Agents, Beliefs, and Plausible Behavior in a Temporal Setting

Nils Bulling and Wojciech Jamroga
Department of Informatics, Clausthal University of Technology, Germany
{bulling,wjamroga}@in.tu-clausthal.de

ABSTRACT

Logics of knowledge and belief are often too static and inflexible to be used on real-world problems. In particular, they usually offer no concept for expressing that some course of events is *more likely to happen* than another. We address this problem and extend **CTLK** (computation tree logic with knowledge) with a notion of *plausibility*, which allows for practical and counterfactual reasoning. The new logic **CTLKP** (**CTLK** with plausibility) includes also a particular notion of belief. A plausibility update operator is added to this logic in order to change plausibility assumptions dynamically. Furthermore, we examine some important properties of these concepts. In particular, we show that, for a natural class of models, belief is a **KD45** modality. We also show that model checking **CTLKP** is **PTIME**-complete and can be done in time linear with respect to the size of models and formulae.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*; I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods—*Modal logic*

General Terms

Theory

Keywords

multi-agent systems, temporal logic, plausibility, beliefs

1. INTRODUCTION

Notions like *time*, *knowledge*, and *beliefs* are very important for analyzing the behavior of agents and multi-agent systems. In this paper, we extend modal logics of time and knowledge with a concept of *plausible behavior*: this notion is added to the language of **CTLK** [19], which is a straightforward combination of the branching-time temporal logic **CTL** [4, 3] and standard epistemic logic [9, 5].

In our approach, plausibility can be seen as a temporal property of behaviors. That is, some behaviors of the system can be assumed plausible and others implausible, with the underlying idea that the latter should perhaps be ignored in practical reasoning about possible future courses of action. Moreover, behaviors can be formally understood as *temporal paths* in the Kripke structure modeling a multi-agent system. As a consequence, we obtain a language to reason about what can (or must) plausibly happen. We propose a particular notion of beliefs (inspired by [20, 7]), defined in terms of epistemic relations and plausibility. The main intuition is that beliefs are facts *that an agent would know if he assumed that only plausible things could happen.*

We believe that humans use such a concept of plausibility and "practical beliefs" quite often in their everyday reasoning. Restricting one's reasoning to plausible possibilities is essential to make the reasoning feasible, as the space of *all* possibilities is exceedingly large in real life. We investigate some important properties of plausibility, knowledge, and belief in this new framework. In particular, we show that knowledge is an **S5** modality, and that beliefs satisfy axioms **K45** in general, and **KD45** for the class of *plausibly serial models*. Finally, we show that the relationship between knowledge and belief for plausibly serial models is natural and reflects the initial intuition well. We also show how plausibility assumptions can be specified in the object language via a *plausibility update operator*, and we study properties of such updates. Finally, we show that model checking of the new logic is no more complex than model checking **CTL** and **CTLK**.

Our ultimate goal is to come up with a logic that allows the study of strategies, time, knowledge, and plausible/rational behavior under both perfect and imperfect information. As combining all these dimensions is highly nontrivial (cf. [12, 14]) it seems reasonable to split this task. While this paper deals with knowledge, plausibility, and belief, the companion paper [11] proposes a general framework for multi-agent systems that regard game-theoretical rationality criteria like Nash equilibrium, Pareto optimality, etc. The latter approach is based on the more powerful logic ATL [1].

The paper is structured as follows. Firstly, we briefly present branching-time logic with knowledge, **CTLK**. In Section 3 we present our approach to plausibility and formally define **CTLK** with plausibility. We also show how

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

temporal formulae can be used to describe plausible paths, and we compare our logic with existing related work. In Section 4, properties of knowledge, belief, and plausibility are explored. Finally, we present verification complexity results for **CTLKP** in Section 5.

2. BRANCHING TIME AND KNOWLEDGE

In this paper we develop a framework for agents' beliefs about how the world can (or must) evolve. Thus, we need a notion of time and change, plus a notion of what the agents are supposed to know in particular situations. **CTLK** [19] is a straightforward combination of the computation tree logic **CTL** [4, 3] and standard epistemic logic [9, 5].

CTL includes operators for temporal properties of systems: i.e., path quantifier E ("there is a path"), together with temporal operators: $\bigcirc$ ("in the next state"), $\Box$ ("always from now on") and $\mathcal{U}$ ("until").[1] Every occurrence of a temporal operator is preceded by exactly one path quantifier in **CTL** (this variant of the language is sometimes called "vanilla" **CTL**). Epistemic logic uses operators for representing agents' knowledge: $K_a\varphi$ is read as "agent a knows that φ".

Let Π be a set of atomic propositions with a typical element p, and $\mathbb{A}\mathrm{gt} = \{1, ..., k\}$ be a set of agents with a typical element a. The language of **CTLK** consists of formulae φ, given as follows:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{E}\gamma \mid K_a\varphi$$
$$\gamma ::= \bigcirc\varphi \mid \Box\,\varphi \mid \varphi\mathcal{U}\,\varphi.$$

We will sometimes refer to formulae φ as ("vanilla") *state* formulae and to formulae γ as ("vanilla") *path* formulae.

The semantics of **CTLK** is based on Kripke models $\mathcal{M} = \langle Q, R, \sim_1, ..., \sim_k, \pi\rangle$, which include a nonempty set of states Q, a state transition relation $R \subseteq Q \times Q$, epistemic indistinguishability relations $\sim_a \subseteq Q \times Q$ (one per agent), and a valuation of propositions $\pi : \Pi \to \mathcal{P}(Q)$. We assume that relation R is serial and that all $\sim_a$ are equivalence relations. A *path* λ in $\mathcal{M}$ refers to a possible behavior (or computation) of system $\mathcal{M}$, and can be represented as an infinite sequence of states that follow relation R, that is, a sequence $q_0q_1q_2...$ such that q_iRq_{i+1} for every $i = 0, 1, 2, ...$ We denote the ith state in λ by $\lambda[i]$. The set of all paths in $\mathcal{M}$ is denoted by $\Lambda_{\mathcal{M}}$ (if the model is clear from context, $\mathcal{M}$ will be omitted). A *q-path* is a path that starts from q, i.e., $\lambda[0] = q$. A *q-subpath* is a sequence of states, starting from q, which is a subpath of some path in the model, i.e. a sequence $q_0q_1...$ such that $q = q_0$ and there are $q^0, ..., q^i$ such that $q^0...q^iq_0q_1... \in \Lambda_{\mathcal{M}}$.[2] The semantics of **CTLK** is defined as follows:

$\mathcal{M}, q \models p$ iff $q \in \pi(p)$;

$\mathcal{M}, q \models \neg\varphi$ iff $\mathcal{M}, q \not\models \varphi$;

$\mathcal{M}, q \models \varphi \wedge \psi$ iff $\mathcal{M}, q \models \varphi$ and $\mathcal{M}, q \models \psi$;

$\mathcal{M}, q \models \mathsf{E}\bigcirc\varphi$ iff there is a q-path λ such that $\mathcal{M}, \lambda[1] \models \varphi$;

$\mathcal{M}, q \models \mathsf{E}\Box\,\varphi$ iff there is a q-path λ such that $\mathcal{M}, \lambda[i] \models \varphi$ for every $i \geq 0$;

$\mathcal{M}, q \models \mathsf{E}\varphi\mathcal{U}\,\psi$ iff there is a q-path λ and $i \geq 0$ such that $\mathcal{M}, \lambda[i] \models \psi$, and $\mathcal{M}, \lambda[j] \models \varphi$ for every $0 \leq j < i$;

$\mathcal{M}, q \models K_a\varphi$ iff $\mathcal{M}, q \models \varphi$ for every q' such that $q \sim_a q'$.

