A Probabilistic Doxastic Temporal Logic for Reasoning about Beliefs in Multi-agent Systems

Karsten Martiny[1] and Ralf Möller[2]
[1]*Institute for Software Systems, Hamburg University of Technology, Hamburg, Germany*
[2]*Institute of Information Systems, University of Lübeck, Lübeck, Germany*
karsten.martiny@tuhh.de, moeller@ifis.uni-luebeck.de

Keywords: Multi-Agent Systems, Epistemic Logic, Doxastic Logic, Dynamic Belief Change

Abstract: We present *Probabilistic Doxastic Temporal (PDT)* Logic, a formalism to represent and reason about probabilistic beliefs and their evolution in multi-agent systems. It can quantify beliefs through probability intervals and incorporates the concepts of frequency functions and epistemic actions. We provide an appropriate semantics for PDT and show how agents can update their beliefs with respect to their observations.

1 Introduction

When logically analyzing knowledge and belief in realistic scenarios, an agent usually has only incomplete and inaccurate information about the actual state of the world, and thus considers several worlds as being possible. As it receives new information, it has to update its beliefs about possible worlds. These updates can for example result in regarding some worlds as impossible or judging some worlds to be more likely than before. Thus, in addition to analyzing the set of worlds an agent believes to be possible, it is also useful to quantify these beliefs in terms of probabilities. This provides means to specify fine-grained distinctions within the range of worlds that an agent considers possible.

When multiple agents are involved in such a setting, an agent may not only have varying beliefs regarding the facts of the actual world, but also regarding the beliefs of other agents. In many scenarios, the actions of one agent will not only depend on its belief of ontic facts (i.e., facts of the actual world), but also on its beliefs in some other agent's beliefs.

To formalize reasoning about such beliefs in multi-agent settings, we present *Probabilistic Doxastic Temporal (PDT) Logic*. PDT Logic builds upon recent work on Annotated Probabilistic Temporal (APT) Logic and provides a formalism which enables the representation of and reasoning about dynamically changing quantified temporal multi-agent beliefs through probability intervals. In this formalism, analyses are intended to be carried out offline by an external observer. In contrast to related work, PDT Logic employs an explicit notion of time and thereby facilitates the expression of richer temporal relations.

The remainder of this work is structured as follows: The next section presents related work about knowledge in multi-agent systems and APT Logic. Then, in Section 3, the syntax of PDT Logic is introduced, followed by the definition of formal semantics in Section 4. The evolution of multi-agent beliefs over time is analyzed in Section 5. Finally, the paper concludes with Section 6.

2 Related Work

Approaches to formalize reasoning about knowledge and belief date back to Hintikka's work on epistemic logic (Hintikka, 1962). Classical forms of epistemic logic do not allow for a quantification of an agent's degree of belief in certain facts; it can only be specified whether an agent does or does not know (resp. believe) some fact. To remove this limitation, several approaches have been proposed to combine logics of knowledge and belief with probabilistic quantifications. For instance, (Fagin and Halpern, 1994) and (van der Hoeck, 1997) define a belief operator to quantify lower bounds on the probabilities that an agent assigns to a formula.

To reason about dynamically changing beliefs, extensions to epistemic logics have been proposed, e.g., (Scherl and Levesque, 2003). In these works only the single-agent case is considered, and therefore they do not provide for representations of nested beliefs.

Multi-agent extensions to these approaches can be found for example in (van Ditmarsch et al., 2007). A common limitation of these works is that they are only able to reason about step-by-step changes and therefore explicit reasoning about time is difficult in these frameworks. (Renne et al., 2009) alleviates these limitations by combining Dynamic Epistemic Logic (van Ditmarsch et al., 2007) with temporal modalities.

(Shakarian et al., 2011) introduce APT Logic, a framework to represent probabilistic temporal evolutions of worlds in threads. APT Logic assigns prior probabilities to every thread and uses these probabilities to determine probabilities of events occurring in specific threads. To represent temporal relationships between events, APT Logic introduces the concept of frequency functions. We utilize the approach of APT Logic to create a doxastic multi-agent framework that can explicitly reason about temporal relationships through the adoption of frequency functions.

3 PDT Logic programs: Syntax

In this section, we start with defining the syntax of PDT Logic programs, and then give a definition of the formal semantics in the next section.

We assume the existence of a function-free first order logic language with finite sets of constant symbols $\mathcal{L}_{cons}$ and predicate symbols $\mathcal{L}_{pred}$, and an infinite set of variable symbols $\mathcal{L}_{var}$. Every predicate symbol $p \in \mathcal{L}_{pred}$ has an *arity*. A *term* is any member of the set $\mathcal{L}_{cons} \cup \mathcal{L}_{var}$. A term is called a *ground term* if it is a member of $\mathcal{L}_{cons}$. If $t_1,..,t_k$ are (ground) terms, and p is a predicate symbol in $\mathcal{L}_{pred}$ with arity n, then $p(t_1,...,t_k)$ with $k \in \{0,...,n\}$ is a (ground) atom. If a is a (ground) atom, then a and $\neg a$ are (ground) *literals*. The set of all ground literals is denoted by $\mathcal{L}_{lit}$. As usual, $\mathcal{B}$ denotes the Herbrand Base of $\mathcal{L}$.

