International Journal of Approximate Reasoning 172 (2024) 109212

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

A possible worlds semantics for trustworthy non-deterministic computations

Ekaterina Kubyshkina *, Giuseppe Primiero

University of Milan, Logic, Uncertainty, Computation and Information Group (LUCI), Research Center for the Philosophy of Technology (PhilTech), Italy

ARTICLE INFO

Keywords:
Trustworthy AI
Probabilistic processes
Possible worlds semantics
Typed natural deduction

ABSTRACT

The notion of trustworthiness, central to many fields of human inquiry, has recently attracted the attention of various researchers in logic, computer science, and artificial intelligence (AI). Both conceptual and formal approaches for modeling trustworthiness as a (desirable) property of AI systems are emerging in the literature. To develop logics fit for this aim means to analyze both the non-deterministic aspect of AI systems and to offer a formalization of the intended meaning of their trustworthiness. In this work we take a semantic perspective on representing such processes, and provide a measure on possible worlds for evaluating them as trustworthy. In particular, we intend trustworthiness as the correspondence within acceptable limits between a model in which the theoretical probability of a process to produce a given output is expressed and a model in which the frequency of showing such output as established during a relevant number of tests is measured. From a technical perspective, we show that our semantics characterizes the probabilistic typed natural deduction calculus introduced in D'Asaro and Primiero (2021)[12] and further extended in D'Asaro et al. (2023) [13]. This contribution connects those results on trustworthy probabilistic processes with the mainstream method in modal logic, thereby facilitating the understanding of this field of research for a larger audience of logicians, as well as setting the stage for an epistemic logic appropriate to the task.

1. Introduction

During the last decades, AI systems have been developed and deployed on a massive scale, not only in scientific research, but also in daily life. In this context, as noted in the Ethics Guidelines for Trustworthy AI [29], trustworthiness is a prerequisite. For this reason, providing a formal account of the complex notion of trust has been a major task for many researchers in logic, computer science, and AI, see, e.g., Ferraiolo et al. [21], Liau [33], Demolombe [15], Primiero [37] just to name some approaches.

From the perspective of AI, modeling trust requires systematic and operational tools for representing both artificial and human agents' mental states, as well as communication means between them, see, e.g., Castelfranchi & Falcone [6]. In this sense, trust can be seen as an attitude of agents, or as a relation between them. A different way of approaching the issue, both conceptually (see, e.g., [34]) and formally (see e.g. [39]) is to understand trustworthiness as a property of a relation between agents or processes. The literature on Trustworthy AI is focusing on trustworthiness of computational processes in non-deterministic systems in this second sense, and on reasoning about such property. Trustworthiness in this context is an especially complex concept, which, depending

* Corresponding author.
E-mail address: ekaterina.kubyshkina@unimi.it (E. Kubyshkina).

https://doi.org/10.1016/j.ijar.2024.109212
Received 19 March 2024; Accepted 11 May 2024

Available online 18 May 2024

on the application, can be understood as predictability ("what will this AI system do?", see, e.g., [41]), explicability ("can one explain what this AI system does when acting instead of a human?", see, e.g., [23]), human agency ("how is the human involved in and aware of the effects of this AI system's results?", see, e.g., [35]), and safety ("how safe are this AI system's results?", see, e.g., [2]). Despite this variety of theoretical and conceptual problems being addressed, depending on whether we look at AI systems pre- or post-deployment, or whether we are interested in their ontological or epistemological status, a common technical aspect is represented by verification methods, which, depending on their formulation, can address these different aspects, see, e.g., the recent [40]. A number of approaches are being explored: symbolic verification [1,30,43]; verification of linear temporal logic properties defined over Markov decision processes, e.g., with reinforcement learning methods [24] or with imprecise probabilities [42]; proof-checking techniques like untyped λ-calculi [7,10,3], probabilistic event λ-calculi [4], calculi for Bayesian learning [11], and calculi with types or natural deduction systems [16,5,26]. In this latter tradition, [12] and [13] provide an example of a formal verification method which aims to automatize the task of inferential reasoning about probabilistic computational processes to verify properties of AI systems such as trustworthiness. The system is designed to capture the inferential steps that an agent reasoning over the behavior of a non-deterministic system of interest needs to perform in order to evaluate it as trustworthy. Such evaluation eventually consists in checking that the observed behavior of the system diverges only within acceptable limits from the expected, intended or desirable performance. The system and its methodology are further applied to bias detection for classification methods in [38], and in [25] with a λ-calculus simulating the logical behavior of a program performing such a formal check, with an additional measure of confidence being defined to evaluate its workings. All these approaches have a procedural nature, in that they are inspired by the formal verification tradition and they aim at producing in principle implementable tools. Possible applications include, for example, the verification of the behavior of classifiers in use in default risk assessment by credit institutions, or in insurance fraud risk, to establish whether predictions made for a given protected category (defined for example by gender, race or education) are beyond the values assigned by a given model of reference, like statistically available data or desirable distributions. For an early version of a tool that implements this strategy, see https://github.com/DLBD-Department/BRIO_x_Alkemy.

In the present work, we take a slightly different perspective and we aim at describing semantically the sufficient and necessary conditions for evaluating whether a process can be considered trustworthy in the sense mentioned above. As it will be shown later, the procedural and the descriptive approaches are not incompatible. They shed different lights on the same evaluation process. The former provides agents – who possibly have only partial knowledge of the system under observation, as it is the case for AI-style black-boxes – with instructions on how to evaluate its performance given a transparent or known counterpart; the latter, on the other hand, allows one to describe the overall setting in which the agent can follow the instructions, and moreover provides the basis on which to build further formal tools, in particular to describe the agent's knowledge as it evolves during such a verification process.

The semantics introduced in this paper is a variation on the standard possible worlds semantics adapted to probabilistic reasoning. The idea of combining logic with probability theory is not new (see Demey et al. [14] and references therein). Nilsson [36] points out the conceptual usefulness of possible worlds analysis of probabilistic reasoning (see also Fagin & Halpern [19] for further technical development of this work). Fagin & Halpern [18,20] relate probabilistic and epistemic logics by providing a modification of relational semantics for knowledge representation with probability spaces associated with each world. These ideas are further developed in various works on epistemic logic (see, e.g., Kooi [31,32], van Benthem [9], Baltag & Smets [8], Gierasimszuk [27]) which, however, are not focused on trustworthiness, and do not provide a unified framework for treating idealized probabilities, actual frequencies, and expected probabilities as we do here. Moreover, the novelty of our approach consists in keeping the underlying semantic structures as simple as possible. We define sentences involving probabilities and frequencies as syntactic constructions describing events in models, and not via special functions from worlds to positive rationals, as it is the case in the aforementioned literature. The language presented here recalls the style of term-modal logics [22], in which formulas have terms expressing idealized processes (or random variables) and empirical processes, and corresponding outputs (or values). In the present work, we do not quantify over such terms. Instead, trustworthiness can be seen as a measure over sets of worlds in different models: a trustworthy process for a given output is one in which the frequency of such output over a given number of trials does not diverge beyond an acceptable threshold from its expected probability. By connecting the study on trustworthy probabilistic processes with the mainstream methods in modal logic, we also facilitate the understanding of this area of research for a larger audience. The present work focuses on a logic-based theoretical approach to trustworthiness in non-deterministic systems. Clearly, the field of Trustworthy AI currently offers a very large variety of other approaches, too many to be mentioned here, with a number of important venues where new models are presented every year. Several quantitative approaches and tools are being developed, see, e.g., https://oecd.ai/en/catalogue/metrics.

The remainder of this article is as follows. In Section 2 we start with a toy example to illustrate the main intuition behind the working of our logic. In Section 3 we provide a formal semantics for representing trustworthy processes. Section 4 briefly reviews the formal system TPTND introduced in [13]. In Sections 5 and 6, we show that TPTND is characterized by our semantics, which indicates that our descriptive approach matches its procedural counterpart. Section 7 provides insights on further extension of our semantics with an epistemic modality. We conclude by summing up the results and describing further stages of this research.

2. A toy example

To address our problem, let us consider the following toy example.

Example 1. Consider a non-deterministic system which simulates throwing a die. Assume that this system is launched 18 times, we know it has produced output "3" three times, output "1" one time, and output "5" eight times. We know nothing on the remaining outputs, nor on the order of these outputs. On the basis of this distribution, one may conclude that the output "3" was received a fair number of times, while the outputs “1” and “5” were received an unexpected few number of times and too many times, respectively. In this respect, one could say that the system has a trustworthy behavior with respect to output “3”, and is untrustworthy with respect to output “1” and “5”.

What are the main ingredients for evaluating trustworthiness in this sense?

First, notice that our analysis assumes the modularity of a system with respect to its outputs: a system may be considered trustworthy when observed relatively to a given output, and untrustworthy when a different output is under consideration. This is not too strange, as partial knowledge may be involved concerning the possible outputs, or because the system may be affected in its behaviors by the circumstances of its execution (e.g., by an unbalanced dataset in input for a classifier). Obviously, this does not prevent a system to be considered trustworthy only if all of its possible outputs are known and evaluated as trustworthy in the above sense.

Second, we assume knowledge of the theoretical probability of each output (for a fair die to land on one of its 6 sides is $\frac{1}{6}$), and we use this as a reference (we expect that after 18 throws of an ideal die under ideal experimental conditions, the die should land on each side around three times). This can also be interpreted as knowledge of the theoretical distribution describing the desirable behavior of the system under observation.

Third, we have the empirical data obtained by observing the system at work (“3” was received three times, “1” was received one time, “5” was received eight times). This aspect requires us to consider as different observations of the system in which a different number of trials is involved.

Thus, when aiming at an adequate representation for the trustworthiness evaluation of our system, we should model two levels: a *theoretical* and an *empirical* one. The theoretical level aims to capture an ideal or desirable behavior of the system when working under idealized conditions for each of its possible outputs. The empirical level represents the data obtained during real executions of a process for each observed output. From this perspective, trustworthiness can be formulated in terms of a particular correlation between theoretical and empirical levels: a process is trustworthy with respect to a given output if the data represented at the empirical level for the output at hand matches within acceptable error limits the expectation for that output based on the theoretical level (theoretical or desirable distribution).

Our main objective is to introduce a logic which permits one to model this kind of examples, thus capturing this notion of trustworthiness formally. In particular, in the present work we provide a possible world semantics and show how it relates to the proof system presented in [13].

3. Formal semantics

The main idea of the proposed semantics is to provide a unified framework for evaluating the empirically observed behavior of a non-deterministic system against what is expected of it, in terms of the theoretical or desirable distribution on the probabilities of its outputs. Such evaluation provides a measure of trustworthiness for the system. In particular, we use the tools of possible worlds semantics to construct two types of models. A *theoretical model* is meant to represent an ideal or theoretical distribution of possible outputs by a process. An *empirical model* represents a series of executions of the process and their outputs. Trustworthiness is then evaluated on a fusion of the theoretical and empirical models.

The section is structured as follows. First, we define the language for the logic and explain its informal interpretation in natural language. Second, we define theoretical models, as random variables and theoretical probabilities of their values. Third, we define empirical models, as real-world experiments on processes associated to random variables. Then, we combine these models and provide operations defined over both theoretical and empirical models.

3.1. Syntax

The language should be expressive enough to represent: idealized processes (or random variables) and their outputs (or values); executed processes and their outputs; judgments about theoretical probabilities and frequencies of an output to be produced by a (resp. ideal and empirical) process. In order to distinguish propositions about idealized processes and their concrete executions, we introduce two kinds of elementary statements[1]:

(ES^t) $\mathtt{x} : \alpha$, to be read as “the idealized process x produces output α.”[2]
(ES^e) $\mathtt{t} : \alpha$, to be read as “the empirical process t produces output α.”

We also need to express that a given process x with output α is the idealized counterpart of an empirical one t, for which we write $\mathtt{x}_{\mathtt{t}} : \alpha$. Next, we decorate expressions with theoretical probabilities and frequencies as follows:

- $\mathtt{x} : \alpha_a$ stands for “the theoretical probability of idealized process x to produce output α is a.” This can be easily read also as “the theoretical probability of random variable x to have value α is a.”

[1] The syntax of these expressions is chosen to match precisely the syntax of TPTND in [13].

[2] This can be easily read also as “the random variable x has value α.”

- $\mathtt{x}_\mathtt{t} : \alpha_a$ stands for "the theoretical probability of idealized process $\mathtt{x}$ associated with the empirical process $\mathtt{t}$ to produce output α is a." For example, $x_d : 1_{1/6}$ stands for "the theoretical probability of random variable $\mathtt{x}$ associated with die d to have value 1 is 1/6."
- $\mathtt{t}_n : \alpha_{\tilde{a}}$ stands for "the expected probability of empirical process $\mathtt{t}$ to produce α over n executions is $\tilde{a}$." For example, $d_{10} : 1_{1.6}$ stands for "the expected probability of die d to have value 1 over 10 throws is 1.6."
- $\mathtt{t}_{\{w',\dots,w'^n\}} : \alpha_f$ stands for "after executions $w',\dots,w'^n$ of empirical process $\mathtt{t}$, output α has been produced with frequency f." For example, $d_{\{w',\dots,w'^{10}\}} : 1_{3/10}$ stands for "considering the launches $w',\dots,w'^{10}$ of a die d, the output 1 has resulted 3 times out of 10."

We now define the alphabet useful to construct expressions about idealized processes.

Definition 1 *(Alphabet of $\mathcal{L}^{theo}$).*

$$\mathtt{X} := \mathtt{x} \mid \langle \mathtt{X}, \mathtt{X} \rangle \mid \mathit{fst}(\mathtt{X}) \mid \mathit{snd}(\mathtt{X}) \mid [\mathtt{X}]\mathtt{X} \mid \mathtt{X}.\mathtt{X}$$
$$\mathtt{O} := \alpha \mid \alpha_r \mid \neg \mathtt{O}_r \mid (\mathtt{O} \times \mathtt{O})_r \mid (\mathtt{O} + \mathtt{O})_r \mid (\mathtt{O} \rightarrow \mathtt{O})_r$$

The domain $\mathtt{X}$ is constituted of a finite number of idealized processes, or random variables. The variable $\mathtt{x}$ stands for an idealized process x, y, z, etc. A construction $\langle \mathtt{X}, \mathtt{Y} \rangle$ denotes an ordered pair of idealized processes: this construction is required to express a pair of idealized processes, each with its own output. The terms $\mathit{fst}(\mathtt{X})$ and $\mathit{snd}(\mathtt{X})$ denote the first and the second element of a pair $\langle \mathtt{Y}, \mathtt{Z} \rangle = \mathtt{X}$, respectively. The construction $[\mathtt{X}]\mathtt{Y}$ denotes the dependency of an idealized process from another: this construction is required to express that a given idealized process produces a given output, provided the output of another process. The construction $\mathtt{Y}.\mathtt{X}$ denotes the process resulting from the construction $[\mathtt{X}]\mathtt{Y}$, when condition $\mathtt{X}$ obtains.

The variable $r \in \mathbb{Q}$ denotes here the theoretical probability of the value α of a random variable x (or output of an idealized process) and will be instantiated as a in $\mathtt{x} : \alpha_a$.[3]

The domain $\mathtt{O}$ is constituted of a finite number of possible outputs, a set for each (idealized) process. These outputs are always assumed to be exclusive and exhaustive for each process, and we will define our models in a way that satisfies these constraints. The variable α denotes an output, α_r denotes that α obtains with a theoretical probability r. $\neg\alpha_r$ denotes that output α is not valid with theoretical probability r (and as we shall define its validity, it implies that output α has probability $1 - r$ to occur). The construction $(\alpha \times \beta)_r$ denotes that the joint probability of two independent outputs α and β is r. The construction $(\alpha + \beta)_r$ denotes that the joint probability of obtaining output α or output β is r. The construction $(\alpha \rightarrow \beta)_r$ expresses that the probability of obtaining β under condition that α is obtained, is r. Note that, in general, the condition that α is obtained might itself have a probabilistic value a. In such cases, we keep track of such value with the notation $[a]b$ for expressing the probability b of obtaining β under the probability a of obtaining α. When $a = 1$, that is α is a determined output, $r = b$, which means that the probability of β coincides with the probability of β under condition that α is obtained.

With this building blocks we construct formulae of $\mathcal{L}^{theo}$:

Definition 2 *(Language $\mathcal{L}^{theo}$).*

$$\mathcal{L}^{theo} := \mathtt{X} : \alpha \mid \mathtt{X} : \alpha_r \mid \mathtt{X} : \neg\mathtt{O}_r \mid \langle \mathtt{X}, \mathtt{X} \rangle : (\mathtt{O} \times \mathtt{O})_r \mid \mathit{fst}(\langle \mathtt{X}, \mathtt{X} \rangle) : \mathtt{O}_r \mid \mathit{snd}(\langle \mathtt{X}, \mathtt{X} \rangle) : \mathtt{O}_r \mid$$
$$[\mathtt{X}]\mathtt{X} : (\mathtt{O} \rightarrow \mathtt{O})_r \mid \mathtt{X}.\mathtt{X} : \mathtt{O}_r$$

Similarly, we define the syntax of the language $\mathcal{L}^{emp}$ to form statements about tests or executions of empirical processes.

Definition 3 *(Alphabet of $\mathcal{L}^{emp}$).*

$$\mathtt{T} := \mathtt{t} \mid \mathtt{T}_{\{w',\dots,w'^n\}} \mid \langle \mathtt{T}, \mathtt{T} \rangle \mid \mathit{fst}(\mathtt{T}) \mid \mathit{snd}(\mathtt{T})$$
$$\mathtt{O} := \alpha \mid \alpha_r \mid \neg\mathtt{O}_r \mid (\mathtt{O} \times \mathtt{O})_r \mid (\mathtt{O} + \mathtt{O})_r$$

The natural language reading of expressions in $\mathcal{L}^{emp}$ is similar to those of $\mathcal{L}^{theo}$. The main difference is that now we are speaking about empirical processes whose execution is observed in the real world and their outputs produced with a given frequency. Thus, the domain $\mathtt{T}$ is constituted of a finite number of processes; $\mathtt{t}$ stands for an executed process; $\mathtt{T}_{\{w',\dots,w'^n\}}$ denotes executions $w',\dots,w'^n$ of the (possibly complex) process $\mathtt{T}$; $\langle \mathtt{T}, \mathtt{U} \rangle$ stands for the joint execution of two independent processes; $\mathit{fst}(\mathtt{T})$ and $\mathit{snd}(\mathtt{T})$ denote respectively the first and second process of such a joint execution.

The domain for outputs can be interpreted as before, except $r \in \mathbb{Q}$ in α_r may now indicate two distinct parameters: the expected probability of an output α for a process $\mathtt{t}$ (as determined by the theoretical probability a of the corresponding random variable $\mathtt{x}_\mathtt{t}$ to get assigned value α) which will be denoted as $\tilde{a}$ in $\mathtt{t} : \alpha_{\tilde{a}}$; or the frequency of a given output α after n executions of process $\mathtt{t}$, which will be denoted as f in $\mathtt{t}_{\{w',\dots,w'^n\}} : \alpha_f$.

[3] We follow here [13] in expressing probabilities in terms of real numbers. As it will be clear later, the probabilities are in fact restricted to rational numbers. However, this does not affect the resulting system and semantics, as the models remain constrained to a finite number of worlds.

With these building blocks we construct formulae of $\mathcal{L}^{emp}$ (where the use of $r \in \mathbb{Q}$ can be replaced with either $\tilde{a}$ or f as appropriate):

Definition 4 *(Language $\mathcal{L}^{emp}$).*

$$\begin{aligned}\mathcal{L}^{emp} := {} & \mathtt{T} : \alpha \mid \mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_r \mid \mathtt{T}_{\{w',\ldots,w'^n\}} : \neg \mathtt{O}_r \mid \mathtt{T}_{\{w',\ldots,w'^n\}} : \mathtt{O}_{\tilde{a}} \mid \\ & \mathtt{T}_{\{w',\ldots,w'^n\}} : \mathtt{O}_f \mid \langle \mathtt{T},\mathtt{T}\rangle_{\{w',\ldots,w'^n\}} : (\mathtt{O}\times\mathtt{O})_r \mid \mathit{fst}(\langle \mathtt{T},\mathtt{T}\rangle)_{\{w',\ldots,w'^n\}} : \mathtt{O}_r \mid \\ & \mathit{snd}(\langle \mathtt{T},\mathtt{T}\rangle)_{\{w',\ldots,w'^n\}} : \mathtt{O}_r.\end{aligned}$$

Note that in this fragment we do not have →- formulae, which are instead re-introduced in the joint language as the validity of a term $\mathtt{t}$ from its corresponding variable $\mathtt{x}_{\mathtt{t}}$.