[1] Additional operators A ("for every path") and $\Diamond$ ("sometime in the future") are defined in the usual way.

[2] For **CTLK** models, λ is a q-subpath iff it is a q-path. It will not always be so when plausible paths are introduced.

3. EXTENDING TIME AND KNOWLEDGE WITH PLAUSIBILITY AND BELIEFS

In this section we discuss the central concept of this paper, i.e. the concept of plausibility. First, we outline the idea informally. Then, we extend **CTLK** with the notion of plausibility by adding *plausible path operators* $\mathbf{Pl}_a$ and *physical path operator* $\mathbf{Ph}$ to the logic. Formula $\mathbf{Pl}_a\varphi$ has the intended meaning: *according to agent a, it is plausible that φ holds*; formula $\mathbf{Ph}\,\varphi$ reads as: *φ holds in all "physically" possible scenarios* (i.e., even in implausible ones). The plausible path operator restricts statements only to those paths which are defined to be "sensible", whereas the physical path operator generates statements about all paths that may theoretically occur. Furthermore, we define beliefs on top of plausibility and knowledge, as the facts *that an agent would know if he assumed that only plausible things could happen.* Finally, we discuss related work [7, 8, 20, 18, 16], and compare it with our approach.

3.1 The Concept of Plausibility

It is well known how knowledge (or beliefs) can be modeled with Kripke structures. However, it is not so obvious how we can capture knowledge *and* beliefs in a sensible way in *one* framework. Clearly, there should be a connection between these two notions. Our approach is to use the notion of plausibility for this purpose. Plausibility can serve as a primitive concept that helps to define the semantics of beliefs, in a similar way as indistinguishability of states (represented by relation $\sim_a$) is the semantic concept that underlies knowledge. In this sense, our work follows [7, 20]: essentially, beliefs are what an agent would know if he took only plausible options into account. In our approach, however, plausibility is explicitly seen as a *temporal property.* That is, we do not consider states (or possible worlds) to be more plausible than others but rather define some behaviors to be plausible, and others implausible. Moreover, behaviors can be formally understood as temporal paths in the Kripke structure modeling a multi-agent system.

An actual notion of plausibility (that is, a particular set of plausible paths) can emerge in many different ways. It may result from observations and learning; an agent can learn from its observations and see specific patterns of events as plausible ("a lot of people wear black shoes if they wear a suit"). Knowledge exchange is another possibility (e.g., an agent a can tell agent b that "player c always bluffs when he is smiling"). Game theory, with its rationality criteria (undominated strategies, maxmin, Nash equilibrium etc.) is another viable source of plausibility assumptions. Last but not least, folk knowledge can be used to establish plausibility-related classifications of behavior ("players normally want to win a game", "people want to live").

In any case, restricting the reasoning to plausible possibilities can be essential if we want to make the reasoning feasible, as the space of *all* possibilities (we call them "physical" possibilities in the rest of the paper) is exceedingly large in real life. Of course, this does not exclude a more extensive analysis in special cases, e.g. when our plausibility assumptions do not seem accurate any more, or when the cost of inaccurate assumptions can be too high (as in the case of high-budget business decisions). But even in these cases, we usually do not get rid of plausibility assumptions completely – we only revise them to make them more cautious.[3]

To formalize this idea, we extend models of **CTLK** with *sets of plausible paths* and add plausibility operators $\mathbf{Pl}_a$, physical paths operator $\mathbf{Ph}$, and belief operators B_a to the language of **CTLK**. Now, it is possible to make statements that refer to plausible paths only, as well as statements that regard all paths that may occur in the system.

3.2 CTLK with Plausibility

In this section, we extend the logic of **CTLK** with plausibility; we call the resulting logic **CTLKP**. Formally, the language of **CTLKP** is defined as:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathsf{E}\gamma \mid \mathbf{Pl}_a\varphi \mid \mathbf{Ph}\,\varphi \mid K_a\varphi \mid B_a\varphi$$
$$\gamma ::= \bigcirc\varphi \mid \Box\,\varphi \mid \varphi\mathcal{U}\,\varphi.$$

For instance, we may claim it is plausible to assume that a shop is closed after the opening hours, though the manager may be physically able to open it at any time: $\mathbf{Pl}_a\mathsf{A}\Box\,(\mathsf{late} \rightarrow \neg\mathsf{open}) \wedge \mathbf{Ph}\,\mathsf{E}\Diamond\,(\mathsf{late} \wedge \mathsf{open})$.

The semantics of **CTLKP** extends that of **CTLK** as follows. Firstly, we augment the models with *sets of plausible paths*. A *model with plausibility* is given as

$$\mathcal{M} = \langle Q, R, \sim_1, ..., \sim_k, \Upsilon_1, ..., \Upsilon_k, \pi\rangle,$$

where $\langle Q, R, \sim_1, ..., \sim_k, \pi\rangle$ is a **CTLK** model, and $\Upsilon_a \subseteq \Lambda_{\mathcal{M}}$ is the set of paths in $\mathcal{M}$ that are plausible according to agent a. If we want to make it clear that Υ_a is taken from model $\mathcal{M}$, we will write $\Upsilon_a^{\mathcal{M}}$. It seems worth emphasizing that this notion of plausibility is *subjective* and *holistic*. It is subjective because Υ_a represents *agent a's subjective view on what is plausible* – and indeed, different agents may have different ideas on plausibility (i.e., Υ_a may differ from Υ_b). It is holistic because Υ_a represents agent a's idea of the plausible behavior *of the whole system* (including the behavior of other agents).

Remark 1. *In our models, plausibility is also* global*, i.e., plausibility sets do not depend on the state of the system. Investigating systems, in which plausibility is relativized with respect to states (like in [7]), might be an interesting avenue of future work. However, such an approach – while obviously more flexible – allows for potentially counterintuitive system descriptions. For example, it might be the case that path λ is plausible in $q = \lambda[0]$, but the set of plausible paths in $q' = \lambda[1]$ is empty. That is, by following plausible path λ we are bound to get to an implausible situation. But then, does it make sense to consider λ as plausible?*

Secondly, we use a non-standard satisfaction relation $\models_P$, which we call *plausible satisfaction*. Let $\mathcal{M}$ be a **CTLKP** model and $P \subseteq \Lambda_{\mathcal{M}}$ be an arbitrary subset of paths in $\mathcal{M}$ (not necessarily any $\Upsilon_a^{\mathcal{M}}$). $\models_P$ restricts the evaluation of temporal formulae to the paths given in P only. The "absolute" satisfaction relation $\models$ is defined as $\models_{\Lambda_{\mathcal{M}}}$.

