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
1.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—Multiagent Systems; 1.2.4 [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 non-trivial (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
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 $\text{CTLKP}$ 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. $\text{CTLK}$ [14] is a straightforward combination of the computation tree logic $\text{CTL}$ [4, 3] and standard epistemic logic [9, 5].

$\text{CTL}$ includes operators for temporal properties of systems: i.e., path quantifier $\exists$ ("there is a path"); together with temporal operators: $\bigcirc$ ("in the next state"); $\Box$ ("always from now on") and $\bigcup$ ("until"). Every occurrence of a temporal operator is preceded by exactly one path quantifier in $\text{CTL}$ (this variant of the language is sometimes called "vanilla" $\text{CTL}$). Epistemic logic uses operators for representing agents’ knowledge: $K\alpha$ is read as "agent $a$ knows that $\varphi$".

Let $\Pi$ be a set of atomic propositions with a typical element $p$, and $\text{Ag} \triangleq \{1, \ldots, k\}$ be a set of agents with a typical element $a$. The language of $\text{CTLK}$ consists of formulae $\varphi$, given as follows:

$$\varphi ::= p \mid \neg \varphi \mid \varphi \land \varphi \mid E\gamma \mid K\alpha \varphi$$

$$\gamma ::= \bigcirc \varphi \mid \Box \varphi \mid A\varphi \varphi.$$  

We will sometimes refer to formulae $\varphi$ as ("vanilla") state formulae and to formulae $\gamma$ as ("vanilla") path formulae.

The semantics of $\text{CTLK}$ is based on Kripke models $M = \langle Q, R, \sim_1, \ldots, \sim_n, \pi \rangle$, which include a nonempty set of states $Q$, a state transition relation $R \subseteq Q \times Q$, epistemic indistinguishability relations $\sim_1 \subseteq Q \times Q$ (one per agent), and a valuation of propositions $\pi: \Pi \rightarrow P(Q)$. We assume that relation $R$ is serial and that all $\sim$ are equivalence relations. A path $\lambda$ in $M$ refers to a possible behavior (or computation) of system $M$, and can be represented as an infinite sequence of states that follow relation $R$, that is, a sequence $q_0, q_1, \ldots$ such that $q_i, R, q_{i+1}$, for every $i \geq 0$. We denote the $i$th state in $\lambda$ by $\lambda[i]$. The set of all paths in $M$ is denoted by $\Lambda_M$ (if the model is clear from context, $M$ 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 $q_0, q_1, \ldots$ such that $q_0 = q_0$ and there are $q_0, q_1, \ldots, q_i$ such that $q_0, q_1, \ldots, q_i \in \Lambda_M$. The semantics of $\text{CTLK}$ is defined as follows:

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

$$M, q \models \neg \varphi \iff M, q \not\models \varphi;$$

$$M, q \models \varphi \land \psi \iff M, q \models \varphi \text{ and } M, q \models \psi;$$

$$M, q \models \bigcirc \varphi \iff \text{there is a } q \text{-path } \lambda \text{ such that } M, \lambda[i] \models \varphi;$$

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

3. EXTENDING TIME AND KNOWLEDGE WITH PLASIBILITY 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 $\text{CTLK}$ with the notion of plausibility by adding plausible path operators $Pl_1$ and physical path operator $Ph$ to the logic. Formula $Pl_1 \lambda \varphi$ has the intended meaning: according to agent $a$, it is plausible that $\varphi$ holds; formula $Ph \lambda \varphi$ reads as: $\varphi$ 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 $\sim$) is the semantic concept that underlies knowledge. In this sense, our work follows [7, 8, 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, maximin, 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.\footnote{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.}

To formalize this idea, we extend models of CTL with 
sets of plausible paths and add plausibility operators $P_\alpha$, physical paths operator $P_\phi$, and belief operators $B_\alpha$ to the language of CTL. 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:

$$
\begin{align*}
\varphi &::= p | \neg \varphi | \varphi \land \varphi | E\gamma | P_\alpha \varphi | P_\phi \varphi | K_\alpha \varphi | B_\alpha \varphi \\
\gamma &::= O \varphi | O \varphi | \varphi U \varphi.
\end{align*}
$$

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: $P_\alpha (\text{late} \rightarrow \neg \text{open}) \land P_\phi (\text{late} \land \text{open})$.