Hence, the two languages, $\mathcal{L}^{theor}$ and $\mathcal{L}^{emp}$ can now be combined and enriched with expressions to make statements about relations between theoretical probabilities and empirical frequencies:

Definition 5 *(Alphabet of $\mathcal{L}$).*

$$\begin{aligned}&\mathtt{X} := \mathtt{x} \mid \mathtt{x}_{\mathtt{T}} \mid \langle \mathtt{X},\mathtt{X}\rangle \mid \mathit{fst}(\mathtt{X}) \mid \mathit{snd}(\mathtt{X}) \mid [\mathtt{X}]\mathtt{X} \mid \mathtt{X}.\mathtt{X}\\ &\mathtt{T} := \mathtt{t} \mid \mathtt{T}_{\{w',\ldots,w'^n\}} \mid \langle \mathtt{T},\mathtt{T}\rangle \mid \mathit{fst}(\mathtt{T}) \mid \mathit{snd}(\mathtt{T}) \mid [\mathtt{X}]\mathtt{T} \mid \mathtt{T}.\mathtt{T}\\ &\mathtt{F} := \mathit{Trust}(\mathtt{T}) \mid \mathit{UTrust}(\mathtt{T})\\ &\mathtt{O} := \alpha \mid \alpha_r \mid \neg\mathtt{O}_r \mid (\mathtt{O}\times\mathtt{O})_r \mid (\mathtt{O}+\mathtt{O})_r \mid (\mathtt{O}\rightarrow\mathtt{O})_r\end{aligned}$$

Let us shortly discuss the natural language interpretation of the new elements of the language with respect to the grammar of $\mathcal{L}^{theo}$ and of $\mathcal{L}^{emp}$. As mentioned above, the new variable $\mathtt{x}_{\mathtt{T}}$ denotes the idealized process $\mathtt{x}$ (with its output and the theoretical probability attached to it) associated with an empirical one $\mathtt{T}$ (or the random variable corresponding to a given event). A construction of the form $[\mathtt{X}]\mathtt{T}$ denotes the execution of a process $\mathtt{T}$ with its own expected probability or observed frequency on the assumption that the corresponding idealized process is $\mathtt{X}$ with theoretical probability a (or the event, given the corresponding random variable – the inverse of the previous construction $\mathtt{X}_{\mathtt{T}}$). This construction will be used therefore to express the expected probability or frequency assigned to the output of a process $\mathtt{T}$ assuming the theoretical probability assigned to $\mathtt{X}$. A construction of a form $\mathtt{U}.\mathtt{T}$ expresses the result of process $\mathtt{T}$ when the corresponding random variable is instantiated as some (possibly distinct) $\mathtt{U}$ (i.e. when the theoretical probability of the latter is replaced by a given frequency or expected probability). We finally introduce terms for (un)trustworthy processes: $\mathit{Trust}(\mathtt{T})$ (resp. $\mathit{UTrust}(\mathtt{T})$) is used to express the fact that the frequency of process $\mathtt{T}$ is considered trustworthy (resp. untrustworthy) with respect to the theoretical probability of its output. The interpretation of the elements of the domain $\mathtt{O}$ is as in $\mathcal{L}^{theo}$, if the expression is preceded by "X:"; and it is as in $\mathcal{L}^{emp}$, if the expression is preceded by "T:".

Formulae of the language $\mathcal{L}$ are all the formulae in $\mathcal{L}^{theo}$, all those in $\mathcal{L}^{emp}$ and the new formulas where a is a theoretical probability and r is either an expected probability of the output of an empirical process, or the observed frequency (in the latter case, the corresponding term is indexed with n the number of executions):

Definition 6 *(Language $\mathcal{L}$).*

$$\begin{aligned}\mathcal{L} := {} & \mathcal{L}^{theo} \cup \mathcal{L}^{emp} \cup \\ & \{[\mathtt{X}]\mathtt{T}_n : (\mathtt{O}\rightarrow\mathtt{O})_{[a]\tilde{b}}, \mathtt{T}_n.\mathtt{T}' : \mathtt{O}_{\tilde{b}}, \\ & \mathit{Trust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \mathtt{O}_f), \mathit{UTrust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \mathtt{O}_f)\}\end{aligned}$$

Formulae of the form $[\mathtt{x}]\mathtt{t}_n : (\alpha \rightarrow \beta)_{[a]\tilde{b}}$ should be read as 'under theoretical probability a of receiving an output α, the process $\mathtt{t}$ should produce output β with an expected probability $\tilde{b}$ over n executions of $\mathtt{t}$.' Formulae of the form $\mathtt{t}_n.\mathtt{u} : \alpha_{\tilde{b}}$ should be read as 'the expected probability of receiving α after an independent execution of $\mathtt{t}$ n times, distinct from $\mathtt{u}$, is $\tilde{b}$.' Formulae of the form $\mathit{Trust}(\mathtt{t}_{\{w',\ldots,w'^n\}} : \alpha_f)$ should be read as 'the process $\mathtt{t}$ producing α with a frequency f is considered trustworthy on the interval of tests $w',\ldots,w'^n$.' Formulae of the form $\mathit{UTrust}(\mathtt{t}_{\{w',\ldots,w'^n\}} : \alpha_f)$ should be read as 'the process $\mathtt{t}$ producing α with a frequency f is considered untrustworthy on the interval of tests $w',\ldots,w'^n$.'

3.2. Theoretical models

We propose a variant of standard possible worlds semantics without accessibility relations between worlds. In this sense, it is closer to Carnap's usage, than Kripke's one. As we will discuss it in Section 7, accessibility relations can be introduced to our semantics to model epistemic operators. However, this is not necessary for our current purposes.

Theoretical models are used to express idealized processes as events at worlds, and to evaluate their theoretical probabilities as measures across worlds. We thus associate the possible outcomes of an idealized process with possible worlds and then compute their probability in a set of worlds. For instance, the theoretical probability associated with the outcome "1" of a fair die is $\frac{1}{6}$. To represent it, the corresponding model will contain 6 worlds in which only one world shows output 1. Let us now introduce this setting formally.

Definition 7 *(Theoretical models).* Let

$$\mathcal{M}^{theor} = (W^{theor}, v^{theor})$$

such that

- W^{theor} is a non-empty set of worlds $w_1, ..., w_n$ such that $w_1, ..., w_n$ are sets of formulas ES^t,
- $v^{theor} : \mathtt{x} : \alpha \rightarrow P(W^{theor})$ is a valuation function, such that:
  - $v^{theor}(\mathtt{x} : \alpha) \cap v^{theor}(\mathtt{x} : \beta) = \emptyset$, whenever $\alpha \neq \beta$;
  - for all w there exists v^{theor} such that $w \in v^{theor}(ES^t)$.

  We call $\mathcal{M}^{theor}_s$ a submodel of $\mathcal{M}^{emp}$ iff $\mathcal{M}^{theor}_s = (W^{theor}_s, v^{theor}_s)$, where $W^{theor}_s \subseteq W^{theor}$ and v^{theor}_s is v^{theor} restricted to the worlds of W^{theor}_s.

In this model we can evaluate ES^t formulas as follows:

Definition 8 *(Satisfiability in $\mathcal{M}^{theor}$).* Given a theoretical model $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$, truth conditions for formulas of $\mathcal{L}^{theo}$ are defined as follows:

1. $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{x} : \alpha$ iff $w_i \in v(\mathtt{x} : \alpha)$;
2. $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : (\alpha + \beta)$ iff $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha$ or $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \beta$;
3. $\mathcal{M}^{theor}, w_i \vDash_t \langle \mathtt{X}, \mathtt{Y} \rangle : (\alpha \times \beta)$ iff $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha$ and $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{Y} : \beta$;
4. $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha_a$ iff
   - $\mid W^{theor} \mid = n$;
   - $b = \mid \{w_i \in W^{theor} \mid \mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha\} \mid$;
   - $a = \frac{b}{n}$;
5. $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \neg\alpha_a$ iff $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha_{1-a}$;
6. $\mathcal{M}^{theor} \vDash_t \mathtt{X} : (\alpha + \beta)_a$ iff $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha_b$, $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \beta_c$, and $a = b + c$;
7. $\mathcal{M}^{theor} \vDash_t \langle \mathtt{X}, \mathtt{Y} \rangle : (\alpha \times \beta)_a$ iff $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha_b$, $\mathcal{M}^{theor} \vDash_t \mathtt{Y} : \beta_c$, and $a = b \cdot c$;
8. $\mathcal{M}^{theor} \vDash_t [\mathtt{X}]\mathtt{Y} : (\alpha \rightarrow \beta)_a$ iff whenever $\mathcal{M}^{theor}_s = (W^{theor}_s, v^{theor}_s)$ is a submodel of $\mathcal{M}^{theor}$ s.t. $W^{theor}_s = \{w_i \in W^{theor} \mid \mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha\}$, then $\mathcal{M}^{theor}_s \vDash_t \mathtt{Y} : \beta_a$.

According to this definition, the valuation function is defined for each elementary statement $\mathtt{x} : \alpha$ at a world, i.e., worlds contain expressions about elementary idealized processes and their outcomes. Clause 2 defines disjunctive formulas: $\mathtt{X} : (\alpha + \beta)$ should be read as "the process $\mathtt{X}$ produces an output α or an output β." For example, the clause $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : (1 + 2)$ states that in the world w_i of the model $\mathcal{M}^{theor}$ the variable $\mathtt{X}$ produces in this world either output 1 or output 2. Clause 3 defines the case of conjunctive outputs: $\langle \mathtt{X}, \mathtt{Y} \rangle : (\alpha \times \beta)$ should be read as "the joint processes $\mathtt{X}$ and $\mathtt{Y}$ produce outputs α and β." For example, the clause $\mathcal{M}^{theor}, w_i \vDash_t \langle \mathtt{X}, \mathtt{Y} \rangle : (1 \times 2)$ states that in the world w_i of model $\mathcal{M}^{theor}$ two variables $\mathtt{X}$ and $\mathtt{Y}$ considered together produce in this world output 1 and output 2. A formula "$\mathtt{X} : \alpha_a$" means that the theoretical probability of α to be produced by process $\mathtt{X}$ is a. We evaluate these formulas with probabilistic outputs in the model (rather than at worlds). The semantic clause 4 states therefore that $\mathtt{X} : \alpha_a$ holds in a model if and only if the number of worlds of this model is n ($\mid W^{theor} \in \mathcal{M}^{theor} \mid = n$), the number of worlds in which $\mathtt{X} : \alpha$ holds is b ($b = \mid \{w_i \in W^{theor} \mid \mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha\} \mid$, and the probability a is calculated according to standard probability theory as $a = \frac{b}{n}$. For example, the clause $\mathcal{M}^{theor} \vDash_t \mathtt{X} : 1_{\frac{1}{6}}$ states that if the model contains 6 worlds, only one world among them validates $\mathtt{X} : 1$. According to this definition, it is evident that a cannot be an irrational number. Clause 5 provides the evaluation condition for $\mathtt{X} : \neg\alpha_a$: the probability of an output different from α is a for a process $\mathtt{X}$ which produces α with probability $1 - a$. For example, the clause $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \neg 1_{\frac{5}{6}}$ states that the probability of receiving an output different from 1 is $\frac{5}{6}$. Clauses 6 and 7 establish conditions for producing disjunctive and conjunctive outputs. For example, $\mathcal{M}^{theor} \vDash_t \mathtt{X} : (1 + 2)_{\frac{1}{3}}$ states that $\mathtt{X}$ produces output 1 or 2 with probability $\frac{1}{3}$; $\mathcal{M}^{theor} \vDash_t \langle \mathtt{X}, \mathtt{Y} \rangle : (1 \times 2)_{\frac{1}{36}}$ states that two variables (for distinct processes) considered together produce outputs 1 and 2 simultaneously with a probability $\frac{1}{36}$. Clause 8 provides the evaluation condition for dependent processes: the expression "$[\mathtt{X}]\mathtt{Y} : (\alpha \rightarrow \beta)_a$" should be read as "process $\mathtt{Y}$ has output β under condition that process $\mathtt{X}$ has output α holds with probability a." The clause thus states that $[\mathtt{X}]\mathtt{Y} : (\alpha \rightarrow \beta)_a$ holds in a model $\mathcal{M}^{theor}$ iff in all its worlds in which $\mathtt{X} : \alpha$ holds, the probability of $\mathtt{Y} : \beta$ calculated over information contained only in these worlds is a. Notice, that in this case the probability of $\mathtt{Y} : \beta$ in all worlds of $\mathcal{M}^{theor}$ could be not equal to a. For example, $\mathcal{M}^{theor} \vDash_t [\mathtt{X}]\mathtt{Y} : (2 \rightarrow 1)_{\frac{1}{36}}$ states that $\mathtt{Y}$ produces output 1 under condition that $\mathtt{X}$ produces output 2 with probability $\frac{1}{36}$.

Example 2. A theoretical model of a fair die would satisfy the following formulae:

$$\mathcal{M}^{theor} \vDash_t x_d : 1_{1/6}$$

$$\mathcal{M}^{theor} \vDash_t x_d : (1 + 3)_{0.33}$$

A theoretical model of two fair dice would satisfy the following formula:

$$\mathcal{M}^{theor} \vDash_t \langle x_d, y_{d'} \rangle : (1 \times 3)_{0.33}$$

Definition 9 *(Satisfiability (Further clauses)).* Let $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$. Then,

- $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha$ iff $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X} : \alpha$, $\forall w_i \in \mathcal{M}^{theor}$;
- $\mathcal{M}^{theor}, w_i \vDash_t \Gamma$ where $\Gamma = \{\mathtt{X}^1 : \alpha^1, ..., \mathtt{X}^n : \alpha^n\}$ iff $\mathcal{M}^{theor}, w_i \vDash_t \mathtt{X}^j : \alpha^j$ for all $j \in \{1, ..., n\}$;
- $\mathcal{M}^{theor} \vDash_t \Gamma$ where $\Gamma = \{\mathtt{X}^1 : \alpha^1, ..., \mathtt{X}^n : \alpha^n\}$ iff for $\Gamma' \subseteq \Gamma$ such that Γ' contains all and only non-probabilistic formulae of Γ, then $\mathcal{M}^{theor}, w_i \vDash_t \Gamma'$ for all $w_i \in W^{theor}$, and the model satisfies all probabilistic formulas occurring in Γ as for Definition 8.

The following observations further clarify the design of this semantics.[4]

Observation 1. *For any $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$ and any $\mathtt{X} : \alpha$, there exists a value a, s.t. $\mathcal{M}^{theor} \vDash \mathtt{X} : \alpha_a$.*

Observation 1 makes it explicit, that there is a theoretical probability for each output of each process.

Observation 2. *Let $\alpha, ..., \nu$ be mutually exclusive and exhaustive outputs of $\mathtt{X}$. Then, for any $\mathcal{M}^{theor}$ and any $\mathtt{X} : \alpha_i$ such that $\mathcal{M}^{theor} \vDash \mathtt{X} : \alpha_a, \ldots, \mathcal{M}^{theor} \vDash \mathtt{X} : \nu_n$, $\sum_{i=a}^{n}(\alpha_i) = 1$.*

Observation 2 clarifies that any model $\mathcal{M}^{theor}$ satisfies the standard requirement on the additivity of theoretical probabilities within a probability distribution.

Observation 3. *For any $\mathcal{M}^{theor}$ and for any $\mathtt{X} : \alpha_a$ such that $\mathcal{M}^{theor} \vDash \mathtt{X} : \alpha_a$, $a \in [0, 1]$.*

Observation 3 states that the probability of producing some output α in a model is always in a range between 0 and 1.

By definitions of terms $\mathit{fst}(\mathtt{X})$ and $\mathit{snd}(\mathtt{X})$ denoting respectively the first and the second process from a pair of processes $\mathtt{X} = \langle \mathtt{Y}, \mathtt{Z} \rangle$ and clause 7 above, we have the following truth conditions for these expressions:

Proposition 1.

- *$\mathcal{M}^{theor} \vDash_t \mathit{fst}(\langle \mathtt{Y}, \mathtt{Z} \rangle) : \alpha_a$ iff there exist b and c such that $\mathcal{M}^{theor} \vDash_t \langle \mathtt{Y}, \mathtt{Z} \rangle : (\alpha \times \beta)_c$, $\mathcal{M}^{theor} \vDash_t \mathtt{Z} : \beta_b$ and $a = \frac{c}{b}$.*
- *$\mathcal{M}^{theor} \vDash_t \mathit{snd}(\langle \mathtt{Y}, \mathtt{Z} \rangle) : \beta_b$ iff there exist b and c such that $\mathcal{M}^{theor} \vDash_t \langle \mathtt{Y}, \mathtt{Z} \rangle : (\alpha \times \beta)_c$ and $\mathcal{M}^{theor} \vDash_t \mathtt{Y} : \alpha_a$, and $b = \frac{c}{a}$.*

From the evaluation clause of term $[\mathtt{X}]\mathtt{Y}$, we have the following evaluation conditions for term $\mathtt{Y}.\mathtt{X}$:

Proposition 2.

- *$\mathcal{M}^{theor} \vDash_t \mathtt{Y}.(\mathtt{X} : \alpha) : \beta_b$ iff there exist a and c such that $\mathcal{M}^{theor} \vDash_t [\mathtt{X}]\mathtt{Y} : (\alpha \to \beta)_c$, $\mathcal{M}^{theor} \vDash \mathtt{X} : \alpha_a$ and $b = a \cdot c$.*

Proof. We need to define the probability b of the output β to be produced by the process $\mathtt{Y}$ under condition that $\mathtt{X}$ produced α with probability a. Taking into account that any expression in $\mathcal{M}^{theor}$ has a probability value attached (Obs. 1), we have

1. $\mathcal{M}^{theor} \vDash_t [\mathtt{X}]\mathtt{Y} : (\alpha \to \beta)_c$ and
2. $\mathcal{M}^{theor} \vDash \mathtt{X} : \alpha_a$.

Let us consider a model $\mathcal{M}_s^{theor} = \langle W_s^{theor}, v_s^{theor} \rangle$ s.t. it is a submodel of $\mathcal{M}^{theor}$ and $W_s^{theor} = \{w \mid \mathcal{M}^{theor}, w \vDash \mathtt{X} : \alpha\}$. By (2) we have $a = \frac{m}{i}$, where m is the size of $\mathcal{M}_s^{theor}$ (i.e., the number of worlds where $\mathtt{X} : \alpha$ holds) and i is the size of $\mathcal{M}^{theor}$. From (1) we have $\mathcal{M}_s^{theor} \vDash \mathtt{Y} : \beta_c$, and thus $c = \frac{i}{m}$, where i is the number of worlds of $\mathcal{M}_s^{theor}$ in which $\mathtt{Y} : \beta$ holds. The dependent probability of $\mathtt{Y} : \beta$ with respect to $\mathtt{X} : \alpha$ (i.e., $\mathtt{Y}.(\mathtt{X} : \alpha) : \beta$) is thus $b = \frac{i}{j} = a \cdot c$. □

Definition 10 *(Semantic consequence).* A statement $\mathtt{X} : \alpha_a$ is a semantic consequence of Γ, denoted by $\Gamma \vDash_t \mathtt{X} : \alpha_a$, if $\mathcal{M}^{theor} \vDash_t \Gamma$ implies $\mathcal{M}^{theor} \vDash_t \mathtt{X} : \alpha_a$.

Let us conclude this subsection with an example of a theoretical model.

[4] In particular, note these observations express properties of our theoretical models that precisely correspond to the construction rules for distribution in TPTND as illustrated below in Fig. 4.

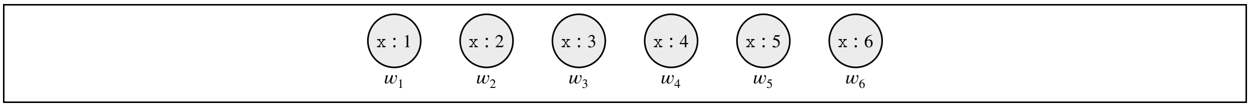

Fig. 1. Model $\mathcal{M}^{theor}$ for fair die.

