Probabilistic Rule-Based Argumentation for Norm-Governed Learning Agents*

Régis Riveret[1], Antonino Rotolo[2], and Giovanni Sartor[2,3]

[1] Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, London, United Kingdom
[2] CIRSFID, University of Bologna, Bologna, Italy
[3] European University Institute, Florence, Italy

Abstract. This paper proposes an approach to investigate norm-governed learning agents which combines a logic-based formalism with an equation-based counterpart. This dual formalism enables us both to describe the reasoning of such agents and their interactions using argumentation, and, at the same time, to capture systemic features using equations. The approach is applied to norm emergence and internalisation in systems of learning agents. The logical formalism is rooted into a probabilistic defeasible logic instantiating Dung's argumentation framework. Rules of this logic are attached with probabilities to describe the agents' minds and behaviours as well as uncertain environments. Then, the equation-based model for reinforcement learning, defined over this probability distribution, allows agents to adapt to their environment and self-organise.

1 Introduction

Research on agent-based social simulation results from the intersection of agent-based computing, social sciences, and computer simulation: it designs social simulations that are based on agent-based modeling, and implements them using artificial agent technologies. A more specific development of this approach is multi-agent based simulation (MABS) which is a research effort which brings together researchers within the agent-based social simulation community and the multi-agent systems community. Multi-agent based simulation has also addressed normative phenomena: the emergence, spreading, and internalisation of norms, the dynamics of institutions and organisations, the development of co-operation and competition in norm-governed frameworks, etc. [7].

MABS typically differs from macro simulation techniques, since it aims at explicitly modeling the individual behaviour of agents. In contrast, macro simulations are instead oriented to exploiting powerful mathematical models (e.g., equation-based and stochastic methods) where the characteristics of a large population are averaged together to model systemic features of the whole population

* Pre-final Version. Part of this work has been carried out in the scope of the EC co-funded project SMART (FP7-287583).

[6]. Macro simulations may have some computational advantages, but they fail, in comparison to MABS, to provide a qualitative internal description of the agents.

In this paper, we address such issues by investigating a probabilistic rule-based argumentation which presents a *dual formalism* since it is logic-based and, at the same time, an equation-based representation can be extracted from it. So, we have a formal framework that allows us to describe entities and their interaction at the micro level using a logic, and, at the same time, we can extract from this logical description an equation-based representation useful to investigate systemic properties at the macro level.

Moreover, since social-cognitive models are encoded using a defeasible logic, they have a formal representation which allows us to obtain faithful operational simulations: the specifications of a social-cognitive model are directly executable by a machine.

We illustrate our framework with two simple examples regarding norm emergence and internalisation in agent societies. We will describe agents in a logical way and consider how they can acquire mental states, engage in individual and social behaviour and adapt to an uncertain environment. Agents operate in a setting where uncertain states and their dynamics are handled in a probabilistic fashion. In particular, we assume that agents revise probabilities to deal with uncertainty: agents' mental states and actions are thus probable, and their dynamics is non-deterministic.

Following the view of J. Pollock [15], the motivation for defeasible reasoning is related to the limitations of the cognitive capacities of agents. Indeed, a bounded agent must be able to adopt beliefs on the basis of incomplete information and to reach provisional conclusions while continuing its inquiries. In our approach, defeasible reasoning is formalised via a common rule-based defeasible logic instantiating Dung's abstract argumentation framework, which provides for "ergonomic" dialectical constructions, particularly useful when experts from different disciplines or opinions are modelling or building a system. However, defeasible logic is insufficient for the purpose of non-deterministic simulation, since iteratively running an inference engine on the same knowledge base leads always to the same outcome. For this purpose we shall combine defeasible reasoning with probabilities, using the framework presented in [18], where a logic instantiating Dung's argumentation framework is coupled with probabilities.

On the basis of an integration of rule-based argumentation and probabilities, we address agents' social adaptation to their uncertain environment under the following assumptions:

- the environment is non-deterministic;
- agents do not have an explicit model of the environment (but they are allowed to build by themselves such a model);
- the adaptation process is based on learning mechanisms;
- agents have bounded rationality;
- they are fallible.

We also assume that agents have only an approximate valuation of their action's utilities: they have limited cognitive abilities, limited access to information and a finite amount of time for deliberating and acting. They are rational but fallible: they can be practically mistaken, in the sense that, given the utility distribution of possible actions, they may not choose the best action. At the same time, those mistakes allow agents to explore new combinations that may turn out to be very useful in some circumstances.

To meet all these assumptions, we shall propose a learning mechanism based on the idea of reinforcement learning: an agent chooses its action based on a probability distribution over the rules controlling its mental states and this probability distribution is updated at each instant based on the feedback from the environment.

The overall stack of techniques we shall deploy in our approach is given in Figure 1.

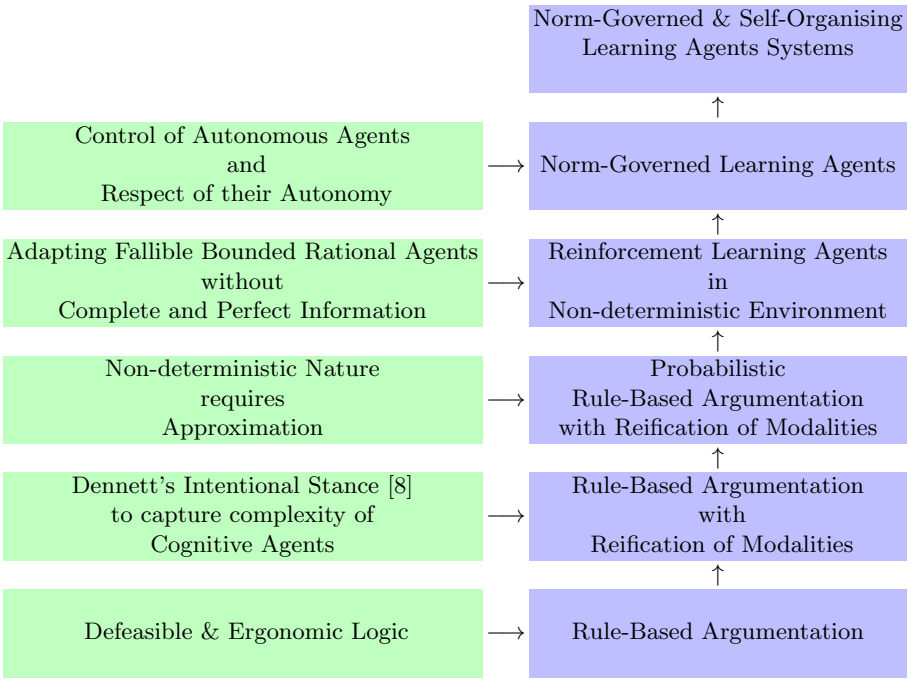

Fig. 1. This diagram shows the layered architecture of our approach where each layer addresses some requirements by integrating techniques layer-by-layer.

The proposed method is applied to two normative phenomena: norm internalisation (see [2]) and norm emergence (see [22]). The first phenomenon is captured by representing explicit and exogenous norms adopted by the agents' though de-

feasible rules. Such norms may be set by the designer of the MABS, but their probabilities change as the agents interact with other agents and with the environment. Such probabilities identify different degrees of norm internalisation (e.g.: if a norm is in the mental description of agents and the probability is 1, the corresponding norm is fully internalized; if the value is 0 the norm is deactivated in the mental profile of the agent, and so it is known but not internalized by the agent).

However, we want to model also emerging and unplanned social norms. For this purpose, we shall show how norms and conventions result from the agents' mental and behavioural adaptation over time, since the logic also represents the cognitive profile of agents. In this perspective, we are aligned to [1]'s view in that emerging norms "rather than mere behavioral regularities" are "behaviors spreading to the extent that and because the corresponding commands and beliefs do spread as well".

Furthermore, while rules emerge because they reduce the cost involved in face-to-face personal influence, we also want to cater for the emergence of inefficient norms. For this reason, our learning mechanism does not necessary lead the system to converge towards an optimal equilibrium; it rather simulates a "realistic" non-deterministic agents adaptation. Although our learning mechanism is not aimed at making the system converge towards a social optimal equilibrium, it remains interesting to know the social optimum. This social optimum can be computed via common algebra or using algorithms from artificial evolution.

The paper is organized as follows. After an introductory example in Section 2, the rule-based argumentation framework is given in Section 3, and its combination with probabilities is proposed in Section 4. In Section 5, we propose a minimal structure for any theory catering for multi-agent systems and discuss computational perspectives leading us to consider learning agents. We illustrate the animation of agents societies by two examples in Section 6. Finally, we relate our results with relevant approaches and work in Section 7 before concluding.

2 An example

In this section we present an introductory an example which illustrates the basic intuitions behind our probabilistic framework. The example addresses civil liability, assuming that there is uncertainty both about the facts and about the applicable rules.

Let us start with the rules. Assume that compensation for economic damages is always granted (as indicated by probability 1 of rule r_1), while compensation is acknowledged only sometimes for health damages (as indicated by probability 0.8 of rule r_2), and no compensation is usually given for moral damages (as indicated by probability 0.2 of rule r_3).

$$\begin{array}{ll} r_1, 1: & economicDamage \Rightarrow economicCompensation \\ r_2, 0.8: & healthDamage \Rightarrow healthCompensation \\ r_3, 0.2: & moralDamage \Rightarrow moralCompensation \end{array}$$

Assume that the following forecasts can be made with regard to the judicial assessment of the facts in a case concerning a car accident: it will almost certainly be established that there was an economic damage (that the car of the plaintiff was destroyed as a consequence of the incident), there is some chance that it will be established that there was a health damage (that the persisting backache of the plaintiff was caused by the accident), and some chance that it will be established that there was a moral damage (that the plaintiff has lost the opportunity of taking a holiday because of the accident). The following facts encode these assumptions:

$$\begin{array}{ll} f_1, 0.9 : & \Rightarrow economicDamage \\ f_2, 0.2 : & \Rightarrow healthDamage \\ f_3, 0.6 : & \Rightarrow moralDamage \end{array}$$

Now the issue we want to address is the following: how likely is that the plaintiff will get those different kinds of compensation when going to court?

We shall handle this problem by considering all possible theories that can be constructed by using these rules, and assigning a probability to each of such theories. The probability that a conclusion ϕ holds will then be the sum of the probability of all theories defeasibly entailing ϕ.

The probability of a theory containing rules $\{q_1, \ldots q_m\}$ and not containing rules $\{q_{m+1}, \ldots q_n\}$ is a joint probability. If we consider that rules are independent, then this joint probability will be determined by multiplying the probabilities of rules in $\{q_1, \ldots q_m\}$ and the complements of the probabilities of the rules in $\{q_{m+1}, \ldots q_n\}$ (namely, the probability that such rules are not adopted). Thus, for instance, the probability that economic damage is accepted and that compensation is given on this basis is the probability of the set of theories with the following set of rules $\{r_1, f_1\}, \{r_1, f_1, r_2\}$, $\{r_1, f_1, f_2\}, \{r_1, f_1, r_2, f_2\}, \{r_1, f_1, r_3\}, \ldots$, that is, the set of all theories including the rules $\{r_1, f_1\}$.

According to the above information, *economicCompensation* is justified with probability 0.9, which is the product of 1×0.9. The probability of having moral compensation is instead: $0.6 \times 0.2 = 0.12$. These results are obtained since in the given set of rules, there is no conflict, and therefore whenever a possible theory contains an argument for ϕ, it will justify ϕ regardless of the inclusion or exclusion of other rules.

The situation will be different if there is a 0.5 probability that the theory of contributory negligence is adopted, according to which no compensation is due if the plaintiff contributes to the damage, as expressed by the following rules

$$\begin{array}{l} r_4, 0.5 : contributionToDamage \Rightarrow \neg economicalDamage \\ r_5, 0.5 : contributionToDamage \Rightarrow \neg healthDamage \\ r_6, 0.5 : contributionToDamage \Rightarrow \neg moralDamage \end{array}$$

which prevail over all other rules r_1, r_2, r_3 and which are coupled with the fact

$$f_4, 0.5 : \Rightarrow contributionToDamage$$

Then the relevant probabilities are the probabilities of the theories that do contain arguments justifying compensation for a certain kind of dosage, but do not

contain r_4, r_5, or r_6 together with f_4. For instance, given these additional rules, the probability of obtaining compensation for economic damage, will go down to $1 \times 0.9 \times (1 - 0.5 \times 0.5) = 0.675$.

Note that the calculations we have made assume that the rules are stochastically independent. This is appropriate only in some contexts. In other contexts the choice is not among single rules, but rather among alternative clusters of interconnected rules. In the following we shall consider both possibilities, namely, that rules are stochastically independent and that they are not. The difference is that stochastic independence allows us to assign probabilities to any possible theory constructible out of a set of probable rules as a function of the probability of the rules in the theory. This is not allowed when the rules are not stochastically independent, occurring in related clusters. Thus, in the remainder we shall both provide a general theory for addressing such cases and provide a method for computing probabilities.

A further aspect we shall address concerns the evolution of rules' probabilities. The probability that a rule is applied can change dramatically over time, even without a legislative intervention: for instance in Italy, until the beginning of the 1980's one could most probably get only compensation for economic damage, while at a later point, while still being possible that only economic damages were obtained, it was more probable to get also health damages. The evolution of rules on civil liability took place in a very complex and highly institutionalised domain, through a learning process involving citizens, lawyers and judges. In our analysis of normative evolution we shall use much simpler examples, which do not include these complex institutional and social dimensions. This choice is motivated by the need of keeping the framework manageable and the presentation clear.

However, we think that our model provides the basis for further developments. Our basic assumption is that agents experiment with rules, and that this experimentation is guided by the fact that certain rules (or rule clusters) are more successful than others, providing these agents with a higher utility (a utility for the individual agents or for their society), which increases the probability of those rules. We think that this idea, while being directly applicable to simple scenarios, can also be applied to more complex situations, when expanded with further learning mechanisms and normative structures (such as goals and values). But this will be left to future research.