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

$$
M = \langle Q, R, \sim_1, \ldots, \sim_k, T_1, \ldots, T_k, \pi \rangle,
$$

where $\langle Q, R, \sim_1, \ldots, \sim_k, \pi \rangle$ is a CTLK model, and $T_a \subseteq \Lambda_M$ is the set of paths in $M$ that are plausible according to agent $\alpha$. If we want to make it clear that $T_a$ is taken from model $M$, we will write $T_a^M$. It seems worth emphasizing that this notion of plausibility is subjective and holistic. It is subjective because $T_a$ represents agent $\alpha$’s subjective view on what is plausible – and indeed, different agents may have different ideas on plausibility (i.e., $T_a$ may differ from $T_b$). It is holistic because $T_a$ represents agent $\alpha$’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 $?[\pi]$), 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 $\lambda$ is plausible in $q = \lambda[0]$, but the set of plausible paths in $q' = \lambda[1]$ is empty. That is, by following plausible path $\lambda$ we are bound to get to an implausible situation. But then, does it make sense to consider $\lambda$ as plausible?

Secondly, we use a non-standard satisfaction relation $\models_\rho$, which we call plausible satisfaction. Let $M$ be a CTLK model and $P \subseteq \Lambda_M$ be an arbitrary subset of paths in $M$ (not necessarily any $\Gamma^*_M$). $\models_\rho$ restricts the evaluation of temporal formulae to the paths given in $P$ only. The “absolute” satisfaction relation $\models$ is defined as $\models_\Lambda_M$.

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

$$
\begin{align*}
M, q &\models_\rho p \ \text{iff} \ q \in \pi(p); \\
M, q &\models_\rho \neg \varphi \ \text{iff} \ M, q \not\models_\rho \varphi; \\
M, q &\models_\rho \varphi \land \psi \ \text{iff} \ M, q \models_\rho \varphi \ \text{and} \ M, q \models_\rho \psi; \\
M, q &\models_\rho E \varphi \ \text{iff there is a } q\text{-subpath } \lambda \in P \ \text{such that } M, \lambda[i] \models_\rho \varphi; \\
M, q &\models_\rho P_\phi \varphi \ \text{iff there is a } q\text{-subpath } \lambda \in P \ \text{such that } M, \lambda[i] \models_\rho \varphi \ \text{for every } i \geq 0; \\
M, q &\models_\rho K_\alpha \varphi \ \text{iff there is a } q\text{-subpath } \lambda \in P \ \text{and } i \geq 0 \ \text{such that } M, \lambda[i] \models_\rho \psi, \ \text{and } M, \lambda[j] \models_\rho \varphi \ \text{for every } 0 \leq j < i; \\
M, q &\models_\rho B_\alpha \varphi \ \text{iff for all } q' \ \text{such that } q \sim_\alpha q'; \\
M, q &\models_\rho B_\alpha \varphi \ \text{iff for all } q' \in \mathcal{P}(\alpha) \ \text{with } q \sim_\alpha q', \ \text{we have that } M, q' \models_\rho \varphi.
\end{align*}
$$

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 $\varphi$, 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 $\varphi$, 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 $\varphi$ is $M$-true ($M \models \varphi$) if $M, q \models_\rho \varphi$ for all $q \in Q_M$. $\varphi$ is valid ($\models_\rho \varphi$) if $M, q \models_\rho \varphi$ for all models $M$. $\varphi$ is $M$-strongly true ($M \models_\rho \varphi$) if $M, q \models_\rho \varphi$ for all $q \in Q_M$ and all $P \subseteq \Lambda_M$. $\varphi$ is strongly valid ($\models_\rho \varphi$) if $M, q \models_\rho \varphi$ for all models $M$.