Time is modeled in discrete steps and we assume that all agents reason about an arbitrarily large, but fixed size window of time. The set of time points is given by $\tau = \{1,...,t_{max}\}$. The set of agents is denoted by $\mathcal{A}$. The number of agents ($|\mathcal{A}|$) is denoted by n. To describe what agents observe, we define observation atoms as follows:

Definition 1 (Observation atoms)**.** For any group of agents $\mathcal{G} \subseteq \mathcal{A}$ and ground literal $l \in \mathcal{L}_{lit}$, $Obs_{\mathcal{G}}(l)$ is an *observation atom*. The set of all observation atoms is denoted by $\mathcal{L}_{obs}$.

Intuitively, the meaning of a statement of the form $Obs_{\mathcal{G}}(l)$ is that all agents in the group $\mathcal{G}$ observe that the fact l holds. We assume that the agents in $\mathcal{G}$ not only observe that l holds, but that each agent in $\mathcal{G}$ is also aware that all other agents in $\mathcal{G}$ make the same observation.

Definition 2 (Formulae)**.** Atoms and observation atoms are formulae. If F and G are formulae, then so are $F \wedge G$, $F \vee G$, and $\neg F$. A formula is ground if all atoms of the formula are ground.

To describe observations at a specific time, we furthermore define *time-stamped observation atoms*:

Definition 3 (Time-stamped observation atoms)**.** If $Obs_{\mathcal{G}}(l) \in \mathcal{L}_{obs}$ is an observation atom, and $t \in \tau$ is a time point, then $[Obs_{\mathcal{G}}(l) : t]$ is a time-stamped observation atom.

To express temporal relationships, we define temporal rules following the approach of APT rules from (Shakarian et al., 2011).

Definition 4 (Temporal rules)**.** Let F, G be formulae, Δt a time interval, and fr a name for a so-called frequency function (as defined below in Definition 11). Then $r^{\mathrm{fr}}_{\Delta t}(F,G)$ is called a temporal rule.

The meaning of such an expression is "F is followed by G in Δt time units w.r.t. fr".

Now, we can define the belief operator $B^{\ell,u}_{i,t'}$ to express agents' beliefs. Intuitively, $B^{\ell,u}_{i,t'}(\cdot)$ means that at time t', agent i believes that some fact $(\cdot)$ is true with a probability $p \in [\ell,u]$. We call the probability interval $[\ell,u]$ the *quantification* of agent i's belief. We use F_t to denote that formula F holds at time t.

Definition 5 (Belief formulae)**.** Let i be an agent, t' a time point, and $[\ell,u] \subseteq [0,1]$. Then, *belief formulae* are inductively defined as follows:

1. If F is a formula and t is a time point, then $B^{\ell,u}_{i,t'}(F_t)$ is a belief formula.
2. If $r^{\mathrm{fr}}_{\Delta t}(F,G)$ is a temporal rule, then $B^{\ell,u}_{i,t'}(r^{\mathrm{fr}}_{\Delta t}(F,G))$ is a belief formula.
3. If F and G are belief formulae, then so are $B^{\ell,u}_{i,t'}(F)$, $F \wedge G$, $F \vee G$, and $\neg F$.

4 Semantics

In this section, we will provide a formal semantics that captures the intuitions explained above. We start with the introduction of an example, which we will return to repeatedly when introducing the various concepts of the semantics.

Example 1 (Trains)**.** *Let Alice and Bob be two agents living in two different cities C_A and C_B, respectively. Suppose that Alice wants to take a train to visit Bob* *and has to change trains at a third city C_C. We assume that train T_1 connects C_A and C_C, and train T_2 connects C_C and C_B. Both trains usually require 2 time units for their trip, but they might be running late and arrive one time unit later than scheduled. Alice requires one time unit to change trains at city C_C. If T_1 runs on time, she has a direct connection to T_2, otherwise she has to wait for two time units until the next train T_2 leaves at city C_C. If a train is running late, she can call Bob to let him know. These calls can be modeled as shared observations between Alice and Bob. For instance, if Alice wants to tell Bob that train T_1 is running late (i.e., T_1 does not arrive at C_C at the expected time), this can be modeled as $Obs_{AB}(\neg at(T_1, C_C))$ at the expected arrival time.*

4.1 Possible Worlds

Ontic facts and according observations form *worlds*. A world w consists of a set of ground atoms and a set of observation atoms, i.e., $w \in 2^{\mathcal{B}} \times 2^{\mathcal{L}_{obs}}$. With a slight abuse of notation, we use $a \in w$ and $Obs_{\mathcal{G}}(l) \in w$ to denote that an atom a (resp. observation atom $Obs_{\mathcal{G}}(l)$) holds in world w. Since agents can only observe facts that actually hold in the respective world, we can define consistency of worlds w.r.t. the set of observations:

Definition 6 (World consistency)**.** A world w is consistent, iff for every observation atom $Obs_{\mathcal{G}}(l) \in w$, the observed fact holds, i.e., $x \in w$ if l is a positive literal x, $x \notin w$ if l is a negative literal $\neg x$.

The set of all possible worlds is denoted by W. For the following discussion we assume a manual succinct specification of possible worlds depending on the respective domain. Especially, we assume in the following discussion that W does not contain any inconsistent worlds according to Definition 6.