3 Argumentation framework

In this section, a rule-based argumentation framework is introduced. We see how arguments are built from defeasible theories. Then a fixed-point semantics defines justified arguments. Finally, the dialogue game used to compute justified arguments is provided at the end of the section.

First of all, let us introduce some basic language and terminology. We start with literals, and then define rules, rule preference and rule theories.

Definition 1 (Language). *Let* Atoms *be a set of atomic formulas and* Lbl *a set of labels.*

Literals *The set* Lit *of literals consists of all atoms and their negations (we use* $\pm\phi$ *to cover positive and negative literals, namely,* ϕ *and* $\neg\phi$*):*

$$\mathsf{Lit} = \{\pm p | p \in \mathsf{Atoms}\}$$

Pure defeasible rules *A pure defeasible rule has the form*

$$r : \phi_1, \ldots, \phi_n \Rightarrow \phi$$

where $r \in \mathsf{Lbl}$*, and* $\phi_1, \ldots, \phi_n, \phi \in \mathsf{Lit}$*. Informally, this is a rule with identifier* r*, stating that if* $\phi_1, \ldots, \phi_n$ *hold then* ϕ *presumably holds. A rule with no antecedent, is written* $r :\Rightarrow \phi$ *or simply* $r : \phi$*.*

Preference ordering *Let* R *be a set of rules. Then* $\succ$ *is an antisymmetric partial order over* R*, i.e., if* $r \succ r'$ *then* $r' \not\succ r$*. Informally, a rule preference* $r_1 \succ r_2$ *states that rule* r_1 *prevails over rule* r_2*.*

Pure defeasible theories *A* pure defeasible theory *is a tuple* $\langle R, S\rangle$ *where* R *is a set of pure defeasible rules, and* S *is a set of preferences.*

By combining the pure defeasible rules in a theory, we can build arguments (we use the definition in [16], simplified to take into account that we just have one type of premises, namely, rules). In what follows, for a given argument, Conc returns its conclusion, Sub returns all its sub-arguments, Rules returns all the rules in the argument and, finally, TopRule returns the last inference rule in the argument.

Definition 2 (Argument). *An argument* A *constructed from a pure theory* $\langle R, \succ\rangle$ *has the form* $A_1, \ldots A_n \Rightarrow_r \phi$*, where* $A_1, \ldots, A_n$ *are arguments constructed from* $\langle R, \succ\rangle$ *and* $r : \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \phi$ *is a rule in* R*. With regard to argument* A*, the following holds:*

$$\begin{array}{l}
\mathsf{Conc}(A) = \phi \\
\mathsf{Sub}(A) = \mathsf{Sub}(A_1) \cup \ldots \cup \mathsf{Sub}(A_n) \cup A \\
\mathsf{TopRule}(A) = r : \mathsf{Conc}(A_1), \ldots, \mathsf{Conc}(A_n) \Rightarrow \phi \\
\mathsf{Rules}(A) = \mathsf{Rules}(A_1) \cup \ldots \cup \mathsf{Rules}(A_n) \cup \{\mathsf{TopRule}(A)\}
\end{array}$$

The following example illustrates the notions just introduced.

Example 1. Given a rule set

$$R = \{r_1 : \Rightarrow a; \quad r_2 : \Rightarrow b; \quad r_3 : a, b \Rightarrow c\}$$

we have arguments:

$$\begin{array}{l}
A_1 : \Rightarrow_{r_1} a \\
A_2 : \Rightarrow_{r_2} b \\
A_3 : A_1, A_2 \Rightarrow_{r_3} c
\end{array}$$

Let us now consider conflicts between arguments. In some argumentation systems, two kinds of conflict are considered, rebuttal (clash of incompatible conclusions) and undercutting (attacks on inferences). For our purposes, only rebuttals are needed. To introduce rebuttals we need to define incompatible literals. Let Lit be a set of literals, ψ an atom. We assume a function $-$, which works with Lit and returns the set of literals which are incompatible for a given literal. For the moment we assume that $-$ returns a singleton: $-\psi = \neg\psi$ and $-\neg\psi = \psi$. Conflicts between contradictory argument conclusions are resolved on the basis of preferences over arguments using a simple last-link ordering according which an argument A is preferred over another argument B, denoted as $A \succ B$, if, and only if, the rule $\mathsf{toprule}(A)$ is preferred to the rule $\mathsf{toprule}(B)$ (i.e. $\mathsf{toprule}(A) \succ \mathsf{toprule}(B)$).

Definition 3 (Defeats). *An argument B defeats an argument A if, and only if, $\exists A' \in \mathsf{Sub}(A)$, such that $\mathsf{Conc}(B) = -\mathsf{Conc}(A')$, and $A' \not\succ B$.*

Our semantics for argumentation frameworks is based on the Dung's grounded semantics [9] though other semantics could certainly be investigated.

Definition 4 (Argumentation framework and semantics).

Argumentation framework *An argumentation framework is a pair $\langle \mathcal{A}, \gg \rangle$ where $\mathcal{A}$ is a set of arguments, and $\gg \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of defeat. In particular we assume that for any arguments A and B, $B \gg A$ if, and only if, B defeats A according to Definition 3.*

Conflict-free set *A set $\mathcal{S}$ of arguments is said to be* conflict-free *if, and only if, there is no argument A and B in $\mathcal{S}$ such that B defeats A.*

Acceptable argument *An argument A is acceptable w.r.t. a set of arguments $\mathcal{S}$ if and only if any argument defeating A is defeated by an argument in $\mathcal{S}$.*

Characteristic function *The characteristic function, denoted F_{AF}, of an argumentation framework $AF = \langle \mathcal{A}, \gg \rangle$, is defined as $F_{AF} : 2^{\mathcal{A}} \Rightarrow 2^{\mathcal{A}}$ and $F_{AF}(\mathcal{S}) = \{A |$ A is acceptable w.r.t. $\mathcal{S} \subseteq \mathcal{A}\}$.*

Admissible set *A conflict-free set $\mathcal{S}$ of arguments is admissible if and only if $\mathcal{S} \subseteq F_{AF}(\mathcal{S})$. If a set $\mathcal{S}$ is admissible then we write $\mathsf{adm}(\mathcal{S})$. The family of the admissible sets of an argumentation framework AF_T built from a pure theory T is denoted as $\mathsf{Adms}(T)$.*

Grounded extension *A grounded extension $GE(AF)$ of an argumentation framework AF is the least fixed-point of F_{AF}. The grounded extension of an argumentation framework AF_T built from a pure theory T is also denoted as $GE(T)$. If there exists an argument A in $GE(T)$ such that $\mathsf{Conc}(A) = \phi$ we say that T defeasibly entails ϕ and write $T \vdash_{GE} \phi$.*

Justified argument and conclusion *An argument A and its conclusion are justified with regard to an argumentation framework AF if, and only if, $A \in GE(AF)$.*

The following example illustrates the notions just introduced.

Example 2. Given a theory $T = (R, \succ)$ where

$$\begin{aligned} R = \{\ & r_1 : \quad \Rightarrow a, \\ & r_2 : \quad \Rightarrow b, \\ & r_3 : a, b \Rightarrow c, \\ & r_4 : \quad \Rightarrow d, \\ & r_5 : d \Rightarrow \neg c\ \} \\ \succ = \{\ & r_3 \succ r_5\ \} \end{aligned}$$

The grounded extension $GE(T)$ contains the following arguments:

$$\begin{aligned} A_1 &: \ \Rightarrow_{r_1} a \\ A_2 &: \ \Rightarrow_{r_2} b \\ A_3 &: A_1, A2 \Rightarrow_{r_3} c \\ A_4 &: \ \Rightarrow_{r_4} d \end{aligned}$$

To establish whether an argument is provably justified, we shall use the dialogue game of [17], according to which a dialogue between a proponent pro and a opponent opp is a finite nonempty sequence of moves move$(i) = (Player_i, A_i)$ with $0 < i$ such that

- $Player_i =$ pro if i is odd, and $Player_i =$ opp if i is even,
- the argument A_i defeats the argument A_{i-1},
- if $Player_i = Player_j =$ pro then A_i is different than A_j.

A dialogue tree is a finite tree of moves such that each branch is a dialogue and if $Player_i =$ pro then the children of $move_i$ are all the defeaters of A_i. A player wins a dialogue tree if the other player cannot move, and a player wins a dialogue tree if it wins all the branches of the dialogue tree. An argument A is provably justified if, and only if, there is a dialogue tree with A as its root, and won by the proponent, else A is rejected. All provable justified argument are justified., i.e. belongs to the grounded extension. An argument is justified w.r.t. the grounded extension if and only if it belongs to the grounded extension, it is rejected if and only if it does not belong to it.

A more efficient dialogue in terms of computation is obtained by prohibiting the proponent from moving an argument A that is itself attacked by the arguments that the proponent moves against.

4 Probabilistic framework

In this section, we integrate probabilistic reasoning with the rule-based argumentation framework presented so far, developing the framework sketched in [19]. We combine argumentation and probabilities in two ways:

– When rules are stochastically independent, each of them may individually receive a probability. A probabilistic defeasible theory is thus defined as theory consisting of a set of probabilistic rules.
– When rules are not independent, then we will assign probability values to the theories where such rules occur, and the probability of a rule is thus a marginal probability (i.e. the probability of one rule irrespective of the probability of other rules).

Next, we will offer an analysis of the intuitive meaning of probability of rules and theories. To do so, we first consider empirical probabilities to set a clear interpretation of rules' probabilities (Section 4.1), then, we deal with the related theoretical probabilities using a Kolmorogov setting (Section 4.2). Finally, an intensional computational technique tracking dialogue trees is proposed in case of stochastically independent rules (Section 4.4). Notice that the method proposed in Section 4.2 works fine even when rules are not independent: however, there is a computational cost.

4.1 Empirical probabilities

In this section we provide an empirical foundation for the initialisation of rule probabilities. We start considering a collection of sample theories obtained through empirical inquiries or experiments (e.g. the theories that agents in a certain population are likely to endorse or practice).

Given a multiset Γ of pure defeasible theories, each theory T_i consisting in a set of rules R_i and a set of preferences among such rules S_i (where same sample theory may occur multiple times) we collect all rules and preferences in such theories into two sets, denoted by $\mathsf{rul}(\Gamma)$ and $\mathsf{sup}(\Gamma)$. If $\Gamma = \{\langle R_1, S_1\rangle, \ldots, \langle R_n, S_n\rangle\}$ then

$$\mathsf{rul}(\Gamma) = \bigcup_{i=1}^{n} R_i \qquad\qquad \mathsf{sup}(\Gamma) = \bigcup_{i=1}^{n} S_i$$

For simplicity, let us assume that the preference set is fixed, so that the preference set of each sample theory coincides with $\mathsf{sup}(\Gamma)$.

We can now proceed to assign probabilities to every rule in $\mathsf{rul}(\Gamma)$, by considering the number of sample theories in Γ which contain r. The empirical probability $\pi(r)$ that a rule r appears in a sample set Γ of theories is:

$$\pi(r) = |\Gamma_r|/|\Gamma|$$

where $\Gamma_r = \{T | T \in \Gamma, r \in \mathsf{rul}(T)\}$. Rules with probability 1 would appear in every theory whereas rules with probability 0 would appear in no theory. It is on this basis that we can introduce the notions of probabilistic defeasible rule and probabilistic defeasible theory.

Definition 5 (Probabilistic defeasible rule and theories). *A probabilistic defeasible rule has the form*

$$\pi, r: \quad \phi_1, \ldots, \phi_n \Rightarrow \phi$$

where π is a probability assignment, $r \in \mathsf{Lbl}$, and $\phi_1, \ldots, \phi_n, \phi \in \mathsf{Lit}$. A probabilistic defeasible theory is a tuple $\langle R, \succ \rangle$ of a set of defeasible rules and a set of preferences over them.

Let us now consider the rules extracted from Γ, and the corresponding empirical probabilistic defeasible theory.

Definition 6 (Empirical probabilistic defeasible theory). *The set $\mathsf{probrul}(\Gamma)$ of the probabilistic rules from Γ, contains any rule in $\mathsf{rul}(\Gamma)$ expanded with the appropriate probability:*

$$\mathsf{probrul}(\Gamma) = \{(\pi, r) | r \in \mathsf{rul}(\Gamma) \wedge \pi = \pi(r)\}$$

The empirical probabilistic defeasible theory of a sample multiset Γ is the probabilistic defeasible theory $\langle R, S \rangle$ such that $R = \mathsf{probrul}(\Gamma)$ is the set of probabilistic defeasible rules of Γ and $S = \mathsf{sup}(\Gamma)$ is the corresponding set of preferences.

Here is an example illustrating these concepts.

Example 3. Let us have the following sample multiset (with no preferences):

$$\Gamma = \{\langle \{r_1, r_2, r_4\}, \emptyset \rangle, \langle \{r_1, r_2, r_4\}, \emptyset \rangle, \langle \{r_2, r_3, r_4\}, \emptyset \rangle, \langle \{r_2, r_3, r_4\}, \emptyset \rangle\}$$

Though this multiset is not large enough to be statistically relevant, we use it to illustrate the concepts presented so far: $\mathsf{rul}(\Gamma) = \{r_1, r_2, r_3, r_4\}$, $\mathsf{sup}(\Gamma) = \emptyset$ and $\mathsf{probrul}(\Gamma) = \{(0.5, r_1), (1, r_2), (0.5, r_3), (1, r_4)\}$. The probabilistic theory of Γ is $\langle \mathsf{probrul}(\Gamma), \emptyset \rangle\rangle$.

Finally, the empirical probability of the admissibility of a set $\mathcal{S}$ of arguments is simple: count the number of sample theories where the set $\mathcal{S}$ is admissible, and divide this number by the size of the multiset Γ of sample theories:

$$P(\mathsf{adm}(\mathcal{S})) = |\Gamma_{\mathsf{adm}(\mathcal{S})}|/|\Gamma|$$

where $\Gamma_{\mathsf{adm}(\mathcal{S})} = \{T | \mathcal{S} \in \mathsf{adm}(T), T \in \Gamma\}$. Accordingly, the empirical evaluation of $P(Just(A))$ is:

$$P(Just(A)) = |\Gamma_A|/|\Gamma|$$

where $\Gamma_A = \{T_i | T_i \in \Gamma_A, A \in GE(T_i)\}$.

