Towards a Formalisation of Justification and Justifiability*

Willem Hagemann
Department of Computing Science
University of Oldenburg
Germany
willem.hagemann@uol.de

We introduce the logic QK_S^D which is a normal multi-modal logic over finitely many modalities that additionally supports bounded quantification of modalities. An important feature of this logic is that it allows to quantify over the information components of systems and, hence, can be used to derive justifications. We compare the proposed logic with Artemov's justification logic and also report on a prototypical implementation of a satisfiability solver of this logic and show some examples.

1 Introduction

We are in the process of building complex autonomous systems that perceive their environment, exchange information and decide on their actions based on their understanding of the world. The hope is that the multitude of perspectives will lead to a more complete view of the world and enable systems to act more optimally. The problem is that information from different sources is often inconsistent. Inconsistencies are one of the major concern of formal methods, as the presence of an inconsistency usually means that a contradiction is derivable, and thus — ex falso quodlibet — any proposition is derivable.

A wide variety of formal methods have been developed for the analysis and verification of complex cyber-physical systems (CPS). A central approach is to create a formal model of the system (and its environment) and then to prove certain properties using rigorous logical-algorithmic methods such as model checking [1, 11, 10]. However, model checking of cyber-physical systems is often limited to the verification of temporal properties. Common assumptions are that the internal state and control laws of the system are precisely known and that such systems are deployed in limited environmental scenarios. Hence, the temporal behaviour of the system and its environment can be predicted within quantifiable uncertainties [18].

Especially with regard of autonomous CPS, such simplifying assumptions must be revised or — at least — be re-examined [28, 6]. Autonomous systems are supposed to operate in complex and open environments with a high degree of independence and self-determination. An autonomous system knows its action alternatives and independently decides for one of the possible actions [15]. Such a decision, however, shall then be justifiable and explainable — a requirement that research projects like XAI [12] or Perspicuous Computing [27] address. Moreover, a justification shall run through the entire perception and decision-making process of the system. Ideally, a justification keeps record of the sources from which the information originates that was used for decision-making. Such sources of information are, for example, external agents, but also internal subsystems, such as basic control laws, neural networks for image recognition/classification [28], or a catalogue of traffic rules [2].

Increasingly, modal logic concepts are proposed to formally describe autonomous systems as epistemic agents [28, 13, 15]. The numerous epistemic logics dealing with belief and knowledge (as true belief) in

*This work is partly funded by the German Research Council (DFG) as part of the project "Integrated Socio-technical Models for Conflict Resolution and Causal Reasoning", grant no. FR 2715/5-1, DA 206/12-1.

M. Farrell & M. Luckcuck (Eds.):
Third Workshop on Formal Methods
for Autonomous Systems (FMAS2021)
EPTCS 348, 2021, pp. 136–149, doi:10.4204/EPTCS.348.10

multi-agent systems [14, 36] are joined by other modal logics. E.g., the Center for Perspicuous Computing [27] proposes to use the knowledge representation of description logics [35, 31, 20], whereas [28] proposes to use differential dynamic logics, and [15] refers to the belief–desire–intention framework [30]. Another important candidate is Artemov's justification logic [5], as this logic promises for the first time to add a justification component to previous epistemic logics, thus satisfying Plato's characterisation of knowledge as justified true belief.

This paper deals with justification in a fallibilist process of cognition and decision making. Instead of belief and knowledge, it is based on the weaker notion of information, see Barwise [7, Ch. 9] and Gerbrandy [19] for a short discussion of the contrast of knowledge and information. Unlike belief or knowledge, information need not be consistent in its entirety. Consequently, we abandon destructive techniques that contract the current set of beliefs as proposed in the AGM framework[1] of belief revision [36]. Rather, we stockpile and maintain each source of information regardless of whether or not it contributes to a consistent worldview. A statement is justifiable in this setting if and only if we find a group of information sources that is free of contradiction in its totality and from which the statement can be substantiated.

In Sec. 2, I propose to use the well-known multimodal logic $K_{\mathcal{S}}^{D}$ — also referred to as K_n^D in [14] — to model the information distribution within an information processing system $\mathcal{S}$, where $\mathcal{S} = \{s_1, \ldots, s_n\}$ is a the set of atomic information components $s_1, \ldots, s_n$ of a system. In $K_{\mathcal{S}}^{D}$ the *principle of information distribution* — referred to as *distributed knowledge* in [14] — holds. That is, we consider arbitrary nonempty subsets of $\mathcal{S}$, called instances of $\mathcal{S}$. The information of an instance is given as the deductive closure of the information of its atomic components. Hence, $K_{\mathcal{S}}^{D}$ allows us to reason about the information of arbitrary combinations of information components. This in turn yields an analogous notion of justifications as in Artemov's justification logic [5, 4]. I gradually introduce basic features of $K_{\mathcal{S}}^{D}$ along an application example of automated driving. Besides proposing $K_{\mathcal{S}}^{D}$ as a justification logic and comparing it with Artemov's justification logic, this section does not contain any new results, but compiles those results which will be used in the rest of the paper.

In Sec. 3, I propose the extension of $K_{\mathcal{S}}^{D}$ by quantification over instances yielding the logic $QK_{\mathcal{S}}^{D}$, and propose a tableau calculus for $QK_{\mathcal{S}}^{D}$. With the help of quantifiers, the justifiability of a statement ϕ is defined formally as an existential quantified formula running over consistent instances of the systems $\mathcal{S}$. Afterwards, I revisit the running example and show some results with the help of a prototypical implementation of this tableau.

Finally, in Sec. 4, I summarise the presented work, and address a still outstanding comparison of the solver with other solvers. I also address possible extensions of $QK_{\mathcal{S}}^{D}$ and mention conceivable philosophical positions that are touched by this approach.

2 A Normal Logic of Justification

The following running example will guide us to through this section. A vehicle with an automated driving system supporting conditional driving automation (SAE Level 3, [26]) is driving on a multi-lane road. The vehicle observes that $[o_1]$ the current lane is blocked and that $[o_2]$ there is a gap on the neighbouring lane that is just large enough for a lane change. The car has to decide whether it should brake or use the small gap for the lane change. In addition to the observation data, the driving system of the vehicle uses the following information in its decision-making process:

[1] The framework is named after its proponents, Alchourrón, Gärdenfors, and Makinson.

- The driving system adheres to the following rules: $[r_1]$ Whenever the lane is blocked, brake or change lane. $[r_2]$ Never change lane when no gap is available. $[r_3]$ Never brake when the goal is to drive fast. $[r_4]$ When the goal is to drive considerately, then change lane only if there is a large gap.
- Since the user is in a hurry, for this trip he set the goal $[g_1]$ to drive fast. During the initial start-up dialog, the user has chosen $[g_2]$ a considerate driving behaviour as default.

