When Laws Collide
Automated Reasoning About Conflicting Obligations in Multi-Jurisdictional Environments

Adam Pease[1] and Pantelimon Stănică[2]

[1]Department of Computer Science, [2]Department of Applied Mathematics
Naval Postgraduate School, Monterey, CA 93943, USA
{adam.pease, pstanica}@nps.edu

Abstract

An agent subject to several authorities may receive incompatible obligations while lacking knowledge of which authorities apply. We give a Kripke model that combines knowledge, belief, authority activation, obligations, and weak permission, together with a finite priority policy for selecting effective requirements. Issued obligations and effective obligations are represented separately, so selecting a requirement does not erase the conflict that prompted the selection. For a binary decision, the policy examines the maximal authorities that issue either of the two opposing requirements. Agreement among those authorities determines the effective requirement, whereas disagreement is recorded as requiring external mediation. We give an explicit axiomatization and prove soundness and strong completeness by canonical-model construction and definitional elimination. We also prove that the resolution predicates form a conservative definitional extension of the underlying model, that binary resolution never selects both alternatives, and that selection takes quadratic time once the relevant obligations and priorities are supplied. We distinguish this selection cost from the cost of detecting logical incompatibility and from finite-model checking. A typed first-order encoding makes the correspondence between the model and the implementation explicit. A simplified naval example illustrates the treatment of uncertain applicability, priority, and incomparable authorities. Vampire proves a twelve-part conjecture for the finite policy instance in about 175 seconds, and an explicit satisfying model independently checks its consistency. Analysis of the separate scenario input distinguishes an assumed conclusion from a derived one and shows that its unresolved-conflict query is not determined by the premises. Z3 verifies a grounded finite countermodel to that query.

1 Introduction

Individuals and organizations can be subject to several authorities whose requirements cannot all be met. An additional difficulty arises when the information needed to determine applicability is incomplete. A commander may know that a threat is present without knowing whether a vessel is in territorial waters. An organization may know which operation it intends to perform without knowing every jurisdiction in which that operation has legal consequences. Reasoning about the requirements therefore involves both normative information and information about the circumstances in which the norms apply.

We study a finite collection of authorities equipped with a priority order. The order is part of the policy supplied to the reasoner. It may vary between possible worlds, but the logic does not determine which legal or institutional source should have priority. Our motivating example uses three named sources and a small set of stipulated requirements. It illustrates a formal decision problem rather than a legal interpretation of maritime law.

The reasoner must determine which authorities issue applicable requirements at the current world, which requirements survive the chosen priority policy, and what the agent knows or believes about these conclusions. In particular, an issued requirement can remain a fact about an authority even when it does not become effective. If one authority requires p and a higher authority requires $\neg p$, selecting $\neg p$ does not justify denying

that the first authority issued its requirement. This distinction permits the system to retain an explanation of the conflict and the priority decision.

We use ordinary Kripke semantics for knowledge, belief, and an authority's normative standards. Applicability guards the issued-obligation predicate. The guard prevents an inactive authority from issuing obligations through vacuous truth. Priority resolution is then defined over a finite set of requests. For a binary topic, the relevant authorities are those issuing either p or $\neg p$. The policy retains the maximal authorities in this set. Several maximal authorities may agree on the same requirement. If maximal authorities disagree, the policy returns an unresolved status rather than choosing an authority arbitrarily.

The contributions are a precise separation of issued and effective obligations, a declarative priority policy with a corresponding algorithm, and a full axiomatization with proofs of soundness, consistency, and strong completeness. The canonical construction verifies the frame properties and the truth lemma without modifying issued requirements during resolution. Separate results establish the computational properties of selection and finite-model evaluation. We also describe a typed first-order translation that separates accessibility clauses from the definitions of conflict and effectiveness. The computational results concern explicit finite models and supplied obligation data. They do not assert a complexity classification for unrestricted first-order reasoning or use a small benchmark to infer a general satisfiability bound.

Section 2 presents the naval example, and Section 3 places the construction in relation to modal and deontic reasoning. Section 4 defines the language, Kripke model, and priority policy. Section 5 gives the full axiomatization and inference rules. Section 6 constructs the canonical model, proves consistency and completeness, and establishes selection and finite-model checking results. Section 7 describes the first-order encoding and the benchmark queries. Section 8 discusses extensions, including permissions and integration with the Suggested Upper Merged Ontology (SUMO) [12, 14], and Section 9 concludes.

2 Motivating example in naval operations

Example 1. A commander encounters a suspicious vessel near a strategic asset. The commander knows that a threat has been detected and that the asset is nearby. Territorial location is uncertain to the commander, even though it has a definite truth value at the actual world.

The example has three authority labels, UNCLOS, ROE, and NSD. We stipulate that UNCLOS applies when the vessel is in territorial waters and requires a warning-first response. ROE applies throughout the example and, when a threat is detected, requires an immediate response without a preceding warning. NSD applies when a strategic asset is nearby and requires protection of that asset.

The action labels denote choices at a single decision point. A warning is an action in ordinary language. Here, however, `issue_warning` denotes choosing to warn before intervention, while `immediate_action` denotes choosing to intervene without first warning. The two choices are stipulated to be incompatible. Neither label describes the whole subsequent sequence of operations.

The supplied policy gives ROE and NSD priority over UNCLOS and leaves ROE and NSD incomparable. Protecting the asset is not stipulated to conflict with immediate intervention. Thus incomparability alone does not create a conflict.

These requirements and priorities are assumptions of the example. They are not claims that UNCLOS prescribes this response or that military rules generally override international law. The reasoner determines consequences of the supplied policy, and the suitability of that policy remains a separate domain question.

Let p denote the immediate response and q the warning-first response. The scenario background contains $\neg(p \wedge q)$. For binary resolution of the intervention topic, an explicit policy can classify q as the alternative $\neg p$. This classification treats the two labels as the complete response choices for that topic, excluding a third option such as postponing both responses. Mere mutual exclusion, by itself, would not imply that q and $\neg p$ are equivalent. Protection of the asset belongs to a separate topic.

The implementation discussion examines four queries. Q1 asks whether the two response requirements conflict. Q2 asks whether immediate action is effective. Q3 asks whether the commander knows the proposition `issue_warning`. Q4 asks whether no unresolved conflict exists. Q3 must be read literally. Knowledge that a warning is issued as a requirement would instead be expressed by $K_a\mathsf{Issued}_{\mathrm{UNCLOS}}(q)$, and knowledge that it is effective would be $K_a\mathsf{Effective}(q)$. These formulas have different meanings from knowledge of the action proposition q itself.