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 $T_a$ into the semantics of nested formulae. The importance of strong validity, however, lies in the fact that $\models_\rho \varphi \iff \psi$ makes $\varphi$ and $\psi$ completely interchangeable, while the same is not true for normal validity.

Proposition 3. Let $\Phi[\varphi/\psi]$ denote formula $\Phi$ in which every occurrence of $\psi$ was replaced by $\varphi$. Also, let $\models_\rho \varphi \iff \psi$. Then for all $M, q, P$: $M, q \models_\rho P \varphi$ if and only if $M, q \models_\rho \Phi[\varphi/\psi]$ (in particular, $M, q \models_\rho \Phi[\varphi/\psi]$).

Note that $\models_\rho \varphi \iff \psi$ does not even imply that $M, q \models_\rho \Phi[\varphi/\psi]$. 
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 (and collect EUR 1,000,000) if both chose 1, otherwise they lose. Formally, we model the game with a CTLKP model \( M \), in which the set of states \( Q \) includes \( q_r \) for the initial situation, states \( q_r, r \in [0, 1] \), for the situations after a has chosen number \( r \), and “final” states \( q_w, q_1 \) for the winning and the losing situation, respectively. The transition relation is as follows: \( q_r R q_r \) for all \( r \in [0, 1] \); \( q_1 R q_w, q_w R q_w, \) and \( q_r R q_r \). Moreover, \( \pi(\text{one}) = \{ q_1 \} \) and \( \pi(\text{win}) = \{ q_w \} \). Player a has perfect information in the game (i.e., \( q \sim q' \) iff \( q = q' \)), but player b does not distinguish between states \( q_r \) (i.e., \( q \sim_b q_{r'} \) 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 \( Y_a = Y_b = \{ q_1 q_w q_w \ldots \} \).

Note that, in principle, the outcome of the game is uncertain: \( \models M, q_r \models \neg A \Box \oplus \neg A \Box \). However, assuming rationality of the players makes it only plausible that the game must end up with a win: \( \models M, q_r \models P L a A \Box \oplus A \Box \). The agents believe that this will be the case: \( \models M, q_r \models B a A \Box \). Note also that, in any of the states \( q_r \), agent b believes that a (being rational) has played 1: \( \models M, q_r \models B a \). 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 \( \gamma \). 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 \( \Box \neg \text{snows} \). Formally, let \( \gamma \) be a CTLK path formula. We define \( \models M, q \models \gamma \gamma_{1,\ldots,n} \) to be the set of paths that satisfy \( \gamma \) in model \( M \):

\[
\begin{align*}
|\Box \varphi|_M & = \{ \lambda \mid M, \lambda[1] \models \varphi \} \\
|\varphi |_M & = \{ \lambda \mid \forall \lambda (M, \lambda[1] \models \varphi) \} \\
|\varphi_1 U \varphi_2 |_M & = \{ \lambda \mid \exists \lambda_j (M, \lambda[j] \models \varphi_2 \land \forall j(0 \leq j < i \Rightarrow M, \lambda[j] \models \varphi_1)) \}.
\end{align*}
\]

Moreover, we define the plausible paths model update as follows. Let \( M = (Q, R, \sim_1, \ldots, \sim_k, T_1, \ldots, T_k, \pi) \) be a CTLKP model, and let \( P \subseteq \Lambda_M \) be a set of paths. Then \( M^{P} = (Q, R, \sim_1, \ldots, \sim_k, T_1, \ldots, T_k, P, Y_{a+1}, \ldots, Y_k, \pi) \) denotes model \( M \) with a’s set of plausible paths reset to \( P \).