Let Σ be a formalisation[2] of all given information as follows:

$$\begin{aligned}
&[o_1]\texttt{laneBlocked}, \qquad [o_2](\texttt{gapAvailable} \land \neg\texttt{largeGapAvailable}),\\
&[r_1](\texttt{laneBlocked} \to \texttt{brake} \lor \texttt{changeLane})\\
&[r_2](\texttt{changeLane} \to \texttt{gapAvailable})\\
&[r_3](\texttt{beFast} \to \neg\texttt{brake})\\
&[r_4](\texttt{beConsiderate} \land \texttt{changeLane} \to \texttt{largeGapAvailable})\\
&[g_1]\texttt{beFast}, \qquad [g_2]\texttt{beConsiderate}.
\end{aligned}$$

The given information is contradictory as there is no satisfying truth assignment of the propositional variables of the formalisation. Hence, the automated driving system sends out a request to intervene to the user. The surprised user takes control, gets an overview but misses to use the small gap. Since this situation occurs several times during his trip, he finally misses his appointment and sends a disappointed experience report to the manufacturer.

2.1 Labels are Normal Modalities

We put ourselves in the role of the manufacturer, analyse the formalisation of the situation described above, and immediately identify the unsatisfiability of Σ as a cause of the request to intervene. Since we already structured the information by labels, it is quite natural to ask what labels contribute to the contradiction.

To this end, let us discuss the role of the labels first. The labels refer to various information components. For example, $[g_1]$ refers to the initial start-up dialog of the car, $[r_4]$ refers to the engineer who added comfort functions to the vehicle, and $[o_1]$ refers to the front radar of the vehicle. The notation $[o_1]\phi$ denotes that ϕ belongs to the information of $[o_1]$, and — contrary to the given situation above — $[o_1]\neg\texttt{laneBlocked}$ describes a situation where the front radar has the information that the lane is not blocked. Moreover, we use negation of labelled formulae to denote the absence of information: $\neg[o_1]\texttt{laneBlocked}$ formalises that the front radar does not have the information that the lane is blocked, i.e., the front radar considers it possible that the lane is not blocked. Often, we write $\langle o_1\rangle\phi$ instead of $\neg[o_1]\neg\phi$. Readers familiar with modal logic immediately recognise that $[o_1]$ and $\langle o_1\rangle$ are a pair of dual modal operators. In the following, any label s will be a modality and $[s]$ the corresponding normal modal operator. A normal modal operator $[s]$ has access to all logical tautologies via the necessitation rule

$$\text{from} \vdash \phi \text{ infer} \vdash [s]\phi, \tag{\textbf{Nec}}$$

[2] A hint on the formalisation of rule $[r_4]$ seems appropriate. The rule consists of two nested conditionals, the latter being "change lane only if there is a large gap". The correct formalisation of the "only if"-expression is $\texttt{changeLane} \to \texttt{largeGapAvailable}$, just as the popular mathematical jargon "A if and only if B" decomposes into "A if B", formalised as $B \to A$, and "A only if B", formalised as $A \to B$. Hence, the whole rule is formalised as $\texttt{beConsiderate} \to (\texttt{changeLane} \to \texttt{largeGapAvailable})$. Finally, the listed form is obtained by a simple equivalence transformation.

and its information is closed under logical consequence. The latter is obtained by the Kripke schema

$$\vdash [s](\phi \to \psi) \to [s]\phi \to [s]\psi. \tag{K}$$

We denote the set of all labels by $\mathcal{S}$ and we will refer to elements of $\mathcal{S}$ as atomic modalities, or — for short — as atoms.

2.2 The Principle of Information Distribution

In order to identify those atoms of $\mathcal{S}$ that contribute to a contradiction, it is advantageous to consider arbitrary *nonempty* sets of atomic modalities, to which in turn we also assign modal operators. We will refer to such sets as instances. E.g., the instance $\{o_1, o_2\}$ consists of the atoms o_1 and o_2 and $[o_1, o_2]$ denotes its assigned normal modal operator. Further, we stipulate that for all instances the following principle of *information distribution* — in [14] referred to as *distributed knowledge* — holds: For all instances $s = \{s_1, \dots, s_n\} \subseteq \mathcal{S}$ and $t = \{t_1, \dots, t_m\} \subseteq \mathcal{S}$ it holds

$$\vdash [s](\phi \to \psi) \to [t]\phi \to [s,t]\psi \tag{Dist$'$}$$

where $s,t = \{s_1, \dots, s_n, t_1, \dots, t_m\}$.[3] Hence, $[o_1, o_2]$ has access to all information that o_1 or o_2 have, and also to all logical consequences that can be obtained by combining the information of o_1 and o_2, as $[o_1, o_2]$ is again closed under logical consequence.

2.3 Syntax and Axiomatisation of $K^D_\mathcal{S}$

We formally introduce the logic $K^D_\mathcal{S}$. It turns out that $K^D_\mathcal{S}$ is actually a normal multi-agent logic with an axiom for information distribution over (nonempty) groups of agents. The axiom of information distribution (or an equivalent form thereof) is often found in multi-agent epistemic logics, sometime referred to as group information or distributed knowledge [21, 14, 19]. However, the presented logic lags behind most epistemic logics, as it does not make any stronger requirements on the modality of information. Information neither has to be factually true, like it is the case for epistemic logics dealing with knowledge, nor be free of contradictions, as it is usually stipulated in logics of belief. The main difference lies in the designation of the modalities. While the usual epistemic logics identify the modalities with agents, we use modalities for fairly liberally chosen information sources.

The following two definitions give the language and an axiomatisation of $K^D_\mathcal{S}$.

Definition 1. *Let $\mathcal{V}$ be a enumerable set of propositional variables and $\mathcal{S}$ be a finite set of atomic modalities. A formula ϕ of $K^D_\mathcal{S}$ is any expression that satisfies the Backus-Naur form*

$$\phi ::= \bot \mid A \mid \phi \to \phi \mid [s]\phi,$$

where $A \in \mathcal{V}$, and $s \subseteq \mathcal{S}$ is an instance, i.e., nonempty.

We make use of additional logical operators that are defined as abbreviations in the usual way, e.g., $\neg\phi$ abbreviates $\phi \to \bot$ and $\langle s\rangle\phi$ abbreviates $\neg[s]\neg\phi$. The logical connective $\to$ is treated as right-associative and the operator precedence is defined according the descending sequence $([s], \langle s\rangle, \neg, \vee, \wedge, \to, \leftrightarrow)$.

[3] Note that we deliberately avoid the obvious use of the set operation $s \cup t$ and prefer to write s,t here, as this would yield confusion with the reversed notational conventions in Boolean modal logic or description logic $\mathcal{ALB}$. Indeed, in these logics the proper notation of s,t would be rendered by $s \cap t$.