Example 2 (Trains continued)**.** *For Example 1, we have ground terms A, B, C_A, C_B, C_C, T_1, and T_2, representing Alice, Bob, three cities, and two trains. Furthermore, we have atoms $on(x,y)$ indicating that person y is on train x, and $at(y,z)$ indicating that train y is at city z. Finally, we have observation atoms of the kind $Obs_{\mathcal{G}}(at(y,z))$, indicating that the agents in $\mathcal{G}$ observe that train y is at station z. Thus, a possible world can for example be $w_1 = \{at(T_1, C_A), on(T_1, A), Obs_A(at(T_1, A))\}$, indicating that train T_1 is at city C_A and A has boarded that train.*

We define satisfaction of a ground formula F by a world w, in the usual way (Lloyd, 1987):

Definition 7 (Satisfaction of ground formulae)**.** Let F, F', F'' be ground formulae and w a world. Then, F is satisfied by w (denoted $w \models F$)
- If $F = a$ for some ground atom a, then $a \in w$.
- If $F = \neg F'$, then $w \not\models F'$.
- If $F = F' \wedge F''$, then $w \models F'$ and $w \models F''$.
- If $F = F' \vee F''$, then $w \models F'$ or $w \models F''$.

4.2 Threads

We use the definition of *threads* from (Shakarian et al., 2011) (equivalent to the concept of *runs* in (Fagin et al., 1995)):

Definition 8. A *thread* is a mapping $Th : \tau \rightarrow W$

Thus, a thread is a sequence of worlds and $Th(i)$ identifies the actual world at time i according to thread Th. The set of all possible threads is denoted by $\mathcal{T}$. Again, we refrain from using $\mathcal{T}$ as the set of all possible sequences constructible from τ and W, and instead assume that any meaningful problem specification gives information about possible temporal evolutions of the system. For notational convenience, we assume that there is an additional prior world $Th(0)$ for every thread.

Example 3 (Trains continued)**.** *The description from Example 1 yields the set of possible threads $\mathcal{T}$ depicted in Figure 1.*

4.3 Kripke Structures

With the definition of threads, we can use a slightly modified version of Kripke structures (Kripke, 1963). As usual, we define a Kripke structure as a tuple $\langle W, \mathcal{K}_1, ..., \mathcal{K}_n \rangle$, with the set of possible worlds W and binary relations $\mathcal{K}_i$ on W for every agent $i \in \mathcal{A}$. Intuitively, $(w, w') \in \mathcal{K}_i$ specifies that in world w, agent i considers w' as a possible world.

We initialize the Kripke structure such that the set of possible worlds contains exactly the worlds that occur at time $t = 1$ in some thread Th':

$$\forall Th \in \mathcal{T} : \ \mathcal{K}_i(Th(0)) := \bigcup_{Th' \in \mathcal{T}} \{Th'(1)\}, \ i = 1, ..., n$$

With the evolution of time, each agent can eliminate the worlds that do not comply with its respective observations. Through the elimination of worlds, an agent will also reduce the set of threads it considers possible. We assume that agents have perfect recall and therefore will not consider some thread possible again if it was considered impossible at one point. Thus, $\mathcal{K}_i$ is updated w.r.t. the agent's respective observations, such that it considers all threads possible that both comply with its current observations and were

	t	1	2	3	4	5	6	7	8	9	10
	Th_1	$at(T_1,C_A)$ $on(T_1,A)$		$at(T_1,C_C)$ $on(T_1,A)$	$at(T_2,C_C)$ $on(T_2,A)$		$at(T_2,C_B)$ $on(T_2,A)$				
②	Th_2	$at(T_1,C_A)$ $on(T_1,A)$		$at(T_1,C_C)$ $on(T_1,A)$	$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T2,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$			
②	Th_3	$at(T_1,C_A)$ $on(T_1,A)$		$at(T_1,C_C)$ $on(T_1,A)$	$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_2,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$			
①	Th_4	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$at(T_2,C_B)$ $on(T_2,A)$		
①	Th_5	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$at(T_2,C_B)$ $on(T_2,A)$		
① ②	Th_6	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T_1,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$	
① ②	Th_7	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T_1,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$	
① ②	Th_8	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{A}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_1,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$	
① ②	Th_9	$at(T_1,C_A)$ $on(T_1,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_1,C_C))$	$at(T_1,C_C)$ $on(T_1,A)$		$at(T_2,C_C)$ $on(T_2,A)$		$Obs_{\mathbf{AB}}$ $(\neg at(T_1,C_B))$	$at(T_2,C_B)$ $on(T_2,A)$	

Figure 1: Visualization of the possible threads Th_k from Example 1: $at(T_i,C_j)$ denotes that train T_i is currently at city C_j, $on(T_i,A)$ that Alice is currently on train T_i, and $Obs_{AG}(\neg at(T_i,C_j))$ denotes a call from Alice to inform Bob that train T_i is currently not at city C_j. For the sake of simplicity, facts irrelevant to the analysis (such as $on(T_i,A)$ for time points 2 and 5) are omitted from the presentation. Note that if a train is running late (the respective threads are marked with according circles), there are always two possible threads: one where only A observes this and one where both share the observation. For an easier distinction, we have marked the according group of an observation with boldface indices.

considered possible at the previous time point:

$$\mathcal{K}_i(Th(t)) := \{Th'(t) : (Th'(t-1) \in \mathcal{K}_i(Th(t-1)) \wedge \{Obs_{\mathcal{G}}(l) \in Th(t) : i \in \mathcal{G}\} = \{Obs_{\mathcal{G}}(l) \in Th'(t) : i \in \mathcal{G}\})\} \quad (1)$$

The next lemmata describe key properties of $\mathcal{K}_i$ following immediately from the above definitions.