The empirical construction of a probabilistic theory from a sample multiset provides us with a frequentist semantics for rules' probabilities. Though a probabilistic theory does not cater for rules' conditional probabilities, marginal probabilities give important indications on the frequency of rules in a multiset.

4.2 Theoretical approach

We base our theoretical approach on Kolmorogov's framework where the sample space is Ω, an algebra on Ω is a set $F(\Omega)$ of all subsets of Ω (Ω belongs to

$F(\Omega)$), (F is closed under union and complementation w.r.t. Ω). and, finally we assume a probability function P be a probability function from $F(\Omega)$ to $[0,1]$:

$$P(\mathcal{T}) = \sum_{T \in A} P(\mathcal{T}) \tag{1}$$

A sample space can be build from a multiset Γ by gathering all rules and preferences in Γ. Let Ω_Γ denote all pure theories constructible from Γ:

$$\Omega_\Gamma = \{\langle R, S\rangle \,|R \subseteq \mathsf{rul}(\Gamma) \wedge S = \mathsf{sup}(\Gamma)\}$$

In the remainder, we call these theories constructible theories.

On this basis, and before moving on the proper probability of admissible sets and thus justification, we need to deal with the probability $P(T)$ of a pure theory T which is the following joint probability:

$$P(T) = P(\bigwedge_{r_i \in \mathsf{Rul}(T)} r_i \in \mathsf{Rul}(T) \bigwedge_{r_j \in \mathsf{Rul}(\Omega_\Gamma) \setminus \mathsf{Rul}(T)} r_j \notin \mathsf{Rul}(T)) \tag{2}$$

If the rules are stochastically independent, then:

$$P(T) = \prod_{r \in \mathsf{Rul}(T)} \pi(r). \prod_{r \in \mathsf{Rul}(\Omega) \setminus \mathsf{Rul}(T)} [1 - \pi(r)] \tag{3}$$

Unless otherwise specified, we do not assume in the remainder that rules are stochastically independent. Of course, if the rules are not assumed independent then we need rules' conditional probabilities, and that information may not be easily available. For this reason, given a partition $\{\mathcal{T}_1, \ldots, \mathcal{T}_n\}$ of the set Ω of pure theories, any set $\mathcal{T}_j$ of pure theories is attached with a potential $Q(\mathcal{T}_j)$, and its probability is defined using an exponential model (here, a Boltzmann distribution) of the form:

$$P(\mathcal{T}_j) = e^{Q(\mathcal{T}_j)} / \sum_{\mathcal{T}_i} e^{Q(\mathcal{T}_i)} \tag{4}$$

So, $\sum_{\mathcal{T}_i} P(\mathcal{T}_i) = 1$. Roughly speaking, the quality of a set of pure theories can be thought as a glue for rules forming arguments: the higher is the quality of a set, the more the set elements are stuck together, the higher the probability of the set. We encode in Section 5 some utilities into defeasible theories themselves so that we can attach some potentials to pure theories.

4.3 Probability of justification, of conclusions, and of admissible sets

Let us now define the probability of justification (or the rejection) of arguments, and consequently of related conclusions, as well as the probability of admissible sets. Since the grounded extension is based on the least complete extension, we begin our presentation with the probability of an admissible set of arguments.

Given a probability space $(\Omega, F(\Omega), P)$, *the probability that a set $\mathcal{S}$ of arguments is admissible* is the probability of the set of constructible theories in which $\mathcal{S}$ is admissible:

$$P(\mathsf{adm}(\mathcal{S})) = P(\bigcup_{T \in \Omega : \mathcal{S} \in \mathsf{Adms}(T)} T) \tag{5}$$

using the Kolmorogov finite additivity axiom, we obtain:

$$P(\mathsf{adm}(\mathcal{S})) = \sum_{T \in \Omega : \mathcal{S} \in \mathsf{Adms}(T)} P(T) \tag{6}$$

This result answers the question: what is the probability of an admissible set $\mathcal{S}$? Or if the sample space has been built from a multiset Γ, what is the probability to pick up a theory from Γ where the set $\mathcal{S}$ is admissible? Next, we will investigate the case of a particular case of admissible set, the grounded extension.

What is the probability of justification of an argument or a conclusion w.r.t. a probabilistic theory? To answer this question, we need to consider the unique admissible set which is the grounded extension and which contains the argument we are looking at. Thus, *the probability of justification of an argument A* is $P(Just(A)) = P(\mathsf{adm}(\mathcal{S}))$ where $\mathcal{S}$ is the grounded extension.

The probability of a literal ϕ being justified/rejected w.r.t. a universe Ω, written $P(Just(\phi))/P(Rej(\phi))$, is the probability of the set of possible constructible theories in which ϕ is justified/rejected:

$$P(Just(\phi)) = P(\{T | T \in \Omega, T \vdash_{GE} \phi\})$$

$$P(Rej(\phi)) = P(\{T | T \in \Omega, T \nvdash_{GE} \phi\})$$

or, equivalently using equation 1:

$$P(Just(\phi)) = \sum_{T \in \Omega : T \vdash_{GE} \phi} P(T)$$

$$P(Rej(\phi)) = \sum_{T \in \Omega : T \nvdash_{GE} \phi} P(T).$$

Thus, the larger the proportion of constructible theories in Ω where ϕ is justified, the higher the probability that ϕ is justified.

Example 4 (Running example). In this example we use atoms having the form $\mathsf{E}_a(b)$, indicating that agent a performs action b. In particular $\mathsf{E}_a(left)$ (or $\mathsf{E}_a(right)$) states that a drives on the left (or the right) side of the street. Let Γ be a multiset of pure defeasible theories from which we build a sample space Ω_Γ with $\mathsf{rul}(\Omega_\Gamma) = \{r_1, r_2, r_3, r_4\}$, and $\mathsf{sup}(\Omega_\Gamma) = \emptyset$ such that:

$0.5, r_1 : \Rightarrow \mathsf{E}_{Tom} left$	$1, r_3 : \mathsf{E}_{Tom} left \Rightarrow \neg \mathsf{E}_{Tom} right$
$0.5, r_2 : \Rightarrow \mathsf{E}_{Tom} right$	$1, r_4 : \mathsf{E}_{Tom} right \Rightarrow \neg \mathsf{E}_{Tom} left$

Thus, rules r_1 and r_2 appear in half of the theories in this sample set. The sample space Ω_Γ can be represented by a table where each column is a constructible theory:

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	T_{13}	T_{14}	T_{15}	T_{16}
r_1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
r_2	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
r_3	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
r_4	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
$Just(E_{Tom}left)$	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0
$Just(E_{Tom}right)$	0	1	0	0	1	1	0	0	0	1	0	0	1	1	0	0
$\emptyset$	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1

We can compute, among others, the probability of the set $\mathcal{T}$ of constructible theories in which $E_{Tom}left$ is justified. Let's assume stochastic independent rules:

$$\begin{aligned}
P(\mathcal{T}) &= P(T_3) + P(T_7) + P(T_9) + P(T_{11}) + P(T_{13}) + P(T_{15}) \\
&= \pi(r_1)\pi(r_3)\pi(r_4)[1-\pi(r_2)] + \pi(r_1)\pi(r_4)[1-\pi(r_2)][1-\pi(r_3)] \\
&\quad +\pi(r_1)\pi(r_3)[1-\pi(r_2)][1-\pi(r_4)] + \pi(r_1)[1-\pi(r_3)][1-\pi(r_2)][1-\pi(r_4)] \\
&\quad +\pi(r_1)\pi(r_2).\pi(r_3)[1-\pi(r_4)] + \pi(r_1)\pi(r_2)[1-\pi(r_3)][1-\pi(r_4)] \\
&= \pi(r_1)[1-\pi(r_2)]
\end{aligned}$$

$E_{Tom}left$ is justified in the set $\mathcal{T}$ of theories, thus $P(Just(E_{Tom}left) = P(\mathcal{T})$, that is, $P(Just(E_{Tom}left) = \pi(r_1)[1-\pi(r_2)]$ or $1/4$.

Suppose now that we arbitrarily attach the following potentials to theories denoted by their entailed justified actions $just(E_{Tom}left)$ and $just(E_{Tom}right)$, and the theories denoted $\emptyset$ to cater for agents' inaction: $Q(just(E_{Tom}left)) = 10$, $Q(just(E_{Tom}right)) = 10$, $Q(\emptyset) = 0$. Using a Boltzmann distribution, we have:

$$\begin{aligned}
P(just(E_{Tom}right)) &= 6.e^{10}/(6.e^{10} + 6.e^{10} + 5.e^{0}) \\
&\approx 0.5
\end{aligned}$$

Similarly, we can compute that $P(just(E_{Tom}left)) \approx 0.5$ and $P(\emptyset) \approx 0$.

In the remaining part of this section, we illustrate some results we can prove from the framework just provided.

Theorem 1. *Let $\langle \mathcal{A}, \gg \rangle$ be an argumentation framework. For any literal ϕ and $-\phi$ supported by some arguments in $\mathcal{A}$,*

$$P(Just(\phi)\, and\, Just(-\phi)) = 0 \qquad and\ P(Just(\phi)) + P(Just(-\phi)) \leq 1$$

Proof. Let $(\Omega, F(\Omega), P)$ a probability space. Consider the maximal sets of constructible theories $\mathcal{T} \in F(\Omega)$ such that $\forall T \in \mathcal{T}, T \vdash_{GE} \phi$ and $\mathcal{T}' \in F(\Omega)$ such that $\forall T' \in \mathcal{T}', T' \vdash_{GE} -\phi$. $\mathcal{T} \cap \mathcal{T}' = \emptyset$, thus $P(\mathcal{T} \cap \mathcal{T}') = 0$. Since $P(Just(\phi)\, and\, Just(-\phi)) = P(\mathcal{T} \cap \mathcal{T}')$, we obtain $P(Just(\phi)\, and\, Just(-\phi)) = 0$.

Proof. Let $(\Omega, F(\Omega), P)$ a probability space. Consider the maximal sets of constructible theories $\mathcal{T} \in F(\Omega)$ such that $\forall T \in \mathcal{T}, T \vdash_{GE} \phi$ and $\mathcal{T}' \in F(\Omega)$ such that $\forall T' \in \mathcal{T}', T' \vdash_{GE} -\phi$. We know that for any set of constructible theories

$A \subseteq \Omega$, $P(A) \leq P(\Omega)$. Since $\mathcal{T} \cup \mathcal{T}' \subset \Omega$, we have $P(\mathcal{T} \cup \mathcal{T}') \leq P(\Omega)$. However, $P(\Omega) = 1$, thus $P(\mathcal{T} \cup \mathcal{T}') \leq 1$. Furthermore, $\mathcal{T} \cap \mathcal{T}' = \emptyset$, thus $P(\mathcal{T} \cup \mathcal{T}') = P(\mathcal{T}) + P(\mathcal{T}')$. Consequently, $P(\mathcal{T}) + P(\mathcal{T}') \leq 1$. Since, $P(\mathcal{T}) = P(Just(\phi))$ and $P(\mathcal{T}') = P(Just(-\phi))$, we have $P(Just(\phi)) + P(Just(-\phi)) \leq 1$.

Notice that the concept of plausibility (c.f. Dempster-Shafer theory) is interpretable as $Plau(\phi) = 1 - P(Just(-\phi))$.

Theorem 2. *Let $\langle \mathcal{A}, \gg \rangle$ be an argumentation framework. For any literals ϕ supported by any argument in $\mathcal{A}$,*

$$P(Just(\phi)) \leq Plau(\phi)$$

Proof. Let $(\Omega, F(\Omega), P)$ a probability space. Since $P(Just(\phi)) + P(Just(\neg\phi)) \leq 1$ and $P(Plau(\phi)) = 1 - P(Just(\neg\phi))$, we obtain $P(Just(\phi)) \leq P(Plau(\phi))$

Theorem 3. *Let $\langle \mathcal{A}, \gg \rangle$ be an argumentation framework. For any literal ϕ supported by any argument in $\mathcal{A}$,*

$$P(Plau(\phi)) + P(Plau(-\phi)) \geq 1$$

Proof. Let $(\Omega, F(\Omega), P)$ a probability space.

$$\begin{aligned} & P(Plau(\phi)) + P(Plau(-\phi)) \\ & = 1 - P(Just(-\phi)) + 1 - P(Just(\phi)) \\ & = 2 - [P(Just(-\phi)) + P(Just(\phi))] \end{aligned}$$

However, $P(Just(\phi)) + P(Just(\neg\phi)) \leq 1$, consequently $P(Plau(\phi)) + P(Plau(-\phi)) \geq 1$.

Theorem 4. *Let $\langle \mathcal{A}, \gg \rangle$ be an argumentation framework. For any literal ϕ supported by any argument in $\mathcal{A}$,*

$$P(Just(\phi)) + P(Rej(\phi)) = 1$$

Proof. Let $(\Omega, F(\Omega), P)$ a probability space. Consider the maximal sets of constructible theories $\mathcal{T} \in F(\Omega)$ such that $\forall T \in \mathcal{T}, T \vdash_{GE} \phi$ and $\mathcal{T}' \in F(\Omega)$ such that $\forall \mathsf{T}' \in \mathcal{T}', T' \nvdash_{GE}$. Since $\mathcal{T} \cup \mathcal{T}' = \Omega$, we have $P(\mathcal{T} \cup \mathcal{T}') = P(\Omega)$. However, $P(\Omega) = 1$, thus $P(\mathcal{T} \cup \mathcal{T}') = 1$. Furthermore, $\mathcal{T} \cap \mathcal{T}' = \emptyset$, thus $P(\mathcal{T} \cup \mathcal{T}') = P(\mathcal{T}) + P(\mathcal{T}')$. Consequently, $P(\mathcal{T}) + P(\mathcal{T}') = 1$. Since, $P(\mathcal{T}) = P(Just(\phi))$ and $P(\mathcal{T}') = P(Rej(\phi))$, we have $P(Just(\phi)) + P(Rej(\phi)) = 1$.