Definition 2. *Let $\mathcal{S}$ be a finite set of atoms and $\mathcal{V}$ be an enumerable set of propositional variables. The logic $K_{\mathcal{S}}^{D}$ is given by the rules and axioms for all nonempty subsets s and t of $\mathcal{S}$*

$$\text{from} \vdash \phi \text{ and} \vdash \phi \rightarrow \psi \text{ infer} \vdash \psi; \qquad \textbf{(MP)}$$
$$\text{from} \vdash \phi \text{ infer} \vdash [s]\phi; \qquad \textbf{(Nec)}$$
$$\vdash \phi \quad \text{if } \phi \text{ is a substitution instance of a propositional tautology;} \qquad \textbf{(0)}$$
$$\vdash [s](\phi \rightarrow \psi) \rightarrow [s]\phi \rightarrow [s]\psi; \qquad \textbf{(K)}$$
$$\vdash [s]\phi \rightarrow [t]\phi \quad \text{if } s \subseteq t. \qquad \textbf{(Dist)}$$

Note that $K_{\mathcal{S}}^{D}$ is the simplest normal multi-modal logic in which the principle of information distribution **Dist**$'$ holds, and in [14] it is referred to as K_n^D where n denotes the number of atoms.

2.4 Semantics of $K_{\mathcal{S}}^{D}$

It is a well-known result (e.g. [21, 14, 19]) of modal logic that $K_{\mathcal{S}}^{D}$ is sound, complete, and also decidable for the following interpretation over certain Kripke structures.

Definition 3. *Let $\mathcal{S}$ be a nonempty set of modalities. A* multi-modal Kripke structure for $K_{\mathcal{S}}^{D}$ *is the tuple $M = (\Omega, \mapsto, \pi)$, where*

(i) Ω *is the nonempty set of* possible worlds*;*

(ii) $\mapsto$ *is a function that maps any atom $s \in \mathcal{S}$ to a binary relation $\mapsto_s$ on Ω, called the* accessibility relation *of s. The extension of $\mapsto$ to all instances is given by the intersection of their atomic components, i.e., for any instance $s \subseteq \mathcal{S}$ we have $\mapsto_s := \bigcap_{t \in s} \mapsto_t$;*

(iii) π *is a function that maps any possible world ω of Ω to a subset of $\mathcal{V}$, π is called the* truth assignment.

For any Kripke structure M as above, $\omega \in \Omega$, $A \in \mathcal{V}$, and any instance $s \subseteq \mathcal{S}$ we define:

$$(M,\omega) \not\models \bot,$$
$$(M,\omega) \models A \iff A \in \pi(\omega),$$
$$(M,\omega) \models \phi \rightarrow \psi \iff (M,\omega) \models \phi \text{ implies } (M,\omega) \models \psi,$$
$$(M,\omega) \models [s]\phi \iff \text{for all } \omega' \in \Omega \text{ with } \omega \mapsto_s \omega' \text{ it holds } (M,\omega') \models \phi.$$

Whenever the model relation $(M,\omega) \models \phi$ holds, (M,ω) is a said to be a pointed model *of ϕ. Whenever there exists ω with $(M,\omega) \models \phi$, we say that ϕ is* satisfiable *in M and write $M \models \phi$ for short. We say that ϕ is* valid*, denoted by $\models \phi$, if and only if $M \models \phi$ holds for all Kripke structures M as above.*

Let M be a Kripke structure as above. The definition captures the following properties of information. An instance s has information ϕ at ω, i.e., $(M,\omega) \models [s]\phi$, if and only if ϕ holds at all possible worlds that s has access to from ω. An instance s considers an information ϕ as possible at ω, i.e., $(M,\omega) \models \langle s \rangle \phi$, if only if s has access from ω to at least one possible world ω' where ϕ holds. As $\bot$ cannot be assigned to true, there is no possible world at which $\bot$ holds and $\top$ holds at any possible world. It follows that an instance s has inconsistent information at ω, formally $(M,\omega) \models [s]\bot$, if and only if there is no possible world that is accessible from ω for s. Now, we specify exactly what it means when we speak of consistent or inconsistent instances. An instance s is called *inconsistent* in the model M if and only if $M \models [s]\bot$ holds. Otherwise, s is called *consistent* and the dual statement $M \models \langle s \rangle \top$ holds.

2.5 Running Example Revisited

We return to the running example and consider the following derivation from Σ that shows that K_S^D indeed allows us to identify those atoms that contribute to a contradiction, i.e., there is an instance $s \subseteq \mathcal{S}$ such that $\Sigma \vdash [s]\bot$ holds.[4]

$$\begin{aligned}
&[r_3, g_1]\neg\texttt{brake} && \text{from } [r_3] \text{ and } [g_1], && (1)\\
&[o_1, r_1](\texttt{brake} \vee \texttt{changeLane}) && \text{from } [o_1] \text{ and } [r_1], && (2)\\
&[o_1, r_1, r_3, g_1]\texttt{changeLane} && \text{from (1) and (2)}, && (3)\\
&[r_4, g_2](\texttt{changeLane} \rightarrow \texttt{largeGapAvailable}) && \text{from } [r_4] \text{ and } [g_2], && (4)\\
&[o_1, r_1, r_3, r_4, g_1, g_2]\texttt{largeGapAvailable} && \text{from (3) and (4)}, && (5)\\
&[o_1, o_2, r_1, r_3, r_4, g_1, g_2]\bot && \text{from } [o_2] \text{ and (5)}. && (6)
\end{aligned}$$

Eventually, in (6) we derived that the instance $\{o_1, o_2, r_1, r_3, r_4, g_1, g_4\}$ has contradicting information. We are tempted to say that the entirety of its atoms o_1, o_2, r_1, r_3, r_4, g_1, g_4 provides a justification for a contradiction. Let us give in to temptation and look at the intermediate derivation steps. In line (1) we see that rule r_3 and goal g_1 justify not to brake, in line (3) we see that o_1, r_1, r_3, g_1 together justify to change lane. However, line (5) is a bit out of the ordinary as we usually would not accept rules and goals as a justification for an observation. For the present we leave it that way and take it as an indicator to introduce an additional order over the atoms later on.

2.6 Relation to Justification Logic

The reference to justifications in the previous section was given with full intent. Indeed, for each line (1)–(6) an analogous derivation is also possible in justification logic [5, 4] (with axiomatically appropriate constant specification). To see this, we briefly outline Artemov's justification logics.

Justification logics [4] are variants of epistemic modal logics where the modal operators of knowledge and belief are unfolded into justification terms. Hence, justification logics allow a complete realisation of Plato's characterisation of knowledge as justified true belief. A typical formula of justification logic has the form $s\colon\phi$, where s is a justification term built from justification constants, and it is read as "ϕ is justified by s". The basic justification logic J0 results from extending propositional logic consisting of **0** and **MP** by the application axiom and the sum axioms

$$\vdash s\colon(\phi \rightarrow \psi) \rightarrow t\colon\phi \rightarrow (s \cdot t)\colon\psi, \qquad \textbf{(Appl)}$$

$$\vdash s\colon\phi \rightarrow (s+t)\colon\phi \text{ and } \vdash s\colon\phi \rightarrow (t+s)\colon\phi, \qquad \textbf{(Sum)}$$

where s, t, $[s \cdot t]$, $[s+t]$, and $[t+s]$ are justification terms that are assembled from justification constants using the operators $+$ and $\cdot$ according to the axioms. Justification logics tie the epistemic tradition together with proof theory. Justification terms are reasonable abstractions for constructions of proofs. If s is a proof of $\phi \rightarrow \psi$ and t is a proof of ϕ, then the application axiom postulates that there is a common proof, namely $s \cdot t$, for ψ. Moreover, if we have a proof s for ϕ and some proof t, then the concatenations of both proofs, $s+t$ and $t+s$, are still proofs for ϕ. An important predecessor of justification logic is the logic of explicit proofs, *LP* [3], which from today's perspective turns out to be an extension of $J0$.