Lemma 1. *$\mathcal{K}_i$ defines an equivalence relation over the possible worlds $\mathcal{K}_i(Th(t))$ at time t.*

Lemma 2. *The set of threads Th' considered possible w.r.t. $\mathcal{K}_i$ is narrowing to a smaller and smaller subset over time, i.e., $\{Th' : Th'(t) \in \mathcal{K}_i(Th(t))\} \subseteq \{Th' : Th'(t-1) \in \mathcal{K}_i(Th(t-1))\}$ for all $Th \in \mathcal{T}$ and $t \in \tau$.*

Example 4 (Trains continued)**.** *From Figure 1, we obtain that at time 1, the only possible world is $\{\{at(T_1,C_A), on(T_1,A)\}\}$, which is contained in all possible threads. Thus, $\mathcal{K}_i(Th_j(1))$ contains exactly this world for all agents i and threads j. Consequently, both agents consider all threads as possible at time* 1.

Now, assume that time evolves for two steps and the actual thread is Th_4 (i.e., train T_1 is running late, but A does not inform B about this). Both agents will update their possibility relations accordingly, yielding $\mathcal{K}_1(Th_4(3)) = \{\{Obs_A(\neg at(T_1,C_C))\}\}$ and $\mathcal{K}_2(Th_4(3)) = \{\{at(T_1,C_C), on(T_1,A)\}, \{Obs_A(\neg at(T_1,C_C))\}\}$, i.e., A knows that $T1$ is not on time, while B is unaware of this.

4.4 Subjective Posterior Temporal Probabilistic Interpretations

Each agent has probabilistic beliefs about the expected evolution of the world over time. This is expressed through subjective temporal probabilistic interpretations:

Definition 9 (Subjective posterior probabilistic temporal interpretation)**.** Given a set of possible threads $\mathcal{T}$, some thread $Th' \in \mathcal{T}$, a time point t and an agent i, $I_{i,t}^{Th'} : \mathcal{T} \to [0,1]$ specifies the *subjective posterior probabilistic temporal interpretation* from agent i's point of view at time t in thread Th', i.e., a probability distribution over all possible threads: $\sum_{Th \in \mathcal{T}} I_{i,t}^{Th'}(Th) = 1$. We call Th' the *point of view (pov) thread* of interpretation $I_{i,t}^{Th'}$.

The prior probabilities of each agent for all threads are then given by $I_{i,0}^{Th'}(Th)$. Since all threads are indistinguishable a priori, there is only a *single* prior distribution for each agent (i.e., $\forall Th, Th', Th'' \in \mathcal{T} : I_{i,0}^{Th'}(Th) = I_{i,0}^{Th''}(Th)$). Furthermore, in order to be able to reason about nested beliefs (as discussed below), we assume that the prior probability assessments of all agents are commonly known (i.e., all agents know how all other agents assess the prior probabilities of each thread). This in turn requires that all agents have exactly the same prior probability assessment over all possible threads: if two agents have different, but commonly known prior probability assessments, we essentially have an instance of Aumann's well-known problem of "agreeing to disagree" (Aumann, 1976). Intuitively, if differing priors are commonly known, it is common knowledge that (at least) one of the agents is at fault and should revise its probability assessments. As a result, we have only one prior probability distribution which is the same from all viewpoints, denoted by I. Note that I directly corresponds to the concept of temporal probabilistic interpretations in (Shakarian et al., 2011).

Example 5 (Trains continued)**.** *A meaningful interpretation is*

$I = \begin{pmatrix}0.7 & 0.02 & 0.09 & 0.02 & 0.09 & 0.01 & 0.02 & 0.02 & 0.03\end{pmatrix}$,

which assigns the highest probability to Th_1 (no train running late), lower probabilities to the threads where one train is running late and A informs B (Th_3 and Th_5), even lower probabilities to the events that either both trains are running late and A informs B (Th_7, Th_8, and Th_9) or that one train is running late and A does not inform B (Th_2 and Th_4), and lowest probability to the thread where both trains are running late and A does not inform B (Th_6).

Even though we only have a single prior probability distribution over the set of possible threads, it is still necessary to distinguish the viewpoints of different agents in different threads, as the definition of interpretation updates shows:

Definition 10 (Interpretation update)**.** Let i be an agent, t a time point, and Th' a pov thread. Then, if the system is actually in thread Th' at time t, agent i's probabilistic interpretation over the set of possible threads is given by the update rule:

$$I_{i,t}^{Th'}(Th) = \begin{cases} \frac{1}{\alpha_{it}^{Th'}} \cdot I_{i,t-1}^{Th'}(Th) & \text{if } Th(t) \in \mathcal{K}_i(Th'(t)) \\ 0 & \text{if } Th(t) \notin \mathcal{K}_i(Th'(t)) \end{cases} \tag{2}$$

with $\frac{1}{\alpha_{it}^{Th'}}$ being a normalization factor:

$$\alpha_{it}^{Th'} = \sum_{Th \in \mathcal{T}, Th(t) \in \mathcal{K}_i(Th'(t))} I_{i,t-1}^{Th'}(Th) \tag{3}$$

The invocation of $\mathcal{K}_i$ in the update rule yields obvious ramifications about the evolution of interpretations, as stated in the following lemma:

Lemma 3. *The subjective temporal probabilistic interpretation $I_{i,t}^{Th'}$ of an agent i assigns nonzero probabilities exactly to the set of threads that i still considers possible at time t, i.e., $I_{i,t}^{Th'}(Th) > 0 \Leftrightarrow \mathcal{K}_i(Th(t), Th'(t))$*

Essentially, the update rule assigns all impossible threads a probability of zero and scales the probabilities of the remaining threads such that they are proportional to the probabilities of the previous time point.