Before moving on the multi-agent system framework, we will give in the next section a more compact form of probabilities computation based on dialogue trees, which is only valid for stochastically independent rules.

4.4 Dialogue-based computation

In the previous section, we defined the probability of justification w.r.t. a probability space $(\Omega, F(\Omega), P)$, as the probability of the set of possible theories in which an argument or a conclusion is justified. This definition is extensional in the sense that the probability is defined by enumerating the possible theories in which a literal is justified. This method is thus inefficient because the number of possible theories grows exponentially with the number of rules being considered. As an alternative, we propose another approach, which was originally sketched in [20], and which is here presented in more detail. The method consists in computing intensionally the probability of justification by tracking the dialogue for the justification of arguments.

Definition 7 (Constructible theories). *The set* $\mathsf{Th}(A)$ *of constructible theories containing the rules of the argument* A *is defined as follows:*

$$\mathsf{Th}(A) = \{T | T \in \Omega, \mathsf{rul}(A) \subseteq \mathsf{rul}(T)\}.$$

Thus, the probability of having a theory including A is obviously $P(\mathsf{Th}(A))$. Note that the constructible theories $\mathsf{Th}(A)$, in which the rules of an argument A are included, may be different from the constructible theories in which the argument A is justified.

A useful result regards the construction chance of a set of arguments. In case rules are independent, given a set $\mathcal{A}$ of arguments, then $P(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A))$ is the product of the probability of the rules with which the arguments in $\mathcal{A}$ are built.

Theorem 5. *Let* $(\Omega, F(\Omega), P)$ *a probability space, and let* $\mathcal{A}$ *be a set of arguments, if rules are stochastically independent then*

$$P(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A)) = \prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r)$$

Proof. We prove the theorem by induction over sample spaces $(\Omega_n, F(\Omega), P)$ with $\mathsf{rul}(\Omega_n) = \{r_1, r_2, \ldots, r_n\}$. Induction base: suppose the set $\mathsf{rul}(\Omega_n) = \bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)$ and the set $\mathsf{rul}(\Omega_{n+1}) = \mathsf{rul}(\Omega_n) \cup \{r_{n+1}\}$, we have:

$$\begin{aligned}
&P_{\Omega_{n+1}}(\textstyle\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A)) \\
&= \pi(r_{n+1}) \textstyle\prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r) + [1 - \pi(r_{n+1})] \prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r) \\
&= \textstyle\prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r)
\end{aligned}$$

Suppose the following induction hypothesis for Ω_{n+m}:

$$P_{\Omega_{n+m}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A)) = \prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r)$$

We have for Ω_{n+m+1}:

$$P_{\Omega_{n+m+1}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A))$$
$$= \pi(r_{n+m+1})P_{\Omega_{n+m}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A)) + [1-\pi(r_{n+m+1})]P_{\Omega_{n+m}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A))$$
$$= P_{\Omega_{n+m}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A))$$

Using the induction hypothesis, we obtain:

$$P_{\Omega_{n+m+1}}(\bigcap_{A\in\mathcal{A}} \mathsf{Th}(A)) = \prod_{r\in\bigcup_{A\in\mathcal{A}} \mathsf{rul}(A)} \pi(r)$$

Definition 8 (Construction chance). *The construction chance of an argument A is the probability that the argument is found in a constructible theory, namely the probability $P(\mathsf{Th}_A)$ of the set of theories containing the rules of the argument.*

It follows from the previous theorem that, when the rules are stochastically independent, the construction chance $P(\mathsf{Th}(A))$ of an argument A is:

$$P(\mathsf{Th}(A)) = \prod_{r\in\mathsf{rul}(A)} \pi(r)$$

On the basis of the idea of the construction chance of an argument we want now to characterise the idea of its security chance, namely, the probability that the constructing argument is justified.

Consider the simple case in which an argument B defeats an argument A. What is the set of world theories where A is justified? The set of world theories in which A is justified is the world theories in which A can be constructed and B cannot, that is, the set of worlds theories including the set of rules in A and excluding those of B. Formally, the probability of justification of A is

$$P(Just(A)) = P(\mathsf{Th}(A)\backslash\mathsf{Th}(B)).$$

From set theory, we know that $\mathsf{Th}(A)\backslash\mathsf{Th}(B) = \mathsf{Th}(A)\cap\mathsf{Th}(B)^c$ where c indicates the complement set. Thus, we obtain:

$$P(Just(A)) = P(\mathsf{Th}(A)\cap\mathsf{Th}(B)^c)$$

or

$$P(Just(A)) = P(\mathsf{Th}(A)) - P(\mathsf{Th}(B)\cap\mathsf{Th}(A))$$

Let $\langle\{A\},\emptyset\rangle$ be another argumentation framework in which we associate the probability $P'(Just(A))$. It is worth noting that $P(Just(A)) \leq P'(Just(A))$ expresses the common sense observation that the probability of the justification of an argument (or its conclusion) decreases in the face of new evidences against this argument (and thus its conclusion). Moreover, the probability of justification of A remains stable if the probability of construction of its attacking argument B equals 0.

Example 5 (Running example; cont'd). Let Γ be a multiset of pure defeasible theories from which we build a sample space Ω_Γ with $\mathsf{rul}(\Omega_\Gamma) = \{r_1, r_2, r_3, r_4\}$, and $\mathsf{sup}(\Omega_\Gamma) = \emptyset$ such that:

$$\begin{array}{l|l} 0.5, r_1 : \Rightarrow \mathsf{E}_{Tom} left & 1, r_3 : \quad \mathsf{E}_{Tom} left \quad \Rightarrow \neg \mathsf{E}_{Tom} right \\ 0.5, r_2 : \Rightarrow \mathsf{E}_{Tom} right & 1, r_4 : \quad \mathsf{E}_{Tom} right \Rightarrow \neg \mathsf{E}_{Tom} left \end{array}$$

We can build the following arguments L, R, NL and NR:

$$\begin{array}{lll} L: & \Rightarrow_{r_1} \mathsf{E}_{Tom} left & \\ R: & \Rightarrow_{r_2} \mathsf{E}_{Tom} right & \\ NR: & L & \Rightarrow_{r_4} \neg \mathsf{E}_{Tom} right \\ NL: & R & \Rightarrow_{r_5} \neg \mathsf{E}_{Tom} left \end{array}$$

The apriori probability of the justification of L is computed as follows:

$$\begin{array}{l} P(just(L)) \\ = P(\mathsf{Th}(L) \backslash (\mathsf{Th}(NL) \backslash \mathsf{Th}(NR) \backslash \mathsf{Th}(NL))) \\ = P(\mathsf{Th}(L)) - P(\mathsf{Th}(L) \cap \mathsf{Th}(NL) \backslash NR \backslash NL)) \\ = P(\mathsf{Th}(L)) - P(\mathsf{Th}(L) \cap \mathsf{Th}(NL)) + P(\mathsf{Th}(L) \cap \mathsf{Th}(NL) \cap \mathsf{Th}(NR) \backslash NL)) \\ = P(\mathsf{Th}(L)) - P(\mathsf{Th}(L) \cap \mathsf{Th}(NL)) + P(\mathsf{Th}(L) \cap \mathsf{Th}(NL) \cap \mathsf{Th}(NR)) \\ \quad -P(\mathsf{Th}(L) \cap \mathsf{Th}(NL) \cap \mathsf{Th}(NR) \cap \mathsf{Th}(NL)) \end{array}$$

If the rules are assumed stochastically independent then:

$$\begin{array}{l} P(Just(L)) \\ = \pi(r_1) - \pi(r_1)\pi(r_2)\pi(r_4) + \pi(r_1)\pi(r_3)\pi(r_4)\pi(r_2) - \pi(r_1)\pi(r_3)\pi(r_4)\pi(r_2) \\ = \pi(r_1)[1 - \pi(r_2)] \end{array}$$

Thus, as expected, we retrieve the result from Example 4 with independent rules.

Next, we will consider the more general case in which we have a chain of attacking arguments such that an argument A_n is defeated by A_{n-1}. The probability that A_n is successful is:

$$P(Just(A_n)) = P(\mathsf{Th}(A_n) \backslash \mathsf{Th}(A_{n-1}) \ldots \backslash \mathsf{Th}(A_2) \backslash \mathsf{Th}(A_1))$$

From results of probability theories, this can be transformed into the more ergonomic form for computation:

$$P(Just(A_n)) = \sum_{i=1}^{i=n} (-1)^{n+1} P(\bigcap_{j=1}^{j=i} \mathsf{Th}(A_j))$$

What about the case in which we have several attacks on a single argument? Let A be an argument attacked by n arguments $B_1, \ldots, B_n$, we have:

$$p(Just(A)) = P(\mathsf{Th}(A) \backslash (\mathsf{Th}(B_1) \cup \ldots \cup \mathsf{Th}(B_n))$$

This expression can be transformed as follows:

$$\begin{aligned} P(Just(A)) &= P(\mathsf{Th}(A) \cap (\textstyle\bigcup_{i=1}^{n} \mathsf{Th}(B_i))^c) \\ &= P(\mathsf{Th}(A) \cap \mathsf{Th}(B_1)^c \cap \ldots \cap \mathsf{Th}(B_n)^c) \\ &= P(\mathsf{Th}(A)) - P(\mathsf{Th}(A) \cap \textstyle\bigcup_{i=1}^{n} \mathsf{Th}(B_i)) \\ &= P(\mathsf{Th}(A)) - P(\textstyle\bigcup_{i=1}^{n} \mathsf{Th}(A) \cap \mathsf{Th}(B_i)) \end{aligned}$$

The computation can be achieved by developing the above expression using the inclusion-exclusion principle:

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{k=1}^{k=n} (-1)^{k-1} \sum_{I \subset \{1,\ldots,n\}, |I|=k} P(\bigcap_{i \in I} E_i)$$

where $E_i = \mathsf{Th}(A) \cap \mathsf{Th}(B_i)$.

Example 6. Let's $\langle \{A, B, C\}, \{B \succ A, C \succ A\} \rangle$ be an argumentation framework, the probability of justification of A is:

$$P(Just(A)) = P(\mathsf{Th}(A) \backslash [\mathsf{Th}(B) \cup \mathsf{Th}(C)])$$

Using the inclusion-exclusion principle, we obtain:

$$\begin{aligned} P(Just(A)) = P(\mathsf{Th}(A)) &- P(\mathsf{Th}(A) \cap \mathsf{Th}(B)) \\ &- P(\mathsf{Th}(A) \cap \mathsf{Th}(C)) \\ &+ P(\mathsf{Th}(A) \cap \mathsf{Th}(B) \cap \mathsf{Th}(C)). \end{aligned}$$

Now consider a finite dialogue tree where each argument A is attacked by k arguments A_k which are conclusions of k subtrees D_k. The probability of justification of the argument A can be reiteratively determined using the probability of set difference and the inclusion-exclusion principle. Indeed, the probability of justification of A is the following set difference:

$$P(Just(A)) = P(\mathsf{Th}(A) \backslash \bigcup_{i} \mathsf{Th}(D_i))$$

where if A_i is attacked by m arguments A_m, then

$$\mathsf{Th}(D_i) = \mathsf{Th}(A_i) \backslash \bigcup_{j} \mathsf{Th}(D_j)$$

and if A_i is attacked by no argument, then

$$\mathsf{Th}(D_i) = \emptyset$$

Using the probability of set difference, we obtain:

$$P(Just(A)) = P(\mathsf{Th}(A)) - P(\mathsf{Th}(A) \cap [\bigcup_{i} \mathsf{Th}(D_i)])$$

or,

$$P(Just(A)) = P(\mathsf{Th}(A)) - P(\bigcup_i [\mathsf{Th}(A) \cap \mathsf{Th}(D_i)])$$

Finally, the probability $P(\bigcup_i [\mathsf{Th}(A) \cap \mathsf{Th}(D_i)])$ is computed using the inclusion-exclusion principle. Thus we can come up with the following definition.

Definition 9 (Security chance). *The security chance $P(Just(A))$ of an argument A is the probability that A is justified:*

$$P(\mathsf{Th}(A)) - P(\bigcup_i [\mathsf{Th}(A) \cap \mathsf{Th}(D_i)])$$

Next, we will use our probabilistic rule-based argumentation framework as a tool to design, investigate and run multi-agent systems.

5 Multi-agent theory and simulation

We propose now how to represent and structure the content of a multi-agent theory. Then, computational perspectives are discussed and a simple learning mechanism is given to turn out a multi-agent theory into an executable specification.

5.1 Probabilistic multi-agent theory

We need to extend as follows the formal language we have presented in the previous sections: this will allows us to further define multi-agent systems.

Definition 10 (Language for MAS). *Let* Atoms *be a set of atomic formulas of an underlying first order language,* Lbl *a set of labels, obj a parameter for literals*[4]*,* $\mathsf{Ag} = \{i_1, i_2, \ldots\}$ *a finite set of agents,* $\mathsf{Times} = \{t_1, t_2, \ldots\}$ *is a discrete totally ordered set of instants of time.*

Results *Let the set* Results *be defined as follows:*

$$\mathsf{Results} = \{out_i(X) | i \in \mathsf{Ag} \wedge X \in \mathbb{R}\}$$

A proposition $out_i(X)$ affirms that agent i obtains utility X.

Basic literals *Let* Lit *denote the set of basic literals: all atoms in* Atoms *or in* Results *and their negations.*

$$\mathsf{Lit} = \{\pm\gamma | \gamma \in \mathsf{Atoms} \ \vee \gamma \in \mathsf{Results}\}$$

[4] As we shall see in a moment, *obj* is used to denote those factual statements that are objectively true and so independent from the agents' perspective.