3 Related work

Deontic logic studies obligations and their relation to agency and choice. Horty's account of agency and deontic logic [10] provides a broader setting for such questions. Legal reasoning also concerns the interpretation and interaction of norms [15], while Berman [4] examines legal pluralism across jurisdictions. The latter is a jurisprudential account rather than an automated reasoning system. These works explain the roles of agency, norm interpretation, and overlapping jurisdictions in the problem considered here.

Defeasible deontic reasoning distinguishes conclusions supported by norms from conclusions that survive priority comparisons. Governatori et al. [9] study strong and weak permissions in defeasible logic. Our construction shares the need to distinguish a normative source from the outcome of a priority policy. The policy developed here is more restricted than a general defeasible calculus. It selects between two opposing requirements using maximal authorities and records disagreement when those authorities are incomparable.

Kripke models and the correspondence between modal principles and accessibility conditions are standard [5]. Knowledge and belief permit us to represent uncertainty about authority applicability without making the applicability condition itself indeterminate at a world. The knowledge-to-belief principle requires a relation-inclusion condition, which we state explicitly below.

Benzmüller and Paulson [1] give embeddings of quantified multimodal logics in simple type theory. Their work illustrates how modal operators can be interpreted through quantification over worlds. We use the same underlying semantic idea in a typed first-order setting, translating a fixed modal formula recursively rather than quantifying over arbitrary formulas. TPTP supplies the input language [17], and Vampire supplies first-order proof search [11]. A separate encoding with named proposition objects supports the finite scenario but needs explicit interpretation clauses for the propositions it uses.

Our aim is to make the relation between applicability, normative standards, and priority selection precise enough to check in a theorem prover. The axiomatization includes weak permission with a defined dual semantics, while conflict selection concerns issued obligations. The results below establish the semantic and computational properties of this policy within the stated epistemic and deontic model.

4 Formal logical system

4.1 Language and Kripke semantics

Let $\mathcal{A}$ and $\mathcal{AU}$ be finite nonempty sets of agents and authorities, and let $\mathcal{P}$ be a set of propositional atoms. Formulas are built from atoms, Boolean connectives, and the modalities K_a, B_a, and D_h, where $a \in \mathcal{A}$ and $h \in \mathcal{AU}$. We also have applicability atoms Active_h and priority atoms $\mathsf{Less}_{h,k}$. The letter D_h denotes an authority's normative-standard modality. Issued obligations are defined using D_h and applicability, rather than treated as unguarded normal modalities.

Definition 2. *A model is a tuple*

$$\mathcal{M} = \langle W, (R_a^K)_{a\in\mathcal{A}}, (R_a^B)_{a\in\mathcal{A}}, (R_h^D)_{h\in\mathcal{AU}}, (\prec_w)_{w\in W}, \mathcal{I}, V\rangle.$$

Here W is a nonempty set of worlds, R_a^K is an equivalence relation on W, and R_a^B is serial, transitive, and Euclidean. We require $R_a^B \subseteq R_a^K$. Each R_h^D is serial on W. For every w, $\prec_w$ is a strict partial order on $\mathcal{AU}$, and $\mathcal{I}(w,h) \in \{0,1\}$ specifies applicability. The valuation V assigns truth values to propositional atoms at worlds.

A relation R is serial if every world has an R-successor. It is Euclidean if wRu and wRv imply uRv. A strict partial order is irreflexive and transitive. We read $h \prec_w k$ as saying that k has higher priority than h at w. The corresponding non-strict order is $h \preceq_w k$ if $h = k$ or $h \prec_w k$. Distinct authorities are incomparable when neither strict comparison holds.

Boolean truth conditions are classical. For $X \in \{K, B, D\}$ and its appropriate index i, the modal clause is

$$\mathcal{M}, w \models X_i\varphi \quad\Longleftrightarrow\quad \text{for every } u \text{ with } wR_i^X u, \quad \mathcal{M}, u \models \varphi.$$

Applicability and priority atoms satisfy

$$\mathcal{M}, w \models \mathsf{Active}_h \Longleftrightarrow \mathcal{I}(w,h) = 1, \qquad \mathcal{M}, w \models \mathsf{Less}_{h,k} \Longleftrightarrow h \prec_w k.$$

The world is an evaluation parameter in these clauses, not an object-language variable. A conditional priority can therefore be written as $\chi \to \mathsf{Less}_{h,k}$, where χ is an ordinary formula about the circumstances. No nominal or hybrid-logic operator is needed.

Definition 3. *For an authority h, the issued applicable obligation with content φ is the abbreviation*

$$\mathsf{Issued}_h(\varphi) := \mathsf{Active}_h \wedge D_h\varphi.$$

We also write $O_h(\varphi)$ for this same abbreviation.

The relation R_h^D describes the normative alternatives associated with h, including at worlds where h is inactive. Keeping it serial avoids vacuous normative standards. The applicability guard makes every $\mathsf{Issued}_h(\varphi)$ false when h is inactive. Activity is independent of priority unless an application adds a condition connecting them. In particular, a higher-priority authority need not apply merely because a lower-priority one does.

The normative modality is closed under logical consequence. It is therefore an idealized representation of an authority's requirements, not a literal list of its utterances. Priority selection is restricted to a supplied finite set of requests, so it does not attempt to enumerate every consequence of every normative standard.

4.2 Conflict and a finite request vocabulary

Definition 4. *For a formula φ and distinct authorities h, k, define a direct issued conflict by*

$$\mathsf{Conflict}_{h,k}(\varphi) := \mathsf{Issued}_h(\varphi) \wedge \mathsf{Issued}_k(\neg\varphi).$$

The unindexed abbreviation $\mathsf{Conflict}(\varphi)$ is the finite disjunction of these formulas over distinct $h, k \in \mathcal{AU}$.

A direct issued conflict records what the authorities require. It does not assert $\varphi \wedge \neg\varphi$ at the current world, and it is not erased by selecting an effective requirement.

For contents α and β that are not syntactic negations, conflict detection needs a background theory T. We call the contents incompatible relative to T if $T \models \neg(\alpha \wedge \beta)$. This is a question of logical consequence in the chosen content language. A conditional requirement $\chi \to \mathsf{Issued}_h(\alpha)$ contributes a current obligation only when χ holds. Likewise, an implication true only at the actual world does not justify propagating a normative requirement through it. From $D_h\alpha$ and $D_h(\alpha \to \beta)$ one obtains $D_h\beta$, whereas $D_h\alpha$ and an unmodalized $\alpha \to \beta$ do not in general suffice.