Example 6 (Trains continued)**.** *Applying the update rule from (2) to the situation described in Example 4, with I as given in Example 5, yields the updated interpretation for A:*

$$I_{A,3}^{Th_4} = \begin{pmatrix}0 & 0 & 0 & 0.4 & 0 & 0.2 & 0 & 0.4 & 0\end{pmatrix} \tag{4}$$

i.e., A considers exactly those threads possible, where the train is running late and she does not inform B (threads Th_4, Th_6, and Th_8). Due to the lack of any new information, B can only eliminate the situations where A does inform him about being late, and thus B's interpretation is updated to:

$$I_{B,3}^{Th_4} \approx \begin{pmatrix}0.82 & 0.02 & 0.10 & 0.02 & 0 & 0.02 & 0 & 0.02 & 0\end{pmatrix}. \tag{5}$$

4.5 Frequency Functions

To represent temporal relationships within threads, we utilize the concept of *frequency functions* as introduced in (Shakarian et al., 2011). Frequency functions enable us to represent temporal relations between the occurrence of specific events and are defined axiomatically as follows:

Definition 11 (Frequency functions)**.** (Shakarian et al., 2011) Let Th be a thread, F and G be ground formulae, and $\Delta t > 0$ be an integer. A *frequency function* fr maps quadruples of the form $(Th, F, G, \Delta t)$ to $[0, 1]$ such that the following axioms hold:

(FF1) If G is a tautology, then $\mathsf{fr}(Th, F, G, \Delta t) = 1$.

(FF2) If F is a tautology and G is a contradiction, then $\mathsf{fr}(Th, F, G, \Delta t) = 0$.

(FF3) If F is a contradiction, $\mathsf{fr}(Th, F, G, \Delta t) = 1$.

(FF4) If G is not a tautology, and either F or $\neg G$ is not a tautology, and F is not a contradiction, then there exist threads Th_1, $Th_2 \in \mathcal{T}$ such that $\mathsf{fr}(Th_1, F, G, \Delta t) = 0$ and $\mathsf{fr}(Th_2, F, G, \Delta t) = 1$.

To illustrate the concept of frequency functions, we present the point and existential frequency functions from (Shakarian et al., 2011):

The point frequency function *pfr* expresses how frequently some event F is followed by another event G in *exactly* Δt time units:

$$pfr(Th, F, G, \Delta t) = \frac{|\{t : Th(t) \models F \wedge Th(t + \Delta t) \models G\}|}{|\{t : (t \leq t_{max} - \Delta t) \wedge Th(t) \models F\}|} \tag{6}$$

The existential frequency function *efr* expresses how frequently some event F is followed by another event G *within* the next Δt time units:

$$efr(Th, F, G, \Delta t) = \tag{7}$$

$$\frac{efn(Th, F, G, \Delta, 0, t_{max})}{fn(Th, F, \Delta t) + efn(Th, F, G, \Delta t, t_{max} - \Delta t, t_{max})},$$

$$fn(Th, F, \Delta t) := |\{t : (t \leq t_{max} - \Delta t) \wedge Th(t) \models F\}|,$$

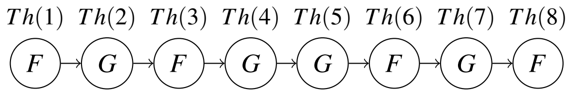

Figure 2: Example thread Th with $\tau = \{1,...,8\}$. This figure shows each world that satisfies formula F or formula G.

$$efn(Th,F,G,\Delta t,t_1,t_2) = |\{t : (t_1 < t \leq t_2) \wedge Th(t) \models F \wedge \exists t' \in [t+1, \min(t_2, t+\Delta t)]\ (Th(t') \models G)\}|$$

To illustrate the concept of frequency functions, we adapt an example from (Shakarian et al., 2011): consider the thread Th depicted in Figure 2. The thread evolves over 8 time steps and in each of the respective worlds, either F or G is satisfied. Suppose that we want to determine how often F is followed by G exactly after two time steps. This can be expressed through a point frequency function: $pfr(Th,F,G,2) = \frac{1}{3}$. If instead we want to know how often F is followed by G within the next two time steps, we can use an existential frequency function: $efr(Th,F,G,2) = \frac{3}{3} = 1$. Note that the frequency functions are defined such that neither of them considers the world at time 8 in the denominator, even though $Th(8) \models F$. This is because there cannot be any world beyond time 8 such that G is satisfied and consequently, considering this world would mitigate any result of the frequency functions.

4.6 Semantics of the Belief Operator

Now, with the definitions of subjective posterior probabilistic temporal interpretations and the introduction of frequency functions, we can build upon the definitions from (Shakarian et al., 2011) for the satisfiability of interpretations to provide formal semantics for the belief operators defined in Section 3:

Definition 12 (Belief in ground formulae)**.** Let $I_{i,t'}^{Th'}$ be agent i's interpretation at time t' in pov thread Th'. Then, it holds w.r.t. this interpretation that agent i believes that some formula F holds at time t with a probability in the range $[\ell,u]$ (denoted by $I_{i,t'}^{Th'} \models B_{i,t'}^{\ell,u}(F_t)$) iff

$$\ell \leq \sum_{Th \in \mathcal{T}, Th(t) \models F} I_{i,t'}^{Th'}(Th) \leq u. \qquad (8)$$

