, that we note $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, where h is an element of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, and $\mathcal{D}$ is a finite distribution over $NF(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$, as specified in Figure 8.*
We define a semantics for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$*, that we denote $[\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ for $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, as:*

$$[\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} = \sup\{\mathcal{D} \mid h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}\}.$$

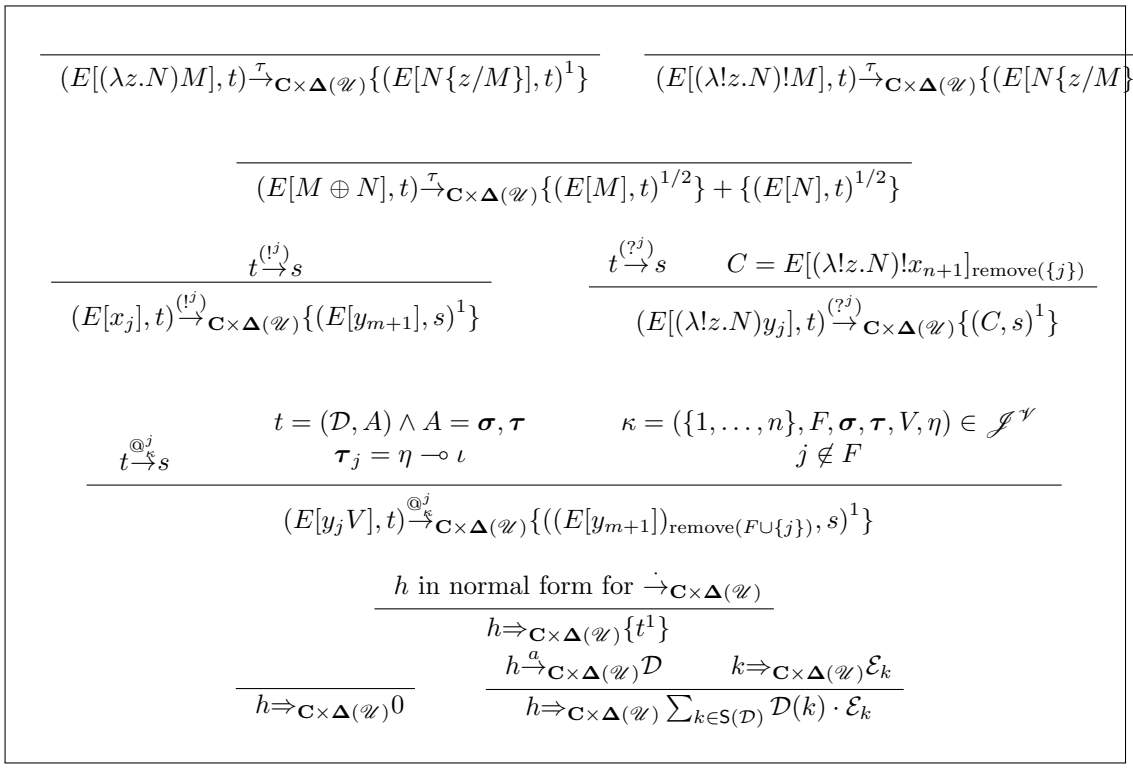


Figure 8: Rules for $\Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$

Relating $[\![]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ and $[\![]\!]$:
We first show that this definition can indeed be related to the usual semantics for terms. This takes the form of the following lemma:

Lemma 4 *Let be $u\in\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})$. Then:*
- $\{\mathcal{D}\mid u\Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}\mathcal{D}\}$ *is a directed set. We define* $[\![u]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ *as its least upper bound;*
- $\boldsymbol{F}(\cdot):\mathit{Distr}(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))\to\mathit{Distr}(\mathscr{T})$ *is continuous;*
- $[\![\boldsymbol{F}(u)]\!]=\boldsymbol{F}([\![u]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$.

Proof. The proof of the first point is exactly the same as the one given in [10] for the operational semantics of a probabilistic λ-calculus. The second and third ones are more involved.

Continuity of $\mathbf{F}()$: We have to show that, for any countable directed subset I of $\mathrm{Distr}(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$:

$$\sup\{\mathbf{F}(\mathcal{D})\mid\mathcal{D}\in I\}=\mathbf{F}(\sup\{\mathcal{D}\mid\mathcal{D}\in I\}),$$

where we have extended in a natural way the definition of $\mathbf{F}(h)$ to $\mathbf{F}(\mathcal{D})$, with $\mathcal{D}\in\mathrm{Distr}(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$.

We can first observe that, since $\mathbf{F}()$ is monotonous, it holds that: $\mathbf{F}(\sup\{\mathcal{D}\mid\mathcal{D}\in I\})\geq\sup\{\mathbf{F}(\mathcal{D})\mid\mathcal{D}\in I\}$. The other inequality is more involved. Let be $\mathcal{E}=\sup\{\mathcal{D}\mid\mathcal{D}\in I\}$. Since I is a countable directed set, there exists an increasing sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ in I such that $\mathcal{E}=\sup\{\mathcal{F}_n\mid n\in\mathbb{N}\}$. For $M\in\mathscr{T}$, we note $\alpha_{M,(C,t)}=\sum_{\{K\mid C[K]=M\}}t(K)$. We use it to express $\mathbf{F}(\mathcal{E})$:

$$\mathbf{F}(\mathcal{E})(M)=\sum_{(C,t)\in\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}\lim_{n\to\infty}\mathcal{F}_n(C,t)\cdot\alpha_{M,(C,t)}.$$

Moreover, please observe that we can restrict ourselves to consider distributions over a countable subset of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ (which consists in the element of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ that can be generated by our language). Since $\mathcal{E}$ is a distribution on a countable set, we know that for every ε there exists a finite subset J_ε of I, such that $\sum_{(C,t)\notin J_\varepsilon} \mathcal{E}(C,t) \leq \varepsilon$. So for every ε, it holds that:

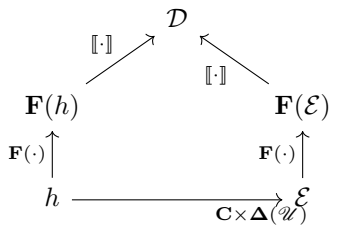
$$\mathbf{F}(\mathcal{E})(M) = \sum_{(C,t)\in J_\varepsilon} \lim_{n\to\infty} \mathcal{F}_n(C,t) \cdot \alpha_{M,(C,t)} + \varepsilon.$$

Since we now have to consider only a finite sum, we can exchange sum and limit:

$$\mathbf{F}(\mathcal{E})(M) = \lim_{n\to\infty} \sum_{(C,t)\in J_\varepsilon} \mathcal{F}_n(C,t) \cdot \alpha_{M,(C,t)} + \varepsilon.$$

And since that's true for every ε, we have the result.

$[\![\mathbf{F}(\cdot)]\!] = \mathbf{F}([\![\cdot]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$**:** We show separately the two inequalities. First, we show that, for every $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, $\mathbf{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}) \leq [\![\mathbf{F}(h)]\!]$. The proof is based on the fact that Diagram (6) commutes:



(6)

The result stated by (6) is formally given by the following lemma.

Lemma 5 *For every $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, it holds that if $h \to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, then $[\![\boldsymbol{F}(h)]\!] = \sum_{k\in\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}(k) \cdot [\![\boldsymbol{F}(k)]\!]$*

Proof. The proof consists in considering every case in the definition of $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, and use continuity result for $[\![\cdot]\!]$. □

We are now able to show that $\mathbf{F}([\![\cdot]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}) \leq [\![\mathbf{F}(\cdot)]\!]$:

Proposition 3 *For every $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, it holds that $\boldsymbol{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}) \leq [\![\boldsymbol{F}(h)]\!]$.*

Proof. Observe that it is sufficient to show that $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$ implies $\mathbf{F}(\mathcal{D}) \leq [\![\mathbf{F}(h)]\!]$. We prove it by induction on the derivation of $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$.

- If this derivation is : $\dfrac{}{h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \emptyset}$, then $\mathcal{D} = \emptyset$, and the result holds.
- If this derivation is $\dfrac{h \text{ in normal form}}{h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{h^1\}}$, then $\mathbf{F}(h)$ is a distribution over values (it can be checked easily by considering the carcaterisation of normal forms given in Equation (5)), and so $[\![\mathbf{F}(h)]\!] = \mathbf{F}(h)$, and the result holds.
- If this derivation is: $\dfrac{h \to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{F} \qquad (k \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}_k)_{k\in\mathsf{S}(\mathcal{F})}}{h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \sum_{k\in\mathsf{S}(\mathcal{F})} \mathcal{F}(k) \cdot \mathcal{E}_k}$, we apply Lemma 5 : $[\![\mathbf{F}(h)]\!] = \sum_{k\in\mathsf{S}(\mathcal{F})} \mathcal{F}(s) \cdot [\![\mathbf{F}(k)]\!]$. Moreover, we can apply the induction hypothesis to $s \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}_s$ for every $k \in \mathsf{S}\mathcal{F}$, which leads to: $\mathbf{F}(\mathcal{E}_k) \leq [\![\mathbf{F}(k)]\!]$. Using it, we obtain:

$$\begin{aligned}\mathbf{F}(\mathcal{D}) &= \sum_k \mathcal{F}(k) \cdot \mathbf{F}(\mathcal{E}_k) \\ &\leq \sum_k \mathcal{F}(k) \cdot [\![\mathbf{F}(k)]\!] \\ &= [\![\mathbf{F}(h)]\!]\end{aligned}$$

, which is the result.

$$\frac{}{h \to^{\star}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{h^1\}} \qquad \frac{h \to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \sum_{i\in I}\alpha_i \cdot \{k_i{}^1\} \qquad h_i \to^{\star}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}_i}{h \to^{+}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \sum_{i\in I}\alpha_i \cdot \mathcal{E}_i} \qquad \frac{h \to^{+}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}}{h \to^{\star}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}}$$

Figure 9: Definition of $h \to^{+}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$

□

We are now going to show the other inequality, namely that, for any $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, $[\![\mathbf{F}(h)]\!] \leq \mathbf{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$. We first define a particular subset of $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, namely those h, such that $\mathbf{F}(h)$ can be seen as a term (remember that in the general case, it is a distribution over terms).

Definition 11 *We define a subset* $Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ *of* $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$*, as:* $Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} = \{(C, (\mathcal{D}, A)) \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U}) \mid \mathcal{D}$ *is a Dirac distribution*$\}$*. We define an operator* $\overline{(\cdot)}^{U}$*:*

$$\overline{(\cdot)}^{U} : Distr(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})) \to Distr(\, Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$$
$$\mathcal{D} \to \sum_{h=(C,(\mathcal{E},A))\in\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}(h) \cdot \sum_{K\in\mathcal{S}(\mathcal{E})} \mathcal{E}(K) \cdot \{(C, (\{K^1\}, A))^1\}.$$

Observe that if $h \in \text{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, $\mathbf{F}(h)$ is a Dirac distribution on terms. By abuse of notation, we'll use $\mathbf{F}(h)$ to denote the term M such that $\mathbf{F}(h) = \{M^1\}$. The operator $\overline{(\cdot)}^{\text{U}}$ preserves $[\![]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, as formally stated in Lemma 6 below:

Lemma 6 *For every* $\mathcal{D} \in Distr(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$*, it holds that:* $\boldsymbol{F}([\![\mathcal{D}]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}) = \boldsymbol{F}([\![\overline{\mathcal{D}}^{U}]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$.

Proof. It is a consequence of the fact that $\mathscr{L}^{!}_{\oplus}$ has the corresponding property: if $t = (\mathcal{D}, A) \in \mathcal{S}_{\mathscr{L}^{!}_{\oplus}}$, $t \xrightarrow{a} (\mathcal{E}, B)$, then for every $K \in \mathsf{S}(\mathcal{D})$ there exists $\mathcal{F}_K$ such that $(\{K^1\}, A) \xrightarrow{a} (\mathcal{F}_K, B)$, and moreover $\mathcal{E} = \sum_{K\in\mathsf{S}(\mathcal{D})} \mathcal{D}(K) \cdot \mathcal{F}_K$. □

We introduce some notations on $\Rightarrow$: we write $M \overset{n}{\Rightarrow} \mathcal{D}$ with $n \in \mathbb{N}$, if there exists a derivation $\Delta : M \Rightarrow \mathcal{D}$ of size at most n. Moreover, we write $\mathcal{D} \overset{n}{\Rightarrow} \mathcal{E}$ if there exists a finite set I such that $\mathcal{D} = \sum_{i\in I} \alpha_i \cdot \{M_i{}^1\}$, and such that moreover $M_i \overset{n}{\Rightarrow} \mathcal{E}_i$, and $\mathcal{E} = \sum_{i\in I} \alpha_i \cdot \mathcal{E}_i$. Please observe that this definition imply that $\mathcal{D}$ is always a *finite* distribution over terms. We also introduce auxiliary notations for $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, as given in Figure 9.

Moreover, if $\mathcal{D}$, $\mathcal{E}$ are distributions over a set, we say that $\mathcal{D}$ is a *finite approximation* of $\mathcal{E}$ if $\mathcal{D}$ is finite, and moreover $\mathcal{D} \leq \mathcal{E}$. We denote it by $\mathcal{D} \overset{f}{\leq} \mathcal{E}$.

We first consider the case where $h \in \text{Unary}^{\mathbf{C}\times\overline{\mathbf{\Delta}}(\mathscr{U})}$ is such that $\mathbf{F}(h)$ is a value. We can see that the following diagram commutes

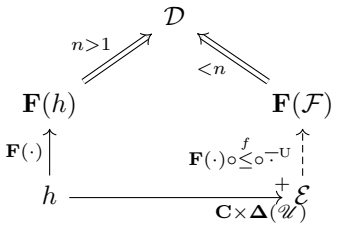
$$\begin{array}{ccc} & \{V^1\} & \\ {\scriptstyle 1}\swarrow & & \searrow \\ \mathbf{F}(h) & & \mathbf{F}(k) \\ {\scriptstyle \mathbf{F}(\cdot)}\uparrow & & {\scriptstyle \mathbf{F}(\cdot)}\uparrow \\ h & \xrightarrow[\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})]{\star} & \{k^1\} \end{array} \tag{7}$$

, as formally stated in Lemma 7 below.