Now we can extend the language of CTLKP with formulae \( (\text{set-pl}_{\gamma} \gamma) \varphi \) with the intuitive reading: “suppose that \( \gamma \) exactly characterizes the set of plausible paths, then \( \varphi \) holds”, and formal semantics given below:

\[
M, q \models (\text{set-pl}_{\gamma} \gamma) \varphi \text{ iff } M^{a_{\gamma}} |_{M, q} \models \varphi.
\]

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 \( \alpha \) (which can be seen as a plausibility assumption expressed in the object language). More precisely, worlds that satisfy \( \alpha \) can be considered as plausible. This concept is expressed via symbols \( B_i^e \varphi \); the index \( i \in \{ 1, 2, 3 \} \) is used to distinguish between three different implementations of beliefs. The first version is given by \( B_i^e \varphi \equiv K(\alpha \rightarrow \varphi) \).

4 A drawback of this version is that if \( \alpha \) is false, then everything will be believed with respect to \( \alpha \). The second version overcomes this problem: \( B_i^e \varphi \equiv K(\alpha \rightarrow \varphi) \land \neg K(\neg \alpha \rightarrow \varphi) \); now \( \alpha \) is only believed if it is known that \( \varphi \) follows from assumption \( \alpha \), and \( \varphi \) must be known if assumption \( \alpha \) is known to be false. Finally, \( B_i^e \varphi \equiv K(\alpha \rightarrow \varphi) \land \neg K(\neg \alpha \rightarrow \varphi) \); if the assumption \( \alpha \) is known to be false, nothing should be believed with respect to \( \alpha \). The strength of these different notions is given as follows: \( B_i^e \varphi \) implies \( B_{e+1}^e \varphi \) and \( B_{e+2}^e \varphi \) implies \( B_{e+3}^e \varphi \). 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 \( P(q, a) = (\Omega, \rho, \zeta) \) to every state, or more generally, every possible world \( q \) and agent \( a \). The plausibility space

\[\text{Unlike in most approaches, } K \text{ is interpreted over all worlds and not only over the indistinguishable worlds.}\]
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 $T$.\footnote{When there are infinite chains \ldots \preceq q_1 \preceq q_2 \preceq q$, the definition is much more sophisticated. An interested reader is referred to \cite{7} for more details.} 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 \cite{6, 15}, on the other hand, offer a quantitative description.

The logic from \cite{7} is defined by the following grammar: $\varphi ::= p | \varphi \land \varphi | \neg \varphi | K_a \varphi | \varphi \rightarrow_a \varphi$, where the semantics of all operators except $\rightarrow_a$ is given as usual, and formulae $\varphi \rightarrow_a \psi$ have the meaning that $\psi$ is true in the most plausible worlds in which $\varphi$ holds. Formally, the semantics for $\rightarrow_a$ is given as: $M,q \models \varphi \rightarrow_a \psi$ iff $S^q_{P(q,a)} \rightarrow P(q,a) S^q_{P(q,a)}$, where $S^q_{P(q,a)} = \{q' \in \Omega(q,a) | M,q' \models \varphi\}$ are the states in $\Omega(q,a)$ that satisfy $\varphi$. 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 $B_a \varphi \equiv K_a(T \rightarrow_a \varphi)$.

Friedman and Halpern have shown that the KD45 axioms are valid for operator $B_a$ if plausibility spaces satisfy consistency (for all states $q \in Q$ that holds that $\Omega(q,a) \subseteq \{q' \in Q | q \sim_a q'\}$) and normality (for all states $q \in Q$ it holds that $\Omega(q,a) \neq \emptyset$).\footnote{Note that this “normality” is essentially seriality of states wrt plausibility spaces.} A temporal extension of the language (mentioned briefly in \cite{7}, and discussed in more detail in \cite{8}) uses the interpreted systems approach \cite{10, 5}. A system $R$ is generated by runs, where a run $r : N \rightarrow Q$ is a function from time moments (modeled by $N$) to global states, and a time point $(r,i)$ is given by a time point $i \in N$ and a run $r$. A global state is a combination of local states, one per agent. An interpreted system $M = (R, \pi)$ is given by a system $R$ and a valuation of propositions $\pi$. 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 $r_a(m)$ are equal. Formulae are interpreted in a straightforward way with respect to interpreted systems, e.g. $M,r,m \models K_a \varphi$ iff $M,r',m' \models \varphi$ for all $(r',m') \sim_a (r,m)$. Now, these are time points that play the role of possible worlds; consequently, plausibility spaces $T(r,m,a)$ are assigned to each point $(r,m)$ and agent $a$.