Let $\mathfrak{on}(P)$ be the set of all states that lie on at least one path in P, i.e. $\mathfrak{on}(P) = \{q \in Q \mid \exists\lambda \in P \exists i\ (\lambda[i] = q)\}$. Now, the semantics of **CTLKP** can be given through the following clauses:

$\mathcal{M}, q \models_P p$ iff $q \in \pi(p)$;

$\mathcal{M}, q \models_P \neg\varphi$ iff $\mathcal{M}, q \not\models_P \varphi$;

$\mathcal{M}, q \models_P \varphi \wedge \psi$ iff $\mathcal{M}, q \models_P \varphi$ and $\mathcal{M}, q \models_P \psi$;

$\mathcal{M}, q \models_P \mathsf{E}\bigcirc\varphi$ iff there is a q-subpath $\lambda \in P$ such that $\mathcal{M}, \lambda[1] \models_P \varphi$;

$\mathcal{M}, q \models_P \mathsf{E}\Box\,\varphi$ iff there is a q-subpath $\lambda \in P$ such that $\mathcal{M}, \lambda[i] \models_P \varphi$ for every $i \geq 0$;

$\mathcal{M}, q \models_P \mathsf{E}\varphi\mathcal{U}\,\psi$ iff there is a q-subpath $\lambda \in P$ and $i \geq 0$ such that $\mathcal{M}, \lambda[i] \models_P \psi$, and $\mathcal{M}, \lambda[j] \models_P \varphi$ for every $0 \leq j < i$;

$\mathcal{M}, q \models_P \mathbf{Pl}_a\varphi$ iff $\mathcal{M}, q \models_{\Upsilon_a} \varphi$;

$\mathcal{M}, q \models_P \mathbf{Ph}\,\varphi$ iff $\mathcal{M}, q \models \varphi$;

$\mathcal{M}, q \models_P K_a\varphi$ iff $\mathcal{M}, q \models \varphi$ for every q' such that $q \sim_a q'$;

$\mathcal{M}, q \models_P B_a\varphi$ iff for all $q' \in \mathfrak{on}(\Upsilon_a)$ with $q \sim_a q'$, we have that $\mathcal{M}, q' \models_{\Upsilon_a} \varphi$.

One of the main reasons for using the concept of plausibility is that we want to define agents' *beliefs* out of more primitive concepts – in our case, these are plausibility and indistinguishability – in a way analogous to [20, 7]. If an agent *knows* that φ, he must be "sure" about it. However, *beliefs* of an agent are not necessarily about reliable facts. Still, they should make sense to the agent; if he believes that φ, then the formula should at least hold in all futures that he envisages as plausible. Thus, beliefs of an agent may be seen as *things known to him if he disregards all non-plausible possibilities.*

We say that φ is $\mathcal{M}$*-true* ($\mathcal{M} \models \varphi$) if $\mathcal{M}, q \models \varphi$ for all $q \in Q_{\mathcal{M}}$. φ is *valid* ($\models \varphi$) if $\mathcal{M} \models \varphi$ for all models $\mathcal{M}$. φ is $\mathcal{M}$*-strongly true* ($\mathcal{M} \equiv\!\!\models \varphi$) if $\mathcal{M}, q \models_P \varphi$ for all $q \in Q_{\mathcal{M}}$ and all $P \subseteq \Lambda_{\mathcal{M}}$. φ is *strongly valid* ($\equiv\!\!\models \varphi$) if $\mathcal{M} \equiv\!\!\models \varphi$ for all models $\mathcal{M}$.

Proposition 2. *Strong truth and strong validity imply truth and validity, respectively. The reverse does not hold.*

Ultimately, we are going to be interested in normal (not strong) validity, as parameterizing the satisfaction relation with a set P is just a technical device for propagating sets of plausible paths Υ_a into the semantics of nested formulae. The importance of strong validity, however, lies in the fact that $\equiv\!\!\models \varphi \leftrightarrow \psi$ makes φ and ψ completely interchangeable, while the same is not true for normal validity.

Proposition 3. *Let $\Phi[\varphi/\psi]$ denote formula Φ in which every occurrence of ψ was replaced by φ. Also, let $\equiv\!\!\models \varphi \leftrightarrow \psi$. Then for all M, q, P: $M, q \models_P \Phi$ iff $M, q \models_P \Phi[\varphi/\psi]$ (in particular, $M, q \models \Phi$ iff $M, q \models \Phi[\varphi/\psi]$).*

Note that $\models \varphi \leftrightarrow \psi$ does not even imply that $M, q \models \Phi$ iff $M, q \models \Phi[\varphi/\psi]$.

[3]That is, when planning to open an industrial plant in the UK, we will probably consider the possibility of our main contractor taking her life, but we will still *not* take into account the possibilities of: an invasion of UFO, England being destroyed by a meteorite, Fidel Castro becoming the British Prime Minister etc. Note that this is fundamentally different from using a probabilistic model in which all these unlikely scenarios are assigned very low probabilities: in that case, they also have a very small influence on our final decision, but we must process the *whole* space of physical possibilities to evaluate the options.

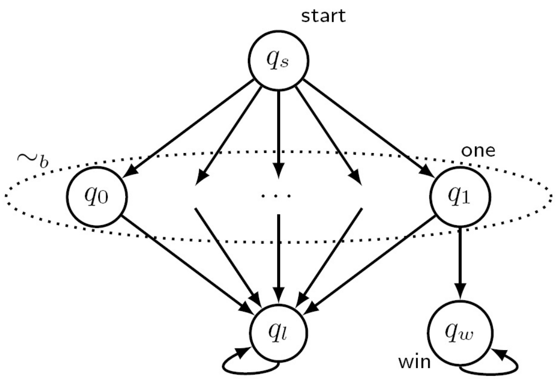

Figure 1: Guessing Robots game

EXAMPLE 1 (GUESSING ROBOTS). *Consider a simple game with two agents a and b, shown in Figure 1. First, a chooses a real number $r \in [0,1]$ (without revealing the number to b); then, b chooses a real number $r' \in [0,1]$. The agents win the game (and collect EUR 1,000,000) if both chose 1, otherwise they lose. Formally, we model the game with a **CTLKP** model $\mathcal{M}$, in which the set of states Q includes q_s for the initial situation, states q_r, $r \in [0,1]$, for the situations after a has chosen number r, and "final" states q_w, q_l for the winning and the losing situation, respectively. The transition relation is as follows: $q_s R q_r$ and $q_r R q_l$ for all $r \in [0,1]$; $q_1 R q_w$, $q_w R q_w$, and $q_l R q_l$. Moreover, $\pi(\mathsf{one}) = \{q_1\}$ and $\pi(\mathsf{win}) = \{q_w\}$. Player a has perfect information in the game (i.e., $q \sim_a q'$ iff $q = q'$), but player b does not distinguish between states q_r (i.e., $q_r \sim_b q_{r'}$ for all $r, r' \in [0,1]$). Obviously, the only sensible thing to do for both agents is to choose 1 (using game-theoretical vocabulary, these strategies are* strongly dominant *for the respective players). Thus, there is only one* plausible *course of events if we assume that our players are rational, and hence $\Upsilon_a = \Upsilon_b = \{q_s q_1 q_w q_w \ldots\}$.*