We now fix a finite set $\mathcal{T}$ of binary decision topics. Each topic t has a designated base-language content φ_t, with opposing content $\neg\varphi_t$. A scenario with distinct action names must state explicitly how those names represent these alternatives. The resolution policy below applies topic by topic. It does not assert joint consistency across different topics.

4.3 Declarative priority resolution

Definition 5. *At a world w and topic t, let*

$$S_t(w) = \{h \in \mathcal{AU} \mid \mathcal{M}, w \models \mathsf{Issued}_h(\varphi_t) \vee \mathsf{Issued}_h(\neg\varphi_t)\}.$$

The maximal relevant authorities are

$$M_t(w) = \{h \in S_t(w) \mid \textit{there is no } k \in S_t(w) \textit{ with } h \prec_w k\}.$$

The status is none *if $S_t(w)$ is empty. It is* positive *if every member of the nonempty set $M_t(w)$ issues φ_t, and* negative *if every member issues $\neg\varphi_t$. It is* unresolved *if some maximal authority issues φ_t and another issues $\neg\varphi_t$.*

Weak permission is defined by $P_h\alpha = \mathsf{Active}_h \wedge \neg D_h\neg\alpha$. It refers to the same normative successor relation and is false for an inactive authority. This extension of the language is given its own axiom in Section 5.

The predicates $\mathsf{Effective}(\varphi_t)$ and $\mathsf{Effective}(\neg\varphi_t)$ mean that the status is positive and negative, respectively. The predicate $\mathsf{Unresolved}(t)$ means that the status is unresolved, and $\mathsf{RequiresMediation}(t)$ is an abbreviation

for $\mathsf{Unresolved}(t)$. Mediation denotes a decision external to this policy. It may add contextual information, modify the supplied priority, or choose between the alternatives.

Maximality is computed among authorities that issue a request on the topic, not among every active authority. Agreement does not require a unique maximal authority. Also, a conflict between two incomparable authorities is not necessarily unresolved after all relevant authorities are considered. A third authority above both may determine the result.

Example 6. Suppose h issues φ_t, k issues $\neg\varphi_t$, and $h \prec_w k$. If these are the only relevant authorities, $M_t(w) = \{k\}$ and the negative requirement is effective. The direct issued conflict remains true. If h and k are incomparable, both are maximal and mediation is required. If a third relevant authority j satisfies $h \prec_w j$ and $k \prec_w j$, only j is maximal, so its requirement determines the outcome. Finally, two incomparable maximal authorities that both issue φ_t yield a positive status.

A requirement from a nonmaximal authority is called defeated for this topic. This term records exclusion by the policy, not the falsity of the issued-obligation formula. It can include a lower-priority request that agrees with the selected content. If explanation requires distinguishing agreement from opposition, that distinction can be recorded separately.

5 Axiomatization

We give a Hilbert-style axiomatization for the modal language and its authority and resolution predicates. It has five groups of axiom schemata. The distinction between issued and effective obligations is part of the syntax, so priority can exclude an issued request from the effective outcome without denying that the request was issued. Some schemata are included to make this intended use explicit even though they are derivable from the defining equivalences.

The underlying propositional vocabulary is countable. It includes ordinary atoms, the distinguished applicability atoms Active_h, and the distinguished priority atoms $\mathsf{Less}_{h,k}$. An authority index is drawn from the fixed finite set $\mathcal{AU}$. A topic index is drawn from the fixed finite set $\mathcal{T}$, and each topic has a fixed base-language content φ_t. Neither authority indices nor topic indices are world variables. The formula $\mathsf{Less}_{h,k}$ can occur inside a modal operator or in a conditional such as $\chi \to \mathsf{Less}_{h,k}$.

For the resolution schemata, write $I_h^+ = \mathsf{Issued}_h(\varphi_t)$, $I_h^- = \mathsf{Issued}_h(\neg\varphi_t)$, $S_h = S_h^t$, $M_h = \mathsf{Max}_h(t)$, and $F_h = \mathsf{Defeated}_h(t)$. Also write $C_t = \mathsf{Conflict}(\varphi_t)$, $U_t = \mathsf{Unresolved}(t)$, and $H_t = \mathsf{RequiresMediation}(t)$. These abbreviations are local to a topic and do not identify requests on distinct topics. For a sign $\sigma \in \{+, -\}$, let E_t^σ denote the corresponding effective requirement. All finite conjunctions and disjunctions below range over the supplied authority set.

5.1 Epistemic axioms

For each agent a, the knowledge and belief schemata are

$$K_a(\alpha \to \beta) \to (K_a\alpha \to K_a\beta), \tag{A1}$$
$$K_a\alpha \to \alpha, \tag{A2}$$
$$K_a\alpha \to K_aK_a\alpha, \tag{A3}$$
$$\neg K_a\alpha \to K_a\neg K_a\alpha, \tag{A4}$$
$$K_a\alpha \to B_a\alpha, \tag{A5}$$
$$B_a(\alpha \to \beta) \to (B_a\alpha \to B_a\beta), \tag{A6}$$
$$B_a\alpha \to \neg B_a\neg\alpha, \tag{A7}$$
$$B_a\alpha \to B_aB_a\alpha, \tag{A8}$$
$$\neg B_a\alpha \to B_a\neg B_a\alpha. \tag{A9}$$

Axiom A1 states closure of knowledge under known implications, and A2 makes knowledge factive. A3 and A4 express positive and negative introspection. A5 connects knowledge to belief. Its frame condition is $R_a^B \subseteq R_a^K$, which must be imposed in addition to the separate conditions on the two relations. A6 is belief distribution, A7 excludes contradictory beliefs, and A8 and A9 express belief introspection.

The doxastic possibility operator is defined by $\langle B_a\rangle\alpha = \neg B_a \neg\alpha$. A7 can then be written as $B_a\alpha \to \langle B_a\rangle\alpha$. The modality and its dual are both indexed by the same agent and refer to the same belief relation.