Lemma 7 *Let be $h \in Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, such that $\boldsymbol{F}(h)$ is a value. There exists $k \in Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ such that $h \to^{\star}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{k^1\}$ and $k \in NF(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$ verifies $\boldsymbol{F}(k) = \boldsymbol{F}(h)$.*

Proof. The proof consists in showing that, if $\mathbf{F}(h)$ is a value, then there is a finite sequence $h = k_0 \ldots, k_n$ with $k_i \to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{{k_{i+1}}^1\}$, every $k_i \in \text{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, $k_n \in NF(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$, and moreover $\mathbf{F}(k_i) = \mathbf{F}(h)$. □

We now consider the case where $\mathbf{F}(h)$ is not a value: it means that we have to consider $\mathcal{D} \in \text{Distr}(\mathscr{T})$, such that $\mathbf{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$, with $n > 1$. Our aim is to show that $\mathcal{D} \leq [\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. In that aim, we show first that Diagram 8 below commutes,



(8)

as formally stated by the following Lemma:

Lemma 8 *For any $h \in Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, such that $\boldsymbol{F}(h)$ is not a value, there exists $\mathcal{E} \in Distr(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$ such that $h \to^{+}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and for every $\mathcal{D} \in Distr(\mathscr{T})$, and $n \geq 1$ such that $\boldsymbol{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$, there exists $\mathcal{F}$ such that $\mathcal{F} \overset{f}{\leq} \overline{\mathcal{E}}^{U}$, and moreover $\boldsymbol{F}(\mathcal{F}) \overset{m}{\Rightarrow} \mathcal{D}$ with $m < n$.*

Before doing the proof, we give some intuitions about this diagram: we first consider these steps of $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ that reduce more that one redex. It forces us to use not the one step transition $\to$ to talk about the semantics for terms, but directly the approximation semantics $\Rightarrow$. Let us consider now the steps of $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ where no redex is reduced: more precisely, we can characterize them as the ones corresponding to actions passing a term that is already a value from the exponential part to the linear part of the inside tuple. The proof is made possible by the fact that we can show that there can only be a finite number of reduction step $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ that don't reduce the underlying term, before there is a step reducing at least a redex in it.

Proof. We are first going to state a lemma about the approximation semantics $\Rightarrow$, which we'll need later in the proof of Lemma 8. Indeed, in order to manage the problem of a step of $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ corresponding to strictly more than one step of $\to$, we need to transform a derivation $\Delta : M \Rightarrow \mathcal{D}$ in derivations corresponding to reduction processes done after having completely reduced part of M. More precisely, the Lemma 9 below gives us a way to extract of a derivation tree the subtrees corresponding to the evaluation after having reduced all the redex in a subterm in evaluation position. Please observe that it talks only about semantics of term, and express a form of compositionnality in the derivation. It will be used to treat the case of applicative actions, where we reduce many redex in one step of reduction for $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$.

Lemma 9 *Let be M a closed term which isn't a value, with $E[M] \overset{n}{\Rightarrow} \mathcal{D}$ a valid derivation of the approximation semantics. Then there exists a finite set $I \subseteq \mathbb{N}$, and for every $i \in I$ a value V_i, a coefficient α_i, and a valid derivation $\Xi_i : E[V_i] \overset{m}{\Rightarrow} \mathcal{E}_i$ such that $n < m$, $\mathcal{D} = \sum_{i\in I} \alpha_i \cdot \mathcal{E}_i$, and moreover for every V, $\sum_{i|V_i=V} \alpha_i \leq [\![M]\!](V)$*

Proof. The proof is by induction on Δ. □

Let us now show the first part of Lemma 8. Let $h \in \text{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ such that $\mathbf{F}(h)$ is not a value. By definition of $\text{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, h is of the form: $h = (C, (\{K^1\}, A))$. We are going to construct the distribution $\mathcal{E}$ by case analysis on C . Please observe that there is at most one possible action a such that $(C, (\{K^1\}, A)) \overset{a}{\to}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, and moreover it depends only on in C and A. Consequently, we do a case analysis on the action a.

- If $a = \tau$: it means that the reduced redex doesn't affect the inside tuple. We take $\mathcal{E}$ such that $h \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E} = \sum_{1\leq i\leq n} \frac{1}{n}\{(C_i, (K, A))^1\}$. Observe that $\mathcal{E}$ is actually a finite distribution. Let be $\mathcal{D} \in \mathrm{Distr}(\mathscr{V})$, and $n \geq 1$, such that $\mathbf{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$. Since $\mathbf{F}(h) \notin \mathscr{V}$, and $n \geq 1$, $\mathbf{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$ is of the form:
$$\frac{\mathbf{F}(h) \to N_1, \ldots, N_p \qquad (N_i \overset{m_i}{\Rightarrow} \mathcal{D}_i)_{1\leq i\leq p}}{\mathbf{F}(h) \Rightarrow \mathcal{D} = \sum \frac{1}{n} \cdot \mathcal{D}_i}$$
, where for every i, $m_i < n$. We take $\mathcal{F} = \mathcal{E}$, $m = \max\{n_i \mid 1 \leq i \leq p\} < n$ and the result is obtained by seeing that $\mathbf{F}(\mathcal{F}) = \sum_{1\leq i\leq p} \frac{1}{n} \cdot \{N_i{}^1\} \overset{m}{\Rightarrow} \mathcal{D}$.
- If $a = (?^j)$: then $C = E[(\lambda!z.M)y_j]$.
It means that the only rule than can be used to show $h \to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \cdot$ is:
$$\frac{t \overset{(?^j)}{\to} s}{h = (E[(\lambda!z.M)y_j], t) \overset{(?^j)}{\to}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{k^1\}}$$
where $k = (E[(\lambda!z.M)!x_{n+1}]_{\mathrm{remove}(\{j\})}, s)$.
Observe that $k \in \{l \in \mathrm{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mid l \xrightarrow{a}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \cdot \text{ implies } a = \tau\}$.
Using the previous case, we know that there exist $\mathcal{E}_k$ verifying the condition of Lemma 8 with respect to k. Now, we are going to show that we can reuse it with respect to h. First, notice that $h \to^+_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$ (since $\to^+_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ is transitive). So we can take $\mathcal{E} = \mathcal{E}_k$. Let be $n \geq 1$, and $\mathcal{D}$ such that $\mathbf{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$. Since $\mathbf{F}(h) = \mathbf{F}(k)$, it holds that $\mathbf{F}(k) \overset{n}{\Rightarrow} \mathcal{D}$. We can conclude by using the properties of $\mathcal{E}_k$ with respect to k.
- If $a = @^j_\kappa$: it means that $C = E[y_j V]$, and the rule used to reduce h by $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ is the applicative one. So we can take $\mathcal{E}$ defined as:
$$\frac{t \overset{@^j_\kappa}{\to} s \qquad \begin{array}{c} t = (\mathcal{D}, A) \wedge A = \boldsymbol{\sigma}, \boldsymbol{\tau} \\ \boldsymbol{\tau}_j = \eta \multimap \iota \end{array} \qquad \kappa = (\{1, \ldots, n\}, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, V, \eta) \in \mathscr{J}^{\mathscr{V}} \quad \begin{array}{c} \\ j \notin F \end{array}}{(E[y_j V], t) \overset{@^j_\kappa}{\to}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E} = \{((E[y_{m+1}])_{\mathrm{remove}(F\cup\{j\})}, s)^1\}}$$
We see that $h \to^+_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$. Let be $\mathcal{D}$, and $n \geq 1$, such that $\mathbf{F}(h) \overset{n}{\Rightarrow} \mathcal{D}$. We need to define $\mathcal{F}$ as specified in Lemma 8. Notice that $\mathcal{E} \notin \mathrm{Distr}(\mathrm{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$, but we are able to express $\overline{\mathcal{E}}^{\mathrm{U}}$ as:
$$\overline{\mathcal{E}}^{\mathrm{U}} = \sum_{W\in\mathscr{V}} [\![\mathbf{F}(y_j V, t)]\!](W) \cdot \{k_W{}^1\},$$
where $\mathbf{F}(k_W) = (\mathbf{F}(E, t))[W]$, and moreover $k_W \in \mathrm{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. We are going to apply Lemma 9 to $(\mathbf{F}(E, t))[\mathbf{F}(y_j V, t)]$: there exist a finite set I, and a family $(W_i)_{i\in I}$ of $\mathscr{V}$, such that $\sum_{i\in I} \alpha_i \cdot \{(\mathbf{F}(E, t))[W_i]^1\} \overset{m}{\Rightarrow} \mathcal{D}$, with $m < n$, and for every $W \in \mathscr{V}$, it holds that $\sum_{i|W_i=W} \alpha_i \leq [\![\mathbf{F}(y_j V, t)]\!](W)$. Now we define $\mathcal{F} = \sum \alpha_i \cdot \{k_{W_i}{}^1\}$, and we can see that $\mathcal{F} \overset{f}{\leq} \overline{\mathcal{E}}^{\mathrm{U}}$ and $\mathbf{F}(\mathcal{F}) \overset{m}{\Rightarrow} \mathcal{D}$, and so we have the result.
- If $a = (!^j)$: it means that $C = E[x_j]$, and that the rule used is:
$$\frac{t \overset{(!^j)}{\to} s}{(E[x_j], t) \overset{(!^j)}{\to}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{(E[y_{m+1}], s)^1\}}$$
We define $\boldsymbol{M}$ and $\boldsymbol{N}$ such that $K = (\boldsymbol{M}, \boldsymbol{N})$. We can distinguish two cases, depending on $\boldsymbol{M}_j$:
  - If $\boldsymbol{M}_j$ is not a value. Then we are going to apply Lemma 9. Indeed, $\mathbf{F}(C, t) = \mathbf{F}(E, t)[\boldsymbol{M}_j]$. The proof is essentially the same that for the applicative case.
  - If $\boldsymbol{M}_j$ is a value. Then we can define $k \in \mathrm{Unary}^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ by $k = (E[y_{m+1}], s)$. It is possible that the only action able to reduce k is of the form $(!^k)$, However, we can iterate the previous reasonning (which intuitively corresponds to do reduction on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, but not in the

actual term) only a finite number of times (because we can have only a finite number of occurrences of values that are in an evaluation position in $M = \mathbf{F}(C,t)$). So after a finite number of steps like that, we arrive to reduce this case to one we have already considered. □

Lemma 10 *For every $h \in Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, if $\boldsymbol{F}(h) \Rightarrow \mathcal{D}$, then $\mathcal{D} \leq \boldsymbol{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$.*

Proof. The proof is by induction on the size n of the derivation $\mathbf{F}(h) \Rightarrow \mathcal{D}$:

- If $n = 0$, then $\mathcal{D} = \emptyset$, and the result is true.
- If $n = 1$, or $\mathcal{D} = \emptyset$, or $\mathcal{D} = \{\mathbf{F}(h)^1\}$, and then $\mathbf{F}(h)$ is a value, and we can apply the second point of Lemma 8.
- If $n > 1$, then it implies that $\mathbf{F}(h)$ is not a value, and so we can apply the second point of Lemma 8, and apply the induction hypothesis on the strictly smaller derivations it gives us.

□

As a corollary, we can see that that $[\![\mathbf{F}(h)]\!] \leq \mathbf{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$ holds in fact for every $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, and not only for those in $Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$:

Proposition 4 *For any $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, it holds that $[\![\boldsymbol{F}(h)]\!] \leq \boldsymbol{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$*

Proof. First, we can see that Lemma 10 imply that $[\![\mathbf{F}(h)]\!] \leq \mathbf{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$ for every $h \in Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. Lemma 6 allows to generalize it to the whole if $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$. □

We have now only to sum up Proposition 3 and 4 to obtain the correspondence between $[\![\cdot]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ and $[\![\cdot]\!]$. □

Equivalence relations on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$: Before proceeding, we need to understand how any reflexive and symmetric relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ can be turned into a relation on *distributions* on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$. If R is a reflexive and symmetric relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, we lift it to distributions over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ by stipulating that $\mathcal{D}\ R\ \mathcal{E}$ whenever there exists a countable set I, a family $(p_i)_{i\in I}$ of positive reals of sum smaller than 1, and families $(h_i)_{i\in I}, (k_i)_{i\in I}$ in $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, such that $\mathcal{D} = \{{h_i}^{p_i}\}_{i\in I}$, $\mathcal{E} = \{{k_i}^{p_i}\}_{i\in I}$, and moreover $h_i\ R\ k_i$ for every $i \in I$.

We now want to precisely capture *when* a relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ can used to evaluate the distance between tuple-context pairs.

Definition 12 *Let R be a reflexive and symmetric relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$.*

- *We say that R is* preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ *if, for any $h, k \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ such that $h\ R\ k$, if $h \xrightarrow{a}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, then there exists $\mathcal{E}$ such that $k \xrightarrow{a}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and $\mathcal{D}\ R\ \mathcal{E}$.*
- *We say that R is* ε-bounding *if $h\ R\ k$, implies $|\sum_{\boldsymbol{F}(h)} - \sum_{\boldsymbol{F}(k)}| \leq \varepsilon$.*
- *Let $\mathscr{C}$ be a set of tuple contexts, and $t, s \in \mathcal{S}_{\mathscr{L}^!_\oplus}$ be two A-states. We say that R is* $\mathscr{C}$-closed with respect to t and s, *if for every C and γ such that $(C, A, \gamma) \in \mathscr{C}$, it holds that $(C,t)\ R\ (C,s)$.*

Please observe how any relation preserving $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ and being ε-bounding can be seen somehow as an ε-bisimulation, but on tuple-context pairs. The way we defined the lifting, however, makes it even a *stronger* notion, i.e., the ideal candidate for an intermediate step towards Soundness.