Note that, in principle, the outcome of the game is uncertain: $\mathcal{M}, q_s \models \neg\mathsf{A}\Diamond\,\mathsf{win} \wedge \neg\mathsf{A}\Box\,\neg\mathsf{win}$. However, assuming rationality of the players makes it only plausible that the game must end up with a win: $\mathcal{M}, q_s \models \mathbf{Pl}_a\,\mathsf{A}\Diamond\,\mathsf{win} \wedge \mathbf{Pl}_b\,\mathsf{A}\Diamond\,\mathsf{win}$, and the agents believe that this will be the case: $\mathcal{M}, q_s \models B_a\mathsf{A}\Diamond\,\mathsf{win} \wedge B_b\mathsf{A}\Diamond\,\mathsf{win}$. Note also that, in any of the states q_r, agent b believes that a (being rational) has played 1: $\mathcal{M}, q_r \models B_b\mathsf{one}$ for all $r \in [0,1]$.

3.3 Defining Plausible Paths with Formulae

So far, we have assumed that sets of plausible paths are somehow given in models. In this section we present a dynamic approach where an actual notion of plausibility can be specified in the object language. Note that we want to specify (usually infinite) sets of infinite paths, and we need a finite representation of these structures. One logical solution is given by using path formulae γ. These formulae describe properties of paths; therefore, a specific formula can be used to characterize a set of paths. For instance, think about a country in Africa where it has never snowed. Then, plausible paths might be defined as ones in which it never snows, i.e., all paths that satisfy $\Box\,\neg\mathsf{snows}$. Formally, let γ be a **CTLK** path formula. We define $|\gamma|_{\mathcal{M}}$ to be the set of paths that satisfy γ in model $\mathcal{M}$:

$$\begin{aligned} |\bigcirc\varphi|_{\mathcal{M}} &= \{\lambda \mid \mathcal{M}, \lambda[1] \models \varphi\} \\ |\Box\,\varphi|_{\mathcal{M}} &= \{\lambda \mid \forall i\ (\mathcal{M}, \lambda[i] \models \varphi)\} \\ |\varphi_1 \mathcal{U}\,\varphi_2|_{\mathcal{M}} &= \{\lambda \mid \exists i\big(\mathcal{M}, \lambda[i] \models \varphi_2 \wedge \\ &\qquad \forall j(0 \le j < i \Rightarrow \mathcal{M}, \lambda[j] \models \varphi_1)\big)\}. \end{aligned}$$

Moreover, we define the *plausible paths model update* as follows. Let $\mathcal{M} = \langle Q, R, \sim_1, ..., \sim_k, \Upsilon_1, ..., \Upsilon_k, \pi\rangle$ be a **CTLKP** model, and let $P \subseteq \Lambda_{\mathcal{M}}$ be a set of paths. Then $\mathcal{M}^{a,P} = \langle Q, R, \sim_1, ..., \sim_k, \Upsilon_1, ..., \Upsilon_{a-1}, P, \Upsilon_{a+1}, ..., \Upsilon_k, \pi\rangle$ denotes model $\mathcal{M}$ with a's set of plausible paths reset to P.

Now we can extend the language of **CTLKP** with formulae $(\mathbf{set\text{-}pl}_a\ \gamma)\varphi$ with the intuitive reading: "suppose that γ exactly characterizes the set of plausible paths, then φ holds", and formal semantics given below:

$$\mathcal{M}, q \models_P (\mathbf{set\text{-}pl}_a\ \gamma)\varphi \text{ iff } \mathcal{M}^{a,|\gamma|_{\mathcal{M}}}, q \models_P \varphi.$$

We observe that this update scheme is similar to the one proposed in [13].

3.4 Comparison to Related Work

Several modal notions of plausibility were already discussed in the existing literature [7, 8, 20, 18, 16]. In these papers, like in ours, plausibility is used as a primitive semantic concept that helps to define beliefs on top of agents' knowledge. A similar idea was introduced by Moses and Shoham in [18]. Their work preceded both [7, 8] and [20] – and although Moses and Shoham do not explicitly mention the term "plausibility", it seems appropriate to summarize their idea first.

Moses and Shoham: Beliefs as Conditional Knowledge

In [18], beliefs are relativized with respect to a formula α (which can be seen as a plausibility assumption expressed in the object language). More precisely, worlds that satisfy α can be considered as plausible. This concept is expressed via symbols $B_i^\alpha\varphi$; the index $i \in \{1,2,3\}$ is used to distinguish between three different implementations of beliefs. The first version is given by $B_1^\alpha\varphi \equiv K(\alpha \to \varphi)$.[4] A drawback of this version is that if α is false, then everything will be believed with respect to α. The second version overcomes this problem: $B_2^\alpha\varphi \equiv K(\alpha \to \varphi) \wedge (K\neg\alpha \to K\varphi)$; now φ is only believed if it is known that φ follows from assumption α, and φ must be known if assumption α is known to be false. Finally, $B_3^\alpha\varphi \equiv K(\alpha \to \varphi) \wedge \neg K\neg\alpha$: if the assumption α is known to be false, nothing should be believed with respect to α. The strength of these different notions is given as follows: $B_3^\alpha\varphi$ implies $B_2^\alpha\varphi$, and $B_2^\alpha\varphi$ implies $B_1^\alpha\varphi$. In this approach, belief is strongly connected to knowledge in the sense that belief is knowledge with respect to a given assumption.

Friedman and Halpern: Plausibility Spaces

The work of Friedman and Halpern [7] extends the concepts of knowledge and belief with an *explicit* notion of plausibility; i.e., some worlds are more plausible for an agent than others. To implement this idea, Kripke models are extended with function P which assigns a *plausibility space* $P(q,a) = (\Omega_{(q,a)}, \preceq_{(q,a)})$ to every state, or more generally every possible world q, and agent a. The plausibility space

[4]Unlike in most approaches, K is interpreted over *all* worlds and not only over the indistinguishable worlds.

is just a partially ordered subset of states/worlds; that is, $\Omega(q, a) \subseteq Q$, and $\preceq_{(q,a)} \subseteq Q \times Q$ is a reflexive and transitive relation. Let $S, T \subseteq \Omega_{(q,a)}$ be finite subsets of states; now, T is defined to be *plausible given S with respect to $P(q, a)$*, denoted by $S \rightarrow_{P(q,a)} T$, iff all minimal points/states in S (with respect to $\preceq_{(q,a)}$) are also in T.[5] Friedman and Halpern's view to modal plausibility is closely related to probability and, more generally, plausibility measures. Logics of plausibility can be seen as a *qualitative* description of agents preferences/knowledge; logics of probability [6, 15], on the other hand, offer a *quantitative* description.