Definition 13 (Belief in rules)**.** Let F and G be ground formulae, fr be a frequency function, and $I_{i,t'}^{Th'}(Th)$ be agent i's interpretation at time t' in pov thread Th'. Then, it holds w.r.t. this interpretation that agent i believes that some rule $r_{\Delta t}^{\mathsf{fr}}(F,G)$ holds with a probability in the range $[\ell,u]$ (denoted by $I_{i,t'}^{Th'} \models B_{i,t'}^{\ell,u}(r_{\Delta t}^{\mathsf{fr}}(F,G))$) iff

$$\ell \leq \sum_{Th \in \mathcal{T}} I_{i,t'}^{Th'}(Th) \cdot \mathsf{fr}(Th,F,G,\Delta t) \leq u. \qquad (9)$$

Definition 14 (Nested beliefs)**.** Let i,j be agents, $B_{k,t}^{\ell_k,u_k}(\cdot)$ be some belief formula, and $I_{i,t'}^{Th'}(Th)$ be agent i's interpretation at time t' in pov thread Th'. Then, it holds w.r.t. this interpretation that agent i believes at time t' that with a probability in the range $[\ell,u]$ agent j has some belief $B_{k,t}^{\ell_k,u_k}(\cdot)$ at time t (denoted by $I_{i,t'}^{Th'} \models B_{i,t'}^{\ell,u}(B_{k,t}^{\ell_k,u_k}(\cdot))$) iff

$$\ell \leq \sum_{Th \in \mathcal{T},\ I_{j,t}^{Th} \models B_{k,t}^{\ell_k,u_k}} I_{i,t'}^{Th'}(Th) \leq u. \qquad (10)$$

Example 7 (Trains continued)**.** *We can use a point frequency function to express beliefs about the punctuality of trains. Assume that both A and B judge the probability of a train running late (i.e., arriving after 3 instead of 2 time units, expressed through the temporal rule r_3^{pfr}) as being at most* 0.4*. This yields the following belief formulae*

$$\begin{array}{l} B_{i,0}^{0,0.4}(r_3^{pfr}(at(T_1,C_A),at(T_1,C_C))) \\ B_{i,0}^{0,0.4}(r_3^{pfr}(at(T_2,C_C),at(T_2,C_B))) \end{array}, \quad i \in \{A,B\}.$$

One can easily verify that these formulae are satisfied by the interpretation given in Example 5.

From the above definitions, we can use the belief about some fact $(\cdot)$ to quantify the belief about the negation of this fact $\neg(\cdot)$:

Lemma 4. $I_{i,t}^{Th'} \models B_{i,t'}^{\ell,u}(\neg(\cdot))$ *iff* $I_{i,t}^{Th'} \models B_{i,t}^{\ell',u'}(\cdot)$ *with* $\ell' = 1-u$ *and* $u' = 1-\ell$.

5 Evolution over Time

In order to completely specify a problem in PDT Logic, we introduce the concept of *doxastic systems*.

Definition 15 (Doxastic system)**.** Let $\mathcal{A}$ be a set of agents, $\mathcal{T}$ be a set of threads, $A_0^{|\mathcal{A}|\times|\mathcal{T}|}$ be a matrix of prior probability distributions across $\mathcal{T}$ for every agent in $\mathcal{A}$, and $\mathcal{F}$ be a set of frequency functions. Then, we call the quadruple $\mathcal{D} = \langle \mathcal{A}, \mathcal{T}, \mathcal{F}, A_0^{|\mathcal{A}|\times|\mathcal{T}|} \rangle$ a *doxastic system*.

Note that several of the parameters discussed before are not explicitly specified in a doxastic system: the set of possible worlds W, the set of ground atoms $\mathcal{B}$, the set of observation atoms $\mathcal{L}_{obs}$, nor the set of time points τ are explicitly specified. However, all relevant information regarding these parameters is already contained in the specification of $\mathcal{T}$.

To identify specific situations in a doxastic system after time has passed and some observations occurred, we furthermore define *pointed doxastic systems*:

Definition 16 (Pointed doxastic system, pds)**.** Let $\mathcal{D} = \langle \mathcal{A}, \mathcal{T}, \mathcal{F}, A_0^{|\mathcal{A}|\times|\mathcal{T}|} \rangle$ be a doxastic system and H be a set of timestamped observation atoms such that all observation atoms from H occur in at least one of the worlds (implicitly) defined in $\mathcal{T}$. Then we call the pair $\langle \mathcal{D}, H \rangle$ a *pointed doxastic system*.

Intuitively, the set of timed observations specified in a pds points to a certain situation in a doxastic system. One could view $\hat{t}(H) = \max\{t : \exists [Obs_{\mathcal{G}}(l) : t] \in H\}$ as the present time in a pds: the most recent observation occurred at $\hat{t}(H)$, all observations that actually occurred in the past ($t < \hat{t}$) are specified in H (and thus deterministic in retrospective), and no further information about future observations $t > \hat{t}$ is given. In this sense, H specifies a certain history up to $\hat{t}(H)$ in a doxastic system and points to the last event of this history.