Action literals *An action literal has the form*

$$\pm \mathsf{E}_i^t \phi$$

where $i \in \mathsf{Ag}$*,* $t \in \mathsf{Times}$*, and* $\phi \in \mathsf{Lit}$*. It affirms that agent* i *attempts to realise (does not attempt) state of affairs* ϕ *at time* t*.*

State literals *An state literal has the form*

$$\pm \mathsf{Hold}_i^t \phi$$

where $i \in \mathsf{Ag} \cup \{obj\}$*,* $t \in \mathsf{Times}$*, and* $\phi \in \mathsf{Lit}$*. It affirms that according to agent* i *state of affairs* ϕ *holds (does not hold) at time* t *(a belief of agent* i*).*

Obligation literals *An obligation literal has the form*

$$\pm \mathsf{Obl}_i^t \phi$$

where $i \in \mathsf{Ag} \cup \{obj\}$*,* $t \in \mathsf{Times}$*, and* $\phi \in \mathsf{Lit}$*. It affirms that the agent* i *has (not) the obligation to bring about* ϕ *at time* t*.*

Rules *The rules of our agents are probabilistic defeasible rules having the form*

$$\pi, r^t : \phi_1, \ldots, \phi_n \Rightarrow \phi$$

where r *is a label (*$r \in \mathsf{Lbl}$*),* t *is a time (*$t \in \mathsf{Times}$*), and each* $\phi_1, \ldots, \phi_n, \phi$ *is a state literal, an action literal, or an obligation literal.*

We assume an incompability function $-$, which returns the set of modal literals which are incompatible for a given modal literal. Let t an instant of time ($t \in \mathsf{Times}$), let i denote an agent ($i \in \mathsf{Ag}$), let M denote a modality Hold_i^t, Hold_{obj}^t, E_i^t, Obl_i^t or Obl_{obj}^t and let ϕ be a literal. Each modal literal $\pm M\phi$ is incompatible both with its complement ($M\phi$ is incompatible with $\neg M\phi$ and vice versa) and with the formula obtained by substituting the embedded literal ϕ with its complement ($\pm M\phi$ is incompatible with $\pm M - \phi$ and vice versa). Let ψ denote an atom, we have:

$$\begin{array}{l|l} -M\psi = \{\neg M\psi, M\neg\psi\} & -M\neg\psi = \{\neg M\neg\psi, M\psi\} \\ -\neg M\psi = \{M\psi, \neg M\neg\psi\} & -\neg M\neg\psi = \{\neg M\psi, M\neg\psi\} \end{array}$$

Let us informally clarify the language. We assume that agents process external inputs from an objective environment into a subjective counterpart, and (successfully or not) act upon this objective environment. We formally cater for basic temporal epistemic features, agency and simple normative aspects by temporalised prefix operators for literals that we present with their meaning in the following list:

$\mathsf{Hold}_i^t\phi$ "It holds, from the viewpoint of an agent i at time t, that ϕ."
$\mathsf{Hold}_{obj}^t\phi$ "It objectively holds at time t that ϕ.",
$\mathsf{E}_i^t\phi$ "The agent i is attempting at time t to bring it about ϕ."
$\mathsf{Obl}_{obj}^t\phi$ "From an objective point of view,
it is obligatory at time t to bring about ϕ."
$\mathsf{Obl}_i^t\phi$ "From the point of view of agent i,
it is obligatory for agent i at time t to bring about ϕ."

Note that we label our formulas with *obj* to indicate that the formula holds objectively, namely, that is the case (rather than being merely believed by an agent). We may say that *obj* embodies the objective point of view, that it only accepts what is true. Thus, any prefixed literal subscripted by *obj* is called an objective environmental literal or environmental literals, while those subscripted with i are called mental literals, or agents literals or subjective literals. Also, notice that $\mathsf{E}_i^t\phi$ does not denote a necessarily successful action of agent i, because agents operate in a probabilistic, non-deterministic framework and their behaviour is governed by defeasible rules: thus, $\mathsf{E}_i^t\phi$ stands, as we mentioned, for i's attempt to bring it about that ϕ at time t.

Theoretically, any sort of rules connecting those prefixed literals can be built (enough to give them some meaningful interpretation), but we would like to propose some typical rules which may be used when specifying a multi-agent system. We need to distinguish environmental rules, holding in the environment, and agent rules, namely, the agent's beliefs in and commitments to rules.

Factual and normative environmental rules constitute of the environment where the agents are located.

Factual environmental rules A factual environmental rule has the form

$$\pi, r^t : \quad \phi_1, \ldots, \phi_n \Rightarrow \pm\mathsf{Hold}_{obj}^t\phi$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}_{obl}^t\psi$, $\pm\mathsf{E}_i^t\psi$ or $\pm\mathsf{Obl}_{obj}^t\phi$. It connects environmental states of affairs (scientific laws and empirical generalisations are factual environmental rules).

Normative environmental rules A normative environmental rule has the form

$$\pi, r^t : \quad \phi_1, \ldots, \phi_n \Rightarrow \pm\mathsf{Obl}_{obj}^t\phi$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}_{obl}^t\psi$ or $\pm\mathsf{E}_i^t\psi$.
It expresses an exogenous norm establishing an obligation or a permission.

Let us now consider the agent rules. For our purposes, any agent is described by the rules it endorses.

Sensor rules A sensor rule has the form

$$\pi, r^t : \quad \phi_1, \ldots, \phi_n \Rightarrow \pm\mathsf{Hold}_i^t\phi$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}^t_{obj}\psi$ or $\pm\mathsf{Hold}^t_i\psi$, but at least one of these has the form $\pm\mathsf{Hold}^t_{obj}\psi$. Thus a sensor rule connects at least an objective environmental literal to a subjective literal, allowing the agent to form beliefs on the basis of perceptual inputs.

Mental rules A mental rule has the form

$$\pi, r^t : \quad \pm\mathsf{Hold}^t_i\phi_1, \ldots, \pm\mathsf{Hold}^t_i\phi_n \Rightarrow \pm\mathsf{Hold}^t_i\phi$$

It relates subjective mental literals to each other.

Action rules An action rule has the form

$$\pi, r^t : \quad \phi_1, \ldots, \phi_n \Rightarrow \pm\mathsf{E}^t_i\phi$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}^t_{obj}\psi$, $\pm\mathsf{Hold}^t_i\psi$, $\pm\mathsf{Obl}^t_i\psi$ or $\pm\mathsf{Hold}^t_i\psi$ It connects some objective or mental literals or action literals to an attempt to bring about a state of affairs or to refrain from it.

Obligation rules An obligation rule has the form

$$\pi, r^t : \quad \phi_1, \ldots, \phi_n \Rightarrow \pm\mathsf{Obl}^t_i\phi$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}^t_i\psi$, $\pm\mathsf{Hold}^t_{obj}\psi$, $\mathsf{Obl}^t_{obj}\psi$ or $\pm\mathsf{E}^t_i\psi$. It allows the agents' attitude to internalise obligations.

Outcome rules An outcome rule has the form

$$\pi, r^t : \phi_1, \ldots, \phi_n \Rightarrow \mathsf{Hold}^t_i out_j(X)$$

where each ϕ_k, $1 \leq k \leq n$, has the form either of $\pm\mathsf{Hold}^t_{obj}\psi$ or of $\pm\mathsf{E}^t_i\psi$, where $i, j \in \mathsf{Ag}$ and X is either a positive real number (benefit) or negative real number (cost). It connects objective states of affairs or actions to an outcome, namely the utility obtained by the agent as a consequence of such states of affairs or actions.

Modeling dynamics of a system via actions is confronted to at least three well-known difficulties, namely the qualification problem, the frame problem and the ramification problem. This paper is not meant to discuss those problems and how they may be dealt in an argumentative rule-based setting, and we assume for our purposes that any attempt to bring about a state of affairs ϕ is defeasible successful, expressed by the following schema rule:

$$\pi, r^t : \quad \mathsf{E}^t_i\phi \Rightarrow \mathsf{Hold}^t_{obj}\phi$$

As we will see in the next sections, we assume that the probabilities of agents rules change on the basis of a reinforcement learning mechanism according to

which an agent seeks for the behaviour with the highest utility (as established by the applicable outcome rules).

A multi-agent system is encoded into a probabilistic theory whose rules are environmental rules or agent rules, as distinguished above. The first hold in the environment, regardless of the beliefs of an individual agent, the second hold in the agent's mind. The fundamental difference between those rules is that the probabilities of environmental rules are assumed to be fixed, whereas probabilities of the agent rules may be changed by the agent itself in order to adapt to the environment.

Thus, we partition the probabilistic theory encoding a multi-agent system into an environmental probabilistic theory (whose rules' probabilities are not meant to change), and probabilistic theories representing agents (whose rules' probabilities may change). So, each agent i is represented by a probabilistic theory $\mathcal{T}_i$ and the environment is given by a probabilistic theory $\mathcal{T}_e$.

Furthermore, we partition an environment $\mathcal{T}_e$ into two sub-environments: a first sub-environmental theory $\mathcal{T}_{e1}$ is used to compute the environmental facts that agents can sense before behaving, and a second sub-environment theory $\mathcal{T}_{e2}$ aimed to compute the outcomes of agents' acts. So we have $\mathcal{T}_e = \mathcal{T}_{e1} \cup \mathcal{T}_{e2}$.

A multi-agent probabilistic theory $\mathcal{T}$ is the union of agents' probabilistic theories and the probabilistic environment theory: $\mathcal{T} = \mathcal{T}_e \bigcup_i \mathcal{T}_i$.

In the remainder, as for notation, since rules r^t are indexed by an instant t indicating their time of application, we shall superscript denotation for theories, arguments or other concepts with t: for example, a theory T^t denotes a theory whose rules hold at time t.

5.2 Computational perspectives

Based on our probabilistic framework, and given a probabilistic multi-agent theory, we can either:

- compute its probabilistic extension, i.e. any probability of justification, or
- pick up a pure theory, and then compute its extension.

This allows us to differentiate:

- *probable behaviours* of agents as probable grounded extensions, and
- *actual behaviours* as the grounded extensions of pure theories.

The distinction between actual behaviour and probable behaviour induces the distinction between the *utility of a behaviour* and the *expected utility of a probable behaviour* (see [18] for the computation of expected utilities in a similar probabilistic framework).

Accordingly, we can take two perspectives:

- a *systemic perspective*: a global point of view, which is focused on global-systemic features and values of the MAS as a whole (e.g. global wealth, index of justice, etc.);

– an *individualistic perspective*: the perspective of a particular agent who is assumed to aim at maximising its utility.

The two perspective are related, in the sense that to compute systemic properties, we can use a probabilistic analysis based on our probabilistic framework, where each agent adopts its individualistic perspective.

Note that a particular global state, as for example the state in which the maximal total wealth is obtained, necessarily corresponds to theories where the probability of rules is either 1 or 0, the assumption according to which rules have probability 0 or 1 is a special case where rules can be assumed independent,

Thus, such particular states we can easily looked for by computing their probability and by assuming that rules are independent (see example 7 for a simple illustration).

Example 7 (Running example; cont'd). In this example, we illustrate how the algebraic expression of the probability of justification can be used to investigate optimum states.

Suppose two agents i and j who can drive either on the left or on the right side:

$$\begin{array}{l|l} \pi(l_k^t), l_k^t : \Rightarrow \mathsf{E}_k^t left & 1, nr_k^t : \mathsf{E}_k^t left \Rightarrow \neg \mathsf{E}_k^t right \\ \pi(r_k^t), r_k^t : \Rightarrow \mathsf{E}_k^t right & 1, nl_k^t : \mathsf{E}_k^t right \Rightarrow \neg \mathsf{E}_k^t left \end{array}$$

At each time t, the agents i and j may be paired in a scene:

$$\gamma, s0_{ij}^t : \Rightarrow \mathsf{Hold}_{obj}^t scene(i,j)$$

In a scene, if the agents both drive left, or right, then all is fine, while if they do not drive on the same side of the road, then an accident may occur (we give here all the grounded combinations):

$$\begin{array}{ll} 0.5, ko.1_i^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t left,\ \ \mathsf{E}_j^t right \Rightarrow \mathsf{Hold}_i^t out_i(out1) \\ 0.5, ko.2_i^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t right,\ \mathsf{E}_j^t left \ \Rightarrow \mathsf{Hold}_i^t out_i(out1) \\ 0.5, ko.1_j^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t left,\ \ \mathsf{E}_j^t right \Rightarrow \mathsf{Hold}_j^t out_j(out1) \\ 0.5, ko.2_j^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t right,\ \mathsf{E}_j^t left \ \Rightarrow \mathsf{Hold}_j^t out_j(out1) \\ 1, ok.1_i^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t left,\ \ \mathsf{E}_j^t left \ \Rightarrow \mathsf{Hold}_i^t out_i(out2) \\ 1, ok.2_i^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t right,\ \mathsf{E}_j^t right \Rightarrow \mathsf{Hold}_i^t out_i(out2) \\ 1, ok.1_j^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t left,\ \ \mathsf{E}_j^t left \ \Rightarrow \mathsf{Hold}_j^t out_j(out2) \\ 1, ok.2_j^t : & \mathsf{Hold}_{obj}^t scene(i,j),\ \mathsf{E}_i^t right,\ \mathsf{E}_j^t right \Rightarrow \mathsf{Hold}_j^t out_j(out2) \end{array}$$

where $out1 < 0$ and $out2 > 0$. Possible arguments are:

$$\begin{array}{lll} L_k^t : & \Rightarrow_{l_k^t} \mathsf{E}_k^t left & NR_k^t : \quad L_k^t \Rightarrow_{nl_k^t} \neg \mathsf{E}_k^t right \\ R_k^t : & \Rightarrow_{r_k^t} \mathsf{E}_k^t right & NL_k^t : \quad R_k^t \Rightarrow_{nr_k^t} \neg \mathsf{E}_k^t left \\ S_{ij}^t : & \Rightarrow_{s1_{ij}^t} \mathsf{Hold}_{obj}^t scene(i,j) & \\ KO1_i^t : & S_{ij}^t, L_i^t, R_j^t & \Rightarrow_{ko.1_i^t} \mathsf{Hold}_i^t out_i(out1) \\ KO2_i^t : & S_{ij}^t, R_i^t, L_j^t & \Rightarrow_{ko.2_i^t} \mathsf{Hold}_i^t out_i(out1) \\ OKL_i^t : & S_{ij}^t, L_i^t, L_j^t & \Rightarrow_{ok.1_i^t} \mathsf{Hold}_i^t out_i(out2) \\ OKR_i^t : & S_{ij}^t, R_i^t, R_j^t & \Rightarrow_{ok.2_i^t} \mathsf{Hold}_i^t out_i(out2) \end{array}$$

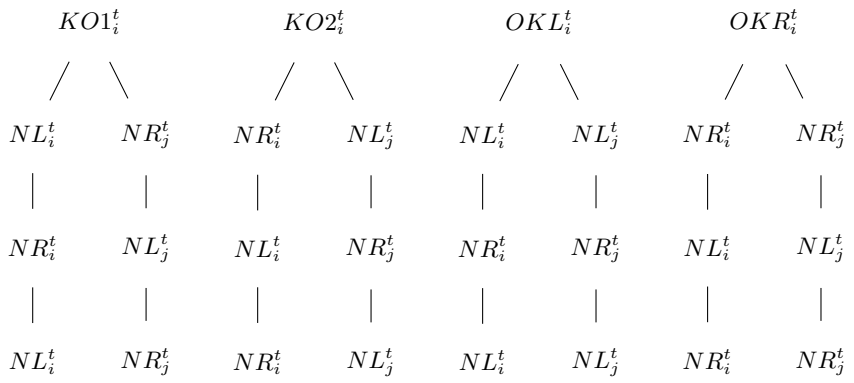

Fig. 2. Possible dialogue trees.