The logic from [7] is defined by the following grammar: $\varphi ::= p \mid \varphi \wedge \varphi \mid \neg\varphi \mid K_a\varphi \mid \varphi \rightarrow_a \varphi$, where the semantics of all operators except $\rightarrow_a$ is given as usual, and formulae $\varphi \rightarrow_a \psi$ have the meaning that ψ is true in the most plausible worlds in which φ holds. Formally, the semantics for $\rightarrow_a$ is given as: $\mathcal{M}, q \models \varphi \rightarrow_a \psi$ iff $S^{\varphi}_{P(q,a)} \rightarrow_{P(q,a)} S^{\psi}_{P(q,a)}$, where $S^{\varphi}_{(q,a)} = \{q' \in \Omega_{(q,a)} \mid \mathcal{M}, q' \models \varphi\}$ are the states in $\Omega_{(q,a)}$ that satisfy φ. The idea of defining beliefs is given by the assumption that an agent *believes* in something if he *knows that it is true in the most plausible worlds of* $\Omega_{(q,a)}$; formally, this can be stated as $B_a\varphi \equiv K_a(\top \rightarrow_a \varphi)$.

Friedman and Halpern have shown that the **KD45** axioms are valid for operator B_a if plausibility spaces satisfy *consistency* (for all states $q \in Q$ it holds that $\Omega_{(q,a)} \subseteq \{ q' \in Q \mid q \sim_a q' \}$) and *normality* (for all states $q \in Q$ it holds that $\Omega_{(q,a)} \neq \emptyset$).[6] A temporal extension of the language (mentioned briefly in [7], and discussed in more detail in [8]) uses the interpreted systems approach [10, 5]. A *system* $\mathcal{R}$ is given by *runs*, where a run $r : \mathbb{N} \rightarrow Q$ is a function from time moments (modeled by $\mathbb{N}$) to global states, and a *time point* (r, i) is given by a time point $i \in \mathbb{N}$ and a run r. A global state is a combination of local states, one per agent. An *interpreted system* $\mathcal{M} = (\mathcal{R}, \pi)$ is given by a system $\mathcal{R}$ and a valuation of propositions π. Epistemic relations are defined over time points, i.e., $(r', m') \sim_a (r, m)$ iff agent a's local states $r'_a(m')$ and $r_a(m)$ of (r', m') and (r, m) are equal. Formulae are interpreted in a straightforward way with respect to interpreted systems, e.g. $\mathcal{M}, r, m \models K_a\varphi$ iff $\mathcal{M}, r', m' \models \varphi$ for all $(r', m') \sim_a (r, m)$. Now, these are time points that play the role of possible worlds; consequently, plausibility spaces $\mathcal{P}_{(r,m,a)}$ are assigned to each point (r, m) and agent a.

Su et al.: KBC Logic

Su et al. [20] have developed a multi-modal, computationally grounded logic with modalities K, B, and C (knowledge, belief, and certainty). The computational model consists of (global) states $q = (q^{vis}, q^{inv}, q^{per}, Q^{pls})$ where the environment is divided into a visible (q^{vis}) and an invisible part (q^{inv}), and q^{per} captures the agent's perception of the visible part of the environment. External sources may provide the agent with information about the invisible part of a state, which results in a set of states Q^{pls} that are plausible for the agent.

Given a global state q, we additionally define $Vis(q) = q^{vis}$, $Inv(q) = q^{inv}$, $Per(q) = q^{per}$, and $Pls(q) = Q^{pls}$. The semantics is given by an extension of interpreted systems [10, 5], here, it is called *interpreted KBC systems*. *KBC* formulae are defined as $\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi \mid B\varphi \mid C\varphi$. The epistemic relation $\sim_{vis}$ is captured in the following way: $(r, i) \sim_{vis} (r', i')$ iff $Vis(r(i)) = Vis(r'(i'))$. The semantic clauses for belief and certainty are given below.

$\mathcal{M}, r, i \models B\varphi$ iff $\mathcal{M}, r', i' \models \varphi$ for all (r', i') with $Vis(r'(i')) = Per(r(i))$ and $Inv(r'(i')) \in Pls(r(i))$

$\mathcal{M}, r, i \models C\varphi$ iff $\mathcal{M}, r', i' \models \varphi$ for all $(r'(i'))$ with $Vis(r'(i')) = Per(r(i))$

Thus, an agent believes φ if, and only if, φ is true in all states *which look like what he sees now and seem plausible in the current state*. Certainty is stronger: if an agent is certain about φ, the formula must hold in all states with a visible part equal to the current perception, regardless of whether the invisible part is plausible or not.

The logic does not include temporal formulae, although it might be extended with temporal operators, as time is already present in KBC models.

What Are the Differences to Our Logic?

In our approach, plausibility is explicitly seen as a temporal property, i.e., it is a property of temporal paths rather than states. In the object language, this is reflected by the fact that plausibility assumptions are specified through path formulae. In contrast, the approach of [18] and [20] is static: not only the logics do not include operators for talking about time and/or change, but these are *states* that are assumed plausible or not in their semantics.

The differences to [7, 8] are more subtle. Firstly, the framework of Friedman and Halpern is static in the sense that plausibility is taken as a property of (abstract) possible worlds. This formulation is flexible enough to allow for incorporating time; still, in our approach, time is *inherent* to plausibility rather than incidental.

Secondly, our framework is more computationally oriented. The implementation of temporal plausibility in [7, 8] is based on the interpreted systems approach with *time points* (r, m) being subject to plausibility. As runs are included in time points, they can also be defined plausible or implausible.[7] However, it also means that time points serve the role of possible worlds in the basic formulation, which yields Kripke structures with *uncountable* possible world spaces in all but the most trivial cases.

Thirdly, [7, 8] build on *linear* time: a run (more precisely, a time moment (r, m)) is fixed when a formula is interpreted. In contrast, we use branching time with explicit quantification over temporal paths.[8] We believe that branching time is more suitable for non-deterministic domains (cf. e.g. [4]), of which multi-agent systems are a prime example. Note that branching time makes our notion of belief different from Friedman and Halpern's. Most notably, property $K\varphi \rightarrow B\varphi$ is valid in their approach, but not in ours: an agent may

[5] When there are infinite chains $\ldots \preceq q_3 \preceq q_2 \preceq_a q_1$, the definition is much more sophisticated. An interested reader is referred to [7] for more details.

[6] Note that this "normality" is essentially seriality of states wrt plausibility spaces.

[7] Friedman and Halpern even briefly mention how plausibility of runs can be embedded in their framework.

[8] To be more precise, time in [7] does implicitly branch at epistemic states. This is because $(r, m) \sim_a (r', m')$ iff a's local state corresponding to both time points is the same ($r_a(m) = r'_a(m')$). In consequence, the semantics of $K_a\varphi$ can be read as "*for every run*, and every moment on this run that yields the same local state as now, φ holds".

know that some course of events is in principle possible, without believing that it can *really* become the case (see Section 4.2). As Proposition 13 suggests, such a subtle distinction between knowledge and beliefs is possible in our approach because branching time logics allow for existential quantification over runs.

Fourthly, while Friedman and Halpern's models are very flexible, they also enable system descriptions that may seem counterintuitive. Suppose that (r, m) is plausible in itself (formally: (r, m) is minimal wrt $\preceq_{(r,m,a)}$), but $(r, m+1)$ is not plausible in $(r, m+1)$. This means that following the plausible path makes it implausible (cf. Remark 1), which is even stranger in the case of linear time. Combining the argument with computational aspects, we suggest that our approach can be more natural and straightforward for many applications.