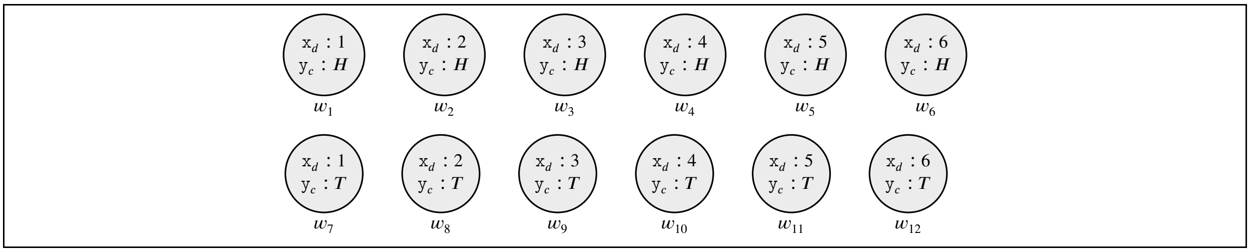
$\mathrm{x}_d : 1$ $\mathrm{y}_c : H$
w_1
$\mathrm{x}_d : 2$ $\mathrm{y}_c : H$
w_2
$\mathrm{x}_d : 3$ $\mathrm{y}_c : H$
w_3
$\mathrm{x}_d : 4$ $\mathrm{y}_c : H$
w_4
$\mathrm{x}_d : 5$ $\mathrm{y}_c : H$
w_5
$\mathrm{x}_d : 6$ $\mathrm{y}_c : H$
w_6
$\mathrm{x}_d : 1$ $\mathrm{y}_c : T$
w_7
$\mathrm{x}_d : 2$ $\mathrm{y}_c : T$
w_8
$\mathrm{x}_d : 3$ $\mathrm{y}_c : T$
w_9
$\mathrm{x}_d : 4$ $\mathrm{y}_c : T$
w_{10}
$\mathrm{x}_d : 5$ $\mathrm{y}_c : T$
w_{11}
$\mathrm{x}_d : 6$ $\mathrm{y}_c : T$
w_{12}

Fig. 2. Model $\mathcal{M}^{theor}$ for fair die and fair coin.

Example 3. Let us consider the die system from Example 1. A theoretical model for this system has to represent the theoretical distribution of outputs of a fair die in which the probability of obtaining each output is $\frac{1}{6}$. The theoretical model depicted in Fig. 1 contains six worlds, in each of which an idealized process x for throwing a die produces six distinct outputs, or the random variable gets one of each possible value, i.e. $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$ such that $W^{theor} = \{w_1, w_2, w_3, w_4, w_5.w_6\}$ and

$$v^{theor}(\mathrm{x} : 1) = \{w_1\}$$
$$v^{theor}(\mathrm{x} : 2) = \{w_2\}$$
$$v^{theor}(\mathrm{x} : 3) = \{w_3\}$$
$$v^{theor}(\mathrm{x} : 4) = \{w_4\}$$
$$v^{theor}(\mathrm{x} : 5) = \{w_5\}$$
$$v^{theor}(\mathrm{x} : 6) = \{w_6\}$$

We might then want to evaluate whether a fair die model evaluates to true the formula

$$\mathrm{x}_d : (1 + 3)_{0.33}$$

i.e., that the theoretical probability to get output 1 or output 3 is 0.33. For this we need to check that $0.33 = a + b$, that $\mathcal{M}^{theor} \vDash_t \mathrm{x}_d : 1_a$ (resp. $\mathcal{M}^{theor} \vDash_t \mathrm{x}_d : 3_b$) and the number of words z in $\mathcal{M}^{theor}$ such that 1 holds is such that $a = z/n$, with n the total number of words (resp. for 3). It is easy to verify that $z = 1$ in both cases (namely w_1 and w_3) and the total number of worlds in the model is 6, hence $a = 1/6$ and $b = 1/6$ and $0.33 = 1/6 + 1/6$.

Example 4. Let us consider a model which represents the theoretical distribution of 2 systems: of a fair die and of a fair coin. This model is depicted in Fig. 2: there is a fair distribution of a die's outputs (x_d) and a fair distribution of a coin's outputs (y_c). In this model we can evaluate the probability of getting output 3 from a die ($\mathrm{x}_d : 3$) and an output $Heads$ from a coin ($\mathrm{y}_c : H$) launched simultaneously:

$$\langle \mathrm{x}_d, \mathrm{y}_c \rangle : (3 \times H)$$

i.e., we can calculate the theoretical probability to get both outputs 3 and H is $\frac{1}{12}$. In this model we have $\mathcal{M}^{theor} \vDash_t \mathrm{x}_d : 3_{\frac{1}{6}}$, $\mathcal{M}^{theor} \vDash_t \mathrm{y}_c : H_{\frac{1}{2}}$, and $\frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$. Thus, $\mathcal{M}^{theor} \vDash_t \langle \mathrm{x}_d, \mathrm{y}_c \rangle : (3 \times H)_{\frac{1}{12}}$.

We can also calculate the probability of getting an output 3 under condition that H was received:

$$[\mathrm{y}_c]\mathrm{x}_d : (H \to 3).$$

For doing so, consider a submodel $\mathcal{M}_s^{theor} = (W_s^{theor}, v_s^{theor})$ of $\mathcal{M}^{theor}$, s.t. $W_s^{theor} = \{w_1, w_2, w_3, w_4, w_5, w_6\}$ (i.e., it contains all and only the worlds validating $\mathrm{y}_c : H$). In this model we have $\mathcal{M}_s^{theor} \vDash_t \mathrm{x}_d : 3_{\frac{1}{6}}$. Thus, $\mathcal{M}^{theor} \vDash_t [\mathrm{y}_c]\mathrm{x}_d : (H \to 3)_{\frac{1}{6}}$.

3.3. Empirical models

Similar to theoretical model, empirical model is defined as a tuple of a set of possible worlds and a valuation function. However, the informal interpretation of the worlds is different and each world is considered as a single test in a series of experiments where we might want to assign either the theoretical probability of a given input on this run of experiments, or express the actual frequency of that output. For instance, consider a test made by throwing a die 100 times, and consider output “3”: if we assume the die to be fair, we expect the probability of this output to be around 16 times while its actual frequency might be what is allowed by standard deviation (assuming the die to be in fact fair), e.g., 20 times. For representing such a test our model will contain 100 worlds, where each world contains information about the output of one throw of the die. Among these worlds, there will be 20 worlds in which the process of launching the die (denoted as $\mathtt{t}$) is labeled by the output “3”.

Definition 11 *(Empirical model).* Let

$$\mathcal{M}^{emp} = (W^{emp}, v^{emp})$$

such that

- W^{emp} is non-empty set of worlds $w_1, ..., w_n$,
- $v^{emp} : ES^e \to P(W^{emp})$ is a valuation function, such that
  - $v^{emp}(\mathtt{t} : \alpha) \cap v^{emp}(\mathtt{t} : \beta) = \emptyset$, whenever $\alpha \neq \beta$;
  - for all w there exists v^{emp} such that $w \in v^{emp}(ES^e)$.

  We call $\mathcal{M}^{emp}_s$ a *submodel* of $\mathcal{M}^{emp}$ iff $\mathcal{M}^{emp}_s = (W^{emp}_s, v^{emp}_s)$, where $W^{emp}_s \subseteq W^{emp}$ and v^{emp}_s is v^{emp} restricted to the worlds of W^{emp}_s.

Definition 12 *(Satisfiability in $\mathcal{M}^{emp}$).* Given an empirical model $\mathcal{M}^{emp} = (W^{emp}, v^{emp})$, truth conditions for formulas of $\mathcal{L}^{emp}$ are defined as follows

1. $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{t} : \alpha$ iff $w \in v^{emp}(\mathtt{t} : \alpha)$;
2. $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \neg\alpha$ iff $\mathcal{M}^{emp}, w_i \nvDash_e \mathtt{T} : \alpha$;
3. $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : (\alpha + \beta)$ iff $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \alpha$ or $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \beta$;
4. $\mathcal{M}^{emp}, w_i \vDash_e \langle \mathtt{T}, \mathtt{U} \rangle : (\alpha \times \beta)$ iff $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \alpha$ and $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{U} : \beta$;
5. $\mathcal{M}^{emp} \vDash_e \mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$ iff there exists a submodel $\mathcal{M}^{emp}_s = (W^{emp}_s, v^{emp}_s)$ of $\mathcal{M}^{emp}$ such that $W^{emp}_s = \{w', ..., w'^n\}$ and
   - $n = \mid W^{emp}_s \mid$;
   - $m = \mid \{w_i \in W^{emp}_s \mid \mathcal{M}^{emp}_s, w_i \vDash_e \mathtt{T} : \alpha\} \mid$;
   - $f = \frac{m}{n}$.

The interpretation of expressions provided in Definition 12 is similar to the one provided in Definition 8, except now we are considering empirically observed executions of processes. Clauses $1 - 4$ express satisfiability of categorical formulae at a world for atomic output, their negation, conjunction and disjunction. Clause 5 introduces the validity conditions for statements over the frequency of a given output: $\mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$ is true in a model if there are n of executions of $\mathtt{T}$ producing α with a frequency $f = \frac{m}{n}$, where f is the number of all worlds among $w', ..., w'^n$ in which α is produced divided by n. For instance, $\mathcal{M}^{emp} \vDash_e \mathtt{T}_{\{w', ..., w'^{100}\}} : 1_{0.1}$ states that the 100 launches $w', ..., w'^{100}$ of the process $\mathtt{T}$ have produced output 1 with frequency 0.1, i.e., 10 out of 100 times. Notice, that if there were 150 launches of $\mathtt{T}$ in total, but only 100 are considered, we have $\mathcal{M}^{emp} \vDash_e \mathtt{T}_{\{w'', ..., w''^{100}\}} : 1_{0.2}$, whenever $\{w', ..., w'^{100}\} \neq \{w'', ..., w''^{100}\}$ and 1 is obtained exactly in 20 worlds belonging to the set $\{w'', ..., w''^{100}\}$. This is due to the fact that an agent is permitted to consider various stages of an experiment, and not only the overall result.

Definition 13 *(Satisfiability (Further clauses)).* Let $\mathcal{M}^{emp} = (W^{emp}, v^{emp})$. Then,

- $\mathcal{M}^{emp} \vDash_e \mathtt{T} : \alpha$ iff $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \alpha$, $\forall w_i \in \mathcal{M}^{emp}$;
- $\mathcal{M}^{emp}, w_i \vDash_e \Gamma$ where $\Gamma = \{\mathtt{T}^1 : \alpha^1, ..., \mathtt{T}^n : \alpha^n\}$ iff $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T}^i : \alpha^i$ for all $i \in \{1, ..., n\}$;
- $\mathcal{M}^{emp} \vDash_e \Gamma$ where $\Gamma = \{\mathtt{T}^1 : \alpha^1, ..., \mathtt{T}^n : \alpha^n\}$ iff for $\Gamma' \subseteq \Gamma$ such that Γ' contains all and only non-probabilistic formulae of Γ, then $\mathcal{M}^{emp}, w_i \vDash_e \Gamma'$ for all $w_i \in W^{emp}$, and the model satisfies all probabilistic formulas occurring in Γ as for Definition 12.

Satisfiability clauses for formulas $fst(\langle \mathtt{T}, \mathtt{U} \rangle)$ and $snd(\langle \mathtt{T}, \mathtt{U} \rangle)$ can be defined in a similar vein to what done in Proposition 1 for their idealized counterparts.

Definition 14 *(Semantic consequence).* A statement $\mathtt{T} : \alpha$ is a semantic consequence of Γ, denoted by $\Gamma \vDash_e \mathtt{T} : \alpha$, if for any $w_i \in \mathcal{M}^{emp}$, $\mathcal{M}^{emp}, w_i \vDash_e \Gamma$ implies $\mathcal{M}^{emp}, w_i \vDash_e \mathtt{T} : \alpha$. A statement $\mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$ is a semantic consequence of Γ, denoted by $\Gamma \vDash_e \mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$, if $\mathcal{M}^{emp} \vDash \Gamma$ implies $\mathcal{M}^{emp} \vDash_e \mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$.

Similarly to the semantics provided in Definition 7, the following observation is useful for understanding its properties.

Observation 4. *For any $\mathcal{M}^{emp} = (W^{emp}, v^{emp})$, for any $\mathtt{T} : \alpha$, given any n there exists an f s.t. $\mathcal{M}^{emp} \vDash \mathtt{T}_{\{w', ..., w'^n\}} : \alpha_f$.*

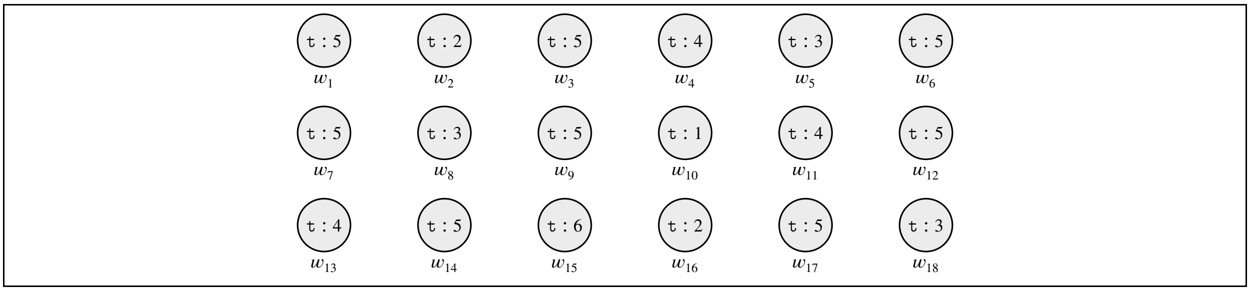

Fig. 3. Model $\mathcal{M}^{emp}$ for a system simulating 18 throws of a die.

Let us now reconsider Example 1 in terms of its empirical model.

Example 5. We construct an empirical model in which a process simulates the throw of a die 18 times ($\texttt{t}_{\{w_1,\dots,w_{18}\}}$) resulting in the output “1” one time, the output “3” three times, and the output “5” eight times (where the remaining outputs might be unknown or irrelevant to the current task). This amounts to the following structure of a $\mathcal{M}^{emp} = (W^{emp}, v^{emp})$, where $W^{emp} = \{w_1, \dots, w_{18}\}$, and there exists exactly one w_i s.t. $\mathcal{M}^{emp}, w_i \vDash_e \texttt{t} : 1$, there exist exactly three w_j s.t. $\mathcal{M}^{emp}, w_j \vDash_e \texttt{t} : 3$, and there exist exactly eight w_k s.t. $\mathcal{M}^{emp}, w_k \vDash_e \texttt{t} : 5$. An example of such a model is provided in Fig. 3.

We might then want to check whether this model evaluates to true the formula

$$\texttt{t}_{\{w_1,\dots,w_{18}\}} : (1+3)_{\frac{2}{9}}$$

i.e., that the frequency of output 1 or output 3 after 18 launches is 0.33. For this we check that $\mathcal{M}^{emp} \vDash_t \texttt{t}_{\{w_1,\dots,w_{18}\}} : 1_{\frac{1}{18}}$ (there is only 1 world, w_{10}, validating $\texttt{t} : 1$), $\mathcal{M}^{emp} \vDash_t \texttt{t}_{\{w_1,\dots,w_{18}\}} : 3_{\frac{1}{6}}$ (there are 3 worlds, w_5, w_8, and w_{18}, validating $\texttt{t} : 3$), and thus $\mathcal{M}^{emp} \vDash_e \texttt{t}_{\{w_1,\dots,w_{18}\}} : (1+3)_{\frac{1}{18}+\frac{1}{6}=\frac{2}{9}}$.

It is also possible that an agent considers less than 18 tests. For this case have $\mathcal{M}^{emp} \vDash_e \texttt{t}_{\{w_1,\dots,w_6\}} : 3_{\frac{1}{6}}$, where the worlds under consideration are only $w_1, w_2, w_3, w_4, w_5, w_6$. If an agent considers other worlds, for instance only $w_5, w_6, w_7, w_8, w_9, w_{10}$, we get $\mathcal{M}^{emp} \vDash_t \texttt{t}_{\{w_5,\dots,w_{10}\}} : 3_{\frac{1}{3}}$. This does not lead to a contradiction, but only indicates that the choice of the tests to consider influences the calculation of the frequency.

3.4. Models for evaluating trustworthiness

The trustworthiness of a non-deterministic process with respect to a given output in a trial is thus the result of comparing the evaluation of the frequency of that output in the empirical model representing the trial with respect to the expected probability as inferred by the theoretical model of the process. For this purpose, we introduce *joint models*, obtained by combining theoretical and empirical models.

Definition 15 *(Joint model).*

$$\mathcal{M} = (\mathcal{M}^{theor}, \mathcal{M}^{emp})$$

where $\mathcal{M}^{theor}$ is as by Definition 7, $\mathcal{M}^{emp}$ is as by Definition 11.

Definition 16 *(Validity in $\mathcal{M}$).*

1. $\mathcal{M}^{emp} \vDash_e \Gamma$ iff $\mathcal{M} \vDash \Gamma$, where Γ contains only expressions of $\mathcal{L}^{emp}$.
2. $\mathcal{M}^{theor} \vDash_t \Gamma$ iff $\mathcal{M} \vDash \Gamma$, where Γ contains only expressions of $\mathcal{L}^{theo}$.
3. $\mathcal{M} \vDash \texttt{T}_n : \alpha_{\tilde{a}}$ iff
   - $\mathcal{M}^{theor} \in \mathcal{M} \vDash_t \texttt{X}_\texttt{T} : \alpha_a$;
   - $\mathcal{M}^{emp} \in \mathcal{M} \vDash_e \texttt{T}_{\{w',\dots,w'^n\}} : \alpha_f$ and $n = \mid \{w', \dots, w'^n\} \mid$;
   - $\tilde{a} = a \cdot n$.
4. $\mathcal{M} \vDash [\texttt{X}]\texttt{T}_n : (\alpha \to \beta)_{\tilde{a}}$ iff whenever there exists $\mathcal{M}^{theor}_s = (W^{theor}_s, v^{theor}_s)$ which is a submodel of $\mathcal{M}^{theor}$ such that $W^{theor}_s = \{w_i \mid \mathcal{M}^{theor}, w_i \vDash_t \texttt{X} : \alpha\}$, we have $\mathcal{M}^{theor}_s \vDash_t \texttt{Y}_\texttt{T} : \beta_a$, $\mathcal{M}^{emp} \vDash_e \texttt{T}_{\{w',\dots,w'^n\}} : \beta_f$, $n = \mid \{w', \dots, w'^n\} \mid$, and $\tilde{a} = a \cdot n$.
5. $\mathcal{M} \vDash \mathit{Trust}(\texttt{T}_{\{w',\dots,w'^n\}} : \alpha_f)$ iff $\mathcal{M}^{theor} \vDash_t \texttt{X}_\texttt{T} : \alpha_a$, $\mathcal{M}^{emp} \vDash_e \texttt{T}_{\{w',\dots,w'^n\}} : \alpha_f$, $n = \mid \{w', \dots, w'^n\} \mid$, and $\mid a - f \mid \leq \epsilon(n)$, where $\epsilon(n)$ is a confidence interval.
6. $\mathcal{M} \vDash \mathit{UTrust}(\texttt{T}_{\{w',\dots,w'^n\}} : \alpha_f)$ iff $\mathcal{M}^{theor} \vDash_t \texttt{X}_\texttt{T} : \alpha_a$, $n = \mid \{w', \dots, w'^n\} \mid$, and $\mid a - f \mid > \epsilon(n)$, where $\epsilon(n)$ is a confidence interval.

Clauses 1 and 2 allow to state that everything valid in an empirical or theoretical model is valid in $\mathcal{M}$. Clause 3 introduces the expected probability of an output given n of executions of a corresponding process: $\mathtt{T}_n : \alpha_{\tilde{a}}$ is valid in $\mathcal{M}$ iff the theoretical probability of the corresponding random variable or idealized process $\mathtt{X}_\mathtt{T}$ to produce output α is a, the expected probability of α over n execution is $\tilde{a} = a \cdot n$. By Clause 4, $[\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{\tilde{b}}$ states the expected probability of output β by $\mathtt{T}$ over n executions when this is considered depending on the theoretical probability of a random variable $\mathtt{X}$ to have value α. By clause 5, $\mathit{Trust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f)$ is valid iff the theoretical probability $\mathtt{X}_\mathtt{T}$ associated with $\mathtt{T}$ producing output α is a, the actual frequency of α by $\mathtt{T}$ in a series of tests $w', \ldots, w'^n$ is f, and their absolute difference is within the confidence interval for n tests. The calculation of the threshold $\epsilon(n)$ depends on the field of study and the level of the required precision. For this reason we do not introduce it in its specific form in our semantics, but leave it as an external parameter, which can be adapted to a concrete example. By Clause 6, the *UTrust* operator is satisfied when the match between theoretical probability and the frequency of the observed output exceeds the given confidence interval. From this perspective *Trust* and *UTrust* are not complementary operators, and thus they express two exclusive, but not exhaustive properties: trustworthiness and untrustworthiness.