5.2 Normative and permission axioms

For each authority h, include

$$D_h(\alpha \to \beta) \to (D_h\alpha \to D_h\beta), \tag{A10}$$

$$D_h\alpha \to \neg D_h\neg\alpha, \tag{A11}$$

$$\mathsf{Issued}_h(\alpha) \leftrightarrow (\mathsf{Active}_h \wedge D_h\alpha), \tag{A12}$$

$$P_h\alpha \leftrightarrow (\mathsf{Active}_h \wedge \neg D_h\neg\alpha). \tag{A13}$$

A10 and A11 give a normal KD modality for an authority's normative alternatives. A12 defines the applicable issued requirement. A13 defines weak permission for an active authority. It is not an independent serial box modality. The formula $P_h\alpha \wedge P_h\neg\alpha$ is satisfiable when the authority is active and its normative successors include both an α-world and a $\neg\alpha$-world. Thus permitting an action and permitting refraining from it need not be contradictory.

Weak permission also differs from absence of an applicable obligation. If h is inactive, A12 makes both issued alternatives false, whereas A13 makes both weak-permission predicates false. The logic thereby distinguishes an inactive source from an active source that leaves a choice open. Explicit permissions granted by norms require additional source information and remain outside the resolution policy. Conflict selection below concerns obligations, while A13 gives a precise meaning to the permission notation used in the language.

5.3 Priority and conflict-selection axioms

For every topic t, include the schemata

$$(S_h \wedge S_k \wedge \mathsf{Less}_{h,k}) \to F_h, \tag{A14}$$

$$\neg(\mathsf{Issued}_h(\alpha) \wedge \mathsf{Issued}_h(\neg\alpha)) \tag{A15}$$

$$S_h \leftrightarrow (I_h^+ \vee I_h^-), \tag{A16}$$

$$M_h \leftrightarrow \left(S_h \wedge \bigwedge_{k\in\mathcal{AU}} \neg(S_k \wedge \mathsf{Less}_{h,k})\right), \tag{A17}$$

$$U_t \leftrightarrow \left(\bigvee_{h\in\mathcal{AU}} (M_h \wedge I_h^+)\right) \wedge \left(\bigvee_{k\in\mathcal{AU}} (M_k \wedge I_k^-)\right). \tag{A18}$$

A14 records exclusion by a higher relevant authority. It does not conclude $\neg I_h^+$ or $\neg I_h^-$, and it does not change the issued-conflict record. A15 is within-authority consistency. A16 identifies the authorities participating in the topic, and A17 identifies its maximal participants. A18 requires disagreement between maximal participants before the topic is declared unresolved.

Two incomparable authorities may issue opposing requirements while a third relevant authority has priority over both. In that case A18 does not trigger mediation. Conversely, two incomparable maximal authorities issuing the same alternative are compatible with an effective outcome. These distinctions require the maximality predicate rather than a rule that equates every pairwise incomparability with an unresolved topic.

5.4 Activation and authority-order axioms

The applicability and order schemata are

$$\neg\mathsf{Active}_h \to \neg\mathsf{Issued}_h(\alpha), \tag{A19}$$

$$\neg\mathsf{Less}_{h,h}, \qquad (\mathsf{Less}_{h,k} \wedge \mathsf{Less}_{k,j}) \to \mathsf{Less}_{h,j}, \tag{A20}$$

$$\mathsf{Issued}_h(\alpha) \to \mathsf{Active}_h. \tag{A21}$$

A20 is one family of strict-order schemata, with an irreflexive instance for every authority and a transitive instance for every triple. It forces no totality condition. Incomparability can be defined as $h \neq k$ together with $\neg\mathsf{Less}_{h,k} \wedge \neg\mathsf{Less}_{k,h}$, where inequality between indices is determined by the fixed authority set.

A19 and A21 spell out consequences of the applicability guard. Priority does not imply activation. A scenario can add a conditional such as $\mathsf{threatDetected} \rightarrow \mathsf{Less}_{\mathrm{UNCLOS,ROE}}$ or a condition connecting applicability to location. Such conditions must be included as part of the scenario theory. A claim intended to hold at every world is treated as a global condition rather than merely a premise about the designated actual world.

5.5 Conflict-processing and effective-obligation axioms

Finally, include

$$C_t \leftrightarrow \bigvee_{\substack{h,k\in\mathcal{AU}\\ h\neq k}} (I_h^+ \wedge I_k^-), \tag{A22}$$

$$F_h \leftrightarrow (S_h \wedge \neg M_h), \tag{A23}$$

$$E_t^\sigma \leftrightarrow \left(\bigvee_{h\in\mathcal{AU}} M_h\right) \wedge \bigwedge_{h\in\mathcal{AU}} (M_h \rightarrow I_h^\sigma), \tag{A24}$$

$$H_t \leftrightarrow U_t. \tag{A25}$$

A24 has one instance for each sign. Its positive and negative instances are also referred to as A24$^+$ and A24$^-$. We use these instances when translating $\mathsf{Effective}(\varphi_t)$ and $\mathsf{Effective}(\neg\varphi_t)$. The existence conjunct excludes vacuous effective requirements when no authority participates. A22 retains the issued conflict, A23 defines defeat for the topic, and A25 identifies the requirement for mediation with unresolved maximal disagreement.

Finiteness and these axioms ensure that every topic receives a unique status. If a relevant issued obligation exists, exactly one of E_t^+, E_t^-, and U_t holds. The uniqueness concerns the effective content or unresolved status, not the authority supporting it. Theorem 18 proves this statement. Axiom A22 is compatible with either effective outcome, so selection does not imply $\neg C_t$.

5.6 Inference rules and definitional status

Let H_0 be the base Hilbert calculus comprising propositional tautologies, A1–A11, and A20. It has modus ponens and necessitation for K_a, B_a, and D_h. Necessitation takes a theorem α to $X\alpha$ for any of these normal modalities. It does not apply to a formula derived only from local scenario premises. Ordinary propositional atoms admit uniform substitution, while distinguished applicability and priority atoms retain their stated roles and fixed indices.

The full calculus H adds the new predicates and A12–A19 and A21–A25. Defining axioms A12, A13, A16–A18, and A22–A25 eliminate issued obligations, weak permission, and the resolution predicates in favor of base formulas. A14, A15, A19, and A21 are included as useful consequences of those definitions. For example, A14 follows from A16, A17, and A23 by propositional reasoning. A15 follows from A11 and A12, and A19 and A21 follow immediately from A12. Retaining these consequences as named axioms makes the source-to-implementation correspondence easier to state without changing the class of models.

The definitions have no circular dependence. First expand issued obligations and weak permission, then relevance, then maximality and issued conflict, and then defeat, unresolved status, effectiveness, and mediation. Expanding a formula commutes with Boolean connectives and with each normal modality. A topic content contains no resolution predicates, so it cannot create a recursive definition. This order gives an elimination translation τ from the full language to the base language.