Example 8 (Trains continued)**.** *A doxastic system for the train example can be specified as*

$$\mathcal{D} = \langle \{A, B\}, \{Th_1, ..., Th_9\}, \{pfr, efr\}, A_0 \rangle,$$

$$A_0 = \begin{pmatrix} 0.7 & 0.02 & 0.09 & 0.02 & 0.09 & 0.01 & 0.02 & 0.02 & 0.03 \\ 0.7 & 0.02 & 0.09 & 0.02 & 0.09 & 0.01 & 0.02 & 0.02 & 0.03 \end{pmatrix}$$

To identify the situation described in Example 4 (T_1 is running late), we can specify the following pointed doxastic system: $\langle \mathcal{D}, [Obs_A(\neg at(T_1, C_C) : 3] \rangle$

5.1 Evolution of Probabilistic Interpretations

In accordance with the prior probability matrix A_0 from Definition 15, we define an interpretation matrix $A_t^{Th'}$ to store the interpretations of all agents $1, ..., n$ across all threads $Th_1, ..., Th_m$ given that the doxastic system is in pov thread Th' at time t:

$$A_t^{Th'} = \begin{pmatrix} I_{1,t}^{Th'}(Th_1) & \dots & I_{1,t}^{Th'}(Th_m) \\ \vdots & \ddots & \vdots \\ I_{n,t}^{Th'}(Th_1) & \dots & I_{n,t}^{Th'}(Th_m) \end{pmatrix} \quad (11)$$

With the definition of $\mathcal{K}_i$ from (1), the update rule from (2), and using the prior probability matrix A_0 from Definition 15, we can provide an update matrix $U_t^{Th'}$ to calculate the interpretation matrix for any pov thread Th' at any time point t ($\circ$ denotes the element-wise multiplication of matrices):

$$A_t^{Th'} = A_{t-1}^{Th'} \circ U_t^{Th'}, \text{ with} \quad (12)$$

$$(u_t^{Th'})_{ij} = \begin{cases} 0 & \text{if } Th_j(t) \notin \mathcal{K}_i(Th'(t)) \\ \frac{1}{\alpha_{it}^{Th'}} & \text{if } Th_j(t) \in \mathcal{K}_i(Th'(t)) \end{cases} \quad (13)$$

and $\alpha_{it}^{Th'}$ a normalization factor as defined in (3).

The timed observations specified in the history H of a pds $\langle \mathcal{D}, H \rangle$ induce an updated set of reachability relations $\mathcal{K}_i(Th(t))$ for every thread Th that complies with the given observations (for threads $Th_\perp$ that do not comply with the given observations $\mathcal{K}_i(Th_\perp(t)) = \emptyset$). These updated reachability relations in turn yield the updated interpretations in $A_t^{Th'}$. The complete state of interpretations at any time point for every possible pov thread $Th_1, ..., Th_m$ can then be specified as a block matrix, which we call the *belief state* (*bs*) of a pds at time t:

$$bs(\langle \mathcal{D}, H \rangle, t) = \left(A_t^{Th_1}, ..., A_t^{Th_m} \right) \quad (14)$$

The belief state can be viewed as a specification of conditional probabilities: the kth entry of $bs(\langle \mathcal{D}, H \rangle, t)$ specifies the interpretations of all agents across all threads at time t given that the system is in pov thread Th_k.

5.2 Evolution of Beliefs

In order to analyze the temporal evolution of beliefs, we use the update rule from (12) to update belief states. Since different possible observations yield different branches in the evolution of beliefs, we have to update every thread in the belief state individually, using the update matrices U_t^{Th} as defined in (13):

$$bs(\langle \mathcal{D}, H \rangle, t) = bs(\langle \mathcal{D}, H \rangle, t-1) \circ (U_t^{Th_1}, ..., U_t^{Th_m}) \quad (15)$$

Furthermore, to analyze satisfiability and validity of arbitrary finite belief expressions $B_{i,t'}^{\ell,u}(\cdot)$ w.r.t. a given pds $\langle \mathcal{D}, H \rangle$, we define an auxiliary belief vector $\vec{b}(\cdot)$ for different beliefs $B_{i,t'}^{\ell,u}(\cdot)$ as follows:

$$B_{i,t'}^{\ell,u}(F_t): \ (\vec{b}(F_t))_j = \begin{cases} 1 \text{ if } Th_j(t) \models F \\ 0 \text{ if } Th_j(t) \not\models F \end{cases} \quad (16)$$

$$B_{i,t'}^{\ell,u}(r_{\Delta t}^{\mathsf{fr}}(F,G)): \ (\vec{b}(r_{\Delta t}^{\mathsf{fr}}(F,G)))_j = \mathsf{fr}(Th_j, F, G, \Delta t)$$

$$B_{i,t'}^{\ell,u}(B_{k,t}^{\ell_k,u_k}(\cdot)): \ (\vec{b}(B_{k,t}^{\ell_k,u_k}(\cdot)))_j = \begin{cases} 1 \text{ if } I_{k,t}^{Th_j} \models B_{k,t}^{\ell_k,u_k}(\cdot) \\ 0 \text{ if } I_{k,t}^{Th_j} \not\models B_{k,t}^{\ell_k,u_k}(\cdot) \end{cases}$$

Using (11) and (16), we can determine a matrix $P_{t'}$ with the probabilities $p_{i,t'}^{Th^k}(\cdot)$ that each agent i assigns at time t' to some event $(\cdot)$, for all possible pov threads $Th^1, ..., Th_m$:

$$P_{t'}(\cdot) = \left(A_t^{Th_1} \cdot \vec{b}(\cdot), ..., A_t^{Th_m} \cdot \vec{b}(\cdot) \right) \quad (17)$$

The rows in $P_{t'}$ can be seen as conditional probabilities: agent i believes at time t' that a fact $(\cdot)$ is true with probability $p_{i,t'}^{Th_k}$ given that the pov thread is Th_k.