Last but not least, our logic provides a mechanism for specifying (and updating) sets of plausible paths in the object language. Thus, plausibility sets can be specified in a succinct way, which is another feature that makes our framework computation-friendly. The model checking results from Section 5 are especially encouraging in this light.

4. PLAUSIBILITY, KNOWLEDGE, AND BELIEFS IN CTLKP

In this section we study some relevant properties of plausibility, knowledge, and beliefs; in particular, axioms **KDT45** are examined. But first, we identify two important subclasses of models with plausibility.

A **CTLKP** model is *plausibly serial* (or *p-serial*) for agent a if every state of the system is part of a plausible path according to a, i.e. $\mathfrak{on}(\Upsilon_a) = Q$. As we will see further, a weaker requirement is sometimes sufficient. We call a model *weakly p-serial* if every state has at least one indistinguishable counterpart which lies on a plausible path, i.e. for each $q \in Q$ there is a $q' \in Q$ such that $q \sim_a q'$ and $q' \in \mathfrak{on}(\Upsilon_a)$. Obviously, p-seriality implies weak p-seriality. We get the following characterization of both model classes.

PROPOSITION 4. *M is plausibly serial for agent a iff formula $\mathbf{Pl}_a \mathsf{E}\bigcirc\top$ is valid in M. M is weakly p-serial for agent a iff $\neg K_a \mathbf{Pl}_a \mathsf{A}\bigcirc\bot$ is valid in M.*

4.1 Axiomatic Properties

THEOREM 5. *Axioms **K**, **D**, **4**, and **5** for knowledge are strongly valid, and axiom **T** is valid. That is, modalities K_a form system **S5** in the sense of normal validity, and **KD45** in the sense of strong validity.*

We do not include proofs here due to lack of space. The interested reader is referred to [2], where detailed proofs are given.

PROPOSITION 6. *Axioms **K**, **4**, and **5** for beliefs are strongly valid. That is, we have:*
$\models\!\!\models (B_a\varphi \wedge B_a(\varphi \to \psi)) \to B_a\psi$, $\models\!\!\models (B_a\varphi \to B_aB_a\varphi)$, *and* $\models\!\!\models (\neg B_a\varphi \to B_a\neg B_a\varphi)$.

The next proposition concerns the "consistency" axiom **D**: $B_a\varphi \to \neg B_a\neg\varphi$. It is easy to see that the axiom is not valid in general: as we have no restrictions on plausibility sets Υ_a, it may be as well that $\Upsilon_a = \emptyset$. In that case we have $B_a\varphi \wedge B_a\neg\varphi$ for all formulae φ, because the set of states to be considered becomes empty. However, it turns out that **D** *is* valid for a very natural class of models.

PROPOSITION 7. *Axiom **D** for beliefs is not valid in the class of all **CTLKP** models. However, it is strongly valid in the class of weak p-serial models (and therefore also in the class of p-serial models).*

Moreover, as one may expect, beliefs do not have to be always true.

PROPOSITION 8. *Axiom **T** for beliefs is not valid; i.e., $\not\models (B_a\varphi \to \varphi)$. The axiom is not even valid in the class of p-serial models.*

THEOREM 9. *Belief modalities B_a form system **K45** in the class of all models, and **KD45** in the class of weakly plausibly serial models (in the sense of both normal and strong validity). Axiom **T** is not even valid for p-serial models.*

4.2 Plausibility, Knowledge, and Beliefs

First, we investigate the relationship between knowledge and plausibility/physicality operators. Then, we look at the interaction between knowledge and beliefs.

PROPOSITION 10. *Let φ be a **CTLKP** formula, and $\mathcal{M}$ be a **CTLKP** model. We have the following strong validities:*

(i) $\models\!\!\models \mathbf{Pl}_a K_a\varphi \leftrightarrow K_a\varphi$

(ii) $\models\!\!\models \mathbf{Ph}\, K_a\varphi \leftrightarrow K_a\mathbf{Ph}\,\varphi$ *and* $\models\!\!\models K_a\mathbf{Ph}\,\varphi \leftrightarrow K_a\varphi$

We now want to examine the relationship between knowledge and belief. For instance, if agent a believes in something, he knows that he believes it. Or, if he knows a fact, he also believes that he knows it. On the other hand, for instance, an agent does not necessarily believe in all the things he knows. For example, we may know that an invasion from another galaxy is in principle possible ($K_a\mathsf{E}\Diamond$ $\mathsf{invasion}$), but if we do not take this possibility as plausible ($\neg\mathbf{Pl}_a\mathsf{E}\Diamond$ $\mathsf{invasion}$), then we reject the corresponding belief in consequence ($\neg B_a\mathsf{E}\Diamond$ $\mathsf{invasion}$). Note that this property reflects the strong connection between belief and plausibility in our framework.

PROPOSITION 11. *The following formulae are strongly valid:*
(i) $B_a\varphi \to K_aB_a\varphi$, *(ii)* $K_aB_a\varphi \to B_a\varphi$,
(iii) $K_a\varphi \to B_aK_a\varphi$.
The following formulae are not *valid:*
(iv) $B_a\varphi \to B_aK_a\varphi$, *(v)* $K_a\varphi \to B_a\varphi$

The last invalidity is especially important: it is *not* the case that knowing something implies believing in it. This emphasizes that we study a specific concept of beliefs here. Note that its specific is not due to the plausibility-based definition of beliefs. The reason lies rather in the fact that we investigate knowledge, beliefs and plausibility in a temporal framework, as Proposition 12 shows.

PROPOSITION 12. *Let φ be a **CTLKP** formula that does not include any temporal operators. Then $K_a\varphi \to B_a\varphi$ is strongly valid, and in the class of p-serial models we have even that $\models\!\!\models K_a\varphi \leftrightarrow B_a\varphi$.*

Moreover, it is important that we use branching time with explicit quantification over paths; this observation is formalized in Proposition 13.

DEFINITION 1. *We define the* universal sublanguage *of* ***CTLK*** *in a way similar to [21]:*

$$\varphi_u ::= p \mid \neg p \mid \varphi_u \wedge \varphi_u \mid \varphi_u \vee \varphi_u \mid \mathsf{A}\gamma_u \mid K_a\varphi_u,$$
$$\gamma_u ::= \bigcirc\varphi_u \mid \Box\, \varphi_u \mid \varphi_u \mathcal{U}\, \varphi_u.$$

We call such φ_u universal formulae, and γ_u universal path formulae.

PROPOSITION 13. *Let φ_u be a universal* ***CTLK*** *formula. Then $\models\!\equiv K_a\varphi_u \rightarrow B_a\varphi_u$.*

The following two theorems characterize the relationship between knowledge and beliefs: first for the class of p-serial models, and then, finally, for all models.

THEOREM 14. *The following formulae are strongly valid in the class of plausibly serial* ***CTLKP*** *models:*
(i) $B_a\varphi \leftrightarrow K_a\mathbf{Pl}_a\varphi$, (ii) $K_a\varphi \leftrightarrow B_a\mathbf{Ph}\,\varphi$.