Preservation by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ leads us to preservation (in a weaker sense) by $\Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. That's the sense of the following lemma.

Lemma 11 *Let R be a reflexive and symmetric relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. Let $h, k \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ be such that hRk. Let $\mathcal{D}$ be such that $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$. Then there exists $\mathcal{E}$ such that $k \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and $\mathcal{D}R\mathcal{E}$.*

Proof. The proof is by induction on the derivation of $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$.

- if the derivation is $\dfrac{}{h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \emptyset}$, the result holds.

□

Lemma 12 *If $h, k \in \mathcal{S}_{\mathscr{L}^!_\oplus}$ are such that there exists a reflexive and symmetric relation R preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, ε-bounding, and containing (h, k), then it holds that $|\sum_{\boldsymbol{F}([\![h]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})} - \sum_{\boldsymbol{F}([\![k]\!]^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})}| \leq \varepsilon$.*

We use Lemma 12 to show the following proposition.

Proposition 5 *Let $\mathscr{C}$ be a set of tuple contexts, t, s two A-states and R a reflexive and transitive relation preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, ε-bounding, and $\mathscr{C}$-closed with respect to t and s. Then it holds that $\delta^c_{\mathscr{C},!}(t, s) \leq \varepsilon$.*

Proof. It's a direct consequence of Lemma 12. □

Moreover, we see that the conditions from Definition 12 are enough to guarantee that two terms are at context distance at most ε.

Proposition 6 *Let M, N be two terms of type σ. Suppose there exists a reflexive and symmetric relation R on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, which is preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, ε-bounding, and $\mathscr{C}^{\mathbf{T}}$-closed with respect to $\hat{s}_\sigma(M)$ and $\hat{s}_\sigma(N)$. Then $\delta^c_{\sigma,!}(M, N) \leq \varepsilon$.*

ε-bisimulation and $\Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$: What remains to be done, then, is to show that if two terms are related by R^ε, then they themselves satisfy Definition 12. Compulsory to that is showing that any ε-bisimulation can at least be turned into a relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$.

We need to do that, in particular, in a way guaranteeing the $\mathscr{C}$-closure of the resulting relation, and thus considering all possible tuple contexts from $\mathscr{C}$:

Definition 13 *Let R be a reflexive and symmetric relation on $\mathcal{S}_{\mathscr{L}^!_\oplus}$. Let be $\mathscr{C}$ a set of tuple contexts. We define its* contextual lifting to $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ with respect to $\mathscr{C}$ *as the following binary relation on $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$:*

$$\widehat{R}^{\mathscr{C}}_A = \bigcup_{(C,A,\sigma)\in\mathscr{C}} \{((C, t), (C, s)) \mid t, s \text{ } A\text{-states}, \text{ } t \text{ } R \text{ } s\}; \qquad \widehat{R}^{\mathscr{C}} = \bigcup_{A\in\mathbf{T}} \widehat{R}^{\mathscr{C}}_A.$$

The following result tells us that, indeed, any ε-bisimulation can be turned into a relation satisfying Definition 12:

Proposition 7 *Let R be a ε-bisimulation. Then $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ is preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ and ε-bounding, and $\mathscr{C}^{\mathbf{T}}$-closed with respect to every t, s such that $t \ R \ s$.*

Proof. Let R be an ε-bisimulation. We obtain that $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ is $\mathscr{C}^{\mathbf{T}}$-closed as a direct consequence of the lifting definition. Similarly, we see that $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ is ε-bounding as a direct consequence of the definition of a ε-bisimulation. Now, let us show that $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ is preserved by $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. Let be A such that $(h, k) \in \widehat{R}^{\mathscr{C}^{\mathbf{T}}}_A$. The proof is by induction on the derivation of $h \Rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$. It is based on the following idea: for every C such that there exist σ for which (C, A, σ) is a tuple context:

- or for every A-state t, (C, t) is in normal form.
- or there exists a family $(C_i)_{1\leq i\leq n}$ such that $(C_i, A, \sigma) \in \mathscr{C}^{\mathbf{T}}$, and moreover for every A-state t, $(C, t) \xrightarrow{\tau}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \sum_{1\leq i\leq n} \{(C_i, t)^1\}$
- or there exists an action a of $\mathscr{L}^!_\oplus$, and D, B such that (D, B, σ), such that for every A-state t, $(C, t) \xrightarrow{a}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \{(D, s)^1\}$, and $t \xrightarrow{a} s$.

□

We are finally ready to give a proof of soundness:

Proof. [of Theorem 4] Consider two terms M and N of type σ. Let ε be $\delta^b_{\sigma,!}(M, N)$. We take R^ε (defined in Definition 3 as the largest ε-bisimulation), and we see that $\hat{s}_\sigma(M) \ R^\varepsilon \ \hat{s}_\sigma(N)$. Proposition 7 tells us that we can apply Proposition 6 to M, N, and $\widehat{(R^\varepsilon)}^{\mathscr{C}^{\mathbf{T}}}$. Doing so we obtain that $\delta^c_{\sigma,!}(M, N) \leq \varepsilon$, which is the thesis. □

5.5.2 Completeness

We can actually show that $\delta^b_{\sigma,!}$ is also complete with respect to the contextual distance: that is the aim of this section.

Theorem 5 (Full Abstraction) *For every σ, $\delta^c_{\sigma,!} = \delta^b_{\sigma,!}$.*

Proof. One inequality is given by Theorem 4.

Please observe that Theorem 4 gives us already half of Theorem 5: indeed, it says that for every M and N of type σ, it holds that: $\delta^c_{\sigma,!}(M,N) \leq \delta^b_{\sigma,!}(M,N)$. Our goal is now to show the other inequality, that is $\delta^c_{\sigma,!}(M,N) \geq \delta^b_{\sigma,!}(M,N)$. The proof of the other one is based on the trace characterization $\delta^b_{\mathscr{L}^!_\oplus}$ given in Section 5.4. More precisely, we use the fact that every trace can be simulated by a context: for every $A \in \mathbf{T}$, we construct a tuple context (C, A, τ), such that for every A-state t, it holds that $\mathsf{PR}_\mathsf{T}(t) = \mathrm{Obs}(\mathbf{F}(C,t))$.

Lemma 13 *For every $\mathsf{T} \in \mathscr{T}$, for every $A \in \mathbf{T}$ such that T is an admissible trace for A, there exists a term C, and a type γ such that:*
- *(C, A, γ) is a tuple context.*
- *for every $t \in \mathcal{S}_{\mathscr{L}^!_\oplus}$ of the form $t = (\mathcal{D}, A)$, it holds that $\sum_{[\![\mathbf{F}(C,t)]\!]} = \mathsf{PR}_\mathsf{T}(t)$.*

Proof. Let be $A = (\boldsymbol{\sigma}, \boldsymbol{\tau})$ a type for tuple. Let be (n, m) the arity of A. We define a context C, and a type γ, by induction on the length of the admissible trace :
- if $\mathsf{T} = \top$: We want a context that terminates with probability 1, whatever term we fill it with. We take:

$$C = \lambda z.zy_1 \dots y_m$$
$$\gamma = (\boldsymbol{\tau}_1 \multimap \dots \multimap \boldsymbol{\tau}_m \multimap \alpha) \multimap \alpha$$

- if $\mathsf{T} = (?^i) \cdot \mathsf{S}$. Let be B such that, whenever we do the action $(?^i)$ from a A-state, we obtain a B-state. Let D and η the open term and type obtained by induction hypothesis applied to S and B. We take $\gamma = \eta$, and:

$$C = (\lambda !z.D\{x_{n+1}/z\}_{\mathrm{add}(\{i\})})y_i$$

- if $\mathsf{T} = (!^i) \cdot \mathsf{S}$. Let be B such that, whenever we do the action $(!^i)$ from a A-state, we obtain a B-state. Let D and η be the open term and type obtained by induction hypothesis applied to S and B. We take $\gamma = \eta$, and $C = D\{x_i/y_{m+1}\}$.
- if $\mathsf{T} = @^i_\kappa \cdot \mathsf{S}$. Since T is an admissible trace for $A = (\boldsymbol{\sigma}, \boldsymbol{\tau})$, κ should be of the form: $\kappa = (E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \iota)$. Let B be the corresponding tuple type obtained after the action $(!^i)$, and let be D and η the term and the type given by the induction hypothesis and B. Then we take $\gamma = \eta$:

$$C = (D_{\mathrm{add}(F \cup \{i\})})\{y_i/y_{n+1}M\}$$

□

When we combine Proposition 13 and Proposition 1, we obtain the following corollary:

Corollary 2 *Let be t and s two A-states.*

$$\delta^b_{\mathscr{L}^!_\oplus}(t,s) \leq \sup\{|Obs(\boldsymbol{F}(C,t)) - Obs(\boldsymbol{F}(C,s))| \;\; s.t. \;\exists \sigma,\, (C, A, \sigma) \in \mathscr{C}^{\mathbf{T}}\}$$

Proof. Let be t and s two A-states that are *not* ε-bisimilar. Then there exists a trace T that shows it, that is verifying $|\mathsf{PR}_t(\mathsf{T}) - \mathsf{PR}_s(\mathsf{T})| > \varepsilon$. Now, we just have to take the open term C such that $(C, A, \cdot) \in \mathscr{C}^{\mathbf{T}}$, corresponding to this trace T. And we have $|\mathrm{Obs}(\mathbf{F}(C,t)) - \mathrm{Obs}(\mathbf{F}(C,s))| > \varepsilon$, and the result folds. □

We can now transform Corollary 2 in a result on terms, which is exactly what we need to end the proof of theorem 5.

Lemma 14 *Let be M, N of type σ. Then:*

$$\delta^b_{\sigma,!}(M,N) \leq \sup\{|Obs(C[M]) - Obs(C[N])| \mid \exists \tau \text{ s.t. } hole : \sigma \vdash C : \tau\}$$

Proof. Let us unfold the definition of $\delta^b_{\sigma,!}(M,N)$: please recall that we defined it as $\delta^b_{\mathscr{L}^!_\oplus}(\hat{s}_\sigma(M), \hat{s}_\sigma(N))$. Let be A such that $\hat{s}_\sigma(\cdot)$ is an A-term. Please recall that $A = ([], [\sigma])$. We can see that for every C, τ such that $(C, A, \tau) \in \mathscr{C}^{\mathbf{T}}$, it holds that C can in fact be see as a context for terms: more precisely, it holds that $y_1 : \sigma \vdash C : \tau$ and moreover, it holds that

$$\mathrm{Obs}(((\lambda y_1.C)\mathrm{hole})[M]) = \mathrm{Obs}(\mathbf{F}(C, \hat{s}_\sigma(M))) \qquad \mathrm{Obs}(((\lambda y_1.C)\mathrm{hole})[N]) = \mathrm{Obs}(\mathbf{F}(C, \hat{s}_\sigma(N))).$$

Since we can transform every tuple of context for A into a context on terms in that way, we obtain the result. □

Theorem 5 is a direct consequence of Lemma 14: indeed, please recall that we had to show that $\delta^c_{\sigma,!}(M,N) \geq \delta^b_{\sigma,!}(M,N)$. Since that's exactly what Lemma 14 says, it ends the proof.

□

5.6 On an Up-to-Context Technique

5.6.1 Up-to ε-bisimulation

As we have just shown, context distance can be characterized as a coinductively defined metric, which turns out to be useful when evaluating the distance between terms. In this section, we will go even further, and show how an *up-to-context* [32] notion of ε-bisimulation is precisely what we need to handle our running example.

We first of all need to generalize our definition of a tuple: an *open tuple* is a pair $(\boldsymbol{M}, \boldsymbol{N})$, where $\boldsymbol{M}$ and $\boldsymbol{N}$ are sequences of (not necessarily closed) typable terms.

Definition 14 *If $K = (\boldsymbol{M}, \boldsymbol{N})$ is an open tuple, and $A = (\boldsymbol{\gamma}, \boldsymbol{\eta})$ is a tuple type, we say that $(\boldsymbol{\sigma}, \boldsymbol{\tau}, K, A)$ is a* substitution judgment *iff:*

- $!\boldsymbol{x} :! \boldsymbol{\sigma} \vdash \boldsymbol{M}_i : \boldsymbol{\gamma}_i$;
- *if n and m are such that $\boldsymbol{\tau}$ is a n-sequence, and $\boldsymbol{N}$ a m-sequence, then there exists a partition $\{E_1, \ldots E_m\}$ of $\{1, \ldots, n\}$ such that $\boldsymbol{y}_{E_j} : \boldsymbol{\tau}_{E_j} \vdash N_j : \eta_j$ for every $j \in \{1, \ldots, m\}$.*

$\mathscr{J}^{subst}$ is the set of all substitution judgments.