For a set Γ of local premises, write $\Gamma \vdash_{\mathsf{H}} \alpha$ if there is a finite subset $\Gamma_0 \subseteq \Gamma$ such that $\vdash_{\mathsf{H}} (\bigwedge \Gamma_0) \rightarrow \alpha$. This definition makes explicit that necessitation does not turn arbitrary local assumptions into global facts. Consistency means that $\Gamma \nvdash_{\mathsf{H}} \bot$. The corresponding definitions for H_0 are the same.

We next prove soundness and construct a canonical model for H_0. The elimination translation then extends the correspondence to H. In particular, maximal consistent sets are built by syntactic consistency alone. There is no operation that inserts $\neg\mathsf{Issued}_h(\alpha)$ to remove a requirement already present in such a set.

6 Consistency, completeness, and computational properties

6.1 Soundness of the axiomatization

Theorem 7. *Every theorem of* H_0 *is valid in the models of Definition* 2*. Every theorem of* H *is valid in their definitional expansions. If* $\Gamma \vdash_{\mathsf{H}} \alpha$*, every expanded model and world satisfying* Γ *satisfies* α*.*

Proof. Propositional tautologies are valid because Boolean truth conditions are classical. For distribution, suppose $X(\alpha \to \beta)$ and $X\alpha$ hold at w. Every X-successor satisfies both $\alpha \to \beta$ and α, hence satisfies β. This proves A1, A6, and A10.

Reflexivity of R_a^K gives A2, since w is among the worlds used to evaluate $K_a\alpha$. For positive introspection, suppose $X\alpha$ holds at w and wR^Xu. If uR^Xv, transitivity gives wR^Xv, so α holds at v. Thus $X\alpha$ holds at every successor u of w, establishing A3 and A8.

For negative introspection, suppose $X\alpha$ fails at w. The modal clause supplies a successor v at which α fails. Let u be any successor of w. Euclideanness gives uR^Xv, so $X\alpha$ fails at u. Therefore $X\neg X\alpha$ holds at w. The belief relation is Euclidean by assumption, and an equivalence relation is Euclidean, so this proves A4 and A9. For A5, each belief successor is a knowledge successor. Truth of α at all knowledge successors therefore entails truth at all belief successors.

Seriality supplies a successor for each belief or normative relation. If both $X\alpha$ and $X\neg\alpha$ held, that successor would satisfy a contradiction. This proves A7 and A11. A20 holds because each world-relative priority is a strict partial order. Hence all base axioms are valid.

The defining equivalences A12, A13, A16–A18, and A22–A25 hold by their interpretations. A14 holds because a higher relevant authority prevents maximality of the lower one, and A23 then marks the lower one as defeated. A15 follows from A11 and the applicability guard. A19 and A21 follow from that same guard. Thus all full-language axioms are valid.

Modus ponens preserves validity. Necessitation preserves validity because a theorem true at every world is true at every successor of every world. This use of necessitation does not assume local scenario premises at successor worlds. Induction on a proof gives the theorem assertions. If $\Gamma \vdash \alpha$, the valid implication $(\bigwedge \Gamma_0) \to \alpha$ for a finite $\Gamma_0 \subseteq \Gamma$ holds at any world satisfying Γ, which proves the final assertion. □

6.2 Maximal consistent sets and modal successors

We construct the canonical model for H_0 before interpreting the defined predicates. A set w of base formulas is maximal consistent if it is consistent and has no proper consistent extension. The construction uses the syntactic notion of local consistency defined in Section 5.

Lemma 8. *Every* H_0*-consistent set extends to a maximal consistent set. Every maximal consistent set contains all theorems, is closed under modus ponens, and contains exactly one of* α *and* $\neg\alpha$ *for each formula* α*. It contains* $\alpha \wedge \beta$ *exactly when it contains both* α *and* β*.*

Proof. Enumerate the base formulas as $\theta_0, \theta_1, \ldots$. Starting from a consistent set Γ_0, form Γ_{i+1} by adding θ_i if the resulting set is consistent, and otherwise leave Γ_i unchanged. Let $w = \bigcup_i \Gamma_i$. If w were inconsistent, a finite inconsistent subset would already lie in some Γ_i, contrary to its construction. Thus w is consistent. Every formula omitted from w was rejected at its enumeration step because adjoining it produced inconsistency. Adding it to the larger set w still produces inconsistency. Hence w is maximal.

A theorem must belong to w, since adding a theorem cannot turn a consistent set into an inconsistent one. If α and $\alpha \to \beta$ belong to w but β does not, adding β is inconsistent. A finite subset of w would then entail $\neg\beta$. Modus ponens with the two premises would make w inconsistent. Thus $\beta \in w$.

The pair $\alpha, \neg\alpha$ cannot both belong to a consistent set. If $\alpha \notin w$, inconsistency of $w \cup \{\alpha\}$ gives a finite subset of w whose conjunction implies $\neg\alpha$. The theorem is in w, so closure under modus ponens gives $\neg\alpha \in w$. Finally, the propositional implications from $\alpha \wedge \beta$ to each conjunct, and from the two conjuncts to their conjunction, establish the stated conjunction property. □

Let W^c be the set of all maximal H_0-consistent sets. For a normal modality X, with its agent or authority index understood, set

$$\Delta_X(w) = \{\alpha \mid X\alpha \in w\}, \qquad wR_X^c u \iff \Delta_X(w) \subseteq u.$$

This definition uses formulas in maximal consistent sets. It does not refer to semantic defeat or to the outcome of a resolution algorithm.

Lemma 9. *If $\neg X\alpha \in w$, there is a maximal consistent set u such that $wR^c_X u$ and $\neg\alpha \in u$. If the seriality axiom $X\beta \to \neg X\neg\beta$ holds for X, every w has an R^c_X-successor.*

Proof. First suppose $\Delta_X(w) \cup \{\neg\alpha\}$ were inconsistent. There would be finitely many $\beta_1, \ldots, \beta_s \in \Delta_X(w)$ for which

$$\vdash_{\mathsf{H}_0} (\beta_1 \wedge \cdots \wedge \beta_s) \to \alpha.$$

Necessitate this theorem and use distribution. A normal modality also preserves finite conjunctions in the forward direction. To derive that fact for two conjuncts, necessitate the tautology $\beta_1 \to (\beta_2 \to \beta_1 \wedge \beta_2)$ and apply distribution twice. Iterating gives $X\beta_1 \wedge \cdots \wedge X\beta_s \to X(\beta_1 \wedge \cdots \wedge \beta_s)$. Therefore the displayed theorem and the formulas $X\beta_i \in w$ entail $X\alpha \in w$. This contradicts $\neg X\alpha \in w$. If $s = 0$, the argument uses necessitation of the theorem α directly. Consequently $\Delta_X(w) \cup \{\neg\alpha\}$ is consistent. Lemma 8 extends it to a maximal consistent set u, which has both required properties.