Similarly to Obs. 1 and 4, the following holds for $\mathcal{M}$:

Observation 5. *For any $\mathcal{M}$, for any $\mathtt{T} : \alpha$, for any n there exists $\widetilde{a}$ s.t. $\mathcal{M} \vDash \mathtt{T}_n : \alpha_{\widetilde{a}}$.*

Observation 6. *For any $\mathcal{M}$, for any $\mathtt{T} : \beta$ and $\mathtt{X} : \alpha$, for any n there exist a and $\widetilde{b}$ s.t. $\mathcal{M} \vDash [\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\widetilde{b}}$.*

By definitions of terms $\mathit{fst}(\langle \mathtt{T}, \mathtt{U} \rangle)$ and $\mathit{snd}(\langle \mathtt{T}, \mathtt{U} \rangle)$ denoting the first and the second process from a pair $\langle \mathtt{T}, \mathtt{U} \rangle$ and clause 6 in Definition 16, we have the following truth conditions for these expressions:

Proposition 3.

- *$\mathcal{M} \vDash \mathit{fst}(\langle \mathtt{T}, \mathtt{U} \rangle)_n : \alpha_{\widetilde{a}}$ iff there exist b and c such that $\mathcal{M} \vDash \langle \mathtt{T}, \mathtt{U} \rangle_{n \times m} : (\alpha \times \beta)_{\widetilde{c}}$, $\mathcal{M} \vDash \mathtt{U}_m : \beta_{\widetilde{b}}$, and $\widetilde{a} = \frac{\widetilde{c}}{\widetilde{b}}$.*
- *$\mathcal{M} \vDash \mathit{snd}(\langle \mathtt{T}, \mathtt{U} \rangle)_m : \beta_{\widetilde{b}}$ iff there exist a and c such that $\mathcal{M} \vDash \langle \mathtt{T}, \mathtt{U} \rangle_{n \times m} : (\alpha \times \beta)_{\widetilde{c}}$, $\mathcal{M} \vDash \mathtt{T}_n : \alpha_{\widetilde{a}}$, and $\widetilde{b} = \frac{\widetilde{c}}{\widetilde{a}}$*

From the definition of term $[\mathtt{X}]\mathtt{T}$ in Definition 16, we have the following condition for the evaluation of the application term $\mathtt{T}.\mathtt{U}$:

Proposition 4. *$\mathcal{M} \vDash \mathtt{T}_n.[\mathtt{U}_n : \alpha] : \beta_{\widetilde{c}}$ iff there exist a and b such that $\mathcal{M} \vDash [\mathtt{X}_u]\mathtt{T}_n : (\alpha \to \beta)_{[a]\widetilde{b}}$, $\mathcal{M} \vDash \mathtt{X}_\mathtt{U} : \alpha_a$, $\mathcal{M}_s^{theor} \vDash \mathtt{Y}_\mathtt{T} : \beta_b$ and $\widetilde{c} = \widetilde{a \cdot b}$.*

Proof. By definition of the application term, we need to define the expected probability of the output β to be produced by the process $\mathtt{T}$ executed n times under condition that $\mathtt{X}_\mathtt{U}$ produces α. By Observations 6 and 1 we have

1. $\mathcal{M} \vDash_t [\mathtt{X}_\mathtt{U}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\widetilde{b}}$;
2. $\mathcal{M} \vDash \mathtt{X}_\mathtt{U} : \alpha_a$.

Let us consider the model $\mathcal{M}_s^{theor} = (W_s^{theor}, v_s^{theor})$ such that $W_s^{theor} = \{w_i \mid \mathcal{M}^{theor}, w_i \vDash_t \mathtt{X}_\mathtt{U} : \alpha\}$. If it is not a submodel of $\mathcal{M}^{theor} \in \mathcal{M}$, then $a = 0$ and thus the expected probability of $\mathtt{T}_n.[\mathtt{U}_n : \alpha] : \beta$ is 0, i.e., $\widetilde{c} = \widetilde{a \cdot b}$.

If $\mathcal{M}_s^{theor}$ is a submodel of $\mathcal{M}$, then, by (1), we have that (3) $\mathcal{M}_s^{theor} \vDash_t \mathtt{Y}_\mathtt{T} : \beta_b$, (4) $\mathcal{M}^{emp} \vDash_e \mathtt{T}_{\{w',\ldots,w'^n\}} : \beta_f$, where $n = \mid \{w', \ldots, w'^n\} \mid$, and (5) $\tilde{b} = n \cdot b$. Let the size of $\mathcal{M}^{theor}$ be i. Then, from (2) we have $a = \frac{m}{i}$, where m is the size of $\mathcal{M}_s^{theor}$. From (3) we have that $b = \frac{j}{m}$, where j is the number of worlds of $\mathcal{M}_s^{theor}$ in which $\mathtt{Y}_\mathtt{T} : \beta$ holds. The dependent probability of $\mathtt{Y}_\mathtt{T} : \beta$ with respect to $\mathtt{X}_\mathtt{U} : \alpha$ in $\mathcal{M}^{theor}$ is thus $\frac{j}{i} = a \cdot b$. To calculate the expected probability of $\mathtt{T}$ producing β with respect to (4), we just need to multiply the theoretical probability and the number n, that is $\widetilde{c} = \frac{j}{i} \cdot n = a \cdot b \cdot n = \widetilde{a \cdot b}$. □

Definition 17 *(Semantic consequence).* A statement $\mathcal{X} \in \mathcal{L}$, is a semantic consequence of Γ, denoted by $\Gamma \vDash \mathcal{X}$, if $\mathcal{M} \vDash \Gamma$ implies $\mathcal{M} \vDash \mathcal{X}$.

Now, we are able to provide a full formal model for Example 1.

Example 6. Consider a non-deterministic system as in Ex. 1. Consider its theoretical model $\mathcal{M}^{theor}$ as defined in Fig. 1 and an empirical model of its execution $\mathcal{M}^{emp}$ as defined in Fig. 3. Then, $\mathcal{M} = (\mathcal{M}^{theor}, \mathcal{M}^{emp})$. In this model we have:

- $\mathcal{M}^{theor} \vDash \mathtt{x}_\mathtt{t} : 3_{\frac{1}{6}}$ and $\mathcal{M}^{emp} \vDash \mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 3_{\frac{3}{18}}$.

Consider that, even for a trustworthy process, the theoretical probability of output 3 to obtain and its frequency may diverge up to a point. This is expressed by fixing for example a 95% confidence level for $\epsilon(18)$ under the normal approximation to the binomial distribution, which results in the interval $[-0.0055, 0.3388]$. Clearly, $\frac{1}{6} \in [-0.0055, 0.3388]$, and thus:

- $\mathcal{M} \vDash Trust(\mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 3_{\frac{3}{18}})$.

As for the outputs “1” and “5” we have the following:

- $\mathcal{M}^{theor} \vDash \mathtt{x}_{\mathtt{t}} : 5_{\frac{1}{6}}$ and $\mathcal{M}^{emp} \vDash \mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 5_{\frac{8}{18}}$;
In this case $\epsilon(18) = [0.2149, 0.6740]$. Having $\frac{1}{6} \notin [0.2149, 0.6740]$ we conclude
- $\mathcal{M} \vDash UTrust(\mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 5_{\frac{8}{18}})$.
- $\mathcal{M}^{theor} \vDash \mathtt{x}_{\mathtt{t}} : 1_{\frac{1}{6}}$ and $\mathcal{M}^{emp} \vDash \mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 1_{\frac{1}{18}}$;
In this case $\epsilon(18) = [0.0502, 0.1614]$. Having $\frac{1}{6} \notin [0.0502, 0.1614]$ we conclude
- $\mathcal{M} \vDash UTrust(\mathtt{t}_{\{w_1,\ldots,w_{18}\}} : 1_{\frac{1}{18}})$.

More complex situations involving trustworthiness can be also represented via joint models. Let us borrow an example from D’Asaro et al. [13, Ex. 13, p. 28].

Example 7. Suppose we are given a commercial, closed-source software to automatically shortlist CVs according to a set of criteria. To this aim, we consider the output of the classification algorithm to fall into one of the following four categories: (1) male, shortlisted, (2) male, not shortlisted, (3) female, short-listed, (4) female, not shortlisted. In accordance with the gender distribution in Italian population, the percentage of female population is 51.28%. This means that we may take a theoretical probability of a process defined on gender distribution to choose female as 0.52. Then, the software has shortlisted 10 candidates, among which there were 3 females. Now we construct a model for evaluating trustworthiness of this software. Let $c : female$ stand for “the software c shortlisted a female candidate,” $x : female$ for “a process selects a female candidate,” and then $x_c : female$ stands for “the idealization of the software c selects female candidate.”

- $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$ s. t. $W^{theor} = \{w_1, \ldots, w_{50}\}$, and there exists exactly 26 w_i s.t. $w_i \in v^{theor}(x_c : female)$;
- $\mathcal{M}^{emp} = (W^{emp}, v^{emp})$ s.t. $W^{emp} = \{w_1, \ldots, w_{10}\}$, and there exists exactly 3 w_i s.t. $w_i \in v^{emp}(c : female)$;
- $\mathcal{M} = (\mathcal{M}^{theor}, \mathcal{M}^{emp})$.

In this model, we have:

- $\mathcal{M}^{theor} \vDash_t x_c : female_{0.52}$, that is the theoretical probability for selecting female candidate;
- $\mathcal{M}^{emp}_e \vDash c_{\{w_1,\ldots,w_{10}\}} : female_{0.3}$, that is the frequency of actually selected female candidates under 10 selected CVs.

Then, we fix a 95% confidence interval based on the number of samples and actual outputs “female”, that is $[0.0667, 0.6525]$. The theoretical probability $0,52$ falls under this interval, which indicates that the theoretical probability of selecting a female candidate matches the actual frequency over 10 samples. Thus, we have that the software c selecting 3 females among 10 samples is trustworthy:

- $\mathcal{M} \vDash Trust(c_{\{w_1,\ldots,w_{10}\}} : female_{0.3})$.

4. The proof theory of $\mathcal{L}$

In this section we briefly present the core of the proof-theoretical system TPTND (*Trustworthy Probabilistic Typed Natural Deduction*) introduced in [13] which, as it will be shown in Sections 5 and 6, is characterized by our semantics. Language $\mathcal{L}$ from Definition 6 interpreted on models $\mathcal{M}$ includes the following kind of formulas:

(S^X) $\mathtt{X} : \alpha$, which means that the idealized process $\mathtt{X}$ has output α.
(S^T) $\mathtt{T} : \alpha$, which means that the empirical process $\mathtt{T}$ has output α.
(S^{Xp}) $\mathtt{X} : \alpha_a$, which means that the theoretical probability of an idealized process $\mathtt{X}$ to have an output α is a.
(S^{Tf}) $\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f$, which means that after n executions $w', \ldots w'^n$ of a process $\mathtt{T}$, output α was produced with frequency f.
(S^{Te}) $\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}}$, where $\mathtt{X}/\mathtt{T}$ stands for $\mathtt{t}$, $\langle \mathtt{T}, \mathtt{U} \rangle$, $fst(\mathtt{T})$, $snd(\mathtt{T})$, or $[\mathtt{X}]\mathtt{T}$, which means that a possibly combined process $\mathtt{X}/\mathtt{T}_n$ produces an output α with the expected probability $\tilde{a}$ over n executions.
(S^{Tr}) $Trust(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f)$, which means that $\mathtt{T}$ producing α with a frequency f over n experiments $w', \ldots, w'^n$ is trustworthy.
(S^{UT}) $UTrust(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f)$, which means that $\mathtt{T}$ producing α with a frequency f over n experiments $w', \ldots, w'^n$ is untrustworthy.

The system TPTND contains four fragments:

- **Distribution construction rules**, which define contexts as lists of assumptions on the probability distributions of processes to have certain outputs (see Fig. 4).

$$\frac{}{\{\} :: \mathit{distribution}}\ \text{base}$$

$$\frac{\Gamma :: \mathit{distribution} \qquad x : \alpha_a \notin \Gamma, \forall a \in [0,1] \qquad \sum_{x:\rho_p \in \Gamma} p \leq 1}{\Gamma, x : \alpha_a :: \mathit{distribution}}\ \text{extend}$$

$$\frac{\Gamma :: \mathit{distribution} \qquad x : \alpha \notin \Gamma}{\Gamma, x : \alpha :: \mathit{distribution}}\ \text{extend det}$$

$$\frac{\alpha :: \mathit{output} \qquad \dots \qquad \omega :: \mathit{output}}{\{x : \alpha_{[0,1]}, \dots, x : \omega_{[0,1]}\} :: \mathit{distribution}}\ \text{unknown}$$

Fig. 4. Distribution construction.

$$\frac{}{\Gamma, x : \alpha \vdash x : \alpha_1}\ \text{identity}_1$$

$$\frac{}{\Gamma, x : \alpha_a \vdash x : \alpha_a}\ \text{identity}_2$$

$$\frac{\Gamma \vdash x : \alpha_a}{\Gamma \vdash x : \neg\alpha_{1-a}}\ \bot$$

$$\frac{\Gamma \vdash x : \alpha_a \qquad \Delta \vdash y : \beta_b \qquad \Gamma \perp\!\!\!\perp \Delta}{\Gamma, \Delta \vdash \langle x, y\rangle : (\alpha \times \beta)_{a \cdot b}}\ \text{I}\times$$

$$\frac{\Gamma \vdash x : (\alpha \times \beta)_c \qquad \Gamma \vdash \mathit{fst}(x) : \alpha_a}{\Gamma \vdash \mathit{snd}(x) : \beta_{c/a}}\ \text{E}\times_L$$

$$\frac{\Gamma \vdash x_n : (\alpha \times \beta)_c \qquad \Gamma \vdash \mathit{snd}(x) : \beta_b}{\Gamma \vdash \mathit{fst}(x) : \alpha_{c/b}}\ \text{E}\times_R$$

$$\frac{\Gamma \vdash x : \alpha_a \qquad \Gamma \vdash x : \beta_b}{\Gamma \vdash x : (\alpha + \beta)_{a+b}}\ \text{I}+$$

$$\frac{\Gamma \vdash x : (\alpha + \beta)_a \qquad \Gamma \vdash x : \alpha_{a'}}{\Gamma \vdash x : \beta_{a-a'}}\ \text{E}+_R$$

$$\frac{\Gamma \vdash x : (\alpha + \beta)_b \qquad \Gamma \vdash x : \beta_{b'}}{\Gamma \vdash x : \alpha_{b-b'}}\ \text{E}+_L$$

$$\frac{\Gamma, x : \alpha \vdash y : \beta_b}{\Gamma \vdash [x]y : (\alpha \to \beta)_b}\ \text{I}\to$$

$$\frac{\Gamma \vdash [x]y : (\alpha \to \beta)_b \qquad \vdash x : \alpha_a}{\Gamma \vdash y.(x : \alpha) : \beta_{a \cdot b}}\ \text{E}\to$$

where we assume, for $E+_L$, $E+_R$, $E\times_L$, $E\times_R$, that $\alpha \neq \beta$

Fig. 5. Rules for random variables.

- **Rules for random variables**, which define operations on the theoretical probability of random variables to produce outputs (see Fig. 5).[5]
- **Sampling rules**, which define operations on observed frequencies of processes to produce outputs and correlate them with their expected probabilities (see Fig. 6).
- **Trust fragment**, which defines a procedure for decision of whether a process is trustworthy (see Fig. 7).

The expressions of the form '$t_n : \alpha_f$' in the rules of TPTND are actually expressions of the form '$t_{w',...,w'^n} : \alpha_f$.' We prefer to keep the original notation of [13], considering that the listing of tests (i.e., $\{w', ..., w'^n\}$) is not crucial for the calculus, but rather the number of these tests (i.e., n). For instance, the rule update (see Fig. 6) provides a clause for merging two series of tests: $\Gamma \vdash t_n : \alpha_f$ and $\Gamma \vdash t_m : \alpha_{f'}$ in the premise, indicating two distinct series of tests, include the case where the number of tests is equal ($m = n$), but the listings are distinct ($\{w', ..., w'^m\} \neq \{w'', ..., w''^n\}$).

Note that in TPTND the right-hand side of a formula is always defined by expressions of the form (S^{Xp}), (S^{Tf}), (S^{Te}), (S^{Tr}), and (S^{UT}), i.e., expressions about theoretical probabilities, expected probabilities and frequencies. Expressions of the form (S^X) and (S^T) are matched in $TPTND$ by formulas where $\mathtt{X} : \alpha$ and $\mathtt{T} : \alpha$ stand for $\mathtt{X} : \alpha_1$ and $\mathtt{T}_n : \alpha_n$, respectively. This provides us with the rules from Fig. 8 and Fig. 9. We rename here the assignment of deterministic values to random variables and the single experiment rules from TPTND with respectively $1x$ and $1t$ annotations.

[5] The side condition $\Gamma \perp\!\!\!\perp \Delta$ in the rule $I\times$ states the independence of the distributions of Γ and Δ, or, in other words, that for all formulas $\mathcal{X}, \mathcal{Y}$, such that $\mathcal{X} \in \Gamma$ and $\mathcal{Y} \in \Delta$, it is not the case that $\Gamma \vdash \mathcal{Y}$ and $\Delta \vdash \mathcal{X}$.

$$\frac{}{x_t : \alpha_a \vdash t_n : \alpha_{\tilde{a}}}\ \text{expectation}$$

$$\frac{\Gamma \vdash t^1 : \alpha^1 \qquad \dots \qquad \Gamma \vdash t^n : \alpha^n}{\Gamma \vdash t_n : \alpha_f}\ \text{sampling}$$

where $f = \frac{|\{i|\alpha^i=\alpha\}|}{n}$

$$\frac{\Gamma \vdash t_n : \alpha_f \qquad \Gamma \vdash t_m : \alpha_{f'}}{\Gamma \vdash t_{n+m} : \alpha_{f\cdot(n/(n+m))+f'\cdot(m/(n+m))}}\ \text{update}$$

$$\frac{\Gamma \vdash t_n : \alpha_{\tilde{a}} \qquad \Gamma \vdash t_n : \beta_{\tilde{b}}}{\Gamma \vdash t_n : (\alpha+\beta)_{\tilde{a}+\tilde{b}}}\ \text{I}+$$

$$\frac{\Gamma \vdash t_n : (\alpha+\beta)_{\tilde{c}} \qquad \Gamma \vdash t_n : \alpha_{\tilde{a}}}{\Gamma \vdash t_n : \beta_{\tilde{c}-\tilde{a}}}\ \text{E}+_L$$

$$\frac{\Gamma \vdash t_n : (\alpha+\beta)_{\tilde{c}} \qquad \Gamma \vdash t_n : \beta_{\tilde{b}}}{\Gamma \vdash t_n : \beta_{\tilde{c}-\tilde{b}}}\ \text{E}+_R$$

$$\frac{\Gamma \vdash t_n : \alpha_{\tilde{a}} \qquad \Delta \vdash u_n : \beta_{\tilde{b}} \qquad \Gamma \perp\!\!\!\perp \Delta}{\Gamma, \Delta \vdash \langle t, u\rangle_n : (\alpha\times\beta)_{\widetilde{a\cdot b}}}\ \text{I}\times$$

$$\frac{\Gamma \vdash t_n : (\alpha\times\beta)_{\tilde{c}} \qquad \Gamma \vdash \mathit{fst}(t)_n : \alpha_{\tilde{a}}}{\Gamma \vdash \mathit{snd}(t)_n : \beta_{\widetilde{(\frac{c}{a})}}}\ \text{E}\times_L$$

$$\frac{\Gamma \vdash t_n : (\alpha\times\beta)_{\tilde{c}} \qquad \Gamma \vdash \mathit{snd}(t)_n : \beta_{\tilde{b}}}{\Gamma \vdash \mathit{fst}(t)_n : \alpha_{\widetilde{(\frac{c}{b})}}}\ \text{E}\times_R$$

$$\frac{\Gamma, x_t : \alpha_a \vdash t_n : \beta_{\tilde{b}}}{\Gamma \vdash [x]t_n : (\alpha\to\beta)_{[a]\tilde{b}}}\ \text{I}\to$$

$$\frac{\Gamma \vdash [x]t_n : (\alpha\to\beta)_{[a]\tilde{b}} \qquad y_u : \alpha_a \vdash u_n : \alpha_{\tilde{a}}}{\Gamma \vdash t_n.[u_n : \alpha] : \beta_{\widetilde{a\cdot b}}}\ \text{E}\to$$