**Su et al.: KBC Logic**

Su et al. \cite{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 $q = (q^{occ}, q^{inv}, q^{per}, Q_{pl})$ where the environment is divided into a visible ($q^{occ}$) 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 $Q_{pl}$ that are plausible for the agent.

Given a global state $q$, we additionally define $Vis(q) = q^{occ}$, $Inv(q) = q^{inv}$, $Per(q) = q^{per}$, and $Pls(q) = Q_{pl}$. The semantics is given by an extension of interpreted systems \cite{10, 5}, here, it is called interpreted KBC systems. KBC formulae are defined as $\varphi ::= p | \neg \varphi | \varphi \land \varphi | K \varphi | B \varphi | C \varphi$. 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.

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

$M,r,i \models C \varphi$ if $M,r',i' \models \varphi$ for all $(r',i')$ with $Vis(r'(i')) = Per(r(i))$

Thus, an agent believes $\varphi$ if, and only if, $\varphi$ 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 $\varphi$, 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 \cite{18} and \cite{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 \cite{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 \cite{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.\footnote{To be more precise, time in \cite{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, $\varphi$ holds”.} 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, \cite{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.\footnote{Friedman and Halpern even briefly mention how plausibility of runs can be embedded in their framework.} We believe that branching time is more suitable for non-deterministic domains (cf. e.g. \cite{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
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. PLASIBILITY, 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. \(\text{sn}(\Upsilon_a) = Q\). 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 \(q' \in \text{sn}(\Upsilon_a)\). 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 \(\text{Pl}_a \beta_a \oplus \top\) is valid in \(M\). \(M\) is weakly p-serial for agent \(a\) iff \(\neg \text{K}_a \text{Pl}_a \bigcirc \bot\) 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: \(\models (B_a \varphi \land B_a (\varphi \rightarrow \psi)) \rightarrow B_a \psi\), \(\models (B_a \varphi \rightarrow B_a B_a \varphi)\), and \(\models (\neg B_a \varphi \rightarrow B_a \neg B_a \varphi)\).

The next proposition concerns the “consistency” axiom D: \(B_a \varphi \rightarrow \neg B_a \neg \varphi\). It is easy to see that the axiom is not valid in general: as we have no restrictions on plausibility sets \(\Upsilon_a\), it may be as well that \(\Upsilon_a = \emptyset\). In that case we have \(B_a \varphi \land B_a \neg \varphi\) for all formulae \(\varphi\), 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., \(\not\models (B_a \varphi \rightarrow \varphi)\). 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 \(\varphi\) be a CTLKP formula, and \(M\) be a CTLKP model. We have the following strong validities:

(i) \(\models \text{Pl}_a K_a \varphi \leftrightarrow K_a \varphi\)

(ii) \(\models \text{Ph} K_a \varphi \leftrightarrow K_a \text{Ph} \varphi\) and \(\models K_a \text{Ph} \varphi \leftrightarrow K_a \varphi\)

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 \((K_a \text{E} \bigcirc \text{invasion})\), but if we do not take this possibility as plausible \((\neg \text{Pl}_a \text{E} \bigcirc \text{invasion})\), then we reject the corresponding belief in consequence \((\neg B_a \text{E} \bigcirc \text{invasion})\). Note that this property reflects the strong connection between belief and plausibility in our framework.

Proposition 11. The following formulae are strongly valid:

(i) \(B_a \varphi \rightarrow K_a B_a \varphi\), (ii) \(K_a B_a \varphi \rightarrow B_a \varphi\), (iii) \(K_a \varphi \rightarrow B_a K_a \varphi\).