For seriality, $X\top$ is a theorem by necessitation. The seriality axiom applied to $\top$ gives $X\top \to \neg X\bot$, using the Boolean equivalence $\neg\top \leftrightarrow \bot$ and normal distribution. Hence $\neg X\bot \in w$. The first part with $\alpha = \bot$ supplies a successor. This proves seriality for belief and normative relations. □

6.3 Canonical frame conditions

Lemma 10. *The canonical knowledge relation is an equivalence relation, the canonical belief relation is serial, transitive, and Euclidean, and the canonical normative relation is serial. Moreover, $R^{B,c}_a \subseteq R^{K,c}_a$ for each agent a.*

Proof. For knowledge reflexivity, $K_a\alpha \in w$ implies $\alpha \in w$ by A2. Thus $\Delta_{K_a}(w) \subseteq w$ and $wR^{K,c}_a w$.

For transitivity of knowledge or belief, suppose $wR^c_X u$ and $uR^c_X v$, and let $X\alpha \in w$. Positive introspection gives $XX\alpha \in w$. Since $wR^c_X u$, this implies $X\alpha \in u$. Since $uR^c_X v$, it follows that $\alpha \in v$. Therefore every member of $\Delta_X(w)$ belongs to v, proving $wR^c_X v$.

For Euclideanness of either relation, suppose $wR^c_X u$ and $wR^c_X v$. To prove $uR^c_X v$, take $X\alpha \in u$. If $\alpha \notin v$, then $X\alpha \notin w$, since $wR^c_X v$ would otherwise put α in v. Maximality gives $\neg X\alpha \in w$, and negative introspection gives $X\neg X\alpha \in w$. From $wR^c_X u$ we then obtain $\neg X\alpha \in u$, contradicting $X\alpha \in u$. Hence $\alpha \in v$, proving Euclideanness.

A reflexive Euclidean relation is symmetric. Indeed, if $wR^{K,c}_a u$, reflexivity also gives $wR^{K,c}_a w$, and Euclideanness applied to these two successors gives $uR^{K,c}_a w$. Together with the transitivity already proved, this makes the knowledge relation an equivalence relation. Seriality of belief and normative relations follows from Lemma 9 using A7 and A11.

Finally, suppose $wR^{B,c}_a u$. If $K_a\alpha \in w$, A5 gives $B_a\alpha \in w$, so $\alpha \in u$. Thus $\Delta_{K_a}(w) \subseteq u$, proving $wR^{K,c}_a u$ and the inclusion claim. □

At a canonical world w, define $h \prec^c_w k$ exactly when $\mathsf{Less}_{h,k} \in w$. Define $\mathcal{I}^c(w,h) = 1$ exactly when $\mathsf{Active}_h \in w$, and let an ordinary atom p be true at w exactly when $p \in w$.

Lemma 11. *Each $\prec^c_w$ is a strict partial order, and the canonical structure satisfies every condition in Definition 2.*

Proof. Every instance of A20 is a theorem and therefore belongs to w. Its irreflexive instances give $\neg\mathsf{Less}_{h,h} \in w$, so $h \not\prec^c_w h$. If $h \prec^c_w k$ and $k \prec^c_w j$, their priority atoms and the transitive instance of A20 yield $\mathsf{Less}_{h,j} \in w$ by modus ponens. Hence $h \prec^c_w j$. This verifies the order conditions from syntactic axioms rather than inserting them as an additional assumption on the canonical model. The accessibility conditions hold by Lemma 10. Applicability and the valuation have the specified Boolean values by construction. The set W^c is nonempty because the base calculus is sound in, for example, a one-world model with reflexive accessibility relations, inactive authorities, and empty strict priority. Hence the empty set is consistent and has a maximal consistent extension. □

6.4 Truth lemma and completeness

Lemma 12. *For every base formula α and $w \in W^c$, one has $\mathcal{M}^c, w \models \alpha$ if and only if $\alpha \in w$.*

Proof. Proceed by induction on the formula. The ordinary atomic, applicability, and priority cases follow from the canonical valuation, applicability function, and order definition. For negation, maximality gives $\neg\beta \in w$ exactly when $\beta \notin w$. The induction hypothesis and the classical truth condition for negation therefore give the desired equivalence. For conjunction, use the conjunction property in Lemma 8 and the induction hypothesis for each conjunct. The remaining Boolean connectives follow from their propositional definitions.

Suppose $X\beta \in w$. Every canonical successor u contains β by the definition of R^c_X. The induction hypothesis makes β true at each such u, so $X\beta$ holds at w. Conversely, if $X\beta \notin w$, maximality gives $\neg X\beta \in w$. Lemma 9 supplies a successor u containing $\neg\beta$. By the induction hypothesis β fails there, so $X\beta$ fails at w. These two directions apply to knowledge, belief, and the normative modality and complete the induction. □

Theorem 13. *The base calculus H_0 and the full calculus H are sound and strongly complete for their stated model classes. In particular, every consistent set of formulas has a model at a designated world, and both calculi are consistent.*

Proof. Let Γ be base-consistent. Extend it to a maximal consistent set w. Lemma 11 makes the canonical structure a model of Definition 2, and Lemma 12 makes every formula of Γ true at w. If $\Gamma \nvdash \alpha$, the local definition of consequence and propositional reasoning show that $\Gamma \cup \{\neg\alpha\}$ is consistent. The same construction supplies a world satisfying Γ and falsifying α. This proves strong completeness for H_0.

For the full calculus, the defining equivalences and propositional reasoning prove $\alpha \leftrightarrow \tau(\alpha)$ for every formula α. The proof is by induction over the definitions and formula structure. At a modal step, necessitate the equivalence already proved for the inner formula and use distribution in each direction to obtain $X\alpha \leftrightarrow X\tau(\alpha)$. This step necessitates a theorem, not a local premise.

Every full axiom translates to a base theorem. Defining axioms translate to Boolean equivalences of a formula with itself. A14 translates to the propositional consequence of relevance, maximality, and defeat. A15 translates to a consequence of A11. A19 and A21 translate to applicability tautologies. The modal and order axioms remain base axiom instances, and translation commutes with modus ponens and necessitation. Thus the full calculus is a conservative extension of H_0 at the level of derivability as well as models.