Fig. 6. Sampling rules.

$$\frac{\Gamma \vdash x : \alpha_a \qquad \Delta \vdash u_n : \alpha_f \qquad |\, a - f \,| \leq \epsilon(n)}{\Gamma, \Delta \vdash \mathit{Trust}(u_n : \alpha_f)}\ \text{IT}$$

$$\frac{\Gamma \vdash \mathit{Trust}(u_n : \alpha_f)}{\Gamma, x_u : \alpha_{[a-\epsilon(n), a+\epsilon(n)]} \vdash u_n : \alpha_f}\ \text{ET}$$

$$\frac{\Gamma \vdash x : \alpha_a \qquad \Delta \vdash u_n : \alpha_f \qquad |\, a - f \,| > \epsilon(n)}{\Gamma, \Delta \vdash \mathit{UTrust}(u_n : \alpha_f)}\ \text{IUT}$$

$$\frac{\Gamma \vdash \mathit{UTrust}(u_n : \alpha_f)}{\Gamma, x_u : \alpha_{[0,1]-[a-\epsilon(n), a+\epsilon(n)]} \vdash u_n : \alpha_f}\ \text{EUT}$$

Fig. 7. Trust fragment.

$$\frac{}{x : \alpha \vdash x : \alpha}\ \text{1x identity}$$

$$\frac{\Gamma \vdash x : \alpha \qquad \Delta \vdash y : \beta}{\Gamma, \Delta \vdash \langle x, y\rangle : (\alpha\times\beta)}\ \text{1x I}\times$$

$$\frac{\Gamma \vdash x : (\alpha\times\beta)}{\Gamma \vdash \mathit{fst}(x) : \alpha}\ \text{1x E}\times_L \quad \frac{\Gamma \vdash x : (\alpha\times\beta)}{\Gamma \vdash \mathit{snd}(x) : \beta}\ \text{1x E}\times_R$$

$$\frac{\Gamma \vdash x : \alpha}{\Gamma \vdash x : (\alpha+\beta)}\ \text{1x I}+$$

$$\frac{\Gamma \vdash x : (\alpha+\beta) \qquad \Gamma \vdash x : \neg\alpha}{\Gamma \vdash x : \beta}\ \text{1x E}+_R$$

$$\frac{\Gamma \vdash x : (\alpha+\beta) \qquad \Gamma \vdash x : \neg\beta}{\Gamma \vdash x : \alpha}\ \text{1x E}+_L$$

Fig. 8. Deterministic rules for $\mathtt{X} : \alpha$.

$$\frac{}{\Gamma, x_t : \alpha_a \vdash t : \alpha}\ \text{1t experiment}$$

$$\frac{\Gamma \vdash t : \alpha \qquad \Delta \vdash u : \beta \qquad \Gamma \perp\!\!\!\perp \Delta}{\Gamma, \Delta \vdash \langle t, u \rangle : (\alpha \times \beta)}\ \text{1t I}\times$$

$$\frac{\Gamma \vdash t : (\alpha \times \beta)}{\Gamma \vdash \mathit{fst}(t) : \alpha}\ \text{1t E}\times_L \quad \frac{\Gamma \vdash t : (\alpha \times \beta)}{\Gamma \vdash \mathit{snd}(t) : \beta}\ \text{1t E}\times_R$$

$$\frac{\Gamma, x_t : \alpha_a, x_t : \beta_b \vdash t : \alpha}{\Gamma \vdash t : (\alpha + \beta)}\ \text{1t I+}$$

$$\frac{\Gamma \vdash t : (\alpha + \beta) \qquad \Gamma \vdash t : \neg\alpha}{\Gamma \vdash t : \beta}\ \text{1t E+}_R$$

$$\frac{\Gamma \vdash t : (\alpha + \beta) \qquad \Gamma \vdash t : \neg\beta}{\Gamma \vdash t : \alpha}\ \text{1t E+}_L$$

$$\frac{\Gamma, x : \alpha_a \vdash t : \beta}{\Gamma \vdash [x]t : (\alpha \to \beta)_a} \qquad \frac{\Gamma \vdash [x]t : (\alpha \to \beta)_a \qquad \vdash u : \alpha}{\Gamma \vdash t.[u : \alpha] : \beta}$$

Fig. 9. Single-experiment rules.

5. Soundness

In this section we prove that TPTND is sound with respect to our semantics. We first show that the construction rules for the probability distribution are sound with respect to the properties of our models. Then for any formula $\mathcal{X}$ derived under such a probability distribution Γ, if $\mathcal{X}$ is derived either by a logical rule applied to an assignment of probability to a random variable, or applied to an assignment of frequency to a given number of processes, or finally obtained by a rule of the trust fragment, such a formula is valid in the joint model.

Theorem 1. *The rules base, extend, extend det, and unknown (see Fig. 4) are sound with respect to $\mathcal{M}$ models.*

Proof. The rule base, which introduces distributions, is semantically modeled by the fact that $\mathcal{M}$ contains $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$, where W^{theor} is a non-empty set, see Definition 7.

The rule extend and extend det claim that any distribution can be extended with a random variable assigning a theoretical probability to outputs as long as this extension respects the standard additivity on probabilities. Semantically this fact is preserved by Observations 1 and 2, which claim that any expression $\mathtt{X} : \alpha$ has a probability attached and that this probability does not exceed 1.

The rule unknown introduces unknown distributions, i.e., lists of random variables for which the full interval is given, which is assured by Observations 1 and 3, that is, any expression $\mathtt{X} : \alpha$ has a probability which is in the interval of $[0, 1]$. □

Now, let us show that TPTND is sound with respect to $\mathcal{M}$ models.

Theorem 2 *(Soundness). If $\Gamma \vdash \mathcal{X}$ then $\Gamma \vDash \mathcal{X}$, where $\mathcal{X} \in \mathcal{L}$.*

Proof. Let $\Gamma \vdash \mathcal{X}$ and $\mathcal{M} \vDash \Gamma$. We prove the theorem by induction on the derivation of $\Gamma \vdash \mathcal{X}$, that is, on structure of the term $\mathcal{X}$.

1. If $\mathcal{X}$ is of a form $x : \alpha$, then $\mathcal{X}$ can be obtained by one of the following rules:
   - identity$_1$. In this case $\mathcal{X}$ is of the form $x : \alpha_1$. From the assumption we have $\mathcal{M} \vDash x : \alpha$, which means that $x : \alpha$ is satisfied in each world of $\mathcal{M}^{theor}$. Thus, $\mid W^{theor} \mid = n$; $n = \mid \{w_i \in W^{theor} \mid \mathcal{M}^{theor}, w_i \vDash_t x : \alpha\} \mid$, $\frac{n}{n} = 1$, which means, by Definition 8, that $\mathcal{M}^{theor} \vDash_t x : \alpha_1$, and, by Definition 16, that $\mathcal{M} \vDash x : \alpha_1$.
   - identity$_2$. $\mathcal{X}$ is of a form $x : \alpha_a$. By assumption we have $\mathcal{M} \vDash x : \alpha_a$.
   - $\bot$. Then, $\mathcal{X}$ is of a form $x : \neg\alpha_{1-a}$. By our assumption and induction hypothesis, we have $\mathcal{M} \vDash x : \alpha_a$, which means $\mathcal{M}^{theor} \vDash_t x : \alpha_a$. Then, by Definition 8, $\mathcal{M}^{theor} \vDash_t x : \neg\alpha_{1-a}$, which means, by Definition 16, that $\mathcal{M} \vDash x : \neg\alpha_{1-a}$.
   - $I+$. In this case $\mathcal{X}$ is of the form $x : (\alpha + \beta)_{a+b}$. Similarly to the previous case, we have $\mathcal{M}^{theor} \vDash_t x : \alpha_a$ and $\mathcal{M}^{theor} \vDash_t x : \beta_b$. Then, by Definition 8, $\mathcal{M}^{theor} \vDash x : (\alpha + \beta)_{a+b}$, that is $\mathcal{M} \vDash x : (\alpha + \beta)_{a+b}$.
   - $E+_R$. Thus, $\mathcal{X}$ is of the form $x : \beta_{a-a'}$. Similarly to the previous cases, we have (i) $\mathcal{M}^{theor} \vDash_t x : (\alpha + \beta)_a$ and (ii) $\mathcal{M}^{theor} \vDash_t x : \alpha_{a'}$. By Definition 8, we have $\mathcal{M}^{theor} \vDash_t x : \beta_{a-a'}$.
   - $E+_L$. This case is similar to the previous one.
2. If $\mathcal{X}$ is of the form $\langle x, y \rangle : (\alpha \times \beta)_{a \cdot b}$, then it is obtained by the rule $I\times$. By our assumption and induction hypothesis, we have $\mathcal{M}^{theor} \vDash_t x : \alpha_a$ and $\mathcal{M}^{theor} \vDash y : \beta_b$. The side condition in the rule $\Gamma \perp\!\!\!\perp \Delta$ for the independence of the distributions corresponds to the condition on the model that $x : \alpha_a$ can be valuated in $\mathcal{M}^{theor}$ without $y : \beta_b$, and vice versa $y : \beta_b$ can be valuated in $\mathcal{M}^{theor}$ without $x : \alpha_a$. By Definition 8, we have $\mathcal{M}^{theor} \vDash_t \langle x, y \rangle : (\alpha \times \beta)_{a \cdot b}$, and thus $\mathcal{M} \vDash \langle x, y \rangle : (\alpha \times \beta)_{a \cdot b}$.

3. If $\mathcal{X}$ is of the form $fst(x) : \alpha_{\frac{c}{b}}$, then it is obtained from the rule $E\times_R$. By our assumption and induction hypothesis, we have $\mathcal{M}^{theor} \vDash_t x : (\alpha \times \beta)_c$ and $\mathcal{M}^{theor} \vDash_t snd(x) : \beta_b$. By Definition 8, this means that $\mathcal{M}^{theor} \vDash_t x : \alpha_{\frac{c}{b}}$. By Proposition 1, we have $\mathcal{M}^{theor} \vDash_t fst(x) : \alpha_{\frac{c}{b}}$, and thus $\mathcal{M} \vDash fst(x) : \alpha_{\frac{c}{b}}$.
4. If $\mathcal{X}$ is of the form $snd(x) : \beta_{\frac{c}{b}}$, then it is obtained from the rule $E\times_L$. The proof is similar to the case of $fst(x)$.
5. If $\mathcal{X}$ is of the form $[x]y : (\alpha \to \beta)_a$, then it is obtained from the rule $I \to$. By our assumption and induction hypothesis, we have that if $\mathcal{M}^{theor} \vDash x : \alpha$, then $\mathcal{M}^{theor} \vDash y : \beta_a$. Let us consider the submodel $\mathcal{M}^{theor}_s$ of $\mathcal{M}^{theor}$ such that $W^{theor}_s = \{w_i \mid \mathcal{M}^{theor}, w_i \vDash_t x : \alpha\}$. It is clear that $\mathcal{M}^{theor}_s \vDash x : \alpha$, and thus $\mathcal{M}^{theor}_s \vDash_t y : \beta_a$, that is the definition of $\mathcal{M}^{theor} \vDash_t [x]y : (\alpha \to \beta)_b$, thus $\mathcal{M} \vDash [x]y : (\alpha \to \beta)_b$.
6. If $\mathcal{X}$ is of the form $y.(x : \alpha)\beta_{a\cdot b}$, then it is obtained from the rule $E \to$. By our assumption and induction hypothesis, this means that (1) $\mathcal{M}^{theor} \vDash_t [x]y(\alpha \to \beta)_b$ and (2) $\mathcal{M}^{theor} \vDash_t x : \alpha_a$. From (1) we have that $\mathcal{M}^{theor}_s \vDash y : \beta_b$, where $\mathcal{M}^{theor}_s$ is a submodel of $\mathcal{M}^{theor}$ s.t. $W^{theor}_s = \{w_i \in W^{theor} \mid \mathcal{M}^{theor}, w_i \vDash x : \alpha\}$. Then, the probability of $y : \beta$ in the same worlds in which x produces α is a multiplication of the probability of α and of $(\alpha \to \beta)$, that is, $a \cdot b$. By Proposition 2, this means that $\mathcal{M}^{theor} \vDash_t y.(x : \alpha)\beta_{a\cdot b}$, that is $\mathcal{M} \vDash y.(x : \alpha)\beta_{a\cdot b}$.
7. If $\mathcal{X}$ is of the form $t : \alpha$, then it can be obtained by one of the following rules:
   (a) expectation. In this case $\mathcal{X}$ is of the form $t_n : \alpha_{\widetilde{a}}$. Let $\mathcal{M} \vDash x_t : \alpha_a$. Then, $\mathcal{M}^{theor} \vDash_t x_t : \alpha_a$ s.t. $\mathcal{M}^{theor} \in \mathcal{M}$. For any $t : \alpha$, we have $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : \alpha_f$ s.t. $n =\mid \{w',\dots,w'^n\} \mid$ and $\mathcal{M}^{emp} \in \mathcal{M}$. Then, the expected probability of $t : \alpha$ after n executions of the process t is $a \cdot n$. Thus, $\mathcal{M} \vDash t_n : \alpha_{\widetilde{a}}$.
   (b) sampling. In this case $\mathcal{X}$ is of the form $t_n : \alpha_f$, where $f = \frac{|\{i|\alpha^i=\alpha\}|}{n}$. By our assumption and induction hypothesis, if $\mathcal{M} \vDash \Gamma, x_t : \alpha_a$ then $\mathcal{M} \vDash t^1 : \alpha^1$, …, if $\mathcal{M} \vDash \Gamma, x_t : \alpha_a$ then $\mathcal{M} \vDash t^n : \alpha^n$, and $\mathcal{M} \vDash \Gamma, x_t : \alpha_a$. Having in mind that the notation $t^1 \dots t^n$ indicates distinct launching of the process t, we have that there exists $\mathcal{M}^{emp}_n = (W^{emp}_n, v^{emp}_n)$ which is a submodel of $\mathcal{M}^{emp}$ s.t. $\mid W^{emp}_n \mid= n$. Among these worlds, there must be $f = \frac{|\{w_i|\mathcal{M}^{emp}_n, w_i \vDash_e t:\alpha\}|}{n}$, which is exactly $f = \frac{|\{i|\alpha^i=\alpha\}|}{n}$. Thus, $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : \alpha_f$, and then $\mathcal{M} \vDash t_{\{w',\dots,w'^n\}} : \alpha_f$.
   (c) update. In this case $\mathcal{X}$ is of the form $t_{n+m} : \alpha_{f\cdot(n/(n+m))+f'\cdot(m/(n+m))}$. As before we have $\mathcal{M} \vDash t_{\{w',\dots,w'^n\}} : \alpha_f$ and $\mathcal{M} \vDash t_{\{w'',\dots,w''^m\}} : \alpha_{f'}$, i.e., $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : \alpha_f$, $\mathcal{M}^{emp} \vDash_e t_{\{w'',\dots,w''^m\}} : \alpha_{f'}$, $n =\mid \{w',\dots,w'^n\} \mid$, and $m =\mid \{w'',\dots,w''^n\} \mid$. By Definitions 12 and 16, this means that there exist submodel $\mathcal{M}^{emp}_s = (W^{emp}_s, v^{emp}_s)$ s.t. $\mid W^{emp}_s \mid= n$ and submodel $\mathcal{M}^{emp}_{s'} = (W^{emp}_{s'}, v^{emp}_{s'})$ s.t. $\mid W^{emp}_{s'} \mid= m$. In the rule update, the premises are obtained in different branches, that is, they are distinct. This means that W^{emp}_s and $W^{emp}_{s'}$ do not share any worlds. Consider $\mathcal{M}^{emp}_{s+s'} = (W^{emp}_{s+s'}, v^{emp}_{s+s'})$, where $W^{emp}_{s+s'} = W^{emp}_s \cup W^{emp}_{s'}$, and $v^{emp}_{s+s'}$ is v^{emp} restricted to the worlds of $W^{emp}_{s+s'}$. The size of $\mathcal{M}^{emp}_{s+s'}$ is $n+m$, and, by standard math, the frequency of $t : \alpha$ is $f'' = f \cdot (n/(n+m)) + f' \cdot (m/(n+m))$, which means that $\mathcal{M}^{emp} \vDash t_{\{w',\dots,w^n\}\cup\{w'',\dots,w''^m\}} : \alpha_{f\cdot(n/(n+m))+f'\cdot(m/(n+m))}$.
   (d) $I+$. In this case $\mathcal{X}$ is of the form $t_n : (\alpha+\beta)_{\widetilde{a}+\widetilde{b}}$. By our hypothesis, we have that if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash t_n : \alpha_{\widetilde{a}}$, if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash t_n : \beta_{\widetilde{b}}$, and $\mathcal{M} \vDash \Gamma$. Thus, $\mathcal{M}^{theor} \vDash_t x_t : \alpha_a$ and $\mathcal{M}^{theor} \vDash_t x_t : b$, that is $\mathcal{M}^{theor} \vDash_t x_t : (\alpha+\beta)_{a+b}$. For any process we have $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : (\alpha+\beta)_f$ and $n =\mid \{w',\dots,w'^n\} \mid$. Thus, $\mathcal{M} \vDash t_n : (\alpha+\beta)_{\widetilde{c}}$, where $\widetilde{c} = (a+b)\cdot n = a\cdot n + b\cdot n = \widetilde{a}+\widetilde{b}$.
   (e) $E+_L$. In this case $\mathcal{X}$ is of the form $t_n : \beta_{\widetilde{c}-\widetilde{a}}$. By our assumption, if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash t_n : (\alpha+\beta)_{\widetilde{c}}$, if $\mathcal{M} \vDash t_n : \alpha_{\widetilde{a}}$, and $\mathcal{M} \vDash \Gamma$. Thus, $\mathcal{M}^{theor} \vDash_t x_t : (\alpha+\beta)_c$, $\mathcal{M}^{theor} \vDash_t x_t : \alpha_a$, and $\mathcal{M}^{theor} \vDash_t x_t : \beta_{c-a}$. We have $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : \beta_f$, $n =\mid \{w',\dots,w'^n\} \mid$ and thus $\mathcal{M} \vDash t_n : \beta_{\widetilde{b}}$, where $\widetilde{b} = (c-a)\cdot n = c\cdot n - a\cdot n = \widetilde{c}-\widetilde{a}$.
   (f) $E+_R$. This case is similar to $E+_L$.
   (g) IT. In this case $\mathcal{X}$ is of the form $t_n : \alpha_f$. By induction hypothesis, we have that if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash Trust(t_{\{w',\dots,w'^n\}} : \alpha_f)$, and $\mathcal{M} \vDash \Gamma$. Thus, $\mathcal{M} \vDash Trust(t_{\{w',\dots,w'^n\}} : \alpha_f)$, and then by Definition 16, $\mathcal{M} \vDash t_{\{w',\dots,w'^n\}} : \alpha_f$.
   (h) EUT. This case is similar to the case IT.
8. If $\mathcal{X}$ is of the form $\langle t,u\rangle_n : (\alpha\times\beta)_{\widetilde{a\cdot b}}$, then it is obtained from the rule I ×. By our hypothesis, we have if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash t_n : \alpha_{\widetilde{a}}$, if $\mathcal{M} \vDash \Delta$ then $\mathcal{M} \vDash u_n : \beta_{\widetilde{b}}$, $\mathcal{M} \vDash \Gamma$, and $\mathcal{M} \vDash \Delta$. Thus, as in previous cases we have $\mathcal{M}^{theor} \vDash_t x_t : \alpha_a$ and $\mathcal{M}^{theor} \vDash_t x_u : \beta_b$. Thus, $\mathcal{M}^{theor} \vDash_t \langle x_t, x_u\rangle : (\alpha\times\beta)_{a\cdot b}$. Having $\mathcal{M} \vDash \langle t,u\rangle_{\{w,\dots,w^n\}} : (\alpha\times\beta)_f$ (and thus $\mid \{w,\dots,w^n\} \mid= n$), we have $\mathcal{M} \vDash \langle t,u\rangle_n : (\alpha\times\beta)_{\widetilde{c}}$, where $\widetilde{c} = n\cdot a\cdot b = \widetilde{a\cdot b}$.
9. If $\mathcal{X}$ is of the form $snd(t)_n : \beta_{\widetilde{(\frac{c}{a})}}$ then it is obtained from the rule E$\times_L$. By our hypothesis, we have if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash t_n : (\alpha\times\beta)_{\widetilde{c}}$, if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash fst(t)_n : \alpha_{\widetilde{a}}$ and $\mathcal{M} \vDash \Gamma$. Then, we have $\mathcal{M}^{theor} \vDash_t x_t : (\alpha\times\beta)_c$ and $\mathcal{M}^{theor} \vDash_t fst(x_t) : \alpha_a$. Thus, $\mathcal{M}^{theor} \vDash_t snd(x_t) : \beta_{\frac{c}{a}}$. Having $\mathcal{M}^{emp} \vDash_e snd(t)_{\{w,\dots,w^n\}} : \beta_f$ and $n =\mid \{w,\dots,w^n\} \mid$, we obtain $\mathcal{M} \vDash snd(t)_n : \beta_{\widetilde{b}}$, where $\widetilde{b} = n \cdot \frac{c}{a} = \widetilde{(\frac{c}{a})}$.
10. If $\mathcal{X}$ is of the form $fst(x)_{\frac{n}{m}}$ then it is obtained from the rule E $\times_R$. This case is similar to the previous one.
11. If $\mathcal{X}$ is of the form $[x]t_n : (\alpha\to\beta)_{[a]\widetilde{b}}$, then it is obtained from the rule I →. By our hypothesis, if $\mathcal{M} \vDash \Gamma$ and $\mathcal{M} \vDash x_t : \alpha_a$ then $\mathcal{M} \vDash t_n : \beta_{\widetilde{b}}$, and $\mathcal{M} \vDash \Gamma$. Then, $\mathcal{M}^{theor} \vDash_t x_t : \beta_b$. Consider $\mathcal{M}^{theor}_s = (W^{theor}_s, v^{theor}_s)$ which is a submodel of $\mathcal{M}^{theor}$ s. t. $W^{theor}_s = \{w_i \mid \mathcal{M}^{theor}, w_i \vDash x_t : \alpha\}$. In $\mathcal{M}^{theor}_s$ we have $\mathcal{M}^{theor}_s \vDash_t y_t : \beta_{[a]b}$, where $[a]b$ stands for the probability of β in the worlds validating $x_t : \alpha$, and thus $\mathcal{M}^{theor} \vDash_t [x_t]y_t(\alpha\to\beta)_{[a]b}$. From $\mathcal{M} \vDash t_{\{w',\dots,w'^n\}} : \beta_b$ we know that the size of $\mathcal{M}^{emp}$ is sufficient for $\mathcal{M} \vDash [x]t_{\{w',\dots,w'^n\}} : (\alpha\to\beta)_f$ for some f, and thus $\mathcal{M} \vDash [x]t_n : (\alpha\to\beta)_{[a]\widetilde{b}}$, where $[a]\widetilde{b}$ stands for $b\cdot n$ under probability a of $x_t : \alpha$.
12. If $\mathcal{X}$ is of the form $t_n.[u_n : \alpha] : \beta_{\widetilde{a\cdot b}}$, then it is obtained from the rule E →. By our hypothesis we have if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash [x]t_n : (\alpha\to\beta)_{[a]\widetilde{b}}$, if $\mathcal{M} \vDash y_u : \alpha_a$ then $\mathcal{M} \vDash u_n : \alpha_{\widetilde{a}}$, and $\mathcal{M} \vDash \Gamma$. By having $\mathcal{M} \vDash [x]t_n : (\alpha\to\beta)_{[a]\widetilde{b}}$, we have that for $\mathcal{M}^{theor}_s = (W^{theor}_s, v^{theor}_s)$ that is a submodel of $\mathcal{M}^{theor}$ s.t. $W^{theor}_s = \{w_i \mid \mathcal{M}^{theor}, w_i \vDash x : \alpha\}$, $\mathcal{M}^{theor}_s \vDash_t y_t : \beta_b$, $\mathcal{M}^{emp} \vDash_e t_{\{w',\dots,w'^n\}} : \beta_f$,