Using Definitions 12 - 14 and (17), we can provide a definition for the satisfiability and validity of beliefs:

Definition 17 (Validity and satisfiability of beliefs)**.** Let F_B be a belief formula as defined in Definition 5. F_B is satisfiable (valid) iff

- For $F_B \equiv B_{i,t'}^{\ell,u}(\cdot)$: $\ell \leq p_{i,t'}^{Th_k}(\cdot)$ and $u \geq p_{i,t'}^{Th_k}(\cdot)$ for at least one (all) $p_{i,t'}^{Th_k}$ in $P_{t'}$.
- For $F_B \equiv \neg B_{i,t'}^{\ell,u}(\cdot)$: $\ell > p_{i,t'}^{Th_k}(\cdot)$ or $u < p_{i,t'}^{Th_k}(\cdot)$ for at least one (all) $p_{i,t'}^{Th_k}$ in $P_{t'}$.
- For $F_B \equiv F_B' \wedge F_B''$: for at least one (all) $p_{i,t'}^{Th_k}$ in $P_{t'}$ both F_B' and F_B'' are satisfied.
- For $F_B \equiv F_B' \vee F_B''$: F_B' is satisfiable (valid) or F_B'' is satisfiable (valid).

To illustrate the evolution of beliefs, we finish the example with an analysis of expected arrival times.

Example 9 (Trains continued)**.** *From $\mathcal{D}$, as specified in Example 8, we can infer that Bob (and of course Alice, too) can safely assume at time 1 that Alice will arrive at time 8 at the latest (i.e., the actual thread is one of $Th_1, \ldots, Th_5$) with a probability in the range* $[0.9, 1]$ *because from Definition 17 we obtain that the following belief is valid w.r.t. $\mathcal{D}$ for $t = 1$:*

$$F_{Bob,t} \equiv B_{B,t}^{0.9,1}(r_8^{efr}(on(T_1,A),(at(T_2,C_B) \wedge on(T_2,A))).$$

Now, consider the previously described situation, where T_1 is running late and A does not inform B about it. This leads to the updated interpretations given in (4) and (5). These updates lead to a significant divergence in the belief of the expected arrival time: Alice's belief exhibits a drastically reduced certainty and changes to

$$B_{A,3}^{0.4,1}(r_8^{efr}(on(T_1,A),(at(T_2,C_B) \wedge on(T_2,A))),$$

while Bob's previous belief remains valid.

Even though Alice's beliefs have changed significantly, she is aware that Bob maintains beliefs conflicting with her own, as is shown by the following valid expression of nested beliefs: $B_{A,3}^{0.6,1}(F_{Bob,3})$

Finally, consider the pointed doxastic system $\langle D, [Obs_{AB}(\neg at(T_1,C_C) : 3] \rangle$, i.e., the same situation as before with the only difference that Alice now shares her observation of the delayed train with Bob. It immediately follows that Bob updates his beliefs in the same way as Alice, which in turn yields an update in Alice's beliefs about Bob's beliefs so that now the following expression is valid (because 0.6 *is not a valid lower bound any longer): $\neg B_{A,3}^{0.6,1}(F_{Bob,3})$. This example shows how Alice can reason about the influence of her own actions on Bob's belief state and therefore she can decide on actions that improve Bob's utility (as he does not have to wait in vain).*

6 Conclusion

In this paper, by extending APT Logic to dynamic scenarios with multiple agents, we have developed a general framework to represent and reason about the belief change in multi-agent systems. Next to lifting the single-agent case of APT Logic to multiple agents, we have also provided a suitable semantics to the temporal evolution of beliefs. The resulting framework extends previous work on dynamic multi-agent epistemic logics by enabling the quantification of agents' beliefs through probability intervals. An explicit notion of temporal relationships is provided through temporal rules building on the concept of frequency functions.

PDT Logic as introduced in this work provides the foundation for future work. While a basic decision procedure can be obtained through a direct application of the given semantics, we will continue to investigate optimized algorithms, using both exact and approximate methods. With a focus on inferring consistent possible threads automatically, this will give rise to a thorough complexity analysis of the decision problems. With efficient algorithms, we can apply PDT Logic to realistic problems.

REFERENCES

Aumann, R. J. (1976). Agreeing to Disagree. *The Annals of Statistics*, 4(6):1236–1239.

Fagin, R. and Halpern, J. Y. (1994). Reasoning about Knowledge and Probability. *Journal of the ACM*, 41:340–367.

Fagin, R., Halpern, J. Y., Moses, Y., and Vardi, M. Y. (1995). *Reasoning About Knowledge*. MIT Press.

Hintikka, J. (1962). *Knowledge and Belief.* Ithaca, N.Y.,Cornell University Press.

Kripke, S. A. (1963). Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, 16(1963):83–94.

Lloyd, J. W. (1987). *Foundations of Logic Programming, 2nd Edition*. Springer.

Renne, B., Sack, J., and Yap, A. (2009). Dynamic Epistemic Temporal Logic. In *Proc. of the Second International Workshop on Logic, Rationality, and Interaction*.

Scherl, R. B. and Levesque, H. J. (2003). Knowledge, Action, and the Frame Problem. *Artificial Intelligence*, 144(1-2):1–39.

Shakarian, P., Parker, A., Simari, G., and Subrahmanian, V. S. (2011). Annotated Probabilistic Temporal Logic. *ACM Trans. Comput. Logic*, 12(2):14:1–14:44.

van der Hoeck, W. (1997). Some considerations on the logics PFD: A logic combining modality and probability. *Journal of Applied Non-Classical Logics*, 7(3).

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Springer, 1st edition.