If a full-language set Γ is consistent, $\tau(\Gamma)$ is base-consistent. Otherwise a finite translated subset would entail $\bot$ in the base calculus, and the proved equivalences would give the same contradiction from the corresponding finite subset of Γ in the full calculus. Choose a canonical world satisfying $\tau(\Gamma)$ and expand its model by the definitions. A structural induction, using their defining equivalences and the unchanged accessibility relations, shows that this world satisfies each formula of Γ. The countermodel argument with $\Gamma \cup \{\neg\alpha\}$ gives strong completeness for H.

Soundness was proved in Theorem 7. A one-world inactive model has a definitional expansion satisfying all full axioms, so $\bot$ is not a theorem of either calculus. This establishes consistency without interpreting defeat as deletion from a maximal consistent set. □

The canonical theorem concerns local consequence. A model satisfying additional global scenario conditions must satisfy them at every world, not merely at its designated world. Such global constraints can be handled in the first-order representation by universal world quantification. The theorem does not classify the complexity of satisfiability with an arbitrary set of global constraints.

6.5 Preservation of issued information and resolution outcomes

Lemma 14. *No authority issues both α and $\neg\alpha$ at a world. An inactive authority issues neither.*

Proof. If $\mathcal{I}(w, h) = 0$, the applicability conjunct in Definition 3 is false. If both issued-obligation formulas hold, then $D_h\alpha$ and $D_h\neg\alpha$ hold. Seriality supplies a world u with $wR^D_h u$. The modal clauses imply both α and $\neg\alpha$ at u, which is impossible under the Boolean truth conditions. □

Theorem 15. *Every model of Definition 2 has a unique expansion interpreting issued obligations, weak permission, and the resolution predicates in Axioms* A12–A25. *The expansion preserves the truth of every formula in the base language. Consequently, adding these definitions to any satisfiable base theory preserves satisfiability.*

Proof. At each world, applicability, the normative relations, and the priority order determine the truth of every I_h^+ and I_h^-. Equation (A16) therefore fixes S_h. Equation (A17) fixes maximality from those truth values and the priority atoms. Equation (A23) then fixes defeat, and Axioms A18 and A24 fix the remaining predicates. No definition refers back to a predicate defined later in this sequence. Assigning these truth values at every world gives an expansion satisfying all the equivalences. Any expansion satisfying the equivalences must make the same assignments, so it is unique.

The weak-permission predicate is fixed by A13, the issued-conflict predicate by A22, and mediation by A25. The remaining consequence axioms are satisfied because A14 follows from maximality and defeat, A15 from serial normative relations, and A19 and A21 from the applicability guard.

The expansion changes neither W, the accessibility relations, the priority orders, applicability, nor the valuation of base atoms. An induction on base formulas proves preservation of truth. The atomic case follows from the unchanged valuation and applicability and priority data. Boolean cases follow from their truth conditions. In a modal case the successor set is unchanged, and the induction hypothesis applies at each successor. Thus every base formula has the same truth value before and after expansion.

If a base theory has a model, expanding that model gives a model of the theory together with all the definitions. Conversely, the reduct of an expanded model satisfies the base theory. This proves the satisfiability assertion without assuming that arbitrary added scenario premises are consistent. □

Proposition 16. *The model conditions are satisfiable. If there are at least two authorities, they permit conflicting issued obligations from distinct authorities.*

Proof. For one authority, a single world with all relations reflexive, the authority inactive, and empty strict priority gives a model. For at least two authorities, let $W = \{u, v\}$ and let p hold at u and fail at v. For every agent let both epistemic relations be $W \times W$. These relations are equivalence relations, are serial, transitive, and Euclidean, and satisfy the inclusion condition. For two authorities h, k, set

$$R_h^D = W \times \{u\}, \qquad R_k^D = W \times \{v\}.$$

Both relations are serial. Activate both authorities everywhere and let $h \prec_w k$ at both worlds. For any further authorities choose serial relations and leave them inactive, with no additional strict priority comparisons. Then $\mathsf{Issued}_h(p)$ and $\mathsf{Issued}_k(\neg p)$ hold at each world. The model contains a direct issued conflict, but no formula and its negation are simultaneously true at a world. For the topic p, k is the sole maximal relevant authority, so its requirement is effective in the unique expansion given by Theorem 15. □

Proposition 17. *For an authority h, the calculus proves $\mathsf{Issued}_h(\alpha) \to P_h\alpha$. It does not prove $\neg(P_h\alpha \wedge P_h\neg\alpha)$. If h is inactive, neither weak-permission formula holds.*

Proof. From $\mathsf{Issued}_h(\alpha)$, A12 yields both Active_h and $D_h\alpha$. A11 then gives $\neg D_h\neg\alpha$. Combining this with applicability and A13 yields $P_h\alpha$.

For the second assertion, consider two worlds u, v with an ordinary atom p true at u and false at v. Activate h and let its normative relation be $W \times W$. At either world $D_h p$ fails because of the successor v, and $D_h\neg p$ fails because of the successor u. Therefore both $P_h p$ and $P_h\neg p$ hold. Choose equivalence knowledge relations, identical belief relations, and an empty strict order to meet the remaining model conditions. By soundness, a formula excluding this conjunction cannot be a theorem. Finally, when h is inactive, the applicability conjunct in A13 is false regardless of the content, which proves the last assertion. □

This proposition gives the usual obligation-to-weak-permission implication without identifying permission with another universal modality. It also shows why permission and obligation play different roles in conflict processing. A weak permission records the absence of a contrary requirement in the active normative standard. It is not itself a requirement that the permitted action be performed, so the current binary selection policy does not insert it into the set of issued obligations.

Theorem 18. *At every world and binary topic, exactly one of none, positive, negative, and unresolved occurs. If mediation is not required and a relevant obligation exists, exactly one of the two opposing requirements is effective. Selecting an effective requirement preserves every issued-conflict formula.*

Proof. If $S_t(w)$ is empty, the status is none and the nonemptiness conjuncts in the two instances of A24 fail. Both existence conjuncts in Axiom A18 also fail.

Suppose $S_t(w)$ is nonempty. A finite nonempty set ordered by a strict partial order has a maximal member. To see this, start at any member and move to a strictly higher member whenever possible. A repeated member would produce a strict cycle, contrary to irreflexivity and transitivity. Finiteness therefore forces the process to stop at a maximal member. Hence $M_t(w)$ is nonempty.