If $\kappa = (\boldsymbol{\sigma}, \boldsymbol{\tau}, K, A) \in \mathscr{J}^{\mathrm{subst}}$, and $H \in \mathscr{U}$ is of type $(\boldsymbol{\sigma}, \boldsymbol{\tau})$, then there is a natural way to form a tuple $\kappa[H]$, namely by substituting the free variables of K by the components of H. In the following, we restrict $\mathscr{J}^{\mathrm{subst}}$ to those judgments κ such that for every H, terms in the linear part of $\kappa[H]$ are values. Observe that we always have $\vdash \kappa[H] : A$. We extend the notation $\kappa[H]$ to distributions over $\mathscr{U}$: if $\mathcal{D}$ is a distribution over tuples of type $(\boldsymbol{\sigma}, \boldsymbol{\tau})$, we note $\kappa[\mathcal{D}] = \{\kappa[H]^{\mathcal{D}(H)}\}_{H \in \mathscr{U}}$, which is a distribution over tuples of type A. Moreover, we want to be able to apply our substitution judgments to the states of $\mathscr{L}^!_\oplus$. If $t = (\mathcal{D}, (\boldsymbol{\sigma}, \boldsymbol{\tau})) \in \mathcal{S}_{\mathscr{L}^!_\oplus}$, and $\kappa = (\boldsymbol{\sigma}, \boldsymbol{\tau}, K, A)$, the state of $\mathscr{L}^!_\oplus$ defined by $(\kappa[\mathcal{D}], A)$ will be often indicated as $\kappa[t]$.

Example 6 *We illustrate on a simple example the use of substitution judgments. Let be τ any type. Consider $\boldsymbol{\sigma} = [\tau \multimap \tau]$, and $\boldsymbol{\tau} = []$. Moreover, let $K = ([x_1], [I])$ and $A = ([\tau \multimap \tau], [\tau \multimap \tau])$. Then $\kappa = (\boldsymbol{\sigma}, \boldsymbol{\tau}, K, A)$ is a substitution judgment. We consider now a tuple of type $(\boldsymbol{\sigma}, \boldsymbol{\tau})$. In fact, we take here a tuple that will be useful in order to analyze our running example: $H = ([\Omega_! \oplus^\varepsilon I], [])$. By substituting H in κ, we obtain $\kappa[H] = ([\Omega_! \oplus^\varepsilon I], [I])$, and we can see easily that we obtain indeed a tuple of type A.*

The main idea behind up-to context bisimulation is to allow for the freedom of discarding any context when proving a relation to be a bisimulation. This is captured by the following definition:

Definition 15 *Let R be a relation on $\mathcal{S}_{\mathscr{L}^!_\oplus}$. R is an ε-*bisimulation up to context *if for every t and s such that $t\ R\ s$, the following holds:*

- *there exists $C \in \mathbf{T}$ such that $t = (\mathcal{D}, C)$, $s = (\mathcal{E}, C)$, and $|\sum_{\mathcal{D}} - \sum_{\mathcal{E}}| \leq \varepsilon$.*
- *for any $a \in \mathcal{A}_{\mathscr{M}^!_\oplus}$, if $t \xrightarrow{a} u = (\mathcal{D}, A)$ and $s \xrightarrow{a} v = (\mathcal{E}, A)$, then there exists a finite set $I \subseteq \mathbb{N}$ such that:*
  - *there is a family of rationals $(p_i)_{i\in I}$ such that $\sum_{i\in I} p_i \leq 1$;*
  - *there are families $\boldsymbol{\sigma}^i$, $\boldsymbol{\tau}^i$, and K^i, such that $\kappa_i = (\boldsymbol{\sigma}^i, \boldsymbol{\tau}^i, K^i, A)$ is a substitution judgment for every $i \in I$;*
  - *there are distributions over tuples $\mathcal{D}_i$, $\mathcal{E}_i$ such that $(\mathcal{D}_i, B_i)\ R\ (\mathcal{E}_i, B_i)$;*

  and moreover $\mathcal{D} = \sum_{i\in I} p_i \cdot \kappa_i[\mathcal{D}_i]$, and $\mathcal{E} = \sum_{i\in I} p_i \cdot \kappa_i[\mathcal{E}_i]$.

The just introduced proof method is indeed quite useful when handling our running example.

Example 7 *We show that up-to bisimulations can handle our running example. Please recall the definition of M_ε given in Example 2. First, we can see that, for every a, for every type τ, $\hat{s}_{!(\tau\multimap\tau)}(M_a) = (\{([\,], [!\Omega_! \oplus^a I])^1\}, ([\,], [!\tau \multimap \tau]))$. We define a relation R on $\mathcal{S}_{\mathscr{L}^!_\oplus}$ containing $(\hat{s}_{!(\tau\multimap\tau)}(M_\varepsilon), \hat{s}_{!(\tau\multimap\tau)}(M_\mu))$, and we show that it is an γ-bisimulation up-to context for an appropriate γ. In order to simplify the notations, we define $B = ([\tau \multimap \tau], [])$, and $t_n, s_n \in \mathcal{S}_{\mathscr{L}^!_\oplus}$ as:*

$$t_n = (\{([(\Omega_! \oplus^\varepsilon I)], [])^{(\varepsilon^n)}\}, B), \qquad s_n = (\{([(\Omega_! \oplus^\mu I)], [])^{(\mu^n)}\}, B).$$

Then, we define the relation R as $R = \{(\hat{s}_\sigma(M), \hat{s}_\sigma(N))\} \cup \{(t_n, s_n) \mid n \in \mathbb{N}\}$. One can check that R is indeed a γ-bisimulation up-to-context (where $\gamma = \sup_{n\in\mathbb{N}} |\varepsilon^n - \mu^n|$) by carefully analysing every possible action. The proof is based on the following observations:

- *The only action starting from $\hat{s}_\sigma(M)$ or $\hat{s}_\sigma(N)$ is $a = (?^1)$, passing a term to the exponential part of the tuple, then we end up in t_0 and s_0 respectively.*
- *If we start from t_n or s_n, the only relevant action is Milner's action $a = (!^1)$, consisting in taking a copy of the term in the exponential part, evaluating it, and putting the result in the linear part. We can see (using the substitution judgment κ defined in Example 6), that $t_n \xrightarrow{a} \kappa[t_{n+1}]$, and similarly $s_n \xrightarrow{a} \kappa[s_{n+1}]$, and the result follows.*

Example 8 *We give now an example illustrating a not-so-obvious constraint in the definition of a substitution judgment. Suppose we take $K = ([x_1], [x_1])$, and $A = ([\sigma], [\sigma])$. Then $\kappa = ([\sigma], [], K, A)$ is a not a substitution judgment, since x_1 appear in the linear part of K, and we need non-linear variables to type it. The constraint is there to avoid the situation in which, $(\{([V], [])^1\}, ([\sigma], []))$ and $(\{([W], [])^1\}, ([\sigma], []))$ would be up-to bisimilar for every V and W of type σ.*

5.6.2 Soundness of up-to technique

Bisimulations up-to contexts would be useless without a correctness result like the following one:

Theorem 6 *If R is an ε-bisimulation up-to context, then $R \subseteq R^\varepsilon$.*

The remaining of this section consists in the proof of Theorem 6. The proof is an extension of that of Theorem 4 (although technically more involved).

We first define a relation $\mathcal{D} \triangleright \mathcal{E}$ between distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, which expresses the fact that $\mathcal{E}$ is obtained from $\mathcal{D}$ by changing the way we split our term into external environment and inside tuple:

Definition 16 *Let be $\mathcal{D}$ and $\mathcal{E}$ two distributions over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$. We write $\mathcal{D} \triangleright \mathcal{E}$ if:*

- *$\mathcal{D} = \sum_{i\in I} \alpha_i \{(C_i, t_i)^1\}$, I countable set.*
- *t_i is a A_i-state.*
- *For every i, there exist an open tuple K_i, a tuple type $B_i = (\boldsymbol{\sigma}^i, \boldsymbol{\tau}^i)$, and s_i a B_i-state, such that:*
  - *$(\boldsymbol{\sigma}^i, \boldsymbol{\tau}^i, K^i, A_i)$ is a substitution judgment,*
  - *$t_i = \kappa_i[s_i]$*

- $\mathcal{E} = \sum_{i \in I} \alpha_i \cdot \{\boldsymbol{F}(C_i, (K_i, A_i)), s_i{}^1\}$

Please observe that there are many possible distributions $\mathcal{E}$ that verify $\mathcal{D} \triangleright \mathcal{E}$. We can express the fact that $\triangleright$ is a congruence relation that doesn't modify the underlying program, and that moreover, the relation $\triangleright$ is designed to preserve normal forms with respect to $\rightarrow_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$, by the following lemma:

Lemma 15 *Let be $\mathcal{D}$, $\mathcal{E}$ such that $\mathcal{D} \triangleright \mathcal{E}$. Then:*
- $\boldsymbol{F}(\mathcal{D}) = \boldsymbol{F}(\mathcal{E})$.
- *If $\mathcal{S}(\mathcal{D}) \subseteq NF(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$, it holds that $\mathcal{S}(\mathcal{E}) \subseteq NF(\mathbf{C} \times \mathbf{\Delta}(\mathscr{U}))$.*

Proof. It is a direct consequence of the unfolding of the definition of $\triangleright$. □

Now, we are ready to define a relation $h \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, that intuitively corresponds to do a step $\rightarrow_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$, and then potentially change the way we split programs between internal part and external part.

Definition 17 *We define a relation $h \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, where $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, and $\mathcal{D}$ a finite distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, by the following rule:*

$$\frac{h \rightarrow_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E} \qquad \mathcal{E} \triangleright \mathcal{D}}{h \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}}$$

We define below the lifting of $\rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$ to distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$:

Definition 18 *We define a one-step relation reduction $\mathcal{D} \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}$ on (finite) distributions over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ as follows:*

$$\frac{\mathcal{D} = \mathscr{V}(\mathcal{D}) + \sum_{i \in I} \alpha_i \cdot \{h_i{}^1\} \qquad h_i \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}_i}{\mathcal{D} \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathscr{V}(\mathcal{D}) + \sum_{i \in I} \alpha_i \cdot \mathcal{E}_i}$$

Since $\rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$ is non-confluent, we cannot use it to define a semantics for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ as we did for $\rightarrow_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$. We are going to define a notion of semantics for $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ depending on what infinite sequance of reduction we are considering: for $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, we call *infinite reduction sequence starting from h* a sequence $s = (\mathcal{D}_n)_{n \in \mathbb{N}}$ of distributions over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, where $\mathcal{D}_0 = \{h^1\}$, and $\mathcal{D}_n \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}_{n+1}$ for every $n \in \mathbb{N}$. The idea is that every reduction sequence gave a different semantics for h, but that we obtain the same distribution over terms for all of them if we use $\mathbf{F}(\cdot)$ on them. Let us formalize this idea.

Definition 19 *If $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, and $s = (\mathcal{D}_n)_{n \in \mathbb{N}}$ a reduction sequence starting from h. We call* semantics of h with respect to s, *and we note $[\![h]\!]_s^{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$ the distribution over $\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$ given by $[\![h]\!]_s^{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} = \sup\{\mathscr{V}(\mathcal{D}_n) \mid n \in \mathbb{N}\}$.*

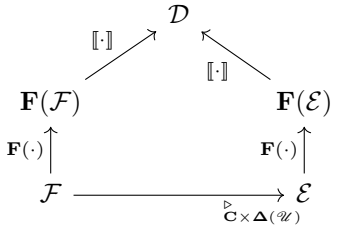
Please observe that Definition 19 is well-posed since the $\mathscr{V}(\mathcal{D}_n)$ are an increasing sequence of distributions over normal forms.

We state below the analogue of the third point of Lemma 4 for up-to bisimulation:

Proposition 8 *Let be $h \in \mathbf{C} \times \mathbf{\Delta}(\mathscr{U})$, and an infinite reduction sequence $s = \mathcal{D}_0 = \{h^1\} \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}_1 \ldots \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{D}_n \ldots$. Then $\boldsymbol{F}([\![h]\!]_s^{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}) = [\![\boldsymbol{F}(h)]\!]$.*

Proof. The proof is similar to the one of Lemma 4, and is based on the following two lemmas, that are a variant of Lemma 5 and 8: the following first lemma is used to show that $\mathbf{F}([\![h]\!]_s^{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}) \leq [\![\mathbf{F}(h)]\!]$.

Lemma 16 *Let be $\mathcal{D}$, $\mathcal{E}$ such that $\mathcal{D} \rightarrow^{\triangleright}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}$. Then $[\![\boldsymbol{F}(\mathcal{D})]\!] = [\![\boldsymbol{F}(\mathcal{E})]\!]$.*

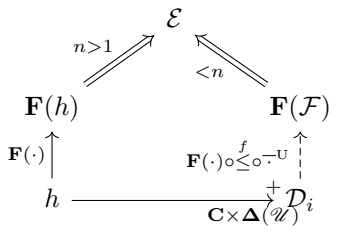


(9)

That other lemma is used to show the other direction, namely that $\mathbf{F}([\![h]\!]_s^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}) \geq [\![\mathbf{F}(h)]\!]$.

Lemma 17 *Let be $h \in \mathbf{C}\times\mathbf{\Delta}(\mathscr{U})$ such that $\boldsymbol{F}(h) \notin \mathscr{V}$, and an infinite reduction sequence $\mathcal{D}_0 = \{h^1\} \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}_1 \ldots \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}_m \ldots$. Then there exists $i \in \mathbb{N}$, such that for every $\mathcal{E}$ and $n \in \mathbb{N}$ with $\boldsymbol{F}(h) \overset{n}{\Rightarrow} \mathcal{E}$, there exists a finite $\mathcal{F} \in Distr(Unary^{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})})$, and an infinite reduction sequence $\mathcal{F}_i = \mathcal{F} \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{F}_{i+1} \ldots \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{F}_m \ldots$, verifying $\forall i$, $\overline{\mathcal{F}_i}^U \leq \overline{\mathcal{D}_i}^U$, and such that moreover $\boldsymbol{F}(\mathcal{F}) \overset{m}{\Rightarrow} \mathcal{E}$, with $m < n$.*

It can be expressed by the following diagram:



(10)

□

Observe that, even if R is an up-to bisimulation, it does not necessarily means that $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ preserves $\rightarrow_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$. We want to express the fact that it preserves $\rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, but we have to be more careful, since this reduction relation is non-deterministic. To deal with it, we introduce a more general preservation notion:

Definition 20 *We say that a relation R over $\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})$ is* weakly preserved by $\rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ *if, for any $h, k \in \mathbf{C}\times\mathbf{\Delta}(\mathscr{U})$ such that $h\ R\ k$, and $h \notin NF(\mathbf{C}\times\mathbf{\Delta}(\mathscr{U}))$, then there exists $\mathcal{D}$ and $\mathcal{E}$ such that $h \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{D}$, $k \rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and $\mathcal{D}\ R\ \mathcal{E}$.*

Our notion of weak preservation is enough to state the following variant of Proposition 5.