$n =\mid \{w', ..., w'^n\} \mid$, and $\tilde{b} = n \cdot b$. Taking into account the size of $\mathcal{M}^{theor}_s$ and $\mathcal{M}^{theor}$, the expected frequency of $t[u : \alpha]\beta$ after n executions is $c = a \cdot b$, and thus $\mathcal{M} \vDash t_n[u_n : \alpha_a] : \beta_{\widetilde{c}}$, where $\widetilde{c} = \widetilde{a \cdot b}$.

13. If $\mathcal{X}$ is of the form *Trust*$(t_n : \alpha_f)$, then it is obtained from the rule IT. By our hypothesis we have if $\mathcal{M} \vDash \Gamma$ then $\mathcal{M} \vDash x : \alpha_a$; if $\mathcal{M} \vDash \Delta$ then $\mathcal{M} \vDash u_{\{w',...,w'^n\}} : \alpha_f$, $\mid a - f \mid \leq \epsilon(n)$; and $\mathcal{M} \vDash \Gamma$, $\mathcal{M} \vDash \Delta$. Thus we have $\mathcal{M}^{theor} \vDash x : \alpha_a$, $\mathcal{M}^{emp} \vDash u_{\{w',...,w'^n\}} : \alpha_f$, and $\mid a - f \mid \leq \epsilon(n)$, that is $\mathcal{M} \vDash$ *Trust*$(t_{\{w',...,w'^n\}} : \alpha_f)$.
14. If $\mathcal{X}$ is of the form *UTrust*$(t_n : \alpha_f)$, then it is obtained from the rule IUT. This case is similar to the previous one. □

6. Completeness

We prove completeness of TPTND by constructing a canonical model. The peculiar feature of this model is that first we define distinct theories for each kind of formula, and then we take their union. The completeness is proved in a standard way with respect to this union. We start by defining the notion of maximal consistent prime TPTND-theory with respect to expressions S^X and S^T, denoting respectively a formula containing a theoretical variable and a formula containing a process. The notion of a TPTND-*theory* $\mathcal{T}$ can be defined in a standard way as a non-trivial set of $TPTND$ formulas, closed under the derivability relation defined by its rules.

(S^X) A theory is (S^X)-*prime*, if it satisfies the following property: if $\mathtt{X} : (\alpha + \beta) \in \mathcal{T}$, then $\mathtt{X} : \alpha \in \mathcal{T}$ or $\mathtt{X} : \beta \in \mathcal{T}$. A theory is ($S^X$)-*consistent* if for no formula $\mathtt{X} : \alpha$, both $\mathtt{X} : \alpha$ and $\mathtt{X} : \neg\alpha \in \mathcal{T}$. A theory is ($S^X$)-*maximal* if for all $\mathtt{X} : \alpha$, either $\mathtt{X} : \alpha \in \mathcal{T}$, or $\mathtt{X} : \neg\alpha \in \mathcal{T}$.

(S^T) A theory is (S^T)-*prime*, if it satisfies the following property: if $\mathtt{T} : (\alpha + \beta) \in \mathcal{T}$, then $\mathtt{T} : \alpha \in \mathcal{T}$ or $\mathtt{T} : \beta \in \mathcal{T}$. A theory is ($S^T$)-*consistent* if for no formula $\mathtt{T} : \alpha$, both $\mathtt{T} : \alpha$ and $\mathtt{X} : \neg\alpha \in \mathcal{T}$. A theory is ($S^T$)-*maximal* if for all $\mathtt{T} : \alpha$, either $\mathtt{T} : \alpha \in \mathcal{T}$, or $\mathtt{T} : \neg\alpha \in \mathcal{T}$.

Proposition 5. *Let $\mathcal{T}$ be a TPTND-theory with respect to (S^X). Then, $\langle \mathtt{X}_x, \mathtt{X}_y \rangle : (\alpha \times \beta) \in \mathcal{T}$ iff $\mathtt{X}_x : \alpha \in \mathcal{T}$ and $\mathtt{X}_y : \beta \in \mathcal{T}$.*

Proof. Let $\langle \mathtt{X}_x, \mathtt{X}_y \rangle : (\alpha \times \beta) \in \mathcal{T}$. By rule 1x E$\times_L$ from Fig. 8, we have *fst*$(\langle \mathtt{X}_x, \mathtt{X}_y \rangle) : \alpha \in \mathcal{T}$, and thus $\mathtt{X}_x : \alpha \in \mathcal{T}$. Similarly, $\mathtt{X}_y : \beta \in \mathcal{T}$. Let $\mathtt{X}_x : \alpha \in \mathcal{T}$ and $\mathtt{X}_y : \beta \in \mathcal{T}$. Then, by introduction of × (1x I× rule), $\langle \mathtt{X}_x, \mathtt{X}_y \rangle : (\alpha \times \beta) \in \mathcal{T}$. □

Proposition 6. *Let $\mathcal{T}$ be a TPTND-theory with respect to (S^T). Then, $\langle \mathtt{T}, \mathtt{U} \rangle : (\alpha \times \beta) \in \mathcal{T}$ iff $\mathtt{T} : \alpha \in \mathcal{T}$ and $\mathtt{U} : \beta \in \mathcal{T}$.*

The proof is similar to the proof of Proposition 5, and thus is omitted.

The canonical model for TPTND is defined as follows.

Definition 18 *(Canonical model).* The canonical model $\mathcal{M}^C$ for TPTND is the tuple $(W^C_t, W^C_e, v^C_t, v^C_e)$, where:

- $W^C_t = \{w \mid w$ is a maximal consistent prime theory with respect to $(S^X)\}$;
- $W^C_e = \{w \mid w$ is a maximal consistent prime theory with respect to $(S^T)\}$;
- $v^C_t(x : \alpha) = \{w \in W^C_t \mid x : \alpha \in w\}$ for all $x : \alpha \in \mathcal{L}$;
- $v^C_e(t : \alpha) = \{w \in W^C_e \mid t : \alpha \in w\}$ for all $t : \alpha \in \mathcal{L}$.

In what follows, it is useful to consider the following theories closed under $\vdash$ of TPTND, defined for the canonical model.

- $\mathcal{T}^1 = \{\mathtt{X} : \alpha \mid \mathtt{X} : \alpha \in w$ for all $w \in W^C_t\}$;
- $\mathcal{T}^2 = \{\mathtt{T} : \alpha \mid \mathtt{T} : \alpha \in w$ for all $w \in W^C_e\}$;
- given a set of all theories $\{\mathcal{T}_1, ..., \mathcal{T}_n\} \in W^C_t$, $\mathcal{T}^3 = \{\mathtt{X} : \alpha_a \mid b =\mid \{\mathcal{T}_i \mid \mathtt{X} : \alpha \in \mathcal{T}_i\} \mid$, and $a = \frac{b}{n}\}$;
- given a set of theories $\{\mathcal{T}_1, ..., \mathcal{T}_n\} \in W^C_e$, $\mathcal{T}^4 = \{\mathtt{T}_n : \alpha_f \mid f = \frac{\mid\{\mathcal{T}_i \mid \mathtt{T}:\alpha \in \mathcal{T}_i\}\mid}{n}\}$;
- $\mathcal{T}^5 = \{\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} \mid \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4$, and $\tilde{a} = a \cdot n\}$;
- $\mathcal{T}^6 = \{$*Trust*$(\mathtt{T}_n : \alpha_f) \mid \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4$, and $|a - f| \leq \epsilon(n)\}$;
- $\mathcal{T}^7 = \{$*UTrust*$(\mathtt{T}_n : \alpha_f) \mid \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4$, and $|a - f| > \epsilon(n)\}$;
- $\mathcal{T}^* = \bigcup(\mathcal{T}^1, \mathcal{T}^2, \mathcal{T}^3, \mathcal{T}^4, \mathcal{T}^5, \mathcal{T}^6, \mathcal{T}^7)$.

The following lemma is a preliminary result for proving the truth lemma. The general idea is to consider each type of expression (S^X), (S^T), (S^{Xp}), (S^{Tf}), (S^{Te}), (S^{Tr}), (S^{UT}) in terms of belonging to a particular TPTND-theory. Once each expression is associated with a particular TPTND-theory, we generalize the result in the truth lemma, and provide the extension lemma, which guarantees the proof of completeness.

Lemma 1 *(Preliminary truth lemma). For all well-formed formulas of $\mathcal{L}$ and all $w \in W^C_t, W^C_e$:*

$$\mathcal{M}^C, w \vDash \mathtt{X} : \alpha \textit{ iff } \mathtt{X} : \alpha \in w, \textit{ s.t. } w \in W_t^C;$$
$$\mathcal{M}^C, w \vDash \mathtt{T} : \alpha \textit{ iff } \mathtt{T} : \alpha \in w, \textit{ s.t. } w \in W_e^C;$$
$$\mathcal{M}^C \vDash \mathtt{X} : \alpha_a \textit{ iff } \mathtt{X} : \alpha_a \in \mathcal{T}^3;$$
$$\mathcal{M}^C \vDash \mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f \textit{ iff } \mathtt{T}_n : \alpha_f \in \mathcal{T}^4;$$
$$\mathcal{M}^C \vDash \mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} \textit{ iff } \mathtt{X}/\mathtt{T} : \alpha_{\tilde{a}} \in \mathcal{T}^5;$$
$$\mathcal{M}^C \vDash \textit{Trust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f) \textit{ iff } \textit{Trust}(\mathtt{T}_n : \alpha_f) \in \mathcal{T}^6;$$
$$\mathcal{M}^C \vDash \textit{UTrust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f) \textit{ iff } \textit{UTrust}(\mathtt{T}_n : \alpha_f) \in \mathcal{T}^7.$$

Proof. The proof is by induction on the length of a formula.

Base case. For the case of $x : \alpha$. By definition of v_t^C, $x : \alpha \in w$ s.t. $w \in W_t^C$ iff $w \in v_t^C(x : \alpha)$, which is by semantics equivalent to $\mathcal{M}^C, w \vDash x : \alpha$.

For the case of $t : \alpha$. By definition of v_e^C, $t : \alpha \in w$, s.t. $w \in W_e^C$ iff $w \in v_e^C(t : \alpha)$, which is equivalent to $\mathcal{M}^C, w \vDash t : \alpha$.

Induction step.

In case of $\mathtt{X} : \alpha_a$, let $\mathtt{X} : \alpha_a \in \mathcal{T}^3$, that is $\mathtt{X} : \alpha_a \in \{\mathtt{X} : \alpha_a \mid b =\mid \{w_i \mid \mathtt{X} : \alpha \in w_i\} \mid, \text{and } a = \frac{b}{n}\}$, provided with a set of all theories $\{w_1, \ldots, w_n\} \in W_t^C$. This means that $\mid W_t^C \mid= n$, $b =\mid \{w_i \mid \mathtt{X} : \alpha \in w_i\} \mid$, and $a = \frac{b}{n}$, which is equivalent to $\mathcal{M}^C \vDash \mathtt{X} : \alpha_a$.

Let $\mathtt{T}_n : \alpha_f \in \mathcal{T}^4$, that is $\mathtt{T}_n : \alpha_f \in \{\mathtt{T}_n : \alpha_f | f = \frac{|\{w_i|\mathtt{T}:\alpha\in w_i\}|}{n}\}$, provided with a set of theories $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\} \in W_e^C$. This means that for a $W_e'^C \subseteq W_e^C$, it is of a size n, and $f = \frac{|\{w_i|\mathtt{T}:\alpha\in w_i\}|}{n}$ s.t. $w_i \in W_e'^C$. This is equivalent to $\mathcal{M}^C \vDash \mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f$ s.t. $n =\mid \{w', \ldots, w'^n\} \mid$ for some $w', \ldots, w'^n$.

Let $\mathtt{T}_n : \alpha_{\tilde{a}} \in \mathcal{T}^5$, that is $\mathtt{T}_n : \alpha_{\tilde{a}} \in \{\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} | \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4, \text{and } \tilde{a} = a \cdot n\}$. This means that $\mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3$, $\mathtt{T}_n : \alpha_f \in \mathcal{T}^4$, and $\tilde{a} = a \cdot n$. By semantics and previous cases we have $\mathcal{M}^C \vDash \mathtt{X}_\mathtt{T} : \alpha_a$, $\mathcal{M}^C \vDash \mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f$ where $\mid \{w', \ldots, w'^n\} \mid= n$, and thus $\mathcal{M}^C \vDash \mathtt{T}_n : \alpha_{\tilde{a}}$.

Let $\mathtt{X} : (\alpha + \beta) \in w$ s.t. $w \in W_t^C$. By the fact that w is a prime theory, this is equivalent to $\mathtt{X} : \alpha \in w$ or $\mathtt{X} : \beta \in w$. By induction hypothesis, this means that $\mathcal{M}^C, w \vDash \mathtt{X} : \alpha$ or $\mathcal{M}^C, w \vDash \mathtt{X} : \beta$, which by semantics is equivalent to $\mathcal{M}, w \vDash \mathtt{X} : (\alpha + \beta)$.

Let $\langle \mathtt{X}_x, \mathtt{X}_y \rangle : (\alpha \times \beta) \in w$ s.t. $w \in W_t^C$. Then, by Proposition 5, $\mathtt{X}_x : \alpha \in w$ and $\mathtt{X}_y : \beta \in w$. By induction hypothesis, this means that $\mathcal{M}^C, w \vDash \mathtt{X}_x : \alpha$ and $\mathcal{M}^C, w \vDash \mathtt{X}_y : \beta$. Thus, by semantics, $\mathcal{M}^C \vDash \langle \mathtt{X}_x, \mathtt{X}_y \rangle : (\alpha \times \beta)$.

Let $\mathtt{X} : \neg\alpha_a \in \mathcal{T}^3$. This means that $b =\mid \{w_i \mid \mathtt{X} : \neg\alpha \in w_i\} \mid_{0 \le i \le n}$, where $\{w_1, \ldots, w_n\}$ is the set of all theories in W_t^C, and $a = \frac{b}{n}$. Having in mind that $w_1, \ldots, w_n$ are maximal consistent theories, this means that the number of w_j, s.t. $\mathtt{X} : \alpha \in w_j$ is $n - b$. This means that $\mathcal{X} : \alpha_{\frac{n-b}{n}} \in \mathcal{T}^3$, and thus, by having $a = \frac{b}{n}$, $\mathtt{X} : \alpha_{1-a} \in \mathcal{T}^3$. By induction hypothesis this means that $\mathcal{M}^C \vDash \mathtt{X} : \alpha_{1-a}$, that is $\mathcal{M}^C \vDash \mathtt{X} : \neg\alpha_a$.

Let $\langle \mathtt{X}_\mathtt{x}, \mathtt{X}_y \rangle : (\alpha \times \beta)_a \in \mathcal{T}^3$. This means that $b =\mid \{w_i \mid \langle \mathtt{X}_\mathtt{x}, \mathtt{X}_y \rangle : (\alpha \times \beta) \in w_i\} \mid_{0 \le i \le n}$, where $\{w_1, \ldots, w_n\}$ is the set of all theories in W_t^C, and $a = \frac{b}{n}$. By Proposition 5, we have that $\mathtt{X}_x : \alpha \in w_i$ and $\mathtt{X}_y : \alpha \in w_i$ for all w_i s. t. $\langle \mathtt{X}_\mathtt{x}, \mathtt{X}_y \rangle : (\alpha \times \beta) \in w_i$ and the number of all w_i is b. By extend rule, we have that $\mathtt{X}_x : \alpha_{c/n} \in \mathcal{T}^3$ and $\mathtt{X}_y : \beta_{d/n} \in \mathcal{T}^3$, and thus the number of all w_j s.t. $\mathtt{X}_x : \alpha \in w_j$ is c, and the number of all worlds w_k s.t. $\mathtt{X}_y : \beta \in w_k$ is d. Having in mind that all w worlds are maximal consistent theories, we can apply standard probabilistic theory and calculate the number of worlds w_i in which both $\mathtt{X}_x : \alpha$ and $\mathtt{X}_y : \alpha$ hold, that is $a = \frac{c}{n} \cdot \frac{d}{n}$. By induction hypothesis, we have $\mathcal{M}^C \vDash \mathtt{X}_x : \alpha_{\frac{c}{n}}$ and $\mathcal{M}^C \vDash \mathtt{X}_y : \beta_{\frac{d}{n}}$. By semantics this means that $\mathcal{M}^C \vDash \langle \mathtt{X}_x, \mathtt{X}_y \rangle (\alpha \times \beta)_a$.