THEOREM 15. *Formula $B_a\varphi \leftrightarrow K_a\mathbf{Pl}_a(\mathsf{E}\bigcirc\top \rightarrow \varphi)$ is strongly valid.*

Note that this characterization has a strong commonsense reading: *believing in φ is knowing that φ plausibly holds in all plausibly imaginable situations.*

4.3 Properties of the Update

The first notable property of plausibility update is that it influences only formulae in which plausibility plays a role, i.e. ones in which belief or plausibility modalities occur.

PROPOSITION 16. *Let φ be a* ***CTLKP*** *formula that does not include operators $\mathbf{Pl}_a$ and B_a, and γ be a* ***CTLKP*** *path formula. Then, we have $\models\!\equiv \varphi \leftrightarrow (\textbf{\textit{set-pl}}_a\ \gamma)\varphi$.*

What can be said about the result of an update? At first sight, formula $(\mathbf{set\text{-}pl}_a\ \gamma)\mathbf{Pl}_a\mathsf{A}\gamma$ seems a natural characterization; however, it is not valid. This is because, by leaving the other (implausible) paths out of scope, we may leave out of $|\gamma|$ some paths that were needed to satisfy γ (see the example in Section 4.2). We propose two alternative ways out: the first one restricts the language of the update similarly to [21]; the other refers to physical possibilities, in a way analogous to [13].

PROPOSITION 17. *The* ***CTLKP*** *formula $(\textbf{\textit{set-pl}}_a\ \gamma)\mathbf{Pl}_a\mathsf{A}\gamma$ is not valid. However, we have the following validities:*

(i) $\models\!\equiv (\textbf{\textit{set-pl}}_a\ \gamma_u)\mathbf{Pl}_a\mathsf{A}\gamma_u$, where γ_u is a universal ***CTLK*** *path formula from Definition 1.*

(ii) If $\varphi, \varphi_1, \varphi_2$ are arbitrary ***CTLK*** *formulae, then:*

$$\models\!\equiv (\textbf{\textit{set-pl}}_a\ \bigcirc\varphi)\mathbf{Pl}_a\mathsf{A}\bigcirc(\mathbf{Ph}\,\varphi),$$
$$\models\!\equiv (\textbf{\textit{set-pl}}_a\ \Box\,\varphi)\mathbf{Pl}_a\mathsf{A}\Box\,(\mathbf{Ph}\,\varphi),\ \textit{and}$$
$$\models\!\equiv (\textbf{\textit{set-pl}}_a\ \varphi_1\mathcal{U}\,\varphi_2)\mathbf{Pl}_a\mathsf{A}(\mathbf{Ph}\,\varphi_1)\mathcal{U}\,(\mathbf{Ph}\,\varphi_2).$$

5. VERIFICATION OF PLAUSIBILITY, TIME AND BELIEFS

In this section we report preliminary results on model checking **CTLKP** formulae. Clearly, verifying **CTLKP** properties directly against models with plausibility does not make much sense, since these models are inherently infinite; what we need is a finite representation of plausibility sets. One such representation has been discussed in Section 3.3: plausibility sets can be defined by path formulae and the update operator $(\mathbf{set\text{-}pl}_a\ \gamma)$.

We follow this idea here, studying the complexity of model checking **CTLKP** formulae *against* ***CTLK*** *models* (which can be seen as a compact representation of **CTLKP** models in which all the paths are assumed plausible), with the underlying idea that plausibility sets, when needed, must be defined explicitly in the object language. Below we sketch an algorithm that model-checks **CTLKP** formulae in time linear wrt the size of the model and the length of the formula. This means that we have extended **CTLK** to a more expressive language with no computational price to pay.

First of all, we get rid of the belief operators (due to Theorem 15), replacing every occurrence of $B_a\varphi$ with $K_a\mathbf{Pl}_a(\mathsf{E}\bigcirc\top \rightarrow \varphi)$. Now, let $\overrightarrow{\gamma} = \langle\gamma_1, ..., \gamma_k\rangle$ be a vector of "vanilla" path formulae (one per agent), with the initial vector $\overrightarrow{\gamma_0} = \langle\top, ..., \top\rangle$, and $\overrightarrow{\gamma}[\gamma'/a]$ denoting vector $\overrightarrow{\gamma}$, in which $\overrightarrow{\gamma}[a]$ is replaced with γ'. Additionally, we define $\overrightarrow{\gamma}[0] = \top$. We translate the resulting **CTLKP** formulae to ones without plausibility via function $tr(\varphi) = tr_{\overrightarrow{\gamma_0},0}(\varphi)$, defined as follows:

$$\begin{aligned}
&tr_{\overrightarrow{\gamma},i}(p) = p,\\
&tr_{\overrightarrow{\gamma},i}(\varphi_1 \wedge \varphi_2) = tr_{\overrightarrow{\gamma},i}(\varphi_1) \wedge tr_{\overrightarrow{\gamma},i}(\varphi_2),\\
&tr_{\overrightarrow{\gamma},i}(\neg\varphi) = \neg tr_{\overrightarrow{\gamma},i}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(K_a\varphi) = K_a\ tr_{\overrightarrow{\gamma},0}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(\mathbf{Pl}_a\,\varphi) = tr_{\overrightarrow{\gamma},a}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}((\mathbf{set\text{-}pl}_a\ \gamma')\varphi) = tr_{\overrightarrow{\gamma}[\gamma'/a],i}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(\mathbf{Ph}\,\varphi) = tr_{\overrightarrow{\gamma},0}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(\bigcirc\varphi) = \bigcirc tr_{\overrightarrow{\gamma},i}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(\Box\,\varphi) = \Box\ tr_{\overrightarrow{\gamma},i}(\varphi),\\
&tr_{\overrightarrow{\gamma},i}(\varphi_1\mathcal{U}\,\varphi_2) = tr_{\overrightarrow{\gamma},i}(\varphi_1)\mathcal{U}\ tr_{\overrightarrow{\gamma},i}(\varphi_2),\\
&tr_{\overrightarrow{\gamma},i}(\mathsf{E}\gamma') = \mathsf{E}(\overrightarrow{\gamma}[i] \wedge tr_{\overrightarrow{\gamma},i}(\gamma')).
\end{aligned}$$

Note that the resulting sentences belong to the logic of **CTLK+**, that is **CTL+** (where each path quantifier can be followed by a *Boolean combination* of "vanilla" path formulae)[9] with epistemic modalities. The following proposition justifies the translation.

PROPOSITION 18. *For any* ***CTLKP*** *formula φ without B_a, we have that $M, q \models_{\textbf{\textit{CTLKP}}} \varphi$ iff $M, q \models_{\textbf{\textit{CTLK+}}} tr(\varphi)$.*

In general, model checking **CTL+** (and also **CTLK+**) is Δ_2^P-complete. However, in our case, the Boolean combinations of path subformulae are always conjunctions of at most two non-negated elements, which allows us to propose the following model checking algorithm. First, subformulae are evaluated recursively: for every subformula ψ of φ, the set of states in M that satisfy ψ is computed and labeled with a new proposition p_ψ. Now, it is enough to define checking $M, q \models \varphi$ for φ in which all (state) subformulae are propositions, with the following cases:

Case $M, q \models \mathsf{E}(\Box\, p \wedge \gamma)$**:** If $M, q \not\models p$, then return **no**. Otherwise, remove from M all the states that do not satisfy p (yielding a sparser model M'), and check the **CTL** formula $\mathsf{E}\gamma$ in M', q with any **CTL** model-checker.