The key observation here is that **Appl** is a variant of the principle of information distribution **Dist**$'$. Hence, many promising features noted so far immediately carry over to K_S^D. However, there is a subtle

[4]The notation $\Sigma \vdash \phi$ means that there exists a finite subset $\{\sigma_1, \ldots, \sigma_m\} \subseteq \Sigma$ such that $\vdash (\sigma_1 \wedge \cdots \wedge \sigma_m) \rightarrow \phi$.

difference. While in $K^D_{\mathcal{S}}$ grouping of modalities is an idempotent, commutative, and associative operation, this is not the case for $\cdot$. The justification terms $s \cdot t$, $s \cdot (t \cdot t)$, and $t \cdot s$ are pairwise different. In $K^D_{\mathcal{S}}$ such a fine grained distinction is not possible. All justification terms noted before coincide to the instance $\{s,t\}$ in $K^D_{\mathcal{S}}$.

All substitution instances of classical logical tautologies, like $A \vee \neg A$ and $s{:}A \vee \neg s{:}A$, are provable in justification logics. But in contrast to modal logics, justification logics do not have a necessitation rule. The lack of the necessitation rule allows justification logics to break the principle of logical awareness, as $s{:}(A \vee \neg A)$ is not necessarily provable for an arbitrary justification term s. Certainly, restricting the principle of logical awareness is attractive to provide a realistic model of restricted logical resources. Since we are also interested in revealing and resolving contradictions, restricting the principle of logical awareness only would hide latent contradictions. E.g., there is no obvious way to derive a justification of a contradiction from $s{:}(A \wedge B)$ and $t{:}\neg A$ without using additional logical resources. Hence, we claim that the principle of logical awareness is indispensable to discover all contradictions that are logically possible.

Nevertheless, justification logic can simulate unrestricted logical awareness by adding proper axiom internalisation rules $\vdash e{:}\phi$ for all axioms ϕ and justification constants e. In such systems a weak variant of the necessitation rule of modal logic holds: for any derivation $\vdash \phi$ there exists a justification term t such that $\vdash t{:}\phi$ holds. Since ϕ was derived using axioms and rules only, also the justification term t is exclusively built from justification constants dedicated to the involved axioms. Beyond that, t is hardly informative as it does not help to reveal *extra-logical sources* of a contradiction.

3 Towards a Logic of Justifiability

Our analysis so far has shown that in $K^D_{\mathcal{S}}$ instances play the role of justifications. Or, to pronounce that an instance is the necessary quality of a justification, an information ϕ *is justifiable* if and only if *there exists an instance* that is consistent and has the information ϕ. Hence, the canonical next step is to introduce quantification over instances into our logic. We introduce a variable modal operators x and the quantifiers $\forall$ and $\exists$ into our logic and call this extension $QK^D_{\mathcal{S}}$. On the justification logic side, a similar approach has been explored in Fitting's paper [17], in which quantification over evidence is introduced for the logic *LP*, yielding the logic *QLP*. However, to the author's best knowledge it is still unknown whether *QLP* is decidable.

Definition 4. *Let $\mathcal{V}$ be an enumerable set of propositional variables and $\mathcal{S}$ be a finite set of atomic modalities. A formula ϕ of $QK^D_{\mathcal{S}}$ is any expression that satisfies the Backus-Naur form*

$$\phi ::= \bot \mid A \mid \phi \to \phi \mid [s]\phi \mid [x]\phi \mid \forall s \subseteq x \subseteq t(\phi) \mid \forall x \subseteq t(\phi),$$

where $A \in \mathcal{V}$, $s \subseteq \mathcal{S}$ and $t \subseteq \mathcal{S}$ are instances, and x is the variable modal operator. In formulae of the form $\forall s \subseteq x \subseteq t(\phi)$ we call s the lower bound and t the upper bound. The formula $\forall x \subseteq t(\phi)$ has only an upper bound. The unbounded quantification in $\forall x(\phi)$ is short for $\forall x \subseteq \mathcal{S}(\phi)$. As usual, $\exists$ denotes the dual of $\forall$.

The variable x is free in the formula ϕ if there is at least one occurrence of x in ϕ that is not within the scope of a quantifier. By $\phi[s/x]$ we denote the formula obtained by replacing all free occurrences of the variable modal operator x by the instance s. A closed formula is a formula where x is not free.

Note that the quantification in $QK^D_{\mathcal{S}}$ is very restrictive, as there is only one variable symbol. However, together with the fact that $\mathcal{S}$ is a finite set this allows for an easy transfer principle: Any closed formula of $QK^D_{\mathcal{S}}$ can be transferred to a formula of $K^D_{\mathcal{S}}$ by substituting $\exists s \subseteq x \subseteq t(\phi)$ with the finite disjunction $\bigvee_{s \subseteq r \subseteq t} \phi[r/x]$, and substituting $\forall s \subseteq x \subseteq t(\phi)$ with the finite conjunction $\bigwedge_{s \subseteq r \subseteq t} \phi[r/x]$. Hence, there is no need to provide an extension of the axiomatisation or the semantics. The soundness, completeness, and decidability results of $K_{\mathcal{S}}^{D}$ carry over to $QK_{\mathcal{S}}^{D}$ directly. Note that the complexity of a decision procedure might change due to the exponential blow up of the formula size by the translation principle given above. However, complexity considerations are beyond the introductory character of this paper.

In the following section, we will report a prototypical implementation of a solver accepting formulae of $QK_{\mathcal{S}}^{D}$. In particular, it avoids the naive explicit enumeration of the instances indicated above. Since the solver searches for a satisfying model of a given formula, we feed it with the negation of a formula and use that the unsatisfiability of the negated formula is equivalent to the provability of the formula under consideration.

3.1 A Tableau for $QK_{\mathcal{S}}^{D}$

We use a prefixed tableau prover [25, 9, 31] for $QK_{\mathcal{S}}^{D}$ and reduce the modal satisfiability problem to a Boolean satisfiability problem by using a similar incremental translation and solving method as has been proposed in [23]. Due to the usage of prefixes, a tableau proof consists of a single tree, see [16]. The tableau starts with a single branch that contains the input formula in negation normal form. The tableau rules are then applied to the branches of the tree until either all branches are closed, i.e., each branch contains a literal and its negation, or there exists at least one branch that remains open as all possible rules have been applied to the formulae of the branch. In the former case, the input formula is unsatisfiable. In the latter case, the input formula is satisfiable and the open branch provides a partial model of the input formula.

Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be a set of n atoms and $\mathbf{p} = (\mathbf{p}_{s_1}, \ldots \mathbf{p}_{s_n})$ a vector of n Boolean variables. Any instance $s \subseteq \mathcal{S}$ corresponds to a particular truth assignment of $\mathbf{p}$, i.e., p_{s_i} is assigned to true if and only if $s_i \in s$. We use Boolean formulae over the variables in $\mathbf{p}$ to constrain the set of feasible truth assignments. E.g., $\bigwedge_{s_i \in s} p_{s_i} \wedge \bigwedge_{s_i \notin s} \neg p_{s_i}$ uniquely determines the assignment corresponding to $s \subseteq \mathcal{S}$ as above, and the formula $\bigvee_{s_i \in \mathcal{S}} p_i$ encodes a nonempty subset of $\mathcal{S}$, i.e., an arbitrary instance of $\mathcal{S}$.

We extend the Boolean satisfiability problem to support several variable assignments. Each possible world ω is assigned with an individual Boolean vector $\mathbf{p} = \mathbf{p}(\omega)$, and Boolean formulae over $\mathbf{p}$ constrain the set of instances that have access to ω. E.g., consider the instance $s = \{s_1, s_n\}$. The formula $p_{s_1} \wedge p_{s_n}$ ensures that any feasible assignment for $\mathbf{p}$ correspond to a superset of s, as at least p_{s_1} and p_{s_n} have to be true under the given constraint. Formulae of this kind are denoted by $s \subseteq \mathbf{p}$ and used to ensure that at least s has access to ω (and possibly also some t with $s \subseteq t$). It follows that the solver only needs to maintain a single general accessibility relation $\mapsto$. The individual accessibility $\omega \mapsto_s \omega'$ is represented by $\omega \mapsto \omega'$ and $\omega' \models s \subseteq \mathbf{p}$. An immediate consequence is that $\omega \mapsto_s \omega'$ is represented if only if $\omega \mapsto_{s_i} \omega'$ is represented for all atoms s_i of s. Hence, the semantic compatibility of this approach is ensured.

In addition, the solver supports an assignment β for the variable modality x. That is, β is an assignment for the Boolean vector $\mathbf{x} = \{x_{s_1}, \ldots, x_{s_n}\}$. E.g., consider again the instance $s = \{s_1, s_n\}$. The formula $\neg x_{s_2} \wedge \neg x_{s_3} \wedge \cdots \wedge \neg x_{s_{n-1}}$ ensures that any assignment of $\mathbf{x}$ corresponds to a subset of s, as only x_{s_1} and x_{s_n} can be true under this assignment. By $\mathbf{x} \subseteq s$ we denote the Boolean formula as before together with the additional constraint that β may not correspond to the empty set. We use $\mathbf{x} \subseteq s$ to encode the upper bound restriction of an existential quantifier.

The prefixes in this tableau consists of pairs (ω, β), where ω denotes a possible world and β denotes an assignment for the variable x. We write $(\omega, \beta) \models \phi$ to denote that (ω, β) is the prefix of ϕ. The tableau makes use of an explicit representation of the accessibility relation, i.e., it maintains a map $\mapsto$ that allows it to look up all possible worlds ω' that are reachable from ω. We write $\omega \mapsto \omega'$ to denote that ω' is reachable from ω.

Note that, for the clarity of presentation, we only give the rules for upper bounded quantifiers.

- The $\wedge$ and $\vee$-rules are classical rules for a Boolean tableau.

$$(\wedge)\ \frac{(\omega,\beta)\models\phi\wedge\psi}{\begin{array}{c}(\omega,\beta)\models\phi\\(\omega,\beta)\models\psi\end{array}} \quad\text{and}\quad (\vee)\ \frac{(\omega,\beta)\models\phi\vee\psi}{(\omega,\beta)\models\phi \quad\mid\quad (\omega,\beta)\models\psi}.$$

- The following $\langle s\rangle$- and $[s]$-rules are adapted from [23]. While the $\langle s\rangle$-rule is transferred to the tableau directly, the $[s]$-rule is extended to supports the principle of information distribution. This extension was found in [9, Ch. 2, pp. 122–123].

  The $\langle s\rangle$-rule introduces a new possible world ω' and adds $\omega\mapsto\omega'$ to the accessibility relation. The constraint $s\subseteq\mathbf{p}$ stipulates that at least s has access to ω'. The $[s]$-rule takes a formula $[s]\phi$ and an accessible world ω' and adds the prefixed formula $(\omega',\beta)\models s\subseteq\mathbf{p}\rightarrow\phi$ to the current branch. Hence, whenever s has access to ω' then ϕ has to hold.

$$(\langle s\rangle)\ \frac{(\omega,\beta)\models\langle s\rangle\phi}{\begin{array}{c}(\omega',\beta)\models s\subseteq\mathbf{p}\wedge\phi\\ \omega\mapsto\omega',\quad \omega'\text{ new}\end{array}} \quad\text{and}\quad ([s])\ \frac{(\omega,\beta)\models[s]\phi,\quad \omega\mapsto\omega'}{(\omega',\beta)\models s\subseteq\mathbf{p}\rightarrow\phi}.$$

- The following $\langle x\rangle$- and $[x]$-rules for variable modal operators are similar to their variants for constant modal operators but depend on the assignment β for $\mathbf{x}$. Readers may convince themselves that the constraint $\mathbf{x}\subseteq\mathbf{p}$ can actually be encoded into a Boolean satisfiability solver using the encoding scheme as indicated above.

$$(\langle x\rangle)\ \frac{(\omega,\beta)\models\langle x\rangle\phi}{\begin{array}{c}(\omega',\beta)\models \mathbf{x}\subseteq\mathbf{p}\wedge\phi\\ \omega\mapsto\omega',\quad \omega'\text{ new}\end{array}} \quad\text{and}\quad ([x])\ \frac{(\omega,\beta)\models[x]\phi,\quad \omega\mapsto\omega'}{(\omega',\beta)\models \mathbf{x}\subseteq\mathbf{p}\rightarrow\phi}.$$

- The $\exists$-rule introduces a new valuation β' for $\mathbf{x}$ and puts the constraint $\mathbf{x}\subseteq s$ on this valuation. The $\forall_0$-rule generates a substitution instance where the variable is substituted by its upper bound.

$$(\exists)\ \frac{(\omega,\beta)\models\exists x\subseteq s(\phi)}{\begin{array}{c}(\omega,\beta')\models \mathbf{x}\subseteq s\wedge\phi\\ \beta'\text{ new}\end{array}} \quad\text{and}\quad (\forall_0)\ \frac{(\omega,\beta)\models\forall x\subseteq s(\phi)}{(\omega,\beta)\models\phi[s/x]}.$$