Let $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_a \in \mathcal{T}^3$. This means that $b =\mid \{w_i \mid [\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta) \in w_i\} \mid_{0 \le i \le n}$, where $\{w_1, \ldots, w_n\}$ is the set of all theories in W_t^C, and $a = \frac{b}{n}$. By the extend rule, we have $\mathtt{X}_x : \alpha_{\frac{c}{n}} \in \mathcal{T}^3$, which means that $c =\mid \{w_j \mid \mathtt{X}_x : \alpha \in w_j\} \mid_{0 \le i \le n}$. The theories w_j constitute a theory $\mathcal{T}^{3'} \subseteq \mathcal{T}^3$ also closed under TPTND principles, s.t. $\mathtt{X}_x : \alpha_1 \in \mathcal{T}^{3'}$. We also have $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_d \in \mathcal{T}^{3'}$, and thus, by $E \to$, we have $\mathtt{X}_y.(\mathtt{X}_x : \alpha) : \beta_d \in \mathcal{T}^{3'}$. We show that $d = a$. Assume that it is not. The statement $\mathtt{X}_y.(\mathtt{X}_x : \alpha) : \beta_d \in \mathcal{T}^{3'}$ means that we have $\mathtt{X}_x : \alpha \vdash \mathtt{X}_y.(\mathtt{X}_x : \alpha) : \beta_d$ in $\mathcal{T}^3$. From $I \to$ we have thus $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_d \in \mathcal{T}^3$. However, if $d \neq a$, this means that for some $w_k \in W_t^C$, $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_d \in w_k$ and $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta) \notin w_k$. Thus, $d = a$. By induction hypothesis, we obtain that for all $w_j \in \mathcal{M}^{C*}$ which is a submodel of $\mathcal{M}^C$, s.t. $\mathcal{M}^C, w_j \vDash \mathtt{X} : \alpha$, we have $\mathcal{M}^{C*} \vDash \mathtt{X}_y : \beta_a$. Then, $\mathcal{M}^c \vDash [\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_a$.

The cases of $\mathtt{T} : (\alpha + \beta)$ and $\mathtt{T} : (\alpha \times \beta)$ are similar to $\mathtt{X} : (\alpha + \beta)$ and $\mathtt{X} : (\alpha \times \beta)$, respectively.

Let $\mathtt{T} : \neg\alpha \in w$, s.t. $w \in W_e^C$. Due to the fact that w is consistent with respect to (S^T), this means that $\mathtt{T} : \alpha \notin w$. By induction hypothesis, we have $\mathcal{M}^C, w \nvDash \mathtt{T} : \alpha$.

Let $[\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}} \in \mathcal{T}^5$, i.e., $[\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}} \in \{\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} | \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4, \text{and } \tilde{a} = a \cdot n\}$. This means that $[\mathtt{X}]\mathtt{X}_{y\mathtt{T}} : (\alpha \to \beta)_{[a]b} \in \mathcal{T}^3$. From the case of $[\mathtt{X}_x]\mathtt{X}_y : (\alpha \to \beta)_a \in \mathcal{T}^3$ we have thus that for all $w \in \mathcal{M}^{C*}$ which is a submodel of $\mathcal{M}^C$ s.t. $\mathcal{M}^C, w \vDash \mathtt{X} : \alpha$, we have $\mathcal{M}^{C*} \vDash \mathtt{X}_{y\mathtt{T}} : \beta_{[a]b}$. By semantics, $\mathcal{M}^{C*} \vDash [\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}}$. This proves the right-to-left part. For the left-to-right, assume that $\mathcal{M}^{C*} \vDash [\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}}$, as shown before, this means that $[\mathtt{X}]\mathtt{X}_{y\mathtt{T}} : (\alpha \to \beta)_{[a]b} \in \mathcal{T}^3$. Then, we observe that $\mathtt{T}_n : (\alpha \to \beta)_f \in \mathcal{T}^4$ (which is true because for any $\mathtt{T} : \alpha$, there is some f s.t. $\mathtt{T}_n : \alpha_f \in \mathcal{T}^4$) and $[a]\tilde{b} = [a](b \cdot n)$ (which is also true, because of the definition of expected frequency, provided with the theoretical probability, $[a]b$, and the number of tests, n). Thus, $[\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}} \in \{\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} | \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4, \text{and } \tilde{a} = a \cdot n\}$, i.e., $[\mathtt{X}]\mathtt{T}_n : (\alpha \to \beta)_{[a]\tilde{b}} \in \mathcal{T}^5$.

Let $\textit{Trust}(\mathtt{T}_n : \alpha_f) \in \mathcal{T}^6$, i.e., $\textit{Trust}(\mathtt{T}_n : \alpha_f) \in \{\textit{Trust}(\mathtt{T}_n : \alpha_f) \mid \mathtt{X}_\mathtt{T} : \alpha_a \in \mathcal{T}^3, \mathtt{T}_n : \alpha_f \in \mathcal{T}^4, \text{and } |a - f| \le \epsilon(n)\}$. By induction hypothesis and the definition of $\mathcal{M}$, this means that $\mathcal{M}^{theor} \vDash \mathtt{X}_\mathtt{T} : \alpha_a$, $\mathcal{M}^{emp} \vDash \mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f$ where $\mid \{w', \ldots, w'^n\} \mid= n$, and $|a - f| \le \epsilon(n)$. Thus, $\mathcal{M}^C \vDash \textit{Trust}(\mathtt{T}_{\{w',\ldots,w'^n\}} : \alpha_f)$.

The case of $\textit{UTrust}(\mathtt{T}_n : \alpha_f) \in \mathcal{T}^8$ is similar to the previous one. □

Lemma 2 *(Truth lemma). For all well-formed formulas $\mathcal{X}$ of $\mathcal{L}$:*
$\mathcal{M}^C \vDash \mathcal{X}$ *iff* $\mathcal{X} \in \mathcal{T}^*$.

Proof. First, we show that $\mathcal{M}^C \vDash \mathtt{X} : \alpha$ iff $\mathtt{X} : \alpha \in \mathcal{T}^1$. From Lemma 1, we have $\mathcal{M}^C, w \vDash \mathtt{X} : \alpha$ iff $\mathtt{X} : \alpha \in w$, s. t. $w \in W_t^C$. $\mathcal{M}^C \vDash \mathtt{X} : \alpha$ iff $\mathcal{M}^C, w \vDash \mathtt{X} : \alpha$ for all $w \in W_t^C$. Thus, $\mathtt{X} : \alpha \in \{\mathtt{X} : \alpha \mid \mathtt{X} : \alpha \in w$ for all $w \in W_t^C\}$, i.e., $\mathtt{X} : \alpha \in \mathcal{T}^1$. Similarly, it is easy to show that $\mathcal{M}^C \vDash \mathtt{T} : \alpha$ iff $\mathtt{T} : \alpha \in \mathcal{T}^2$.

Secondly, we prove the lemma statement. Let $\mathcal{M}^C \vDash \mathcal{X}$. By previous considerations and Lemma 1, this means that $\mathcal{X} \in \mathcal{T}^1$, or $\mathcal{X} \in \mathcal{T}^2$, or $\mathcal{X} \in \mathcal{T}^3$, or $\mathcal{X} \in \mathcal{T}^4$, or $\mathcal{X} \in \mathcal{T}^5$, or $\mathcal{X} \in \mathcal{T}^6$, or $\mathcal{X} \in \mathcal{T}^7$, i.e., $\mathcal{X} \in \mathcal{T}^*$. □

Now we show that v_t^C and v_e^C satisfy the properties of v^{theor} and v^{emp} as in Definitions 7 and 11, namely:

- $\{v^{theor}(\mathtt{x} : \alpha)\} \cap \{v^{theor}(\mathtt{x} : \beta)\} = \emptyset$, whenever $\alpha \neq \beta$;
- for all w there exists v^{theor} such that $w \in v^{theor}(ES^t)$.

Lemma 3. *For all v_t^C, for all v_e^C*

(1) $v_t^C(x : \alpha) \cap v_t^C(x : \beta) = \emptyset$*, whenever* $\alpha \neq \beta$*;*
(2) $v_e^C(t : \alpha) \cap v_e^C(t : \beta) = \emptyset$*, whenever* $\alpha \neq \beta$*;*
(3) *for all w there exists v_t^C such that $w \in v_t^C(x : \alpha)$.*
(4) *for all w there exists v_e^C such that $w \in v_e^C(t : \alpha)$.*

Proof. First we prove (1) and (2). Let $v_t^C(x : \alpha) \cap v_t^C(x : \beta) \neq \emptyset$, i.e., there exists a theory w such that $x : \alpha \in w$ and $x : \beta \in w$. Taking into account that w is a TPTND-theory, $x : \alpha \in w$ is the case only if $x : (\alpha + \beta) \in w$ and $x : \neg\beta \in w$, for any β (rules $1\mathtt{xE}+_L$ or $1\mathtt{xE}+_R$, Fig. 8). However, w is (S^x)-consistent, thus $x : \beta \notin w$. The case (2) can be proven similarly, using the rule $1\mathtt{tE}+_L$.

For the case (3), let w be a theory such that for all $x : \alpha$, $w \notin v_t^C(x : \alpha)$. One can construct a TPTND-theory $\mathcal{T}^+ = \{X : \alpha_a \mid$ if $X : \alpha \in w$, then $a = 1$; if $X : \alpha \notin w$, then $a = 0\}$. Then, $x : \alpha_0 \in \mathcal{T}^+$, which can be obtained only by rules $E+_R$ or $E+_L$ (see Fig. 5). By our assumption and primeness of w, we have $x : (\alpha + \beta) \notin w$, which means that $x : (\alpha + \beta)_0 \in \mathcal{T}^+$. Thus, in accordance with rules $E+_L$ and $E+_R$, $x : \neg\beta_0 \in \mathcal{T}^+$. By construction of $\mathcal{T}^+$ this means that $x : \neg\beta \notin w$, that is, by maximality of w, $x : \beta \in w$, which contradicts our assumption. The case (4) can be proven similarly, by using sampling rules (see Fig. 6). □

Lemma 4 *(Extension lemma). Let $\mathcal{X}$ and $\mathcal{Y}$ be formulae of TPTND s.t. $\mathcal{X} \nvdash \mathcal{Y}$, then there exists a TPTND-theory $\mathcal{T}^*$ such that $\mathcal{X} \in \mathcal{T}^*$ and $\mathcal{Y} \notin \mathcal{T}^*$.*

Proof. In order to prove the lemma, we show that such a theory exists.

The basic step follows closely the theory construction provided by Dunn [17, p. 13, Lemma 8]. Suppose that $\mathcal{X}, \mathcal{Y}$ are of a form (S^X). Then, we enumerate sentences $\mathcal{X}_1, \mathcal{X}_2, \ldots$ and build up a series of theories starting with $\mathcal{T}_0 = \{\mathcal{X}' \mid \mathcal{X} \vdash \mathcal{X}'\}$. Theory $\mathcal{T}_{n+1}$ is obtained from $\mathcal{T}_n$ by adding $\mathcal{X}_{n+1}$ if one can do so while closing the result under the principles of TPTND without getting $\mathcal{Y}$. Theory $\mathcal{T}^1$ is obtained as the union of all the $\mathcal{T}_n$'s, and it is easy to see that it is a (S^X-)maximal theory with respect to not containing $\mathcal{Y}$. It is (S^X-)prime, because from $\mathtt{X} : (\alpha + \beta) \in \mathcal{T}^1$ by $E+_L$ or $E+_R$ we get either $\mathtt{X} : \alpha \in \mathcal{T}^1$, or $\mathtt{X} : \beta \in \mathcal{T}^1$. We can show that $\mathcal{T}^1$ is (S^X-)consistent. Let $\mathtt{X} : \alpha \in \mathcal{T}^1$, which is $\Gamma \vdash \mathtt{X} : \alpha_1$, where $\Gamma \in \mathcal{T}^1$. Then, by $\bot$, we have $\Gamma \vdash \neg\mathtt{X} : \alpha_0$, that is $\mathtt{X} : \alpha \notin \mathcal{T}^1$. Let $\mathtt{X} : \neg\alpha \in \mathcal{T}^1$, that is $\Gamma \vdash \neg\mathtt{X} : \alpha_1$, where $\Gamma \in \mathcal{T}^1$. Then, $\Gamma \vdash \neg\mathtt{X} : \alpha_1$ is derivable iff $\Gamma \vdash \mathtt{X} : \alpha_0$, and thus $\mathtt{X} : \alpha \notin \mathcal{T}^1$.

Where $\mathcal{X}$, $\mathcal{Y}$ are of the form (S^T) we construct an $(S^T$-)maximal with respect to not containing $\mathcal{Y}$, $(S^T$-)consistent, $(S^T$-)prime theory $\mathcal{T}^2$ in a similar way. Assume that $\mathcal{Y}$ is of the form (S^{Xp}). Then, $\mathcal{T}^3 = \{\mathtt{X} : \alpha_a \mid \mathcal{X} \vdash \mathtt{X} : \alpha_a\}$. Clearly, $\mathcal{Y} \notin \mathcal{T}^3$. Similarly, if $\mathcal{Y}$ is of the form (S^{Tf}), then $\mathcal{T}^4 = \{\mathtt{T}_n : \alpha_f \mid \mathcal{X} \vdash \mathtt{T}_n : \alpha_f\}$; if $\mathcal{Y}$ is of the form (S^{Te}), then $\mathcal{T}^5 = \{\mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}} \mid \mathcal{X} \vdash \mathtt{X}/\mathtt{T}_n : \alpha_{\tilde{a}}\}$; if $\mathcal{Y}$ is of the form (S^{Tr}), then $\mathcal{T}^6 = \{\mathit{Trust}(\mathtt{T}_n : \alpha_f) \mid \mathcal{X} \vdash \mathit{Trust}(\mathtt{T}_n : \alpha_f)\}$; if $\mathcal{Y}$ is of the form (S^{UT}), then $\mathcal{T}^7 = \{\mathit{UTrust}(\mathtt{T}_n : \alpha_f) \mid \mathcal{X} \vdash \mathit{UTrust}(\mathtt{T}_n : \alpha_f)\}$. Then, we construct $\mathcal{T}^*$ as a union of all theories $\mathcal{T}^1, \ldots, \mathcal{T}^7$. By the construction, $\mathcal{X} \in \mathcal{T}^*$ and $\mathcal{Y} \notin \mathcal{T}^*$. □

Theorem 3 *(Completeness). If $\Gamma \vDash \mathcal{Y}$, then $\Gamma \vdash \mathcal{Y}$.*

Proof. Let $\Gamma \nvdash \mathcal{Y}$. Then, for all $\mathcal{X} \in \Gamma$, we have $\mathcal{X} \nvdash \mathcal{Y}$. Then, by Lemma 4, there exists $\mathcal{T}^*$ s.t. $\mathcal{X} \in \mathcal{T}^*$ and $\mathcal{Y} \notin \mathcal{T}^*$. Thus, by Lemma 2, $\mathcal{M} \vDash \mathcal{X}$ and $\mathcal{M} \nvDash \mathcal{Y}$, that is $\mathcal{X} \nvDash \mathcal{Y}$, and thus $\Gamma \nvDash \mathcal{Y}$. □

7. Epistemic extension

The semantics introduced in Section 3 provides a descriptive tool for representing trustworthiness as a property of a non-deterministic system. We now turn our attention to the epistemic attitude that an agent can have towards such property. In particular, we are interested in defining conditions for claiming that trustworthiness of an algorithm with respect to one or more of its outputs is *known*. We intend this as a different condition that trustworthiness holds. In our previous analysis, we considered that a process is trustworthy whenever there is a correspondence (within fixed limits) between the theoretical probability of the value of a given random variable as represented in a theoretical model, and the actual frequency of that result as represented in an empirical model. From this perspective, knowledge of the trustworthiness of a process should combine at least two conditions:

1. the process should be evaluated to be trustworthy on a given number n of trials;
2. trustworthiness should be checked for preservation by any further test performed on this process.

To grant these two conditions, we may define a knowledge operator over trustworthiness of formulae as a quantification over new runs of the trial under consideration to express the following difference: a process is trustworthy (respectively untrustworthy) if its trustworthiness (respectively untrustworthiness) holds over a given number of trials, and it *is known* to be trustworthy (respectively untrustworthy) if its trustworthiness (respectively untrustworthiness) is preserved by any increasing number of trials. Hence, the second condition implies monotonicity of the trustworthiness property.

Formally, we extend our semantics with an epistemic operator K defined over formulae $\mathtt{F} = \{S^{Tr}, S^{UT}\}$. Let us first introduce the extended language $\mathcal{L}^K$.

Definition 19 *(Language $\mathcal{L}^K$).*

$$\mathcal{L}^K := \mathcal{L} \cup \{K(\mathtt{F})\}$$

where we abbreviate with $\mathtt{F}$ formulas $\mathit{Trust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f), \mathit{UTrust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f)$.

In order to interpret this new operator semantically, we need to extend our semantics with an accessibility relation. In particular, we propose to extend the empirical models $\mathcal{M}^{emp}$ from Definition 11 with an accessibility relation R^{emp} holding between worlds and where the latter are interpreted as consecutive instances of the same trial run.[6]

Definition 20. Let $\mathcal{M}^{empK} = (W^{emp}, R^{emp}, v^{emp})$ where W^{emp} and v^{emp} are as in Definition 11, and R^{emp} is an accessibility relation $R^{emp} \subseteq \mathcal{P}(W^{emp})$.

In what follows, we will not be interested in any R^{emp}, but in a particular accessibility relation, which we call *test-temporal*.

Definition 21. We call an accessibility relation R *test-temporal*, if it satisfies the following properties:

- reflexivity: $\forall w Rww$;
- transitivity: $\forall w \forall w' \forall w''(Rww' \wedge Rw'w'') \Rightarrow Rww''$;
- anti-symmetry: $\forall w \forall w'(Rww' \wedge Rw'w) \Rightarrow w = w'$;
- linearity: $\forall w \forall w'(w = w' \vee Rww' \vee Rw'w)$;
- beginning*: $\exists w \neg \exists w'(w \neq w' \wedge Rww')$; we call the world w the *beginning-world*.

These conditions restrict the relation R^{emp} with respect to several intuitions on how the tests should be temporally related in an empirical model. Reflexivity states that being in a world in which a test happens, the results of the test are observed. Transitivity means that given a test world w, one observes every previous test world. Anti-symmetry assures that given a test world w, it does not observe the test worlds which temporally follow w. Linearity assures that all test-worlds are connected, and that the flow of time is linear. Beginning* means that there is a test world which is not preceded by any other test world, i.e., there is a first test in a trial. Thus, R^{emp} is the relation that holds at any given moment in time retrospectively: from a given point it directs only to the tests that have been executed until that moment, and it directs to none of those that may occur in the future.

We modify the definition of a joint model as

$$\mathcal{M}^K = (\mathcal{M}^{theor}, \mathcal{M}^{empK})$$

and supplement it with truth conditions for $K(\mathtt{F})$.

Definition 22 *(Satisfiability of K-formulae).*

- $\mathcal{M}^K \vDash K(\mathit{Trust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f))$ iff $\mathcal{M} \vDash \mathit{Trust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f)$ and for any $\mathcal{M}^{empK}_{n+m} = \bigcup(\mathcal{M}^{emp}_m, \mathcal{M}^{emp}_n)$ where all $w_o \in \mathcal{M}^{emp}_m$ are such that $w_o R^{emp} w_n$, $\mathcal{M} \vDash \mathit{Trust}(\mathtt{T}_{\{w',\dots,w'^n\} \cup \{w'',\dots,w''^m\}} : \alpha_{f'})$;
- $\mathcal{M}^K \vDash K(\mathit{UTrust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f))$ iff $\mathcal{M} \vDash \mathit{UTrust}(\mathtt{T}_{\{w',\dots,w'^n\}} : \alpha_f)$ and for any $\mathcal{M}^{empK}_{n+m} = \bigcup(\mathcal{M}^{emp}_m, \mathcal{M}^{emp}_n)$ where all $w_o \in \mathcal{M}^{emp}_m$ are such that $w_o \in \mathcal{M}^{emp}_m$ are $w_o R^{emp} w_n$, $\mathcal{M} \vDash \mathit{UTrust}(\mathtt{T}_{\{w',\dots,w'^n\} \cup \{w'',\dots,w''^m\}} : \alpha_{f'})$.