Case $M, q \models \mathsf{E}(\bigcirc p \wedge \gamma)$**:** Create M' by adding a copy q' of state q, in which only the transitions to states satisfying p are kept (i.e., $M, q' \models r$ iff $M, q \models r$; and $q'Rq''$ iff qRq'' and $M, q'' \models p$). Then, check $\mathsf{E}\gamma$ in M', q'.

[9] For the semantics of **CTL+**, and discussion of model checking complexity, cf. [17].

Case $M, q \models \mathsf{E}(p_1 \mathcal{U}\ p_2 \wedge p_3 \mathcal{U}\ p_4)$**:** Note that this is equivalent to checking $\mathsf{E}(p_1 \wedge p_3)\mathcal{U}\ (p_2 \wedge \mathsf{E}p_3\mathcal{U}\ p_4) \vee \mathsf{E}(p_1 \wedge p_3)\mathcal{U}\ (p_4 \wedge \mathsf{E}p_1\mathcal{U}\ p_2)$, which is a **CTL** formula.

Other cases: The above cases cover all possible formulas that begin with a path quantifier. For other cases, standard **CTLK** model checking can be used.

THEOREM 19. *Model checking* ***CTLKP*** *against* ***CTLK*** *models is PTIME-complete, and can be done in time $O(ml)$, where m is the number of transitions in the model, and l is the length of the formula to be checked. That is, the complexity is no worse than for* ***CTLK*** *itself.*

6. CONCLUSIONS

In this paper a notion of *plausible behavior* is considered, with the underlying idea that *im*plausible options should be usually ignored in practical reasoning about possible future courses of action. We add the new notion of plausibility to the logic of **CTLK** [19], and obtain a language which enables reasoning about what can (or must) plausibly happen. As a technical device to define the semantics of the resulting logic, we use a non-standard satisfaction relation $\models_P$ that allows to propagate the "current" set of plausible paths into subformulae. Furthermore, we propose a non-standard notion of beliefs, defined in terms of indistinguishability and plausibility. We also propose how plausibility assumptions can be specified in the object language via a *plausibility update operator* (in a way similar to [13]).

We use this new framework to investigate some important properties of plausibility, knowledge, beliefs, and updates. In particular, we show that knowledge is an **S5** modality, and that beliefs satisfy axioms **K45** in general, and **KD45** for the class of *plausibly serial models*. We also prove that *believing in φ* is *knowing that φ plausibly holds in all plausibly possible situations*. That is, the relationship between knowledge and beliefs is very natural and reflects the initial intuition precisely. Moreover, the model checking results from Section 5 show that verification for **CTLKP** is no more complex than for **CTL** and **CTLK**.

We would like to stress that we do not see this contribution as a mere technical exercise in formal logic. Human agents use a similar concept of plausibility and "practical" beliefs in their everyday reasoning in order to reduce the search space and make the reasoning feasible. As a consequence, we suggest that the framework we propose may prove suitable for modeling, design, and analysis resource-bounded agents in general.

We would like to thank Juergen Dix for fruitful discussions, useful comments and improvements.

7. REFERENCES

[1] R. Alur, T. A. Henzinger, and O. Kupferman. Alternating-time Temporal Logic. *Journal of the ACM*, 49:672–713, 2002.

[2] N. Bulling and W. Jamroga. Agents, beliefs and plausible behavior in a temporal setting. Technical Report IfI-06-05, Clausthal Univ. of Technology, 2006.

[3] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 995–1072. Elsevier, 1990.

[4] E.A. Emerson and J.Y. Halpern. "sometimes" and "not never" revisited: On branching versus linear time temporal logic. *Journal of the ACM*, 33(1):151–178, 1986.

[5] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. *Reasoning about Knowledge*. MIT Press: Cambridge, MA, 1995.

[6] R. Fagin and J.Y. Halpern. Reasoning about knowledge and probability. *Journal of ACM*, 41(2):340–367, 1994.

[7] N. Friedman and J.Y. Halpern. A knowledge-based framework for belief change, Part I: Foundations. In *Proceedings of TARK*, pages 44–64, 1994.

[8] N. Friedman and J.Y. Halpern. A knowledge-based framework for belief change, Part II: Revision and update. In *Proceedings of KR'94*, 1994.

[9] J.Y. Halpern. Reasoning about knowledge: a survey. In *Handbook of Logic in Artificial Intelligence and Logic Programming. Vol. 4: Epistemic and Temporal Reasoning*, pages 1–34. Oxford University Press, Oxford, 1995.

[10] J.Y. Halpern and R. Fagin. Modelling knowledge and action in distributed systems. *Distributed Computing*, 3(4):159–177, 1989.

[11] W. Jamroga and N. Bulling. A general framework for reasoning about rational agents. In *Proceedings of AAMAS'07*, 2007. Short paper.

[12] W. Jamroga and W. van der Hoek. Agents that know how to play. *Fundamenta Informaticae*, 63(2–3):185–219, 2004.

[13] W. Jamroga, W. van der Hoek, and M. Wooldridge. Intentions and strategies in game-like scenarios. In *Progress in Artificial Intelligence: Proceedings of EPIA 2005*, volume 3808 of *LNAI*, pages 512–523. Springer Verlag, 2005.

[14] W. Jamroga and Thomas Ågotnes. Constructive knowledge: What agents can achieve under incomplete information. Technical Report IfI-05-10, Clausthal University of Technology, 2005.

[15] B.P. Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, 12(4):381–408, 2003.

[16] P. Lamarre and Y. Shoham. Knowledge, certainty, belief, and conditionalisation (abbreviated version). In *Proceedings of KR'94*, pages 415–424, 1994.

[17] F. Laroussinie, N. Markey, and Ph. Schnoebelen. Model checking CTL+ and FCTL is hard. In *Proceedings of FoSSaCS'01*, volume 2030 of *LNCS*, pages 318–331. Springer, 2001.

[18] Y. Moses and Y. Shoham. Belief as defeasible knowledge. *Artificial Intelligence*, 64(2):299–321, 1993.

[19] W. Penczek and A. Lomuscio. Verifying epistemic properties of multi-agent systems via bounded model checking. In *Proceedings of AAMAS'03*, pages 209–216, New York, NY, USA, 2003. ACM Press.

[20] K. Su, A. Sattar, G. Governatori, and Q. Chen. A computationally grounded logic of knowledge, belief and certainty. In *Proceedings of AAMAS'05*, pages 149–156. ACM Press, 2005.

[21] W. van der Hoek, M. Roberts, and M. Wooldridge. Social laws in alternating time: Effectiveness, feasibility and synthesis. *Synthese*, 2005.