  Finally, the $\forall_s$-rule generates a constrained substitution instance $\phi[t/x]$ whenever there occurs another formula of the form $\langle t\rangle\psi$ or $[t]\psi$ that has the same ω-prefix in the tableau. The constraint $t\subseteq s$ ensures that $\phi[t/x]$ has to hold at least if $t\subseteq s$. For the implementation, this constraint can be checked before application of the rule. Similarly, the $\forall_x$-rule takes the valuation β' of any variable modal operator that has the same ω-prefix and adds the constrained formula $\mathbf{x}\subseteq s\rightarrow\phi$. That is, if β' assigns x to some subinstance $t\subseteq s$, then ϕ has to hold, and any free occurrence of x in ϕ is now interpreted as t by β'.

$$(\forall_s)\ \frac{(\omega,\beta)\models\forall x\subseteq s(\phi),\quad (\omega,\beta')\models\langle t\rangle\psi \text{ or } (\omega,\beta')\models[t]\psi}{(\omega,\beta')\models t\subseteq s\rightarrow\phi[t/x]} \quad\text{and}$$

$$(\forall_x)\ \frac{(\omega,\beta) \models \forall x \subseteq s(\phi), \quad (\omega,\beta') \models \langle x \rangle \psi \text{ or } (\omega,\beta') \models [x]\psi}{(\omega,\beta') \models \mathbf{x} \subseteq s \to \phi}.$$

It remains to argue that this tableau is a sound and complete decision procedure for QK_S^D. We use that the rules for $\wedge$, $\vee$, $\langle s \rangle$ and $[s]$ provide a decision procedure for K_S^D, see [9, Ch. 2]. While the existential fragment of this logic is quite directly translated into a Boolean satisfiability problem and not further discussed here, we show that the rules for the universal quantifier are complete for QK_S^D. Clearly, a $\forall$-rule that produced all possible substitution instances would yield a complete decision procedure. However, the $\forall_0$-rule ensures that at least the substitution instance of the upper bound is produced. All other substitution instances are produced only when a corresponding instance occurs in the branch with the same ω-prefix. This approach is sufficient. Assume we found a pointed model (M,ω) for some formula ϕ where s is a least instance occurring in ϕ. Then for any instances $t \subseteq s$ we can substitute t for s in ϕ and construct a pointed model (M',ω) for $\phi[t/s]$ by setting $\omega \mapsto_t \omega'$ in the new model if and only $\omega \mapsto_s \omega'$ in the original model. But this means that either there is a specific instances that contradicts a bounded $\forall$-statement or it is sufficient to show the satisfiability of the substitution instance for the upper bounding instance.

The solver is implemented in C++ and uses minisat to solve the Boolean satisfiability problem. An early prototype is available at `https://vhome.offis.de/~willemh/episat/`. The site will be updated regularly.

3.2 Justifiability in Practice

Let us use quantified expression to analyse our running example with the help of the solver. Note that these examples are also distributed with the prototype. Due to nondeterminism the concrete partial model may change in the satisfiable case.

1. The formula $\exists x([x]\bot)$ holds if and only if there is some instance that has inconsistent information. As we already have seen, the automated driving system of the vehicle has inconsistent information and we can expect that $\Sigma \to \exists x([x]\bot)$ is derivable. Indeed, as we already have shown above, it suffices to choose $x = [o_1, o_2, r_1, r_3, r_4, g_1, g_2]$ to see that the existential formula holds. We use the solver and verify that the negation $\Sigma \wedge \forall x(\langle x \rangle \top)$ is unsatisfiable. Indeed, the solver cannot find a model and returns the unsatisfiability result. Notice the difference between the derivation in a proof calculus as given above and the satisfiability solver. While we could directly read off the instance having inconsistent information from the derivation, the solver provides us no hint in case of unsatisfiability.

2. Is there at least some instance that has consistent information? I.e., can we satisfy $\Sigma \wedge \exists x(\langle x \rangle \top)$? In this case, the solver returns a satisfying model consisting of two possible worlds ω_0 and ω_1, where ω_1 is accessible for o_1 from ω_0. Further, it returns the partial truth assignment $\omega_1 \models \{\neg\mathtt{beConsiderate}, \neg\mathtt{beFast}, \mathtt{laneBlocked}\}$ and the valuation $x = o_1$. This is indeed a model for the formula, as there is only one world ω_1 accessible from the actual world ω_0, and only o_1 can access it. Hence, $\langle o_1 \rangle \top$ holds. Moreover, also the subformula $[o_1]\mathtt{laneBlocked}$ of Σ is satisfied, as `laneBlocked` holds in any o_1-accessible possible world which precisely is ω_1. All other subformulae of Σ are trivially satisfied by the semantics of the modal operator $[.]$. As this is an obvious but somewhat uninformative result, we turn to the next analysis.

3. Do the rules have inconsistent information? Can we satisfy $\Sigma \wedge \forall x \subseteq \{r_1, r_2, r_3, r_4\}(\langle x \rangle \top)$? The solver returns a satisfying model that consists of two possible worlds ω_0 and ω_1 with

$\omega_0 \mapsto_{\{r_1,r_2,r_3,r_4\}} \omega_1$ and the partial truth assignment $\omega_1 \models \{\neg\texttt{changeLane}, \neg\texttt{largeGapAvailable}, \neg\texttt{laneBlocked}, \neg\texttt{brake}, \neg\texttt{beFast}, \neg\texttt{beConsiderate}, \neg\texttt{gapAvailable}\}$. To convince ourselves that this is again a model, we have to take the model properties into account: As ω_1 is accessible for $\{r_1,r_2,r_3,r_4\}$, it is also accessible for all instances $s \subseteq \{r_1,r_2,r_3,r_4\}$.

4. The previous inquiries were mainly motivated to give the reader an overview of the possibilities of the solver and enable him to interpret the output models. Let us now turn to a more interesting inquiry that analyses the role of the user specified goals. Can we find a model where an instance exists that has consistent information and includes all rules and observations? Can we satisfy $\Sigma \wedge \exists \{r_1,r_2,r_3,r_4,o_1,o_2\} \subseteq x \subseteq \mathcal{S}(\langle x \rangle \top)$?
   The solver returns the model $\omega_0 \mapsto_{\{r_1,r_2,r_3,r_4,o_1,o_2\}} \omega_1$ with the truth assignment $\omega_1 \models \{\texttt{changeLane}, \texttt{laneBlocked}, \neg\texttt{largeGapAvailable}, \neg\texttt{beConsiderate}, \texttt{gapAvailable}, \neg\texttt{beFast}, \neg\texttt{brake}\}$ and the valuation $x = \{r_1,r_2,r_3,r_4,o_1,o_2\}$. This is a quite convincing result. If the automated driving system would have given up the user specified goals, it could have found a solution where it restrained from braking and changed the lane.