Proposition 9 *Let $\mathscr{C}$ be a set of tuple contexts, t, s two A-states and R a reflexive and symmetric relation weakly preserved by $\rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$, ε-bounding, and $\mathscr{C}$-closed with respect to t and s. Then it holds that $\delta^c_{\mathscr{C},!}(t,s) \leq \varepsilon$.*

An ε up-to bisimulation verifies the condistions of Proposition9, as stated by the analogue of Proposition 7 below.

Proposition 10 *Let be R an up-to bisimulation. Then $\widehat{R}^{\mathscr{C}^{\mathbf{T}}}$ is weakly preserved by $\rightarrow^{\triangleright}_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$ and ε-bounding, and $\mathscr{C}^{\mathbf{T}}$-closed with respect to every t, s such that $t\ R\ s$.*

We are now ready to do the proof of Theorem 6. Let be R an ε up-to bisimulation, ans $t, s \in \mathcal{S}_{\mathscr{L}^!_{\oplus}}$ such that $t\,R\,s$. Using our completness result stated in Theorem 5, we see that it is enough to show that $\delta^c_{\mathscr{C}^{\mathbf{T}},!}(t,s) \leq \varepsilon$. Propositions 9 and 10 give us immediately the result.

Example 9 *We can exploit the soundness of up-to bisimulation to obtain the contextual distance for our running example, and conclude that $\delta^c_{!(\tau\multimap\tau),!}(M_\varepsilon, M_\mu) = \sup_{n\in\mathbb{N}}|\varepsilon^n - \mu^n|$. The context distance between M_ε and M_μ is thus* strictly *between 0 and 1 whenever $|\varepsilon - \mu|$*

6 Probabilistic λ-Calculi, in Perspective

The calculus $\Lambda^{!,\|}_{\oplus}$ we analyzed in this paper is, at least apparently, nonstandard, given the presence of parallel disjunction, but also because of the linear refinement it is based on. In this section, we will reconcile what we have done so far with calculi in the literature, and in particular with untyped probabilistic λ-calculi akin to those studied, e.g., in [9, 5].

We consider a language $\Lambda_{\oplus}$ defined by the following grammar:

$$M \in \Lambda_{\oplus} ::= x \;\Big|\; MM \;\Big|\; \lambda x.M \;\Big|\; M \oplus M.$$

6.1 On Stable Fragments of $\mathscr{M}^{!}_{\oplus}$.

Our objective in this section is to characterize various notions of context distance for $\Lambda_{\oplus}$ by way of appropriate embeddings into $\Lambda^{!}_{\oplus}$, and thus by the LMC $\mathscr{M}^{!}_{\oplus}$. It is quite convenient, then, to understand when any fragment of $\mathscr{M}^{!}_{\oplus}$ is sufficiently *robust* so as to be somehow self-contained:

Definition 21 *We say that the pair $(\hat{\mathcal{S}}, \hat{\mathcal{A}})$, where $\hat{\mathcal{S}} \subseteq \mathcal{S}_{\mathscr{M}^{!}_{\oplus}}$, and $\hat{\mathcal{A}} \subseteq \mathbf{T} \times \mathcal{A}_{\mathscr{M}^{!}_{\oplus}}$ is a* stable fragment of $\mathscr{M}^{!}_{\oplus}$ *iff for every pair $(A, a) \in \hat{\mathcal{A}}$, for every A-state t, and for every $s \in \mathcal{S}$ such that $\mathcal{P}_{\mathscr{M}^{!}_{\oplus}}(t, a, s) > 0$, it holds that $s \in \hat{\mathcal{S}}$.*

Using a stable fragment of $\mathscr{M}^{!}_{\oplus}$, we can restrict the WLTS $\mathscr{L}^{!}_{\oplus}$ in a meaningful way. The idea is that we only consider some of the states of $\mathscr{L}^{!}_{\oplus}$, and we are able to choose the possible actions depending on the type of the state of $\mathscr{L}^{!}_{\oplus}$ we consider.

Definition 22 *If $\mathscr{F} = (\hat{\mathcal{S}}, \hat{\mathcal{A}})$ is a stable fragment of $\mathscr{M}^{!}_{\oplus}$, we define a WLTS by $\mathscr{L}_{\mathscr{F}} = (\mathcal{S}_{\mathscr{L}_{\mathscr{F}}}, \mathcal{A}_{\mathscr{L}_{\mathscr{F}}}, \dot{\to}_{\mathscr{F}}, w_{\mathscr{F}})$, as*

$$\mathcal{S}_{\mathscr{L}_{\mathscr{F}}} = \bigcup\nolimits_{A \in \mathbf{T}} \mathit{Distr}(\{K \mid (K, A) \in \hat{\mathcal{S}}\}) \times \{A\}; \qquad \mathcal{A}_{\mathscr{L}_{\mathscr{F}}} = \bigcup\nolimits_{(A,a) \in \hat{\mathcal{A}}} \{a\} \cup \mathbf{T};$$

$$\dot{\to}_{\mathscr{F}} = \dot{\to} \cap \{((\mathcal{D}, A), a, s) \mid \mathcal{S}(\mathcal{D}) \subseteq \hat{\mathcal{S}},\ (A, a) \in \hat{\mathcal{A}}\};$$

and $w_{\mathscr{F}}$ is defined as expected.

We want to be able to define a notion of distance on a *fragment* of the original language $\Lambda^{!}_{\oplus}$, so that it verifies the soundness property for a *restricted* set of contexts. To do that, we need the restricted set of contexts $\mathscr{C}$ to be preserved by the stable fragment:

Definition 23 *Let $\mathscr{F} = (\hat{\mathcal{S}}, \hat{\mathcal{A}})$ be a stable fragment of $\mathscr{M}^{!}_{\oplus}$. Let $\mathscr{C}$ be a set of tuple contexts. We say that $\mathscr{C}$* is preserved by *$\mathscr{F}$, if the following holds: for any $(C, A, \gamma) \in \mathscr{C}$ that is not an open value and any A-state t in $\mathcal{S}_{\mathscr{L}_{\mathscr{F}}}$, there exists a such that $(A, a) \in \hat{\mathcal{A}} \bigcup (\mathbf{T} \times \{\tau\})$, $(C, t) \xrightarrow{a}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and moreover:*

$$\mathcal{S}(\mathcal{E}) \subseteq \bigcup\nolimits_{B \in \mathbf{T}} \{(D, s) \mid s \text{ a } B\text{-state} \wedge \exists \eta \text{ s.t. } (D, B, \eta) \in \mathscr{C}\}$$

We are now able to provide guarantees that the contextual distance $\delta^{\varepsilon}_{\mathscr{C}}$ with respect to our restricted set of contexts $\mathscr{C}$ is smaller that the distance defined on the LTS $\mathscr{L}_{\mathscr{F}}$ induced by our stable fragment $\mathscr{F}$. In the following, we assume to have fixed a stable fragment $\mathscr{F} = (\hat{\mathcal{S}}, \hat{\mathcal{A}})$ of $\mathscr{M}^{!}_{\oplus}$, and $\mathscr{C}$ a set of tuple contexts preserved by $\mathscr{F}$.

Proposition 11 *Let be R an ε-bisimulation on $\mathscr{L}_{\mathscr{F}}$. Then $\widehat{R}^{\mathscr{C}}$ is preserved by $\to_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})}$, ε-bounding, and $\mathscr{C}$-closed with respect to every t, s such that $t \, R \, s$.*

Proof. • The fact that $\widehat{R}^{\mathscr{C}}$ is $\mathscr{C}$-closed is a direct consequence of the definition of lifting.
• The fact that it is ε-bounding is a direct consequence of the definition of a ε-bisimulation.

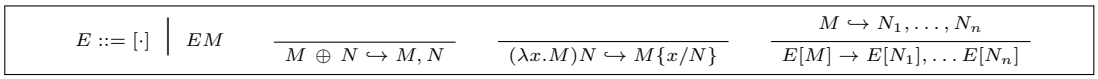


Figure 10: One-step CBN Semantics

- It is preserved by $\to_{\mathbf{C}\times\mathbf{\Delta}(\mathscr{U})}$: it is indeed guarantee by our definition of preservation of $\mathscr{C}$ by $\mathscr{F}$.

□

Proposition 12 *Let $\mathscr{F} = (\hat{\mathcal{S}}, \hat{\mathcal{A}})$ be a stable fragment of $\mathscr{M}^!_\oplus$, $\mathscr{C}$ a set of tuple contexts preserved by $\mathscr{F}$, and $t, s \in \mathcal{S}_{\mathscr{L}_\mathscr{F}}$. Then $\delta^c_\mathscr{C}(t,s) \leq \delta^b_{\mathscr{L}_\mathscr{F}}(t,s)$.*

Proof. Let be $\varepsilon = \delta^b_{\mathscr{L}_\mathscr{F}}(t,s)$. Let be R the biggest ε-bisimulation on $\mathscr{L}_\mathscr{F}$. By lemma 3, it holds that $t\,R\,s$. Using Proposition 11 and 5, we obtain that $\delta^c_{\mathscr{C},!}(t,s) \leq \varepsilon$, which is exactly the result. □

In the following, we make use of Proposition 12 on stable fragments corresponding to embeddings of $\Lambda_\oplus$ into $\Lambda^!_\oplus$. We will consider two different encodings depending on the underlying notion of evaluation.

6.2 Call-by-Name

$\Lambda_\oplus$ can first of all be endowed with call-by-name semantics, as in Figure 10. We use it to define an approximation semantics exactly in the same way as in Figure 1, and we take as usual the semantics of a term to be the least upper bound of its approximated semantics. Moreover, we denote by δ^c_{cbn} the context distance on $\Lambda_\oplus$, defined the natural way. We are going, in the remainder of this section, to use our results about $\Lambda^!_\oplus$ to obtain a characterization of δ^c_{cbn}.

6.2.1 The Call-By-Name Embedding

Girard's translation [19] gives us an embedding $\langle\cdot\rangle^{\text{cbn}} : \Lambda_\oplus \to \Lambda^!_\oplus$, defined as follows:

$$\langle x\rangle^{\text{cbn}} = x \qquad \langle \lambda x.M\rangle^{\text{cbn}} = \lambda !x.\langle M\rangle^{\text{cbn}}$$
$$\langle MN\rangle^{\text{cbn}} = \langle M\rangle^{\text{cbn}} !\langle N\rangle^{\text{cbn}} \qquad \langle M \oplus N\rangle^{\text{cbn}} = \langle M\rangle^{\text{cbn}} \oplus \langle N\rangle^{\text{cbn}}$$

Please observe that $\langle\cdot\rangle^{\text{cbn}}$ respects typing, in the sense that, when we define $\sigma^{\text{cbn}} = \mu\alpha.!\alpha \multimap \alpha$, it holds that for every term M of $\Lambda_\oplus$ whose free variables are in $\{x_1, \ldots, x_n\}$, we can show that $!x_1 :! \sigma^{\text{cbn}}, \ldots, !x_n :! \sigma^{\text{cbn}} \vdash \langle M\rangle^{\text{cbn}} : \sigma^{\text{cbn}}$.

This definition allows us to have some useful properties:

Lemma 18
- *Let be $M, N \in \Lambda_\oplus$, and $z \in \mathscr{X}$. Then $\langle M\{z/N\}\rangle^{cbn} = \langle M\rangle^{cbn}\{z/\langle N\rangle^{cbn}\}$;*
- *Let be M a closed term in $\Lambda_\oplus$. Then $\langle[\![M]\!]\rangle^{cbn} = [\![\langle M\rangle^{cbn}]\!]$.*
- *Let be $M \in \langle\Lambda_\oplus\rangle^{cbn}$, such that there exists an evaluation context E, and $N \in \Lambda_\oplus$, verifying $\langle M\rangle^{cbn} = E[\langle N\rangle^{cbn}]$. There for any term $L \in \Lambda_\oplus$, it holds that $E[\langle L\rangle^{cbn}] \in \langle\Lambda_\oplus\rangle^{cbn}$.*

6.2.2 Metrics for $\Lambda_\oplus$

It is very tempting to define a metric on $\Lambda_\oplus$ just as follows: $\delta^b_{\text{cbn}}(M,N) = \delta^b_{!\sigma^{\text{cbn}},!}(!\langle M\rangle^{\text{cbn}}, !\langle N\rangle^{\text{cbn}})$. We can easily see that it is sound with respect to the context distance for $\Lambda_\oplus$, since any context of this language can be seen, through $\langle\cdot\rangle^{\text{cbn}}$, as a context in $\Lambda^!_\oplus$. However, it is not complete, as shown by the following example:

Example 10 *We consider $M = \Omega \oplus (\lambda x.\Omega)$ and $N = (\lambda x.\Omega)$. We can see that $\delta^b_{!\sigma^{cbn},!}(!\langle M\rangle^{cbn}, !\langle N\rangle^{cbn}) = 1$: indeed, when we define a sequence of $\Lambda^!_\oplus$-contexts by $C_n = \lambda !x.\,((\lambda y_1 \ldots \lambda y_n.(\lambda z.zy_1, \ldots y_n))x \ldots x)\,[]$,*