The overall expected utility EU^t of agents' behaviours is computed as follows:

$$\begin{aligned} EU^t = {} & P(Just(\mathsf{Hold}_i^t out_i(out1))).out1 + P(Just(\mathsf{Hold}_j^t out_j(out1))).out1 \\ & P(Just(\mathsf{Hold}_i^t out_i(out2))).out2 + P(Just(\mathsf{Hold}_j^t out_j(out2))).out2 \end{aligned}$$

such that:

$$\begin{aligned} & P(Just(\mathsf{Hold}_i^t out_i(out1))) \\ & = \mathsf{Th}(KO1_i^t) \backslash [\mathsf{Th}(NL_i^t) \backslash \mathsf{Th}(NR_i^t) \backslash \mathsf{Th}(NL_i^t)] \cup [\mathsf{Th}(NR_j^t) \backslash \mathsf{Th}(NL_j^t) \backslash \mathsf{Th}(NR_j^t)] \\ & \quad + \mathsf{Th}(KO2_i^t) \backslash [\mathsf{Th}(NR_i^t) \backslash \mathsf{Th}(NL_i^t) \backslash \mathsf{Th}(NR_i^t)] \cup [\mathsf{Th}(NL_j^t) \backslash \mathsf{Th}(NR_j^t) \backslash \mathsf{Th}(NL_j^t)] \end{aligned}$$

$$\begin{aligned} & P(Just(\mathsf{Hold}_i^t out_i(out2))) \\ & = \mathsf{Th}(OKL_i^t) \backslash [\mathsf{Th}(NL_i^t) \backslash \mathsf{Th}(NR_i^t) \backslash \mathsf{Th}(NL_i^t)] \cup [\mathsf{Th}(NL_j^t) \backslash \mathsf{Th}(NR_j^t) \backslash \mathsf{Th}(NL_j^t)] \\ & \quad + \mathsf{Th}(OKR_i^t) \backslash [\mathsf{Th}(NR_i^t) \backslash \mathsf{Th}(NL_i^t) \backslash \mathsf{Th}(NR_i^t)] \cup [\mathsf{Th}(NR_j^t) \backslash \mathsf{Th}(NL_j^t) \backslash \mathsf{Th}(NR_j^t)] \end{aligned}$$

To find the maximum of EU^t, we can look at solutions where the rules are independent. The algebraic analysis of EU^t brings us to conclude that the maximum of the expected utility is obtained when:

$$\pi(l_i^t) = 1, \pi(r_i^t) = 0, \ \pi(l_j^t) = 1 \ \text{and} \, \pi(r_j^t) = 0, \text{ or}$$

$$\pi(l_i^t) = 0, \pi(r_i^t) = 1, \ \pi(l_j^t) = 0 \ \text{and} \, \pi(r_j^t) = 1.$$

So, as expected, the maximum is obtained when the agents drive on the same side of the road.

As it is well-known from game-theory, agents optimizing their own utility do not necessarily lead to a global optimum. So, since social systems and agents are often trapped in some local optimum, we are interested in global equilibrium as well as local equilibrium, but we are also interested in studying the dynamics

of the system, for example how conventions evolve. For this reason, we simulate system dynamics by using a subjective perspective where agents optimize their own utility based on reinforcement learning.

5.3 Learning agents

By learning agents, we mean agents that dynamically tend to update their knowledge in such a way as to select actions giving them higher outcomes (utilities). In this sense, they are self-interested, but their utilities can also depend on other agents' utilities.

Since sanctions are expected to play an important role in normative multi-agent systems, we base our multi-agent learning mechanisms on the paradigm of reinforcement learning. At each time t, every agent senses a grounded environment randomly selected by Nature, and, on the basis of those environmental inputs, every agent shall concurrently behave. For every agent, the selection of a behaviour, is simulated by a probability distribution over all possible agent's behaviours. The outcomes of agents' behaviours are then entailed via the grounded extension of agents' pure theories (corresponding to the selected behaviours) and the pure environmental theory. The outcomes obtained at t are then used to compute a time $t+1$ every agent's probability distribution over possible behaviours.

Before moving to the algorithm animating a multi-agent system described in a probabilistic theory, we consider the notions of grounded environment, theory utilities and behaviours.

Grounded environment. At each time t, agents concurrently behave on the basis of a grounded environment. Thus, we assume that Nature plays first, and then, the agents behave. An alternative could have been the concurrent plays of the Nature and agents, but, that would correspond to the case of "blind" agents. Nevertheless, such alternative can be simulated by considering an empty environment.

A grounded environment is built using the reification in a new theory T of the grounded extension of a theory T'.

Definition 11 (Grounded theory). *The grounded theory $T(GE(T'))$ built from the grounded extension of a theory T' is the theory $\langle R, \emptyset \rangle$ such that:*

$$R = \{ \quad \Rightarrow \phi \mid T' \vdash_{GE} \phi\}$$

We partition an environment $\mathcal{T}_e$ into two sub-environments: a first sub-environmental theory $\mathcal{T}_{e1}$ is used to compute the environmental facts that agents can sense before behaving, and a second sub-environment theory $\mathcal{T}_{e2}$, called an outcome theory, is used to compute the outcomes of agents' acts. Accordingly, at each time t, we have $\mathcal{T}_e^t = \mathcal{T}_{e1}^t \cup \mathcal{T}_{e2}^t$, and agents behaves after sensing the grounded environment $T(GE(T_{e1}^t))$, denoted E_{e1}^t in the remainder.

Pure theory utilities and qualities. Once all the agents have behaved in a grounded environment E^t_{e1}, and given a pure outcome environment T^t_{e2}, some outcomes and payoffs are entailed, and thus agents can evaluate the utility and quality of their behaviors, that is, the utility and quality of their pure theories. To deal with the stochastic dependence of agents' rules, any pure theory is associated with a potential (its quality) which is computed by an online weighted average of utilities over time. Formally, the utility for the agent i of a pure theory T^t_i associated to a grounded environment E^t_{e1}, a pure environmental outcome theory T^t_{e2} and other agents' theories T^t_i is:

$$u_i(T^t_i) = \sum_k X_k \tag{7}$$

where for each k, it exists an argument $A_k \in GE(E^t_{e1} \cup T^t_{e2} \bigcup_i T^t_i)$ such that $\mathsf{conc}(A_k) = \mathsf{Hold}^t_i out_i(X_k)$.

Given the utility of a theory T^t_i selected at time t, its quality at time $t+1$ is computed using a weighted average over time:

$$Q(T^{t+1}_i) = Q(T^t_i) + \alpha.[u_i(T^t_i) - Q(T^t_i)] \tag{8}$$

where $\alpha \in [0, 1]$. If T^t_i is not selected, then:

$$Q(T^{t+1}_i) = \beta.Q(T^t_i) \tag{9}$$

where $\beta \in [0, 1]$. The parameter α is the weight of the latest utility in order to keep track of the non-stationary environment, the parameter β allows agents to forget unused theories. At the initialisation of an agent i, say at time t_{init}, we assume an arbitrary value for any $Q(T^{t_{init}}_i)$ (typically 0).

Behaviours. If an agent is represented by n rules, the number of pure theories constructible from them is 2^n. Amongst all these theories, the agent has to search which one is the most adapted to the environment. Such a large search space is in practice very annoying because (i) it may require a lot of memory space, and (ii) many theories may result into no action and so agents remain rather inactive.

We reduce the search space by assuming that every rule with a constant probability is independent. Thus, at each time t, a lottery is performed on agents' theories $\mathcal{T}^t_i$ over the k fixed independent rules, resulting into 2^{n-k} remaining possible pure theories.

Then, to further reduce the search space and to obtain lively agents, we consider the possible arguments and their defeat relations that can be built from the 2^{n-k} theories. We use the notion of *behaviours*: a behaviour is the set of pure theories (within the 2^{n-k} possible theories) supporting the same set of actions. Thus, a behaviour B^t_i is defined by a function $B(\mathcal{E}^t_i, E^t_{e1})$ whose the domain is a grounded environment E^t_{e1} and a (possible empty) set of actions $\mathcal{E}^t_i$, and whose the range is a set of pure theories exactly entailing this set of actions.

Definition 12 (Behaviour). *Let E^t_{e1} be a grounded environment at time t, let i denote an agent and let $\mathcal{T}_i$ be its probabilistic theory, let $\mathcal{E}^t_i$ be a (possibly empty) set of actions $\mathsf{E}^t_i\phi$ and let $\mathcal{T}'_i$ be the set of pure theories resulting from a lottery on $\mathcal{T}_i$ over every rule with a constant probability.*

We define the set of possible behaviours $\mathcal{B}_i(E^t_{e1})$ as a partition of $\mathcal{T}'_i$ such that a behaviour $B(\mathcal{E}^t_i, E^t_{e1}) \in \mathcal{B}_i(E^t_{e1})$ is the set of pure theories exactly entailing the set of actions $\mathcal{E}^t_i$:

$$B(\mathcal{E}^t_i, E^t_{e1}) = \{T^t_i |\ T^t_i \in \mathcal{T}'_i \wedge \\ \{\mathsf{E}^t_i\phi | E^t_{e1} \cup T^t_i \vdash_{GE} \mathsf{E}^t_i\phi\} = \mathcal{E}^t_i\}$$

In the remainder, we shall write B^t_i or $B_i(E^t_{e1})$ as a shortcut notation for $B(\mathcal{E}^t_i, E^t_{e1})$ when the set of actions or the grounded environment has no importance.

Example 8. Let Tom be an agent defined by a probabilistic theory $\mathcal{T}_{Tom} = \langle\{r^t_1, r^t_2, r^t_3, r^t_4\}, \emptyset\rangle$:

$$\begin{array}{ll|ll} \_, r^t_1: & \Rightarrow \mathsf{E}^t_{Tom} left & 1, r^t_3: & \mathsf{E}^t_{Tom} left \Rightarrow \neg\mathsf{E}^t_{Tom} right \\ \_, r^t_2: & \Rightarrow \mathsf{E}^t_{Tom} right & 1, r^t_4: & \mathsf{E}^t_{Tom} right \Rightarrow \neg\mathsf{E}^t_{Tom} left \end{array}$$

An underscore indicates that the probability can be autonomously changed by the agent. At each instant t, the associated sample space can be represented by a table where each column is a constructible theory:

	T^t_1	T^t_2	T^t_3	T^t_4	T^t_5	T^t_6	T^t_7	T^t_8	T^t_9	T^t_{10}	T^t_{11}	T^t_{12}	T^t_{13}	T^t_{14}	T^t_{15}	T^t_{16}
r^t_1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
r^t_2	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
r^t_3	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
r^t_4	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
$Just(\mathsf{E}^t_{Tom} left)$	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0
$Just(\mathsf{E}^t_{Tom} right)$	0	1	0	0	1	1	0	0	0	1	0	0	1	1	0	0
$\emptyset$	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1

The rules r^t_3 and r^t_4 have a fixed probability, and thus, they are considered independent. A lottery is performed on these rules, and since they have a probability equals to 1, then the search space is reduced to $2^{4-2} = 4$ theories:

	T^t_1	T^t_2	T^t_3	T^t_4
r^t_1	1	0	1	0
r^t_2	1	1	0	0
r^t_3	1	1	1	1
r^t_4	1	1	1	1
$Just(\mathsf{E}^t_{Tom} left)$	0	0	1	0
$Just(\mathsf{E}^t_{Tom} right)$	0	1	0	0
$\emptyset$	1	0	0	1

Each theory T_i^t in $\mathcal{T}'_{Tom} = \{T_1^t, T_2^t, T_3^t, T_4^t\}$ is associated with an argumentation framework $AF_{T_i^t}$ from which we can compute the grounded extension. It results that the possible behaviours are:

- $B(\{\mathsf{E}^t_{Tom} left\}, \emptyset) = \{T_3^t\}$,
- $B(\{\mathsf{E}^t_{Tom} right\}, \emptyset) = \{T_2^t\}$,
- $B(\emptyset, \emptyset) = \{T_1^t, T_4^t\}$.

The choice of Tom's behaviour amongst these possible behaviours will be based on a probability distribution taking into account the quality of each behaviour.

Behaviours' probability. In accordance with our probabilistic setting using equation (1), the probability of a behaviour B_i^t is the sum of the probabilities of its constituting pure theories:

$$P(B_i^t) = \sum_{T_i^t \in B_i^t} P(T_i^t) \tag{10}$$

As a direct implication of the Kolmorogov's axioms, we have the following results. For any $B_i^t \in \mathcal{B}_i(E_{e1}^t)$, $P(B_i^t) \geq 0$, for any distinct behaviours B_i^t and B_j^t, since $B_i^t \cap B_j^t = \emptyset$: $P(B_i^t \cup B_j^t) = P(B_i^t) + P(B_j^t)$, and the normalisation of probabilities of behaviours: $P(\bigcup B_i^t) = 1$ and thus, $\sum P(B_i^t) = 1$.