5. We ask whether we can find an instance x that includes the rules, has consistent information and considers braking a possible manoeuvre. Is it possbile to satisfy $\Sigma \wedge \exists \{r_1,r_2,r_3,r_4\} \subseteq x \subseteq \mathcal{S}$ $(\langle x \rangle \texttt{brake} \wedge [x](\top))$?
   The solver returns the model $\omega_0 \mapsto_{\{r_1,r_2,r_3,r_4,o_1,o_2,g_2\}} \omega_1$ with the truth assignment $\omega_1 \models \{\neg\texttt{beFast}, \texttt{beConsiderate}, \neg\texttt{changeLane}, \texttt{laneBlocked}, \texttt{gapAvailable}, \neg\texttt{largeGapAvailable}, \texttt{brake}\}$ and $x = \{r_1,r_2,r_3,r_4,o_1,o_2,g_2\}$. That is, roughly speaking, if braking would be possible, than $\{r_1,r_2,r_3,r_4,o_1,o_2,g_2\}$ can be made a consistent instance with the information as given for ω_1.

4 Conclusion

I presented the logic $QK^D_{\mathcal{S}}$ that has strong connections to justification logic. In contrast to justification logic, $QK^D_{\mathcal{S}}$ supports for the quantification over a variable modality. In combination with the proposed tableau-based satisfiability solver, $QK^D_{\mathcal{S}}$ can be used to identify information components justifying given statements. Moreover, I reported on a prototypical implementation of the satisfiability solver accompanied by some exemplary inquiries along our running example.

An extensive comparison with other solvers remains as future work. To the best of the author's knowledge this is the first proposal of a decidable theory supporting quantification over modalities. Certainly, there is a lack of comparable solvers and benchmarks. An obvious approach would be to use the translation of bounded quantification to finite dis- and conjunctions as indicated above and compare against solvers for K^D_n. Unfortunately, such solvers are rare. Neither InKreSAT [23], nor solvers for description logic like Racer [20] or FaCT++ [35] support the principle of information distribution, as the intersection of roles is not part of the description logics $\mathcal{SHOIQ}$ or $\mathcal{SRIQ}(\mathcal{D})$. The only promising candidate known to the author is the first-order logic theorem prover SPASS that fully supports Boolean modal logic of which K^D_n is a fragment [37]. As the tableau for the multi-modal fragment K_n without the principle of information distribution follows the same rules as InKreSAT, a detailed comparison of this fragment has been postponed so far, as we do not expect any interesting insights of such a comparison beside of potentially revealing weak points of the prototypical implementation.

Another important aspect is to investigate the theoretical complexity of the satisfiability problem for $QK^D_{\mathcal{S}}$ which is at least PSPACE-hard [14], and the complexity of the presented tableau calculus.

The current focus of development is mainly on the extension of this logic. The prototype of the solver already supports a partial order of the atoms in $\mathcal{S}$. Such a partial order implies that the set of instances has the interesting structure of a semilattice presentation [8, 32, 33]. That is, any instance is either irreducible or it can be written as a finite join of entities. However, the semilattice may contain entities that are irreducible but not atomic. Nonatomic entities are of special interest. They inherit all information of smaller instances but allow for the emergence of novel information, e.g., by quantifying over their instances. This gives rise to powerful nonmonotonic capability as entities can mimic reasoning by default, autoepistemic reasoning or counterfactual reasoning [24]. More details are planned for a follow-up paper, until then readers may convince themselves that the partial order of the atoms can simply be added to the Boolean formulae encoding instances.

Any expectation that $QK_{\mathcal{S}}^{D}$ could be a logic that resolves inconsistencies fully automatically has to remain unfulfilled. While this logic provides a rigorous framework in which information atoms can be held and maintained independently, it cannot decide to which atoms should be given preference over others. Nevertheless, $QK_{\mathcal{S}}^{D}$ gives us raise to a weak notion of belief. We may say that a proposition ϕ is believed in a system $\mathcal{S}$ if and only if ϕ is justifiable and $\neg\phi$ is not justifiable, i.e., if and only if $\exists x(\langle x\rangle\top \wedge [x]\phi) \wedge \forall x(\langle x\rangle\top \rightarrow \langle x\rangle\phi)$ holds. To put this definition in legal terms, ϕ is believed if and only if there is a noncontradictory group of witnesses for ϕ, and all other groups of witnesses either contradict themselves or have no argument against ϕ. This concept of belief thus is a reasonable selection of appropriate justifications, and remains consistent in itself. A further observation is that this kind of belief is no longer closed under logical consequence and, hence, is not a normal modality.

Another important extension in this regard is to allow positive or negative introspection axioms [14] to $QK_{\mathcal{S}}^{D}$, i.e., for any ϕ and $s \in \mathcal{S}$

$$\vdash [s]\phi \rightarrow [s][s]\phi \text{ and} \tag{4}$$

$$\vdash \langle s\rangle\phi \rightarrow [s]\langle s\rangle\phi. \tag{5}$$

This would yield the logic $QK_{\mathcal{S}}^{D}45$ of instances with additional epistemic power. In contrast, we are less interested in the extensions

$$\vdash \langle s\rangle\top \text{ or} \tag{D}$$

$$\vdash [s]\phi \rightarrow \phi, \tag{T}$$

as we consider these axioms as too powerful. Logics where at least one of these principles holds do not allow for contradicting instances and are therefore of lesser interest for the analysis of justifiability.

A temporal expansion to better understand the dynamic aspects of information distribution within systems would be very attractive. The interaction of information and time is certainly an appealing epistemic challenge here.

A profound philosophical reflection of the proposed approach would require an intensive examination of Hempel and Oppenheim's concept of a DN-explanation [22], Toulmin's concept of an argument [34], and also the fallibilist position of Popper's critical rationalism [29], just to name a few.

Acknowledgements I would like to thank the anonymous referees for their valuable comments and carefully proofreading my work. I hope that their common desire for better motivation and reflection of related work has been met.

References

[1] Rajeev Alur, Costas Courcoubetis, Nicolas Halbwachs, Thomas A. Henzinger, Pei-Hsin Ho, Xavier Nicollin, Alfredo Olivero, Joseph Sifakis & Sergio Yovine (1995): *The algorithmic analysis of hybrid systems*. Theoretical Computer Science 138(1), pp. 3–34, doi:10.1016/0304-3975(94)00202-T.

[2] Gleifer Vaz Alves, Louise Dennis & Michael Fisher (2021): *A Double-Level Model Checking Approach for an Agent-Based Autonomous Vehicle and Road Junction Regulations*. Journal of Sensor and Actuator Networks 10(3), doi:10.3390/jsan10030041.

[3] Sergei N Artemov (2001): *Explicit provability and constructive semantics*. Bulletin of Symbolic logic 7(1), pp. 1–36, doi:10.2307/2687821.

[4] Sergei N. Artemov (2006): *Justified common knowledge*. Theoretical Computer Science 357(1-3), pp. 4–22, doi:10.1016/j.tcs.2006.03.009.

[5] Sergei N. Artemov (2008): *The Logic of Justification*. The Review of Symbolic Logic 1(4), pp. 477–513, doi:10.1017/S1755020308090060.