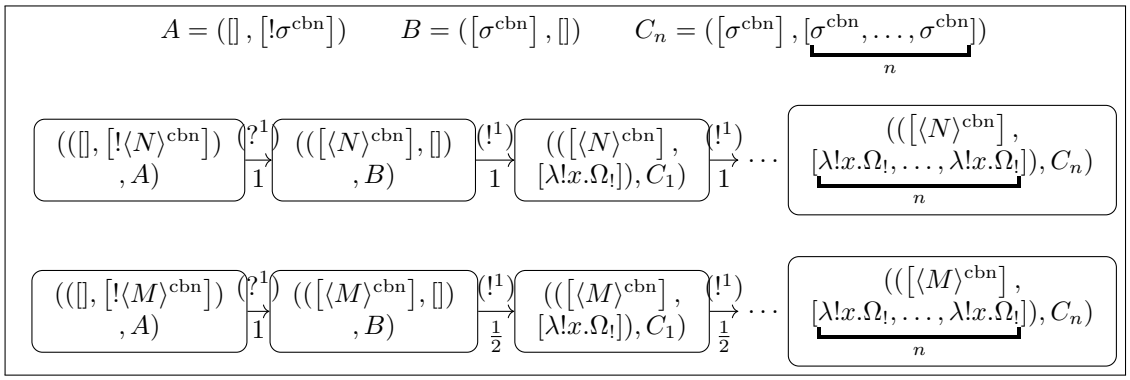


Figure 11: A Fragment of $\mathscr{M}^!_\oplus$

$$A^0 = ([\sigma^{\mathrm{cbn}}], []) \qquad A^1 = ([\sigma^{\mathrm{cbn}}], [\sigma^{\mathrm{cbn}}]) \qquad \hat{U}^{\mathrm{cbn}}(M) = (([\langle M\rangle^{\mathrm{cbn}}], []), A^0)$$
$$\mathcal{S}_{\mathscr{M}^{\mathrm{cbn}}_\oplus} = \left(\left\{\hat{U}^{\mathrm{cbn}}(M) \mid M \in \Lambda_\oplus\right\} \cup \left\{([\langle M\rangle^{\mathrm{cbn}}], [\langle V\rangle^{\mathrm{cbn}}]), A^1 \mid M \in \Lambda_\oplus, V \in \Lambda_\oplus \text{ a normal form}\right\}\right) \bigcap \mathcal{S}_{\mathscr{M}^!_\oplus}$$
$$\mathscr{J}^{\mathscr{V}}_{\mathrm{cbn}} = \mathscr{J}^{\mathscr{V}} \cap \{(E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \gamma)\,,\ M \in \langle\Lambda_\oplus\rangle^{\mathrm{cbn}}\}$$
$$\mathcal{A}_{\mathscr{M}^{\mathrm{cbn}}_\oplus} = \{A^1\} \times \{@^1_\kappa \mid \kappa \in \mathscr{J}^{\mathscr{V}}_{\mathrm{cbn}}\} \cup \{A^0\} \times \mathcal{A}_!$$

Figure 12: The Stable Fragment $\mathscr{F}^{\mathrm{cbn}} = (\mathcal{S}_{\mathscr{M}^{\mathrm{cbn}}_\oplus}, \mathcal{A}_{\mathscr{M}^{\mathrm{cbn}}_\oplus})$.

we see that $Obs(!\langle M\rangle^{cbn}) = 1/2^n$ *while* $Obs(!\langle N\rangle^{cbn}) = 1$. *But those contexts* C_n *have more expressive power that any context in* $\langle\Lambda_\oplus\rangle^{cbn}$, *since they do something that none of the context from* $\Lambda_\oplus$ *can do: they evaluate a copy of the term, and then shift their focus to* another *copy of the term. It can be seen in the embedding: a term in* $\langle\Lambda_\oplus\rangle^{cbn}$ *has never several redexes in linear position. We are now going to explicit this idea, and show that* $\delta^c_{cbn}(M, N) = \frac{1}{2} < \delta^b_{cbn}(M, N)$.

The way out consists in using the notion of stable fragment to refine the Markov Chain $\mathscr{M}^!_\oplus$ by keeping only the states and actions we are interested in.

Definition 24 *We define a stable fragment* $\mathscr{F}^{cbn}$ *as specified in Figure 12, and a distance* δ_{cbn} *on* $\Lambda_\oplus$ *as:*

$$\delta_{cbn}(M, N) = \delta^b_{\mathscr{L}_{\mathscr{F}^{cbn}}}(\hat{s}^{cbn}(M), \hat{s}^{cbn}(N)),$$

where $\hat{s}^{cbn}(M) = (\{([\langle M\rangle^{cbn}], [])^1\}, A^0)$.

We need now to define a set of tuple contexts preserved by $\mathscr{F}^{\mathrm{cbn}}$, the aim of applying Proposition 12.

Definition 25 $\mathscr{C}_{cbn}$ *is the smallest set of tuple contexts such that:*
- *If* $M \in \Lambda_\oplus$ *with* $FV(M) \subseteq \{x_1\}$, *then* $(\langle M\rangle^{cbn}, A^0, \sigma^{cbn}) \in \mathscr{C}_{cbn}$;
- *If* $(C, A^0, \sigma^{cbn}) \in \mathscr{C}_{cbn}$, *and* $C = E[x_1]$, *it holds that* $(E[y_1], A^1, \sigma^{cbn}) \in \mathscr{C}_{cbn}$.

$\mathscr{C}_{\mathrm{cbn}}$ is designed to allow us to link δ^c_{cbn} and $\delta^c_{\mathscr{C}_{\mathrm{cbn}}}$.

Lemma 19 *For* $M, N \in \Lambda_\oplus$ *closed terms,* $\delta^c_{cbn}(M, N) = \delta^c_{\mathscr{C}_{cbn}}(\hat{s}^{cbn}(M), \hat{s}^{cbn}(N))$.

Proof. The proof is based on Lemma 18. □

Lemma 20 $\mathscr{C}_{cbn}$ *is preserved by the stable fragment* $\mathscr{F}^{cbn}$.

Proof. Let be $(C, A, \gamma) \in \mathscr{C}_{\mathrm{cbn}}$, and t a A-state in $\mathscr{L}_{\mathscr{F}^{\mathrm{cbn}}}$. Using the definition of $\mathscr{C}$, we may see that $\gamma = \sigma^{\mathrm{cbn}}$, and moreover we are in one of the following cases:

- Or $A = A^0$, $C = \langle M \rangle^{\mathrm{cbn}}$ with $FV(M) \subseteq \{x_1\}$ and t is of the form $(\mathcal{D}, A^0)$, where $\mathcal{D} = \sum_{i \in \mathbb{N}} p_i \cdot \{([\langle N \rangle_i^{\mathrm{cbn}}], [])^1\}$. Then we still have to consider separately two cases:
  - or $\langle M \rangle^{\mathrm{cbn}} = E[x_1]$. We take $a = (?^1)$. We can easily check that $(A^0, (?^1)) \in \mathcal{A}_{\mathscr{M}_{\oplus}^{\mathrm{cbn}}} \bigcup (\mathbf{T} \times \{\tau\})$. Moreover, we see that there exists $\mathcal{E}$ such that $(C, t) \xrightarrow{a}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and more precisely $\mathcal{E} = \{(E[y_1], s)^1\}$ where s is a A^1-state. Since $(E[y_1], A^1, \sigma^{\mathrm{cbn}}) \in \mathscr{C}_{\mathrm{cbn}}$ by construction, it holds that $\mathsf{S}(\mathcal{E}) \subseteq \bigcup_{B \in \mathbf{T}} \{(D, s) \mid s \text{ a } B\text{-state} \wedge \exists \eta \text{ s.t. } (D, B, \eta) \in \mathscr{C}_{\mathrm{cbn}}\}$, which shows the result.
  - or $\langle M \rangle^{\mathrm{cbn}} = E[R]$, where R is a redex. We take $a = \tau$, and we obtain the result by a similar reasoning as the previous case.
- Or $A = A^1$. Then t is of the form $(\mathcal{D}, A^1)$, where $\mathcal{D} = \sum_{i \in \mathbb{N}} p_i \cdot ([\langle N \rangle^{\mathrm{cbn}}], [\langle V \rangle^{\mathrm{cbn}}])$. Moreover, it implies that $C = E[y_1]$, and there exists L such that $E[x_1] = \langle L \rangle^{\mathrm{cbn}}$. Looking at the cbn encoding, we see that the only way to have $E[x_1] \in \langle \Lambda_{\oplus} \rangle^{\mathrm{cbn}}$ is $E = F[[\cdot]!\langle P \rangle^{\mathrm{cbn}}]$. It means that $C = F[y_1!\langle P \rangle^{\mathrm{cbn}}]$. We take $a = @_{\kappa}^1$ with $\kappa = (\{1\}, \{\}, [\sigma^{\mathrm{cbn}}], [\sigma^{\mathrm{cbn}}], !\langle P \rangle^{\mathrm{cbn}}, \sigma^{\mathrm{cbn}})$. We can easily check that $(A^1, a) \in \mathcal{A}_{\mathscr{M}_{\oplus}^{\mathrm{cbn}}} \bigcup (\mathbf{T} \times \{\tau\})$. Moreover, we see that there exists $\mathcal{E}$ and s a A^1-state such that $(C, t) \xrightarrow{a}_{\mathbf{C} \times \mathbf{\Delta}(\mathscr{U})} \mathcal{E}$, and $\mathcal{E} = \{(F[y_1], s)^1\}$. Using the third point of Lemma 18, and that $F[x_1!\langle P \rangle^{\mathrm{cbn}}] \in \langle \Lambda_{\oplus} \rangle^{\mathrm{cbn}}$, we see that $F[x_1] \in \langle \Lambda_{\oplus} \rangle^{\mathrm{cbn}}$. Looking at the definition of $\mathscr{C}_{\mathrm{cbn}}$, we see that it means that $(F[y_1], A^1, \sigma^{\mathrm{cbn}}) \in \mathscr{C}_{\mathrm{cbn}}$, and so we have the result.

□

Theorem 7 (Full Abstraction for CBN) *δ_{cbn}^{c} and δ_{cbn} coincide.*

Proof. We first show that δ_{cbn} is at least as discriminating $\delta_{\mathrm{cbn}}^{c}$. Let be $M, N \in \Lambda_{\oplus}$. By definition of $\mathscr{L}_{\mathscr{F}^{\mathrm{cbn}}}$, we know that $\hat{s}^{\mathrm{cbn}}(M), \hat{s}^{\mathrm{cbn}}(N) \in \mathcal{S}_{\mathscr{L}_{\mathscr{F}^{\mathrm{cbn}}}}$. Moreover, we know by Lemma 20 that $\mathscr{C}_{\mathrm{cbn}}$ is preserved by $\mathscr{F}^{\mathrm{cbn}}$. So we can apply Proposition 12, and we see that $\delta^{\mathrm{c}}_{\mathscr{C}_{\mathrm{cbn}}}(\hat{s}^{\mathrm{cbn}}(M), \hat{s}^{\mathrm{cbn}}(N)) \leq \delta_{\mathrm{cbn}}(M, N)$. Then, soundness follows using Lemma 19. When proving completeness part, we rely on an "intrinsic" characterization of δ_{cbn}. The details can be found in the next section. □

6.2.3 An intrinsic trace characterization of $\delta_{\mathbf{cbn}}$

Looking at the structure of $\mathscr{F}^{\mathrm{cbn}}$, we see that we can in fact give an intrinsic definition of δ_{cbn}, without considering tuples and exponential constructs. In fact, every sequence of actions in $\mathscr{L}_{\mathscr{F}^{\mathrm{cbn}}}$ can be see as an element of the set $\mathscr{T}_r^{\mathrm{cbn}}$ defined by:

$$\mathsf{T} \in \mathscr{T}_r^{\mathrm{cbn}} ::= \top \;\Big|\; M \cdot \mathsf{T} \text{ ,where } M \in \Lambda_{\oplus} \text{ and } FV(M) \subseteq \{x\}.$$

We can now talk about the probability for a term M to effectuate a trace T defined by:

$$Pr_{\mathrm{cbn}}(M, N_1 \dots N_m \cdot \top) = \sum\nolimits_{[\![M(N_1\{x/M\}) \dots (N_m\{x/M\})]\!]}.$$

and we obtain a simple characterization of δ_{cbn}:

Proposition 13 *Let be $M, N \in \Lambda_{\oplus}$. Then:*

$$\delta_{cbn}(M, N) = \sup \left\{ |\ Pr_{cbn}(M, \mathsf{T}) - Pr_{cbn}(N, \mathsf{T})\ |\ \ \textit{with } \mathsf{T} \in \mathscr{T}_r^{cbn} \right\}.$$

Please observe that we can alternatively see traces as contexts. As a consequence, if we define CIU-contexts in $\Lambda_{\oplus}$ as $E = (\lambda x.xM_1 \dots M_n)\,[]$. We may express δ_{cbn} in a purely contextual way:

Proposition 14 *Let be $M, N \in \Lambda_{\oplus}$. Then $\delta_{cbn}(M, N) = \sup_{E\ CIU\ context} |\sum_{[\![E[M]]\!]} - \sum_{[\![E[N]]\!]}|$.*

Please observe that Proposition 14 allows us to see easily that $\delta_{\mathrm{cbn}}(M, N) \leq \delta_{\mathrm{cbn}}^{c}(M, N)$. As such, it allows us to do the completeness part of the proof Theorem 7.