We assume that the same actions have the same effects. In other words, multi-agent theories are such that, the same sets of actions in a given environment (comprising other agents' behaviours) imply the same outcomes, that is:

$$\forall T_i^t \in B(\mathcal{E}_i^t, E_{e1}^t),\ u_i(E_{e1}^t \cup T_{e2}^t \bigcup_i T_i^t) = u_i(B(\mathcal{E}_i^t, E_{e1}^t))$$

Moreover, the theories T_i^t of a behaviour B_i^t do not necessarily have the same qualities over time (since the selection of a pure theory implies that other pure theories are not selected, see 8 and 9). For this reason, we assume that behaviours' theories have the same qualities. As interpretation, we consider that the selection of a pure theory stands for the selection of any other theory belonging to the same behaviour, or in other words, the selection of a theory implies the co-selection of the theories of the same behaviour. Another possible assumption holds in the definition of sub-behaviours that partition a behaviour, and each sub-behaviour corresponds to a unique theory.

We can now reuse equations (8) and (9) to compute behaviours' quality instead of theories' quality. So, given the utility of a behaviour B_i^t selected at time t, its quality at time $t+1$ is computed using a weighted average over time:

$$Q(B_i^{t+1}) = Q(B_i^t) + \alpha.[u_i(B_i^t) - Q(B_i^t)] \tag{11}$$

where $\alpha \in [0,1]$. If B_i^t is not selected then:

$$Q(B_i^{t+1}) = \beta.Q(B_i^t) \tag{12}$$

where $\beta \in [0,1]$.

The behaviours' distribution is computed via the equation (10):

$$P(B_i^t) = e^{Q(B_i^t)} / \sum_{B_i} e^{Q(B_i^t)} \tag{13}$$

where $Q(B_i^t) = Q(T_i^t) + ln(|B_i^t|)$, such that $T_i^t \in B_i^t$. When $Q(T_i^t) \gg ln(|B_i^t|)$, then $ln(|B_i^t|$ can be omitted to get an approximation of a behaviours' distribution. Instead of using qualities, one can use the differences amongst qualities with respect to the minimum quality in order to compute the probability distribution.

Finally, we add to equation (13) a learning parameter $\tau_i > 0$ modulating the "exploitation" of the pure behaviours that entail the highest outcomes and "exploration" of other behaviours:

$$P(B_i^t) = e^{Q(B_i^t)/\tau_i} / \sum_{B_i} e^{Q(B_i^t)/\tau_i} \tag{14}$$

With this very common distribution in machine learning, we have fallible bounded rational agents who may explore new behaviours that may turned out to be mistakes or, on the contrary, very useful. So, each probability update with respect to behaviours' qualities may not involve new improvements, but each update has a probability of leading to an improvement.

MAS animation algorithm. The algorithm animating a population of learning agents behaving in a non-deterministic environment and described by a probabilistic defeasible theory is sketched in Algorithm 1.

With regard to our introductory assumption on fallible model-free learning agents with bounded-rationality within an non-deterministic environment, the environment is indeed non-deterministic because it is modelled with probabilistic rules, the learning framework is indeed model-free in the sense agents do not need an explicit model of their environment in order to behave rationally. Learned behaviours allow agents to economize on the calculation strategy costs whenever facing a new situation, and thus it enables agents to adapt with "bounded rationality". Agents are fallible because the probability distribution implies that they may choose behaviours appearing as mistakes. Finally, notice that our agents are resilient because, even if some defeasible rules are deleted, then agents may still have the possibility to re-configure themselves by learning new useful behaviours.

Algorithm 1 Animation of a system of learning agents

- Intialise the system with a probabilistic theory $\mathcal{T}_e \bigcup_i \mathcal{T}_i$ with $\mathcal{T}_e = \mathcal{T}_{e1} \cup \mathcal{T}_{e2}$;
for $t = 0$ to t_{end} **do**
  - Do a lottery on independent fixed probabilistic rules (so $\mathcal{T}_e^t$ describing the environment results in one pure defeasible theory T_e^t);
  - Compute the grounded environment $E_{e1}^t = T(GE(T_{e1}^t))$;
  for each agent i **do**
    - Compute the set of possible behaviours $\mathcal{B}_i(E_{e1}^t)$;
    - Compute the probability distribution over the behaviours $\mathcal{B}_i(E_{e1}^t)$, e.g., using a Boltzmann distribution:

$$P(B_i^t) = e^{Q(B_i^t)/\tau_i} / \sum_{B_i^t \in \mathcal{B}_i(E_{e1}^t)} e^{Q(B_i^t)/\tau_i}$$

    where $\tau_i > 0$ is a computational learning parameter which controls the amount of exploration versus exploitation;
    - Do a lottery on the probability distribution of $\mathcal{B}_i(E_{e1}^t)$ resulting in one behaviour B_i^t;
  end for
  - Compute the grounded extension $GE(E_{e1}^t \cup T_{e2}^t \bigcup_i T_i^t)$;
  for each agent i **do**
    for each behaviour B_i^t **do**
      if B_i^t was previously selected **then**

$$Q(B_i^{t+1}) = Q(B_i^t) + \alpha.[u_i(B_i^t) - Q(B_i^t)] \text{ where } \alpha \in [0, 1]$$

      else

$$Q(B_i^{t+1}) = \beta.Q(B_i^t) \text{ where } \beta \in [0, 1].$$

      end if
    end for
  end for
end for

6 Illustration

In this section, we illustrate our framework with two simple case studies of normative phenomena: norm internalisation and norm emergence. Though, the scenarios are quite different, our framework deals with them quite easily, showing thus its expressiveness. The results of the simulations have been achieved via a Prolog proof of concept implementation of the algorithm proposed in 1.

6.1 Case study 1: Norm emergence

In this case study, we simulate the emergence of the convention about driving on either the left side or the right side of the road. This case study is meant to be a simple illustration of our framework rather than a proper analysis of the emergence of a convention (c.f. [3]). For our purposes, each agent i is encoded into few rules of agency (a rule prefixed with an underscore means that its probability can change via learning):

$$\begin{array}{l|l} \_, l_i^t : \Rightarrow \mathsf{E}_i^t left & 1, nr_i^t : \Rightarrow \neg\mathsf{E}_i^t right \\ \_, r_i^t : \Rightarrow \mathsf{E}_i^t right & 1, nl_i^t : \mathsf{E}_i^t right \Rightarrow \neg\mathsf{E}_i^t left \\ \_, int_i^t : \mathsf{Obl}_{obj}^t right \Rightarrow \mathsf{Obl}_i^t right & \_, c_i^t : \mathsf{Obl}_i^t right \Rightarrow \mathsf{E}_i^t right \end{array}$$

An obligation to drive on the right side of the road is introduced at time 200, and its violation is defined. It would be more natural to say that it is forbidden to drive on the left, but we use an obligation to somewhat simplify the demonstration of norm internalisation.

$$\begin{array}{ll} 1, obl^t : \ t > 200 & \Rightarrow \mathsf{Obl}_{obj}^t right \\ 1, vio1_i^t : \mathsf{Obl}_{obj}^t right & \Rightarrow \mathsf{Hold}_{obj}^t viol(i) \\ 1, vio2_i^t : \mathsf{Obl}_{obj}^t right, E_i^t right & \Rightarrow \neg\mathsf{Hold}_{obj}^t viol(i) \end{array}$$

We have a rule encoding the random generation of scenes between agents:

$$\gamma, s_{ij}^t : \Rightarrow \mathsf{Hold}_{obj}^t scene(i,j)$$

We have rules defining the cases where an accident may occur:

$$\begin{array}{lll} 0.5, ko_{ij}^t : \mathsf{Hold}_{obj}^t scene(i,j) & & \Rightarrow \mathsf{Hold}_{obj}^t accident(i,j) \\ 1, ok.1_i^t : \mathsf{Hold}_{obj}^t scene(i,j), & \mathsf{E}_i^t left, \ \mathsf{E}_j^t left & \Rightarrow \neg\mathsf{Hold}_{obj}^t accident(i,j) \\ 1, ok.2_i^t : \mathsf{Hold}_{obj}^t scene(i,j), & \mathsf{E}_i^t right, \mathsf{E}_j^t right & \Rightarrow \neg\mathsf{Hold}_{obj}^t accident(i,j) \end{array}$$

The outcomes rules are defined as follows:

$$
\begin{array}{lll}
1, out1_i^t: & \mathsf{E}_i^t left & \Rightarrow \mathsf{Hold}_i^t out_i(10) \\
1, out2_i^t: & \mathsf{E}_i^t right & \Rightarrow \mathsf{Hold}_i^t out_i(10) \\
1, out3.1_i^t: & \mathsf{Hold}_{obj}^t accident(i,j) & \Rightarrow \mathsf{Hold}_i^t out_i(-500) \\
1, out3.2_j^t: & \mathsf{Hold}_{obj}^t accident(i,j) & \Rightarrow \mathsf{Hold}_i^t out_j(-500) \\
1, out4_i^t: & \mathsf{Hold}_{obj}^t viol(i) & \Rightarrow \mathsf{Hold}_i^t out_i(-200) \\
 & & \\
1, ina_i^t: & & \Rightarrow \mathsf{Hold}_{obj}^t inaction(i) \\
1, na1_i^t: & E_i^t left & \Rightarrow \neg\mathsf{Hold}_{obj}^t inaction(i) \\
1, na2_i^t: & E_i^t right & \Rightarrow \neg\mathsf{Hold}_{obj}^t inaction(i) \\
1, out5_i^t: & \mathsf{Hold}_{obj}^t inaction(i) & \Rightarrow \mathsf{Hold}_i^t out_i(-1000)
\end{array}
$$

The rule $out5_i^t$ is artificially settled to force agents to drive so that we can show the emergence of a convention. Without this rule, agents strongly tend to inhibit themselves in order to avoid accidents.

At each time t, agents are randomly paired in a scene. In a scene, if the agents both drive left, or right, then all is fine, while if they do not drive on the same side of the road, then an accident may occur. An obligation to drive on the right side is introduced at time 200. A violation triggers a fine.

At any time, every agent has the choice amongst the behaviours given in the following table. The last behaviour, that is the compliance via internalisation, is a sub-behaviour, and it is only possible when the obligation exists.

Name of the behaviours	Justified Arguments leading to the actions
Inaction	$\emptyset$
Drive on the left	$L^t: \Rightarrow_{l_i^t} \mathsf{E}_i^t left$
Drive on the right	$R^t: \Rightarrow_{r_i^t} \mathsf{E}_i^t right$
Drive on the right (via internalisation of the obligation)	$C^t: INT^t \Rightarrow_{c_i^t} \mathsf{E}_i^t right$

with the following arguments:

$$
\begin{array}{l}
OBL^t: \quad t > 200 \Rightarrow_{obl^t} \mathsf{Obl}_{obj}^t right \\
INT^t: OBL \Rightarrow_{int_i^t} \mathsf{Obl}_i^t right
\end{array}
$$

A typical simulation run is given in the Figure 3 where the probability of different behaviours is graphed w.r.t. time. At the initial conditions ($t = 0$), the probabilities for any agent to select a behaviour are equal. The graph shows an emergence of a convention: all agents are strongly inclined to drive on one side of the road (but this is a coordination game and so the emergence of the convention deciding to drive on the other side of the road is equiprobably possible). At the instant $t = 200$, an exogenous law introduces the obligation to drive on the other side of the road, and as a result, agents begins to change behaviour. The choice of the different behaviours 'Drive on the right' and 'Drive on the right

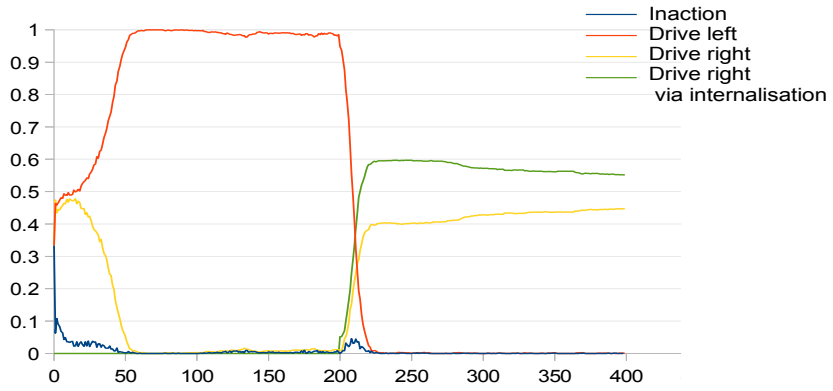

Fig. 3. Emergence of a convention in a population of 1000 agents. Horizontal axis: time. Vertical axis: probability of behaviours.

via internalisation' allows us to show whether agents drive on the right because they have the possibility to do so, or by compliance via the internalisation of the obligation. Interestingly, some agents may continue to drive on the left for few turns: this illustrates "sluggish" norms which, once emerged, tend to guide the behaviour of agents even when they are facing a new environment.