[6] Christel Baier, Clemens Dubslaff, Florian Funke, Simon Jantsch, Rupak Majumdar, Jakob Piribauer & Robin Ziemek (2021): *From Verification to Causality-based Explications*. arXiv preprint arXiv:2105.09533.

[7] Jon Barwise (1989): *The situation in logic*. CSLI Lecture notes 17, Center for the Study of Language and Information (CSLI), Stanford.

[8] G. Bińczak, A.B. Romanowska & J.D.H. Smith (2007): *Poset extensions, convex sets, and semilattice presentations*. Discrete Mathematics 307(1), pp. 1–11, doi:10.1016/j.disc.2006.09.021.

[9] Patrick Blackburn, Johan van Benthem & Frank Wolter (2006): *Handbook of modal logic*. Elsevier.

[10] Edmund M. Clarke Jr., Orna Grumberg, Daniel Kroening, Doron Peled & Helmut Veith (2018): *Model checking*, 2nd edition. Cyper Physical Systems Series, MIT press.

[11] Werner Damm, Henning Dierks, Stefan Disch, Willem Hagemann, Florian Pigorsch, Christoph Scholl, Uwe Waldmann & Boris Wirtz (2012): *Exact and fully symbolic verification of linear hybrid automata with large discrete state spaces*. Science of Computer Programming 77(10-11), pp. 1122–1150, doi:10.1016/j.scico.2011.07.006.

[12] DARPA: *Explainable Artificial Intelligence (XAI) Program*. Available at `https://www.darpa.mil/program/explainable-artificial-intelligence`. Last accessed on 06/10/2021.

[13] Lasse Dissing & Thomas Bolander (2020): *Implementing Theory of Mind on a Robot Using Dynamic Epistemic Logic*. In Christian Bessiere, editor: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, International Joint Conferences on Artificial Intelligence Organization, pp. 1615–1621, doi:10.24963/ijcai.2020/224.

[14] Ronald Fagin, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi (2003): *Reasoning About Knowledge*. MIT Press, Cambridge, MA, USA.

[15] Michael Fisher, Louise Dennis & Matt Webster (2013): *Verifying autonomous systems*. Communications of the ACM 56(9), pp. 84–93, doi:10.1145/2494558.

[16] Melvin Fitting (1983): *Proof methods for modal and intuitionistic logics*. Synthese Library 169, Springer Science & Business Media, doi:10.1007/978-94-017-2794-5.

[17] Melvin Fitting (2008): *A quantified logic of evidence*. Annals of Pure and Applied Logic 152(1), pp. 67–83, doi:10.1016/j.apal.2007.11.003.

[18] Goran Frehse, Colas Le Guernic, Alexandre Donzé, Scott Cotton, Rajarshi Ray, Olivier Lebeltel, Rodolfo Ripado, Antoine Girard, Thao Dang & Oded Maler (2011): *SpaceEx: Scalable Verification of Hybrid Systems*. In Ganesh Gopalakrishnan & Shaz Qadeer, editors: Computer Aided Verification, Lecture Notes in Computer Science 6806, Springer, Heidelberg, pp. 379–395, doi:10.1007/978-3-642-22110-1_30.

[19] Jelle Gerbrandy (1998): *Distributed knowledge*. In: Twendial 1998: Formal Semantics and Pragmatics of Dialogue, 98, pp. 111–124.

[20] Volker Haarslev, Kay Hidde, Ralf Möller & Michael Wessel (2012): *The RacerPro knowledge representation and reasoning system*. Semantic Web Journal 3(3), pp. 267–277, doi:10.3233/SW-2011-0032.

[21] Joseph Y. Halpern & Yoram Moses (1985): *A guide to the modal logics of knowledge and belief: preliminary draft*. In: Proceedings of the 9th international joint conference on Artificial Intelligence – Volume 1, Morgan Kaufmann Publishers Inc., pp. 480–490.

[22] Carl G. Hempel & Paul Oppenheim (1948): *Studies in the Logic of Explanation*. Philosophy of science 15(2), pp. 135–175, doi:10.1086/286983.

[23] Mark Kaminski & Tobias Tebbi (2013): *InKreSAT: modal reasoning via incremental reduction to SAT*. In Maria Paola Bonacina, editor: Automated Deduction – CADE-24, Springer, pp. 436–442, doi:10.1007/978-3-642-38574-2_31.

[24] David Lewis (1973): *Counterfactuals*, reissued 2001 edition. Blackwell.

[25] Fabio Massacci (2000): *Single step tableaux for modal logics*. Journal of Automated Reasoning 24(3), pp. 319–364, doi:10.1023/A:1006155811656.

[26] On-Road Automated Driving (ORAD) committee: *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, 2021-04-30 edition. SAE International, doi:10.4271/J3016_202104.

[27] Foundations of Perspicuous Software Systems: *Center for Perspicuous Computing — Research*. Available at `https://www.perspicuous-computing.science/research/`. Last accessed on 06/10/2021.

[28] André Platzer (2019): *The logical path to autonomous cyber-physical systems*. In: International Conference on Quantitative Evaluation of Systems, Springer, pp. 25–33, doi:10.1007/978-3-030-30281-8_2.

[29] Karl Popper (2002): *The logic of scientific discovery*, 2nd edition. Routledge, doi:10.4324/9780203994627.

[30] Anand S. Rao & Michael P. Georgeff (1995): *BDI agents: From theory to practice.* In: Proceedings of the First International Conference on Multiagent Systems (ICMAS), pp. 312–319.

[31] Roberto Sebastiani & Michele Vescovi (2009): *Automated reasoning in modal and description logics via SAT encoding: the case study of K(m)/ALC-satisfiability*. Journal of Artificial Intelligence Research 35, pp. 343–389, doi:10.1613/jair.2675.

[32] Pedro V. Silva (1993): *On the semilattice of idempotents of a free inverse monoid*. Proceedings of the Edinburgh Mathematical Society 36(2), p. 349–360, doi:10.1017/S0013091500018447.

[33] Pedro V. Silva (2000): *On unique factorization semilattices*. Discussiones Mathematicae-General Algebra and Applications 20(1), pp. 97–120, doi:10.7151/dmgaa.1009.

[34] Stephen Edelston Toulmin (2003): *The Uses of Argument.* Updated ed, Cambridge University Press, doi:10.1017/CBO9780511840005.

[35] Dmitry Tsarkov & Ian Horrocks (2006): *FaCT++ Description Logic Reasoner: System Description*. In Ulrich Furbach & Natarajan Shankar, editors: Automated Reasoning, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 292–297, doi:10.1007/11814771_26.

[36] Hans Van Ditmarsch, Wiebe van Der Hoek & Barteld Kooi (2007): *Dynamic epistemic logic*. 337, Springer Science & Business Media.

[37] Christoph Weidenbach, Renate A. Schmidt, Thomas Hillenbrand, Rostislav Rusev & Dalibor Topic (2007): *System Description: Spass Version 3.0*. In Frank Pfenning, editor: Automated Deduction – CADE-21, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 514–520, doi:10.1007/978-3-540-73595-3_38.