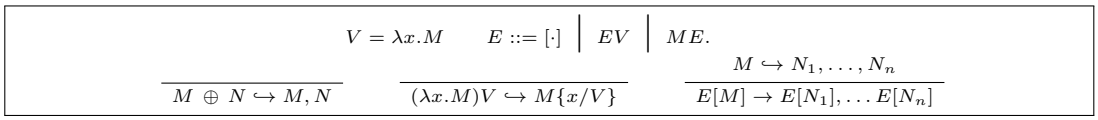


Figure 13: One-step CBV Semantics

6.3 Call-by-Value

In a similar way, we can endow $\Lambda_\oplus$ with a call-by-value semantics, and embedd it into $\Lambda^!_\oplus$. We are then able to define a suitable fragment of $\mathscr{M}^!_\oplus$, a suitable set of tuple contexts preserving it, and a caracterisation of a call-by-value context distance for $\Lambda_\oplus$ follows. While the construction of the stable fragment (and the set of tuple contexts to consider) are more involved than in the call-by-name case, we noticed that the caracterisation we obtain seem to have some similarities with the way environmental bisimulation for a CBV probabilistic λ-calculus was defined in [33]. In this section, we endow $\Lambda_\oplus$ with a call-by-value semantics, as specified in Figure 13. We denote δ^c_{cbv} the contextual metric on $\Lambda_\oplus$ induced by this semantics. We define a new embedding: $\Lambda_\oplus \to \Lambda^!_\oplus$, which respects the CBV semantics, as:

$$\langle x\rangle^{\text{cbv}} = !x \qquad \langle \lambda x.M\rangle^{\text{cbv}} = !\lambda!x.\langle M\rangle^{\text{cbv}}$$
$$\langle MN\rangle^{\text{cbv}} = (\lambda!x.x\langle N\rangle^{\text{cbv}})\langle M\rangle^{\text{cbv}} \qquad \langle M \oplus N\rangle^{\text{cbv}} = \langle M\rangle^{\text{cbv}} \oplus \langle N\rangle^{\text{cbv}}$$

We define σ^{cbv} as the following type : $\sigma^{\text{cbv}} = \mu\alpha.!(\alpha \multimap \alpha)$. We define $\tau^{\text{cbv}} = \sigma^{\text{cbv}} \multimap \sigma^{\text{cbv}}$: it verifies $\sigma^{\text{cbv}} =^{\mathscr{A}} !\tau^{\text{cbv}}$. For any term $M \in \Lambda_\oplus$, if $FV(M) = \{z_1, \ldots, z_n\}$ it holds that $(!z_i : \sigma^{\text{cbv}})_{1\le i\le n} \vdash \langle M\rangle^{\text{cbv}} : \sigma^{\text{cbv}}$.

We will use later the following nice property of the encoding:

Lemma 21 *Let be M and N in $\Lambda_\oplus$, and z a free variable of M. Then $\langle M\rangle^{cbv}\{z/\lambda!z.\langle N\rangle^{cbv}\} \in \langle\Lambda_\oplus\rangle^{cbv}$.*

Proof. The proof is by induction on the form of M. □

Lemma 22 *Let be $M \in \Lambda_\oplus$. Then for any $V \in \mathcal{S}(\llbracket\langle M\rangle^{cbv}\rrbracket)$, it holds that $V \in \langle\Lambda_\oplus\rangle^{cbv}$.*

Proof. The proof uses Lemma 21. □

Again, we are now going to define a stable fragment of $\mathscr{M}^!_\oplus$. However, contrary to the call-by-name case, it does not contain the full call-by-value encoding $\Lambda_\oplus$. It actually contains only the encoding of *values* in $\Lambda_\oplus$, which allows us to have a more tractable characterization. We are able to treat the general case simply by noting that for any term M and N, it holds that $\delta^c_{\text{cbv}}(M, N) = \delta^c_{\text{cbv}}(\lambda z.M, \lambda z.N)$.

Definition 26 *We define a fragment $\mathscr{F}^{cbv} = (\mathcal{S}_{\mathscr{M}^{cbv}_\oplus}, \mathcal{A}_{\mathscr{M}^{cbv}_\oplus})$ as specified in Figure 14, and a metric δ_{cbv} on $\Lambda_\oplus$ specified by:*

$$\delta_{cbv}(M, N) = \delta^b_{\mathscr{L}_{\mathscr{F}^{cbv}}}(\hat{s}_{\sigma^{cbv}}(!\langle\lambda x.M\rangle^{cbv}), \hat{s}_{\sigma^{cbv}}(!\langle\lambda x.N\rangle^{cbv})).$$

Please observe that contrary to the cbn-case, we do not need to restrict action by considering the type of states.

Proposition 15 *$\mathscr{F}^{cbv}$ is a stable fragment of $\mathscr{M}^!_\oplus$.*

Proof. Let be $(K, A) \in \mathcal{S}_{\mathscr{M}^{\text{cbv}}_\oplus}$. Since (K, A) is in $\mathcal{S}_{\mathscr{M}^{\text{cbv}}_\oplus}$, there exist φ, and $\boldsymbol{M}, \boldsymbol{N}, \boldsymbol{L}$ respectively n, m, p term sequences, such that $K = (\lambda!x.\boldsymbol{M}, \lambda!x.\boldsymbol{N}; !(\lambda!x.\boldsymbol{L})_\varphi)$ and $A = A^\varphi_{n,(m,p)}$. Let be a such that $(A, a) \in \mathcal{A}_{\mathscr{M}^{\text{cbv}}_\oplus}$. Let be $s = (H, B)$ such that $\mathcal{P}_{\mathscr{M}^!_\oplus}(t, a, s) > 0$. We want to show that $s \in \mathcal{S}_{\mathscr{M}^{\text{cbv}}_\oplus}$. We distinguish cases depending on a.

$$A^{\varphi}_{n,(m,p)} = ([(\tau^{\mathrm{cbv}})^n], [(\tau^{\mathrm{cbv}})^m, (!\tau^{\mathrm{cbv}})^p])_{\varphi} \qquad \mathscr{J}^{\nu}_{\mathrm{cbv}} = \mathscr{J}^{\nu} \cap \{(E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \sigma^{\mathrm{cbv}}),\ M \in \langle\Lambda_{\oplus}\rangle^{\mathrm{cbv}}\}$$

$$\mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}} = \left\{ \begin{array}{l} (([\lambda!x.\boldsymbol{M}], [\lambda!x.\boldsymbol{N}, !(\lambda!x.\boldsymbol{L})]_{\varphi}), A^{\varphi}_{n,(m,p)}) \\ \text{where } \boldsymbol{M}, \boldsymbol{N}, \boldsymbol{L} \text{ are } n, m, p\text{-sequences in} \langle\Lambda_{\oplus}\rangle^{\mathrm{cbv}}, \text{ and } \forall i, N_i \in \boldsymbol{M} \end{array} \right\} \bigcap \mathcal{S}_{\mathscr{M}^{!}_{\oplus}}$$

$$\mathcal{A}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}} = \mathbf{T} \times \left(\left(\{ @^i_{\kappa} \text{ with } \kappa \in \mathscr{J}^{\nu}_{\mathrm{cbv}} \} \cap \mathcal{A}_{@} \right) \cup \mathcal{A}_{\mathscr{M}^{!}_{\oplus}} \cup \mathcal{A}_{!} \right)$$

Figure 14: The Stable Fragment $\mathscr{F}^{\mathrm{cbv}} = (\mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}, \mathcal{A}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}})$.

- Or $a = (!^i)$. It implies that $H = (\lambda!x.\boldsymbol{M}, \lambda!x.\boldsymbol{N}; !(\lambda!x.\boldsymbol{L})_{\varphi}; V)$ with $\vdash V : \sigma^{\mathrm{cbv}}$ and $B = ([(\tau^{\mathrm{cbv}})^n], [(\tau^{\mathrm{cbv}})^m, (!\tau^{\mathrm{cbv}})^p]_{\varphi}, \sigma^{\mathrm{cbv}})$. Since V is obtained by the evaluation of a term in $\langle\Lambda_{\oplus}\rangle^{\mathrm{cbv}}$, it is also in $\langle\Lambda_{\oplus}\rangle^{\mathrm{cbv}}$. It allows us to see that V is of the form $!\lambda!x.P$. So we see that we can construct a permutation ψ from $\{1, \ldots, m+p+1\}$ such that $s = ((\lambda!x.\boldsymbol{M}, \lambda!x.\boldsymbol{N}; !(\lambda!x.(\boldsymbol{L}; P))_{\psi}))$, and $B = ([(\tau^{\mathrm{cbv}})^n], [(\tau^{\mathrm{cbv}})^m, (!\tau^{\mathrm{cbv}})^{p+1}]_{\psi})$. As a consequence, $s \in \mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}$.
- Or $a = (?^i)$. It implies that $H = (\lambda!x.(\boldsymbol{M}; L_j), \lambda!x.\boldsymbol{N}; !(\lambda!x.\boldsymbol{L}_{\{1,\ldots,p\}\setminus\{j\}})_{\varphi_{\mathrm{remove}(\{j\})}})$, when $\varphi(m+j) = i$, and $B = ([(\tau^{\mathrm{cbv}})^{n+1}], [(\tau^{\mathrm{cbv}})^m, (!\tau^{\mathrm{cbv}})^{p-1}]_{\varphi_{\mathrm{remove}(\{j\})}})$. So we can see that $s \in \mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}$.
- Or $a = @^i_{\kappa}$ with $\kappa \in \mathscr{J}^{\nu}_{\mathrm{cbv}}$. Then $\kappa = (E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, M, \sigma^{\mathrm{cbv}})$ and M is an open value in $\langle\Lambda_{\oplus}\rangle^{\mathrm{cbv}}$. Since there exist s such that $\mathcal{P}_{\mathscr{M}^{!}_{\oplus}}(t, a, s) > 0$, it means that $E \subseteq \{1, \ldots, n\}$, and $F \subseteq \{1, \ldots m+p\}$, and moreover $(\boldsymbol{\sigma}, \boldsymbol{\tau})$ is one of the $A^{m,n,p}_{\varphi}$. verifying $1 \leq \varphi(i) \leq m$ (that is, $\boldsymbol{\tau}_i = \tau^{\mathrm{cbv}} \multimap \tau^{\mathrm{cbv}}$). By a similar reasoning as previous cases, we can see that $s \in \mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}$.

□

Please observe that for any term $M \in \Lambda_{\oplus}$, the associated state $\hat{s}_{\sigma^{\mathrm{cbv}}}(\langle\lambda z.M\rangle^{\mathrm{cbv}})$ is in $\mathcal{S}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}$.

As in the previous section, we define a suitable set $\mathscr{C}_{\mathrm{cbv}}$ of tuple contexts preserving $\mathscr{F}^{\mathrm{cbv}}$, in order to apply Proposition 12.

Definition 27 *We define a set of tuple contexts $\mathscr{C}_{cbv}$ as those of the form $(C_1 \ldots C_k D, (\boldsymbol{\sigma}, \boldsymbol{\tau}), \sigma^{cbv})$ such that:*

- $C_i \in \{y_j \mid \tau_j = \sigma^{cbv} \multimap \sigma^{cbv}\} \cup \{\lambda!z.z\langle M\rangle^{cbv} \mid z \notin FV(M)\} \cup \lambda!z.\langle\Lambda_{\oplus}\rangle^{cbv} \cup \{x_j\}$
- $D \in \langle\Lambda_{\oplus}\rangle^{cbv} \cup \{y_j \mid \tau_j = !(\sigma^{cbv} \multimap \sigma^{cbv})\} \cup \{x_j\}$

Proposition 16 *$\mathscr{C}_{cbv}$ is preserved by $\mathscr{F}^{cbv}$.*

Proof. Let be $(C, A, \gamma) \in \mathscr{C}_{\mathrm{cbv}}$. Let be a such that $(A, a) \in \mathcal{A}_{\mathscr{M}^{\mathrm{cbv}}_{\oplus}}$. Let be t a A-state t in $\mathcal{S}_{\mathscr{L}_{\mathscr{F}^{\mathrm{cbv}}}}$.Let be $\mathcal{E}$ such that $(C, t) \xrightarrow{a}_{\mathbf{C}\times\boldsymbol{\Delta}(\mathscr{U})} \mathcal{E}$. Let be $h \in \mathsf{S}(\mathcal{E})$. We note $h = (F, s)$, and B the tuple type such that s is a B-state. We have to show that $(F, B, \sigma^{\mathrm{cbv}}) \in \mathscr{C}_{\mathrm{cbv}}$.

Please observe that there exists actually only one a such that $(C, t) \xrightarrow{a}_{\mathbf{C}\times\boldsymbol{\Delta}(\mathscr{U})}$, and it depends only on C. Consequently, we do the proof by considering separately the different form of C. We know, by definition of $\mathscr{C}_{\mathrm{cbv}}$, that C is of the form $C_1 \ldots C_k D$.

- If one of the C_i is of the form x_j, we consider the smallest index i for which it happens. Then $a = (!^i)$, and $F = C_1, \ldots, !y_{n+1}, \ldots C_n D$.
- Otherwise, $C_1 \ldots C_k[\cdot]$ is an evaluation context. Then:
  - If $D = x_i$. We have that $a = (!^i)$. It follows that $t \xrightarrow{a}_{\mathbf{C}\times\boldsymbol{\Delta}(\mathscr{U})} s$, and $F = C_1, \ldots, C_k y_{m+p+1}$. So we can see that the result holds.
  - If $D = y_j$, with $\tau_j = !\tau^{\mathrm{cbv}}$. Then it holds that $a = (?^i)$., and $F = C_1, \ldots, C_k !x_{n+1}$.