6.2 Case study 2: Norm internalisation and punishment

Let us consider another case of norm internalisation where agents can punish each other. Here, we do not adopt a rich cognitive model such as the one in [2], where norm internalisation means "a mental process that takes (social) norms as inputs and gives new terminal goals of the internalizing agent". We assume that the presence of the norm in an agent theory and its probability indicate respectively its internalisation and a (dynamic) measure of internalisation. However, this view is simple, so we design below a case where the norm is in principle exogenous and does not belong to the agents' theory. What is endogenous are the divergent mechanisms of deviance and of punishing the other agents for their own violations: both of them are in agent's theories and their probability can change. In that sense, internalisation is here a concern of compliance. Other interesting points of our simulation are that punishments are in fact group compensations to recover from violations (such as in tort law) and can be collective. The environment is possibly represented by the rules:

$$
\begin{array}{lll}
1, obl^t : & & \Rightarrow \mathsf{Obl}^t_{obj} a \\
1, com^t_i : & \mathsf{Obl}^t_{obj} a, \mathsf{E}^t_i a & \Rightarrow \mathsf{Hold}^t_{obj} comply(i) \\
1, vio^t_i : & \mathsf{Obl}^t_{obj} a, \neg \mathsf{E}^t_i a & \Rightarrow \mathsf{Hold}^t_{obj} viol(i) \\
1, dam^t_j : & \mathsf{Hold}^t_{obj} viol_i & \Rightarrow \mathsf{Hold}^t_{obj} damage(j)
\end{array}
$$

Any agent's compliance, deviance or punishment is represented as:

$$\begin{array}{l|l} \_, int_i^t : \mathsf{Obl}_{obj}^t a \Rightarrow \mathsf{Obl}_i^t a & c_i^t : \mathsf{Obl}_i^t a \Rightarrow \mathsf{E}_i^t a \\ \_, v_i^t : \mathsf{Obl}_i^t a \Rightarrow \neg \mathsf{E}_i^t a & p_i^t : \mathsf{Obl}_i^t a, \neg \mathsf{E}_j^t a \Rightarrow \mathsf{E}_i^t punish(j) \end{array}$$

with the following outcomes rules:

$$\begin{array}{l} 1, out1_j^t : \mathsf{Hold}_{obj}^t damage(j) \Rightarrow \mathsf{Hold}_i^t out_j(out1) \\ 1, out2_i^t : \mathsf{Obl}_{obj}^t a, \neg \mathsf{E}_i^t a \Rightarrow \mathsf{Hold}_i^t out_i(out2) \\ 1, out3_i^t : \bigwedge_j^N \mathsf{E}_j^t punish(i) \Rightarrow \mathsf{Hold}_i^t out_i(out3) \\ 1, out4_j^t : \mathsf{E}_j^t punish(i) \Rightarrow \mathsf{Hold}_i^t out_j(out4) \end{array}$$

where $out1, out2, out3$ and $out4$ are real numbers, and the outcomes rules have the following meaning:

- $out1_i^t$: a violation damages any agent j, $out1 < 0$.
- $out2_i^t$: a deviant agent gets some benefits from the violation, $out2 > 0$.
- $out3_i^t$: an agent j punished by a threshold of N agents pays for the violation, typically $out3 = \alpha.out1$, $\alpha \geq 1$.
- $out4_i^t$: a punishing agent i receives a compensation, $out4 = -\alpha.out1$.

For any agent, the possible behaviours are: inaction, punish, comply, violate, punish and comply, punish and violate.

By slightly varying environmental rules and outcome values, it results a multitude of scenarios. Among them, an interesting and typical simulation run is given in the Figure 4 where probabilities of behaviours are graphed w.r.t. time. In this scenario, we can observe periodic picks of compliance. When the compliance probability is at its highest, the probability for deviance and punishment is at its lowest. As time goes by, any attempt for deviance brings benefits without punishment, thus the probability of deviance and punishment increases while the probability of compliance decreases. At a certain point, the probability of punishment attains a level for which collective punishment is realized. It follows a strong punishment of the deviant agents who return to become highly compliant.

7 Related work and discussion

Probabilistic argumentation. Our framework is a more formal development of the probabilistic rule-based setting used in [18,20,21,19]. Other work on probabilistic argumentation is Dung and Thang [10] where the notion of probabilistic argumentation graph is used along assumption-based argumentation. A related approach based on probabilistic argumentation graphs is proposed by Hunter in [13]. This notion of probabilistic argumentation graph is based on the consideration of sub-graphs of an argumentation graph: probabilities are then attached to those sub-graphs. It is easy to see that that each sub-graph corresponds to a set of worlds in our framework. Thus, the three approaches are very related,

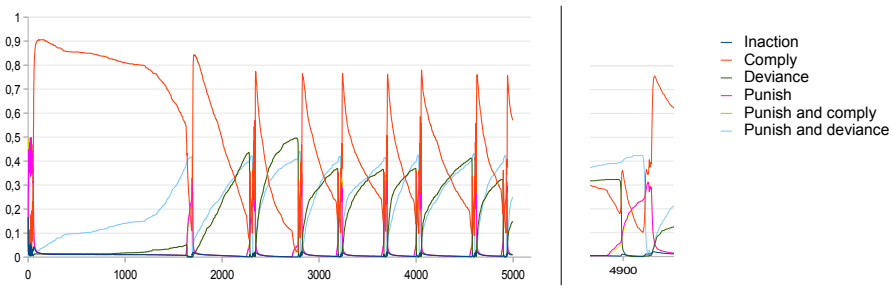

Fig. 4. Picks of compliance in a population of 1000 agents, and zoom on the last pick. Horizontal axis: time. Vertical axis: probability of behaviours.

while we instantiate the Dung's argumentation framework with rule-based arguments. Another line of long-term research is lead by Haenni [12], but how the setting may apply to rule-based argumentation instantiating Dung's argumentation framework is not clear.

Multi-agent learning. Multi-agent learning is a wide topic (see e.g. [23]). We choose the paradigm of reinforcement learning since punishment can be understood as a normative feedback for agents interacting in an environment endowed with normative systems.

Work on multi-agent reinforcement learning (see [4] for a review) present some challenges. Foremost among these is the definition of the goals. Some work put forward the stability of learning dynamics in order to converge to a stationary equilibrium (e.g. a Nash equilibrium), while other goals include the adaptation to the behaviour of the other agents.

The stability requirement is advocated to help for the analysis of performance guarantees. In our case, we are more interested in a relevant model of learning for social simulations than performance guarantees. For example, we do not require that an agent converges to a best-response when the other agents remain stationary. On the contrary, the barriers for the non-convergence to a best-response have to be identified and understood at the level of the encoded social systems. Obviously, a best-response behaviour is fully acceptable but there is overwhelming evidence that such behaviours are not always observed in human societies. So, though the convergence to equilibria is a substantial aspect for analyse, such convergence or stability is not a fundamental requirement at all for our purposes.

The adaptation requirement according to which agents should adapt to the dynamic behaviour of other agents is receivable for some aspects in regard to simulation. However, some usual adaptation requirements seem irrelevant for our case. For example, the notion of no-regret, sometimes but not always requested, implies that any agent has to perform at least as good as in the cases of any stationary strategy. No-regret requirement prevents any learning learner from

being exploited by other agents. Clearly, though no-regret agents are acceptable and a system where no learning agent is exploited is desirable, many agents in real conditions do not perform so well.

To resume, instead of a learning mechanism especially tailored for multi-agents to guarantee some properties such as the convergence to a Nash equilibrium, we assume that each agent is endowed with a reinforcement learning typically used for individual agent learning in order to reproduce the fallible adaptation of individual agents. Nevertheless, our model-free multi-agent learning approach in which agents do not have an explicit model of other agents' strategies is aimed to be improved for future investigations.

Game-theory. In terms of game-theory, our framework is most naturally interpretable as a repeated game played amongst agents within an environment encoded into a defeasible logic. However, our approach based on learning contrasts to the traditional game-theoretical rational choice model of compliance according to which agents calculate the benefit of violations against the cost of norm compliance, and choose actions maximizing the expected utility.

The approximation according to which a mental state is attached to a probability can be re-approached to the use of a mixed strategy in game theory. Indeed, a mixed strategy is usually defined as a probability distribution that assigns to each available action a likelihood of being selected. If the learning parameter τ in the Boltzmann distribution tends to zero, then the behaviour with the highest quality so far is selected with probability 1 and any other behaviour has probability 0. So, one can regard simulations where agents select the best strategy as a degenerate case.

Our agents have incomplete information in the sense they behave simultaneously without any knowledge of each other behaviours. Also, though we do not assume that agents have perfect information. The agents behave without knowledge of previous behaviours of other agents but they may have some imperfect information indirectly provided by the online weighted quality of possible behaviours.

A related game-theoretical framework is evolutionary-game theory interpreted in terms of cultural evolution according which best adapted behaviours propagate in a population (c.f. [24]). An evolutionary-game theoretical approach provides an objective perspective (i.e. a global point of view) on the dynamics of behaviours. So, it would be very interesting to investigate the relationships between the present framework and replicator equations.

Normative multi-agent systems. There is plethora of frameworks for normative multi-agent systems, and many on social simulation works on norm emergence and spreading, which are based on specific mechanisms, such as social power, leadership, sanction, reputation, imitation, off-line design, machine learning, cognitive architectures, emotions, network topologies (see e.g. [22]), etc. Many simulations are not implemented using a logic programming language,

and thus, their operational model is left to the programmer with little possibility to monitor or verify it with respect to the specification. At the opposite, we have models expressed in modal logics aimed at providing a verification semantics of such programs, but, on the assumptions that they are not undermined by so-called paradoxes, they are usually not applied in practice. Between those extremes, we have frameworks on executable logic-based specifications, for instance, [14] modelling self-governed institutions, that have a foundation in logic, with an implementation. But, to our knowledge, they do not offer any integration with probability theory and learning so far. Finally, no system integrating probabilistic rule-based argumentation logic-based agents with reinforcement learning capabilities have been proposed in the literature.

8 Summary and future work

In this paper, we proposed a framework with a dual formalism based on probabilistic rule-based argumentation to model and build norm-governed learning agents. This integration allowed us to cater for:

- an account of environment model-free learning agents with fallible and bounded rationality interacting in non-deterministic environments,
- the integration of quantitative with qualitative reasoning to model a system,
- a better communication among experts from different disciplines or opinions when modelling or building up a (multi-agent) system, thanks to the ergonomic argumentation framework,
- a clear formalisation of multi-agent models with well-known argumentation semantics,
- executable specification to faithfully run norm-governed systems,
- the simulation of norm emergence as well as the expression of exogenous norms.

This framework is also meant to help to fill the gap between efforts in AI&Law and research on multi-agent based simulation (c.f. [5]).

With regard to future work, we would like to test our approach with more advanced and significant simulated scenarios to confirm its potential and feasibility.

Moreover, though argumentation-based reasoning and reinforcement learning are salient features of cognition, agents' behaviours cannot be reduced to them in many (social) activities. We hope that the logic foundation of the framework will help us to develop more sophisticated cognition. On this regard, we have in mind the extensions of Defeasible Logic on temporal, epistemic and deontic aspects (see e.g. [11]) since they are interpretable in a Dung's argumentation framework.

With the integration of a probabilistic rule-based argumentation with learning, we hope to bring a useful tool at the service of the multi-agent systems community, and in particular to investigate norm-governed learning agents systems.

Acknowledgments

This paper extends and revises some preliminary work of [19]. We would like to thank the anonymous reviewers of DEON 2012 for their comments. Our gratitude goes to Giulia Andrighetto and Mario Paolucci for their valuable suggestions. Part of this work has been carried out in the scope of the EC co-funded project SMART (FP7-287583).

References

1. G. Andrighetto, M. Campennì, F. Cecconi, and R. Conte. The complex loop of norm emergence: A simulation model. In *Simulating Interacting Agents and Social Phenomena*. Springer, 2010.
2. G. Andrighetto, D. Villatoro, and R. Conte. Norm internalization in artificial societies. *AI Commun.*, 23(4):325–339, 2010.
3. L. Brooks, W. Iba, and S. Sen. Modeling the emergence and convergence of norms. In T. Walsh, editor, *IJCAI*, pages 97–102. IJCAI/AAAI, 2011.
4. L. Buşoniu, R. Babuška, and B. De Schutter. A comprehensive survey of multi-agent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(2):156–172, Mar. 2008.
5. R. Conte, R. Falcone, and G. Sartor. Introduction: Agents and norms: How to fill the gap? *Artificial Intelligence and Law*, 7(1):1–15, 1999.
6. P. Davidsson. Multi agent based simulation: Beyond social simulation. In *Procs. MABS 2000*, pages 97–107. Springer, 2000.
7. P. Davidsson. Agent based social simulation: A computer science view. *J. Artificial Societies and Social Simulation*, 5(1), 2002.
8. D. C. Dennett. *The Intentional Stance (Bradford Books)*. The MIT Press, Cambridge, MA, 1987.
9. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.
10. P. M. Dung and P. M. Thang. Towards (probabilistic) argumentation for jury-based dispute resolution. In P. Baroni, F. Cerutti, M. Giacomin, and G. R. Simari, editors, *COMMA*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 171–182. IOS Press, 2010.
11. G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in defeasible logic. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, pages 25–34. ACM Press, 2005.
12. R. Haenni. Probabilistic argumentation. *J. Applied Logic*, 7(2):155–176, 2009.
13. A. Hunter. Some foundations for probabilistic abstract argumentation. In *Proceedings of the 4th International Conference on Computational Models of Argument (COMMA 2012)*, 2012.
14. J. Pitt, J. Schaumeier, and A. Artikis. The axiomatisation of socio-economic principles for self-organising systems. *Self-Adaptive and Self-Organizing Systems, IEEE International Conference on*, 0:138–147, 2011.
15. J. L. Pollock. *Cognitive Carpentry*. MIT Press, 1995.
16. H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2011.

17. H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1), 1997.
18. R. Riveret, H. Prakken, A. Rotolo, and G. Sartor. Heuristics in argumentation: A game theory investigation. In *COMMA*, pages 324–335, 2008.
19. R. Riveret, A. Rotolo, and G. Sartor. Norms and learning in probabilistic logic-based agents. In *DEON 2012*. Springer, 2012.
20. R. Riveret, A. Rotolo, G. Sartor, H. Prakken, and B. Roth. Success chances in argument games: a probabilistic approach to legal disputes. In *Proceeding of the 2007 conference on Legal Knowledge and Information Systems: JURIX 2007: The Twentieth Annual Conference*, pages 99–108, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press.
21. B. Roth, R. Riveret, A. Rotolo, and G. Governatori. Strategic argumentation: a game theoretical investigation. In *Proceedings of the 11th international conference on Artificial intelligence and law*, ICAIL '07, pages 81–90, New York, NY, USA, 2007. ACM.
22. B. T. R. Savarimuthu and S. Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011.
23. Y. Shoham, R. Powers, and T. Grenager. If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, 171(7):365 – 377, 2007.
24. K. Tuyls and A. Nowe. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review*, 20(01):63–90, 2005.