The following formulae are not valid:

(iv) \(B_a \varphi \rightarrow B_a K_a \varphi\), (v) \(K_a \varphi \rightarrow B_a \varphi\)

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 \(\varphi\) be a CTLKP formula that does not include any temporal operators. Then \(K_a \varphi \rightarrow B_a \varphi\) is strongly valid, and in the class of p-serial models we have even that \(\models K_a \varphi \leftrightarrow B_a \varphi\).
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]:
\[ \varphi_u := p \mid \neg p \mid \varphi_u \land \varphi_u \mid \varphi_v \lor \varphi_u \mid A\gamma_u \mid K\alpha\varphi_u, \]
\[ \gamma_u := \bigcirc \varphi_u \mid \bigcirc \varphi_u \lor \varphi_u U \varphi_u. \]
We call such \( \varphi_u \) universal formulae, and \( \gamma_u \) universal path formulae.

**Proposition 13.** Let \( \varphi_u \) be a universal CTL formula. Then \( \models K\alpha\varphi_u \rightarrow B\alpha\varphi_u \).

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 two theorems are strongly valid in the class of plausibly serial CTLKP models:
(i) \( B\alpha\varphi \leftrightarrow K\alpha\Pi_\alpha\varphi \),
(ii) \( K\alpha\varphi \leftrightarrow B\alpha\Phi\alpha\varphi \).

**Theorem 15.** Formula \( B\alpha\varphi \leftrightarrow K\alpha\Pi_\alpha\varphi (E \circ T \rightarrow \varphi) \) is strongly valid.

Note that this characterization has a strong commonsense reading: believing in \( \varphi \) is knowing that \( \varphi \) 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 \( \varphi \) be a CTLKP formula that does not include operators \( I_\alpha \) and \( B\alpha \), and \( \gamma \) be a CTLKP path formula. Then, we have \( \models \varphi \leftrightarrow (\text{set-pl}_\alpha \gamma) \varphi \).

What can be said about the result of an update? At first sight, formula \( (\text{set-pl}_\alpha \gamma) \Pi_\alpha A\gamma \) 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 \( \gamma \) (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 is to physical possibilities, in a way analogous to [13].

**Proposition 17.** The CTLKP formula \( (\text{set-pl}_\alpha \gamma) \Pi_\alpha A\gamma \) is not valid. However, we have the following validities:
(i) \( \models (\text{set-pl}_\alpha \gamma) \Pi_\alpha A\gamma_u \), where \( \gamma_u \) is a universal CTL path formula from Definition 1.
(ii) If \( \varphi, \varphi_1, \varphi_2 \) are arbitrary CTLK formulae, then:
\[ \models (\text{set-pl}_\alpha \bigcirc \varphi) \Pi_\alpha A\alpha (\Phi\varphi), \]
\[ \models (\text{set-pl}_\alpha \bigcirc \varphi) \Pi\alpha A\alpha (\Phi\varphi), \]
\[ \models (\text{set-pl}_\alpha \varphi U \varphi_2) \Pi_\alpha A\alpha (\Phi\varphi_1) U (\Phi\varphi_2). \]

### 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 \( \text{set-pl}_\gamma \).

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 \( B\alpha\varphi \) with \( K\alpha\Pi_\alpha (E \circ T \rightarrow \varphi) \). Now, let \( \overrightarrow{\gamma} = \langle \gamma_1, \ldots, \gamma_k \rangle \) be a vector of "vanilla" path formulae (one per agent), with the initial vector \( \gamma_0 = \langle T, \ldots, T \rangle \), and \( \gamma_\gamma /a \) denoting vector \( \gamma \), in which \( \gamma [a] \) is replaced with \( \gamma \). Additionally, we define \( \gamma[0] = T \). We translate the resulting CTLKP formulae to ones without plausibility via function \( tr(\varphi) = tr\gamma_0(\varphi) \), defined as follows:

\[ tr\overrightarrow{\gamma},(p) = p, \]
\[ tr\overrightarrow{\gamma},(\varphi_1 \land \varphi_2) = tr\overrightarrow{\gamma},(\varphi_1) \land tr\overrightarrow{\gamma},(\varphi_2), \]
\[ tr\overrightarrow{\gamma},(\neg \varphi) = \neg tr\overrightarrow{\gamma},(\varphi), \]
\[ tr\overrightarrow{\gamma},(K\alpha\varphi) = K\alpha tr\overrightarrow{\gamma},(\varphi), \]
\[ tr\overrightarrow{\gamma},(\Pi_\alpha \varphi) = tr\overrightarrow{\gamma},(\varphi), \]
\[ tr\overrightarrow{\gamma},((\text{set-pl}_\alpha \gamma) \varphi) = tr\overrightarrow{\gamma}[\gamma/a],(\varphi), \]
\[ tr\overrightarrow{\gamma},(\Phi\varphi) = tr\overrightarrow{\gamma},(\varphi), \]
\[ tr\overrightarrow{\gamma},(\bigcirc \varphi) = \bigcirc tr\overrightarrow{\gamma},(\varphi), \]
\[ tr\overrightarrow{\gamma},(\varphi U \varphi_2) = \bigcirc tr\overrightarrow{\gamma},(\varphi) \land tr\overrightarrow{\gamma},(\varphi_2), \]
\[ tr\overrightarrow{\gamma},(E\gamma) = E(\gamma[0] \land tr\overrightarrow{\gamma},(\gamma)). \]

Note that the resulting sentences belong to the logic of CTLK+, that is CTLK+ (where each path quantifier can be followed by a Boolean combination of "vanilla" path formulae)\(^3\) with epistemic modalities. The following proposition justifies the translation.

**Proposition 18.** For any CTLKP formula \( \varphi \) without \( B\alpha \), we have that \( M, q \models \text{CTLKP} \varphi \) iff \( M, q \models \text{CTLK} \varphi \).

In general, model checking CTLK+ (and also CTLKP+) is \( \Delta^2_2 \)-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 \( \psi \) of \( \varphi \), the set of states in \( M \) that satisfy \( \psi \) is computed and labeled with a new proposition \( p_\psi \). Now, it is enough to define checking \( M, q \models \varphi \) for \( \varphi \) in which all (state) subformulae are propositions, with the following cases:

**Case** \( M, q \models E(\bigcirc p \land \gamma) \): 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 \( E\gamma \) in \( M' \), with any CTL model-checker.

**Case** \( M, q \models E(\bigcirc p \land \gamma) \): 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 \( q' Ra' \) \iff \( q Ra \) and \( M, q' \models p \)). Then, check \( E\gamma \) in \( M', q' \).

\(^3\) For the semantics of CTLK+, and discussion of model checking complexity, cf. [17].
Case $M, q \models \text{E}(p_1 \land p_2 \land p_3 \land p_4)$: Note that this is equivalent to checking $\text{E}(p_1 \land p_3 \land (p_2 \land \text{E} p_4)) \lor \text{E}(p_1 \land p_3 \land p_4 \land (p_2 \land \text{E} p_4))$, which is a $\text{CTL}$ formula.

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

THEOREM 19. Model checking $\text{CTLKP}$ against $\text{CTLK}$ models is PTIME-complete, and can be done in time $O(m^n)$, where $n$ 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 $\text{CTLK}$ itself.

6. CONCLUSIONS

In this paper a notion of plausible behavior is considered, with the underlying idea that implausible 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 $\text{CTLK}$ [19], and obtain a language which enables reasoning about what can (and 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 subformulæ. 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 $\text{S}5$ modality, and that beliefs satisfy axioms $\text{K45}$ in general, and $\text{KD45}$ for the class of plausibly serial models. We also prove that believing in $\varphi$ is knowing that $\varphi$ 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 $\text{CTLKP}$ is no more complex than for $\text{CTL}$ and $\text{CTLK}$.

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