The two clauses define knowledge of the trustworthiness (respectively untrustworthiness) of a process $\mathtt{T}$ after n executions: knowledge of such property holds if the process remains trustworthy (respectively untrustworthy) for any increase of the number of

[6] Basically, we are reconsidering $\mathcal{M}^{emp}$ in terms of instant-based models of time, see [28] for more details on this type of models.

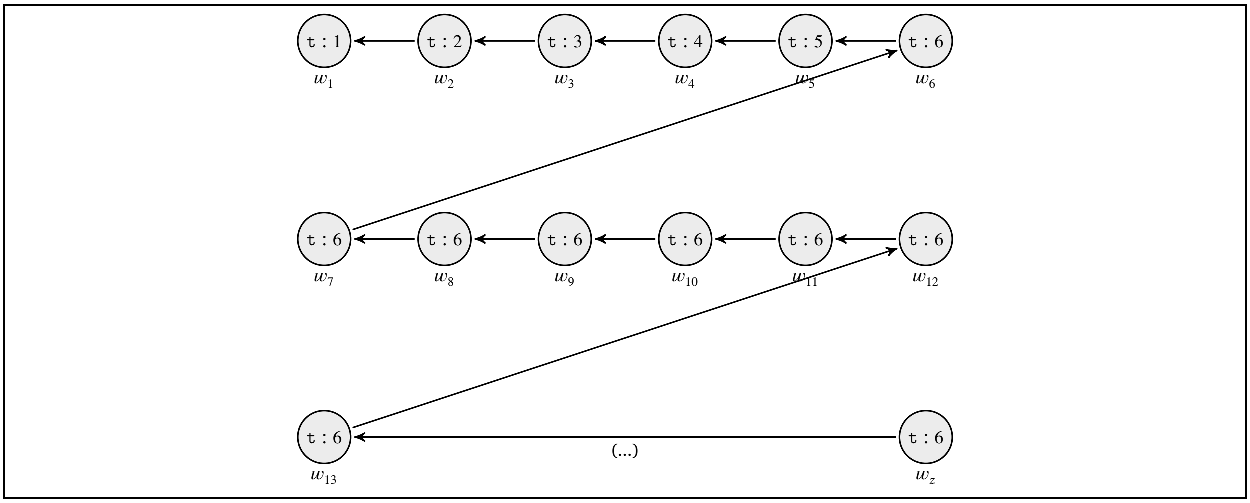

Fig. 10. Model $\mathcal{M}^{empK}$ for an unfair die.

executions. This is expressed by the fact that if we add another series of experiments, that is $\mathcal{M}^{emp}_m$ such that the worlds of $\mathcal{M}^{emp}_m$ test-temporally access the worlds of $\mathcal{M}_n$, the process remains trustworthy (or untrustworthy respectively).

The functioning of the K operator can be exemplified as follows.

Example 8. We construct an $\mathcal{M}^K$ model in which a process simulates the throw of a die ($\mathtt{t}$): $\mathcal{M}^K = (\mathcal{M}^{theor}, \mathcal{M}^{empK})$ such that:

- $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$ where $W^{theor} = \{w_1, ..., w_6\}$ and $v^{theor}(\mathtt{x}_\mathtt{t} : 1) = \{w_1\}$, ..., $v^{theor}(\mathtt{x}_\mathtt{t} : 6) = \{w_6\}$;
- $\mathcal{M}^{empK} = (W^{emp}, R^{emp}, v^{emp})$ where $W^{emp} = \{w_1, w_2, ..., w_z\}$, R^{emp} is a test-temporal relation with the beginning world w_1, and $v^{emp}(\mathtt{t} : 1) = \{w_1\}$, ... , $v^{emp}(\mathtt{t} : 5) = \{w_5\}, v^{emp}(\mathtt{t} : 6) = \{w_i\}$ for all w_i such that $i \geq 6$.

In this model, we have that $\mathcal{M}^K \vDash Trust(\mathtt{t}_{\{w_1,...,w_6\}} : 1_{\frac{1}{6}})$, because the die behaves as a fair die on the interval between the worlds w_6 and w_1. However, starting from the world w_6 it starts to produce only the output 6, which means that there exists $\mathcal{M}^{emp}_{6+m} \subseteq \mathcal{M}^{emp}$, such that $\mathcal{M} \nvDash Trust(\mathtt{t}_{\{w_1,...,w_6\} \cup \{w',...,w'^m\}} : 1_{f'})$. For instance, it can be an interval between the worlds w_1 and w_{12}, i.e., $\mathcal{M}^{emp}_{6+6}$, where $w_7, ..., w_{12} \in \mathcal{M}^{emp}_6$ test-temporally access w_6. Thus, in this model we have $\mathcal{M}^K \nvDash K(Trust(\mathtt{t}_{\{w_1,...,w_6\}} : 1_{\frac{1}{6}}))$, i.e., even though the die behaves as trustworthy on the initial interval of 6 launches, the agent does not know its trustworthiness because it is not preserved over additional series of experiments. The model $\mathcal{M}^{theor}$ is as in Fig. 1, and $\mathcal{M}^{empK}$ is represented in Fig. 10.

Example 9. We construct an $\mathcal{M}^K$ model in which a process simulates the throw of a fair die ($\mathtt{t}$): $\mathcal{M}^K = (\mathcal{M}^{theor}, \mathcal{M}^{empK})$ such that:

- $\mathcal{M}^{theor} = (W^{theor}, v^{theor})$ where $W^{theor} = \{w_1, ..., w_6\}$ and $v^{theor}(\mathtt{x}_\mathtt{t} : 1) = \{w_1\}$, ..., $v^{theor}(\mathtt{x}_\mathtt{t} : 6) = \{w_6\}$;
- $\mathcal{M}^{empK} = (W^{emp}, R^{emp}, v^{emp})$ where $W^{emp} = \{w_1, w_2, ..., w_z\}$, R^{emp} is a test-temporal relation with the beginning world w_1, and $v^{emp}(\mathtt{t} : 1) = \{w_1, w_7, w_{13}...\}$, ... , $v^{emp}(\mathtt{t} : 6) = \{w_6, w_{12}, w_{18}, ...\}$.[7]

As in the previous example we have $\mathcal{M}^K \vDash Trust(\mathtt{t}_{\{w_1,...,w_6\}} : 1_{\frac{1}{6}})$. Moreover, for any submodel $\mathcal{M}^{emp}_{6+m}$ trustworthiness will be preserved, because any world up to w_z will access all the previous worlds which provide us with a trustworthy frequency of getting the output 1, that is $\mathcal{M} \vDash Trust(\mathtt{t}_{\{w_1,...,w_6\} \cup \{w',...,w'^m\}} : 1_{\frac{1}{6}})$. This means that the agent knows that $\mathtt{t}$ producing 1 with frequency $\frac{1}{6}$ is trustworthy, because the trustworthiness is preserved over all series of tests. See also Fig. 11.

8. Conclusion

We have presented a semantics for evaluating the trustworthiness of probabilistic computations. It is constructed as a combination of sets of possible worlds endowed with evaluations. The formulas evaluated at worlds express theoretical probabilities, expected probabilities and frequencies, depending where they are evaluated. Trustworthiness formulas are evaluated by comparison of outputs across distinct subsets of worlds (respectively where expected probabilities and frequencies are computed).

We have shown that this semantics characterizes the calculus TPTND which combines typed natural deduction with probabilistic reasoning. From a theoretical point of view, our results on completeness and soundness relate the constructive approach of TPTND

[7] Clearly, in a real world the die will not produce each output every 6 launches. However, we choose this model to simplify the reading of the model.

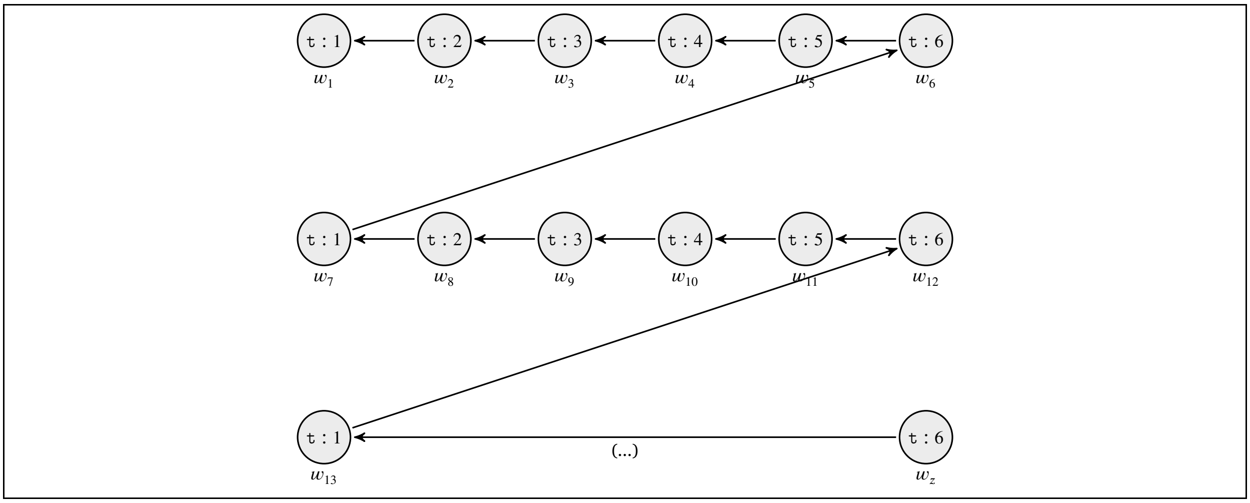

Fig. 11. Model $\mathcal{M}^{empK}$ for a fair die.

and the descriptive tools of possible worlds semantics. In particular, by providing a descriptive semantics for TPTND, we capture the general framework in which agents reason about trustworthiness when the rules are conceived as instructions for reasoning. The adequacy between TPTND and our semantics, as expressed by soundness and completeness results, shows that this line of research permits one not only to make judgments on the trustworthiness of non-deterministic systems, but also to analyze and model conditions for implementing trustworthy systems. The most prominent field of applications where such knowledge would be considered essential is the safe and fair deployment of AI systems.

We further presented a preliminary extension of such semantics with an accessibility relation on worlds of the empirical model and an epistemic operator. The former serves the purpose of modeling temporally ordered test trials, the latter ranges over trustworthiness formulae to quantify over the validity of such property on an increasing number of trials.

This line of research can be considered as a necessary basis for several future developments. First, we aim to extend our epistemic semantics with other more expressive modalities, in order to be able to model cognitive attitudes of group of agents towards AI systems beyond simple knowledge. Secondly, we will formulate the system characterized by it. Third, we plan to enrich this semantics with hyperintensionality, to provide a more fine-grained framework to evaluate trustworthiness of equivalent but non-identical outputs.

CRediT authorship contribution statement

Ekaterina Kubyshkina: Conceptualization, Formal analysis, Investigation, Methodology. **Giuseppe Primiero:** Conceptualization, Formal analysis, Investigation, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Giuseppe Primiero reports financial support was provided by Government of Italy Ministry of Education University and Research. Giuseppe Primiero reports a relationship with Government of Italy Ministry of Education University and Research that includes: funding grants.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors would like to thank the anonymous referees of the International Journal of Approximate Reasoning. The authors acknowledge the support of the Project PRIN2020 “BRIO - Bias, Risk and Opacity in AI” (2020SSKZ7R) awarded by the Italian Ministry of University and Research (MUR). The authors are further funded by the Department of Philosophy “Piero Martinetti” of the University of Milan under the Project “Departments of Excellence 2023-2027” awarded by the Ministry of University and Research (MUR). The research of Ekaterina Kubyshkina is funded under the “Foundations of Fair and Trustworthy AI” Project of the University of Milan.

References

[1] R. Alur, T.A. Henzinger, P.-H. Ho, Automatic symbolic verification of embedded systems, IEEE Trans. Softw. Eng. 22 (1996) 181–201, https://doi.org/10.1109/32.489079.

[2] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, D. Mané, Concrete Problems in AI Safety, arXiv:1606.06565, 2016, https://arxiv.org/abs/1606.06565.

[3] P.H.A. de Amorim, D. Kozen, R. Mardare, P. Panangaden, M. Roberts, Universal semantics for the stochastic λ-calculus, in: 36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021, IEEE, 2021, pp. 1–12.

[4] M. Antonelli, U.D. Lago, P. Pistone, Curry and Howard meet Borel, in: C. Baier, D. Fisman (Eds.), LICS '22: 37th Annual ACM/IEEE Symposium on Logic in Computer Science, Haifa, Israel, August 2 - 5, 2022, ACM, 2022, pp. 45:1–45:13.

[5] M. Boričić, Sequent calculus for classical logic probabilized, Arch. Math. Log. 58 (2019) 119–136.

[6] C. Castelfranchi, R. Falcone, Trust Theory: A Socio-Cognitive and Computational Model, Wiley, 2010.

[7] G. Bacci, R. Furber, D. Kozen, R. Mardare, P. Panangaden, D.S. Scott, Boolean-valued semantics for the stochastic λ-calculus, in: A. Dawar, E. Grädel (Eds.), Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018, ACM, 2018, pp. 669–678.

[8] A. Baltag, S. Smets, Probabilistic dynamic belief revision, Synthese 165 (2010) 179–202.

[9] J. van Benthem, Conditional probability meets update logic, J. Log. Lang. Inf. 12 (2003) 409–421.

[10] J. Borgström, U.D. Lago, A.D. Gordon, M. Szymczak, A lambda-calculus foundation for universal probabilistic programming, in: J. Garrigue, G. Keller, E. Sumii (Eds.), Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming, ICFP 2016, Nara, Japan, September 18-22, 2016, ACM, 2016, pp. 33–46.

[11] F. Dahlqvist, D. Kozen, Semantics of higher-order probabilistic programs with conditioning, Proc. ACM Program. Lang. 4 (2020) 57:1–57:29, https://doi.org/10.1145/3371125.

[12] F.A. D'Asaro, G. Primiero, Probabilistic typed natural deduction for trustworthy computations, in: Dongxia Wang, Rino Falcone, Jie Zhang (Eds.), Proceedings of the 22nd International Workshop on Trust in Agent Societies (TRUST 2021) Co-located with the 20th International Conferences on Autonomous Agents and Multiagent Systems (AAMAS 2021), CEUR Workshop Proceedings, 2021, http://ceur-ws.org/Vol-3022/paper3.pdf.

[13] F.A. D'Asaro, F. Genco, G. Primiero, Checking trustworthiness of probabilistic computations in a typed natural deduction system, CoRR, arXiv:2206.12934 [abs], https://arxiv.org/pdf/2206.12934.pdf, 2023.

[14] L. Demey, B. Kooi, J. Sack, Logic and probability, in: Edward N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2019 edition, 2019.

[15] R. Demolombe, Reasoning about trust: a formal logical framework, in: C.D. Jensen, S. Poslad, T. Dimitrakos (Eds.), Trust Management, iTrust 2004, in: Lecture Notes in Computer Science., vol. 2995, Springer, Berlin, Heidelberg, 2004, pp. 291–303.

[16] A. Di Pierro, A type theory for probabilistic λ-calculus, in: From Lambda Calculus to Cybersecurity Through Program Analysis, 2020.

[17] M.J. Dunn, Partiality and its dual, Stud. Log. 66 (1) (2000) 5–40.

[18] R. Fagin, J.Y. Halpern, Reasoning about knowledge and probability, in: Proceedings of the 2nd Conference on Theoretical Aspects of Reasoning About Knowledge, M. Y. Vardi, 1988, pp. 277–293.

[19] R. Fagin, J.Y. Halpern, Uncertainty, belief, and probability, Int. J. Comput. Intell. (1991) 160–173.

[20] R. Fagin, J.Y. Halpern, Reasoning about knowledge and probability, J. ACM (1994) 340–367.

[21] D.F. Ferraiolo, R. Sandhu, S. Gavrila, D.R. Kuhn, R. Chandramouli, Proposed NIST standard for role-based access control, ACM Trans. Inf. Syst. Secur. (2001) 224–274.

[22] M. Fitting, L. Thalmann, A. Voronkov, Term-modal logics, Stud. Log. (2001) 133–169, https://doi.org/10.1023/A:1013842612702.

[23] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations, Minds Mach. (2018) 689–707, https://doi.org/10.1007/s11023-018-9482-5.

[24] Q. Gao, D. Hajinezhad, Y. Zhang, Y. Kantaros, M.M. Zavlanos, Reduced variance deep reinforcement learning with temporal logic specifications, in: Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems, ICCPS '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 237–248.

[25] F.A. Genco, Giuseppe Primiero, A typed lambda-calculus for establishing trust in probabilistic programs, CoRR, arXiv:2302.00958 [abs], 2023, https://doi.org/10.48550/arXiv.2302.00958.

[26] S. Ghilezan, J. Ivetić, S. Kašterović, Z. Ognjanović, N. Savić, Probabilistic reasoning about simply typed lambda terms, in: International Symposium on Logical Foundations of Computer Science, Springer, 2018, pp. 170–189.

[27] N. Gierasimszuk, Bridging learning theory and dynamic epistemic logic, Synthese 169 (2009) 371–374.

[28] V. Goranko, A. Rumberg, Temporal logic, in: Edward N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2022 edition), 2022.

[29] High-Level Expert Group on Artificial Intelligence set up by the European Commission, Ethics Guidelines for Trustworthy AI, 2018.

[30] M.Z. Kwiatkowska, G. Norman, D. Parker, Prism: probabilistic symbolic model checker, in: Proceedings of the 12th International Conference on Computer Performance Evaluation, Modelling Techniques and Tools, TOOLS '02, Springer-Verlag, Berlin, Heidelberg, 2002, pp. 200–204.

[31] B. Kooi, Knowledge, Chance, and Change, Ph.D. Thesis, Groningen University, 2003.

[32] B. Kooi, Probabilistic dynamic epistemic logic, J. Log. Lang. Inf. 12 (2003) 381–408.

[33] C. Liau, Belief, information acquisition, and trust in multi-agent systems-a modal logic formulation, Artif. Intell. 149 (1) (2003) 31–60.

[34] C. McLeod, Trust, in: Edward N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, 2021 (Fall 2021 Edition), https://plato.stanford.edu/archives/fall2021/entries/trust/.

[35] E. Mosqueira-Rey, E. Hernández = Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, A. Fernández-Leal, Human-in-the-loop machine learning: a state of art, Artif. Intell. Rev. 56 (2023) 3005–3054.

[36] N.J. Nilsson, Probabilistic logic, Artif. Intell. 28 (1) (1986) 71–87.

[37] G. Primiero, A logic of negative trust, J. Appl. Non-Class. Log. 30 (3) (2020) 193–222.

[38] G. Primiero, F.A. D'Asaro, Proof-checking bias in labeling methods, in: Guido Boella, Fabio Aurelio D'Asaro, Abeer Dyoub, Giuseppe Primiero (Eds.), Proceedings of 1st Workshop on Bias, Ethical AI, Explainability and the Role of Logic and Logic Programming (BEWARE 2022) Co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AI*IA 2022), Udine, Italy, December 2, 2022, in: CEUR Workshop Proceedings, vol. 3319, 2022, pp. 9–19, http://ceur-ws.org/Vol-3319/paper1.pdf.

[39] G. Primiero, M. Taddeo, A modal type theory for formalizing trusted communications, J. Appl. Log. 10 (1) (2012) 92–114, https://doi.org/10.1016/j.jal.2011.12.002.

[40] S.A. Seshia, D. Sadigh, S.S. Sastry, Toward verified artificial intelligence, Commun. ACM 65 (2022) 46–55, https://doi.org/10.1145/3503914.

[41] M. Taddeo, M. Ziosi, A. Tsamados, L. Gilli, S. Kurapati, Artificial Intelligence for National Security: The Predictability Problem, CETaS Research Reports, September 2022.

[42] A. Termine, A. Antonucci, G. Primiero, A. Facchini, Logic and model checking by imprecise probabilistic interpreted systems, in: A. Rosenfeld, N. Talmon (Eds.), Multi-Agent Systems - 18th European Conference, EUMAS 2021, Virtual Event, June 28-29, 2021, Revised Selected Papers, in: Lecture Notes in Computer Science, vol. 12802, Springer, 2021, pp. 211–227.

[43] A. Termine, G. Primiero, F.A. D'Asaro, Modelling accuracy and trustworthiness of explaining agents, in: S. Ghosh, T. Icard (Eds.), Logic, Rationality, and Interaction - 8th International Workshop, LORI 2021, Xi'ian, China, October 16-18, 2021, Proceedings, in: Lecture Notes in Computer Science, vol. 13039, Springer, 2021, pp. 232–245.