– If D is in $\langle\Lambda_\oplus\rangle^{\text{cbv}}$. We show the result by induction on the term L such that $D = \langle L\rangle^{\text{cbv}}$.
  * Or $D = \langle M \oplus N\rangle^{\text{cbv}}$. Then $a = \tau$, and $F = C_1, \ldots, C_k G$, with $G \in \{\langle M\rangle^{\text{cbv}}, \langle N\rangle^{\text{cbv}}\} \subseteq \langle\Lambda_\oplus\rangle^{\text{cbv}}$ and so the result holds.
  * Or $D = \langle MN\rangle^{\text{cbv}}$. Then $D = (\lambda!z.z\langle N\rangle^{\text{cbv}})\langle M\rangle^{\text{cbv}}$. We can take $G = \langle M\rangle^{\text{cbv}}$, $C_{n+1} = (\lambda!z.z\langle N\rangle^{\text{cbv}})$, and we can see that $C = C_1, \ldots, C_{n+1}G$, and so we can apply the induction hypothesis to $G = \langle M\rangle^{\text{cbv}}$.
  * If D is an open value. It means that $D = !\lambda!z.\langle M\rangle^{\text{cbv}}$, or $D = !x_i$. Please observe that C_n is always an open value (by construction). It means that the redex $C_n D$ is going to be reduced.
    · If $C_n = y_j$, with $\tau_j = \sigma^{\text{cbv}} \multimap \sigma^{\text{cbv}}$, then $a = @^j_\kappa$, where $\kappa = (E, F, \boldsymbol{\sigma}, \boldsymbol{\tau}, D, \sigma^{\text{cbv}})$, where $E \subseteq \{1, \ldots, n\}$, and $!\boldsymbol{x}_E : !\boldsymbol{\sigma}_E, \boldsymbol{y}_F : \boldsymbol{\tau}_F \vdash D : \sigma^{\text{cbv}}$ and we see that indeed $\kappa \in \mathscr{J}^{\mathscr{V}}_{\text{cbv}}$. Moreover, if we note q the cardinality of E, we can see that $F = C_1, \ldots, C_{n-1} y_{n+1-q+1}$. As a consequence, we have the result.
    · If $C_n = \lambda!z.(z\langle N\rangle^{\text{cbv}})$. Then $a = \tau$. Please observe that we can define C_{n+1} such that $!C_{n+1} = D$. Now, please observe that we should have $F = C_1, \ldots, C_{n-1}C_{n+1}\langle N\rangle^{\text{cbv}}$, and so the result holds.
    · If $C_n = \lambda!z.(\langle N\rangle^{\text{cbv}})$. Then if $D = !x_i$, we have that $F = C_1, \ldots, C_{n-1}\langle M\{z/x_i\}\rangle^{\text{cbv}}$, and we have the result. Similarly, if $D = !\lambda!z.\langle M\rangle^{\text{cbv}}$, it holds that $F = C_1, \ldots, C_{n-1}\langle M\{z/\lambda!z.\langle M\rangle^{\text{cbv}}\}\rangle^{\text{cbv}}$. We use Lemma 21 to obtain the result.

□

Definition 28 *We define a metric δ_{cbv} on $\Lambda_\oplus$ by: $\delta_{cbv}(M,N) = \delta^b_{\mathscr{L}_{\mathscr{J}^{cbv}}}(\hat{s}_{\sigma^{cbv}}(!\langle\lambda x.M\rangle^{cbv}), \hat{s}_{\sigma^{cbv}}(!\langle\lambda x.N\rangle^{cbv}))$.*

Theorem 8 (Full Abstraction for CBV) *δ^c_{cbv} and δ^b_{cbv} coincide.*

The proof can be found in [6]. Again, soundness is based on Proposition 12, while completeness relies on yet another characterization of the context distance.

6.3.1 Intrinsic characterization

We don't have an intrinsic characterization as simpler as in the cbn case. However, we can express the distance δ^c_{cbv} using a LMC $\mathscr{M}^{\text{cbv}}_\oplus$, designed to be a simplified and untyped version of $\mathscr{M}^!_\oplus$, and that doesn't use the embedding.

Definition 29 *We define a labelled Markov chain $\mathscr{M}^{cbv}_\oplus = (\mathscr{S}^{cbv}_\oplus, \mathscr{A}^{cbv}_\oplus, \mathscr{P}^{cbv}_\oplus)$, where:*
- $\mathscr{S}^{cbv}_\oplus = \{(\boldsymbol{V}, n) \mid \boldsymbol{V}$ *a n-sequence of values*$\}$
- $\mathscr{A}^{cbv}_\oplus = \{@^M_i \mid \exists k \in \mathbb{N},\ M = x_k$ *or* $L = \lambda z.M\}$
- $\mathscr{P}^{cbv}_\oplus$ *is defined in Figure 15.*

As for $\Lambda^!_\oplus$, we transform the LMC $\mathscr{M}^{\text{cbv}}_\oplus$ into a purely non-deterministic WLTS. We define $\mathscr{L}^{\text{cbv}}_\oplus = (\mathcal{S}_{\mathscr{L}^{\text{cbv}}_\oplus}, \mathcal{A}_{\mathscr{L}^{\text{cbv}}_\oplus}, \dot{\to}, w)$, where:
- The set of states is $\mathcal{S}_{\mathscr{L}^{\text{cbv}}_\oplus} = \bigcup_{n\in\mathbb{N}} (\text{Distr}(\{\boldsymbol{V} \mid \boldsymbol{V}$ a n-sequence $\}) \times \{n\})$
- The set of actions is $\mathcal{A}_{\mathscr{L}^{\text{cbv}}_\oplus} = \mathscr{A}^{\text{cbv}}_\oplus \cup \mathbb{N}$,
- The transition function $\dot{\to}$ is defined as follows:

$$(\mathcal{D}, n) \xrightarrow{n} (\mathcal{D}, n)$$
$$(\mathcal{D}, n) \xrightarrow{a} \sum_{\boldsymbol{V}} \mathcal{D}(\boldsymbol{V}) \cdot \mathcal{P}_{\mathscr{M}^!_\oplus}((\boldsymbol{V}, n))(a),$$

- The weight w is such that $w(\mathcal{D}, n) = \sum_{\mathcal{D}}$.

Please remember that we give in Section 5.3 a definition of the trace metric $\delta^b_{\mathscr{L}}$ for any WLTS. We can apply it to $\mathscr{L}^{\text{cbv}}_\oplus$, and it gives us a characterization of δ_{cbv}.

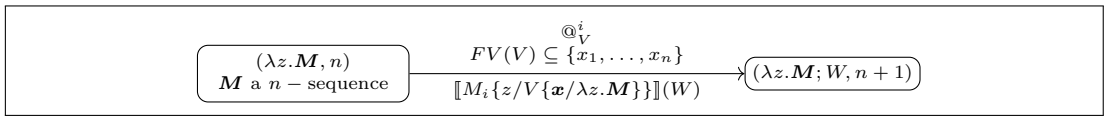


Figure 15: Definition of $\mathscr{M}_{\oplus}^{\mathrm{cbv}}$

Proposition 17 *Let be $M, N \in \Lambda_{\oplus}$. Then:*

$$\delta^b_{\mathscr{L}}(([\lambda x.M], 1), ([\lambda x.N], 1)) = \delta_{cbv}(M, N).$$

7 Related Work

This is definitely *not* the first work on metrics in the context of programming languages semantics. A very nice introduction to the topic, together with a comprehensive (although outdated) list of references can be found in [36]. One of the many uses of metrics is as an alternative to order-theoretic semantics. This has also been applied to higher-order languages, and to *deterministic* PCF [15].

If one focuses on probabilistic programming languages, the first attempts at using metrics as a way to measure "how far" two programs are, algebraically or behaviourally, are due to Giacalone et al. [18], and Desharnais et al. [11, 12], who both consider process algebras in the style of Milner's CCS. Most of further work in this direction has focused on concurrent specifications. Among the recent advances in this direction (and without any hope of being comprehensive), we can cite Gebler et al.'s work on uniform continuity as a way to enforce compositionality in metric reasoning [17, 16].

Finally, great inspiration for this work came from the many contributions on metrics for labelled Markov chains and processes appeared in the last twenty years (e.g. [37, 13]).

8 Conclusions

We have shown *how* the context distance can be characterized so as to simplify concrete proofs, and *to which extent* this metric trivializes. All this has been done in a universal linear λ-calculus for probabilistic computation. This clarifies to which extent refining equivalences into metrics is worth in such a scenario. The tuple-based techniques in Section 5.6 are potentially very interesting in view of possible applications to cryptography, as hinted in [4]. This is indeed what we are working on currently.

References

[1] Hendrik Pieter Barendregt. *The Lambda Calculus – Its Syntax and Semantics*, volume 103 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1984.

[2] Hendrik Pieter Barendregt, Wil Dekkers, and Richard Statman. *Lambda Calculus with Types*. Perspectives in logic. Cambridge University Press, 2013.

[3] Ales Bizjak and Lars Birkedal. Step-indexed logical relations for probability. In *Proc. of FoSSaCS*, pages 279–294, 2015.

[4] Alberto Cappai and Ugo Dal Lago. On equivalences, metrics, and polynomial time. In *Proc. of FCT*, pages 311–323, 2015.

[5] Raphaëlle Crubillé and Ugo Dal Lago. On probabilistic applicative bisimulation and call-by-value λ-calculi. In *Proc. of ESOP*, pages 209–228, 2014.

[6] Raphaëlle Crubillé and Ugo Dal Lago. Metric reasoning about λ-terms: The general case (long version). Available at `http://eternal.cs.unibo.it/mrltgc.pdf`, 2016.

[7] Raphaëlle Crubillé and Ugo Dal Lago. Metric reasoning about λ-terms: The affine case. In *Proc. of LICS*, pages 633–644, 2015.

[8] Raphaëlle Crubillé, Ugo Dal Lago, Davide Sangiorgi, and Valeria Vignudelli. On applicative similarity, sequentiality, and full abstraction. In *Proc. of Correct System Design - Symposium in Honor of Ernst-Rüdiger Olderog on the Occasion of His 60th Birthday*, pages 65–82, 2015.

[9] Ugo Dal Lago, Davide Sangiorgi, and Michele Alberti. On coinductive equivalences for higher-order probabilistic functional programs. In *Proc. of POPL*, pages 297–308, 2014.

[10] Ugo Dal Lago and Margherita Zorzi. Probabilistic operational semantics for the lambda calculus. *RAIRO - Theor. Inf. and Applic.*, 46(3):413–450, 2012.

[11] Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labeled markov systems. In *Proc. of CONCUR*, 1999.

[12] Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proc. of LICS*, pages 413–422, 2002.

[13] Josée Desharnais, François Laviolette, and Mathieu Tracol. Approximate analysis of probabilistic processes: Logic, simulation and games. In *Proc. of QEST*, pages 264–273, 2008.

[14] Thomas Ehrhard, Christine Tasson, and Michele Pagani. Probabilistic coherence spaces are fully abstract for probabilistic PCF. In *Proc. of POPL*, pages 309–320, 2014.

[15] Martin Escardo. A metric model of PCF. Proceedings of the Workshop on Realizability Semantics and Applications. Available at `http://www.cs.bham.ac.uk/~mhe/papers/metricpcf.pdf`, 1999.

[16] Daniel Gebler, Kim Guldstrand Larsen, and Simone Tini. Compositional metric reasoning with probabilistic process calculi. In *Proc. of FoSSaCS*, pages 230–245, 2015.

[17] Daniel Gebler and Simone Tini. SOS specifications of probabilistic systems by uniformly continuous operators. In *Proc. of CONCUR*, pages 155–168, 2015.

[18] Alessandro Giacalone, Chi chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2*, pages 443–458. North-Holland, 1990.

[19] Jean-Yves Girard. Linear logic. *Theor. Comput. Sci.*, 50:1–102, 1987.

[20] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984.

[21] Noah D. Goodman, Vikash K. Mansinghka, Daniel M. Roy, Keith Bonawitz, and Joshua B. Tenenbaum. Church: a language for generative models. In *UAI 2008*, pages 220–229, 2008.

[22] C. Jones and Gordon D. Plotkin. A probabilistic powerdomain of evaluations. In *Prof. of LICS*, pages 186–195, 1989.

[23] Achim Jung and Regina Tix. The troublesome probabilistic powerdomain. *Electr. Notes Theor. Comput. Sci.*, 13:70–91, 1998.

[24] Ugo Dal Lago, Davide Sangiorgi, and Michele Alberti. On coinductive equivalences for higher-order probabilistic functional programs. In *Proc. of POPL*, pages 297–308, 2014.

[25] Christopher D Manning and Hinrich Schütze. *Foundations of statistical natural language processing*, volume 999. MIT Press, 1999.

[26] Radu Mardare. Logical foundations of metric behavioural theory for markov processes. Doctoral Thesis. In Preparation, 2016.

[27] Sungwoo Park, Frank Pfenning, and Sebastian Thrun. A probabilistic language based on sampling functions. *ACM Trans. Program. Lang. Syst.*, 31(1), 2008.

[28] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[29] Gordon D. Plotkin. LCF considered as a programming language. *Theor. Comput. Sci.*, 5(3):223–255, 1977.

[30] Norman Ramsey and Avi Pfeffer. Stochastic lambda calculus and monads of probability distributions. In *Prof. of POPL*, pages 154–165, 2002.

[31] N. Saheb-Djahromi. Probabilistic LCF. In *Proc. of MFCS*, pages 442–451, 1978.

[32] Davide Sangiorgi. On the bisimulation proof method. *Mathematical Structures in Computer Science*, 8:447–479, 1998.

[33] Davide Sangiorgi and Valeria Vignudelli. Environmental bisimulations for probabilistic higher-order languages. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2016, St. Petersburg, FL, USA, January 20 - 22, 2016*, pages 595–607, 2016.

[34] Alex K. Simpson. Reduction in a linear lambda-calculus with applications to operational semantics. In *Proc. of RTA*, pages 219–234, 2005.

[35] Sebastian Thrun. Robotic mapping: A survey. *Exploring artificial intelligence in the new millennium*, pages 1–35, 2002.

[36] Franck van Breugel. An introduction to metric semantics: operational and denotational models for programming and specification languages. *Theor. Comput. Sci.*, 258(1-2):1–98, 2001.

[37] Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331(1):115–142, 2005.