Each member of $M_t(w)$ issues at least one of the two alternatives by the definition of $S_t(w)$ and at most one by Lemma 14. Thus the maximal members either all issue the positive alternative, all issue the negative alternative, or include both signs. These three possibilities are exhaustive and mutually exclusive. The first makes exactly E_t^+ true, the second makes exactly E_t^- true, and the third makes exactly U_t true, using A24 and A18. This proves the classification and excludes simultaneous opposing effective requirements.

Finally, selecting a status only interprets the new predicates. It does not change applicability or D_h, so it does not change either conjunct of any issued-conflict formula. Theorem 15 gives the same conclusion for all base formulas. □

Remark 19. Consistency for each binary topic is not joint consistency of all effective requirements. For example, the contents p, q, and $\neg(p \wedge q)$ are pairwise compatible but jointly incompatible. They can be required on separate topics by different authorities without violating within-authority consistency. A policy that promises an executable set of requirements must additionally test the conjunction of selected contents against the scenario background. The policy here makes no such promise.

6.6 Selection and incompatibility costs

Given the issued-obligation truth values and the strict priority comparison table, the declarative policy is computed as follows.

Resolution of one binary topic

1. Form the set S of authorities issuing either alternative. If it is empty, return none.
2. Form $M = \{h \in S \mid \text{no } k \in S \text{ satisfies } h \prec_w k\}$.
3. If every member of M issues the positive alternative, return positive. If every member issues the negative alternative, return negative. Otherwise return unresolved.

Theorem 20. *With $m = |\mathcal{AU}|$, the algorithm above computes Definition 5 in $O(m^2)$ time, assuming constant-time access to supplied obligation truth values and strict priority comparisons.*

Proof. Two truth-value lookups per authority form S in $O(m)$ time. For each $h \in S$, compare it with every $k \in S$ to determine whether a strictly higher relevant authority exists. This uses at most m^2 comparisons and computes exactly the set $M_t(w)$ in Definition 5. If S is nonempty, M is nonempty by the finite-order argument in Theorem 18. Scanning the signs of the members of M takes $O(m)$ time and returns the corresponding status. The empty case returns immediately. These finite loops give the claimed time bound and termination. □

This is a bound on selection after the relevant information has been obtained. It excludes the cost of proving issued obligations, determining applicability, or deciding logical incompatibility. It also assumes that the input gives the strict order itself. If only immediate priority edges are supplied, reachability must first be computed. For instance, an elementary transitive-closure algorithm takes $O(m^3)$ time.

Proposition 21. *For propositional contents, deciding whether $T \models \neg(\alpha \wedge \beta)$ is coNP-complete, with the finite propositional theory T included in the input. Deciding whether the two contents are jointly satisfiable with T is NP-complete.*

Proof. Represent T by the conjunction of its formulas. Incompatibility is equivalent to unsatisfiability of $T \wedge \alpha \wedge \beta$, so its complement has a polynomially checkable certificate consisting of a satisfying propositional assignment. This proves membership in coNP. For hardness, let θ be any propositional formula and let x be fresh. Set $T = \top$, $\alpha = \theta \vee x$, and $\beta = \neg x$. Then $\alpha \wedge \beta$ is satisfiable exactly when θ is satisfiable. The

construction has linear size. It reduces propositional unsatisfiability to incompatibility and propositional satisfiability to compatibility. Both α and β are individually satisfiable, so the reduction also respects that natural restriction. Propositional satisfiability is NP-complete [7], which gives both completeness statements. □

For unrestricted first-order contents, incompatibility is a first-order validity question after the premises are combined. Under the usual effective assumptions on the signature, proof search can enumerate proofs of valid consequences but does not decide all instances. Thus incompatibility is semidecidable in this setting, and a timeout does not establish compatibility or incompatibility [6]. The number of authorities that conflict does not by itself remove this content-level difficulty.

6.7 Finite-model checking and bounded satisfiability

Theorem 22. *Let an explicit finite model have n worlds, m authorities, and total accessibility size e, where e is the number of relation edges summed over all agents and authorities. Let L be the total length of the modal formulas to be checked, including the contents of r supplied binary topics. Evaluating those formulas and all topic statuses at every world takes $O(L(n+e)+nrm^2)$ time, with applicability and priority tables supplied.*

Proof. First evaluate the base-language topic contents from the inside outward, storing their truth values at every world. Each atomic or Boolean step uses $O(n)$ time. For a modal subformula $X_i\alpha$, scan the edges of R_i^X and mark a world false if any successor fails α. Worlds with no such failure satisfy the universal clause. This uses $O(n+|R_i^X|)$ time. There are at most L base subformula occurrences, so $O(L(n+e))$ bounds their evaluation. For each topic, scan the normative edges to obtain the truth values of $D_h\varphi_t$ and $D_h\neg\varphi_t$ for all authorities, and conjoin applicability. This costs $O(r(e+nm))$. At each world, apply Theorem 20 to each topic, at cost $O(nrm^2)$. The resulting status predicates now have stored truth values at every world. Evaluate the remaining checked formulas, including formulas such as $K_aE_t^+$, using these values and the same inside-out procedure. This adds at most $O(L(n+e))$. Since $r \le L$ and $m \ge 1$, the edge-scan and table costs are absorbed by $O(L(n+e)+nrm^2)$, proving the assertion. Verification that the supplied relations satisfy the frame conditions is a separate finite computation. □

Proposition 23. *For fixed finite agent and authority sets, a finite input theory, and a supplied bound N on the number of worlds, satisfiability within the model class of Definition 2 with at most N worlds is decidable.*

Proof. Only finitely many propositional atoms and topic contents occur in the input. For each $1 \le n \le N$, enumerate valuations of these atoms on $\{1,\dots,n\}$, applicability tables, all accessibility relations, and all world-relative strict priority relations on the fixed authority set. Each is a choice from a finite set. Reject choices failing the frame conditions, which can be checked by finite loops. Evaluate the input theory in each remaining model, adding the uniquely determined resolution predicates when needed. If satisfaction is required at a designated world, test each possible designation. The enumeration terminates and answers yes exactly when a model of the stipulated size exists. Atoms not occurring in the input can receive arbitrary values and do not affect the test. □

The supplied bound in Proposition 23 is essential to this proof. The enumeration proves decidability only for the supplied world bound and gives no complexity classification for unrestricted satisfiability in the combined language. In particular, finite nesting of knowledge operators does not define common knowledge, which requires a separate semantic clause. The benchmark uses a finite request vocabulary, whereas an open first-order encoding can still quantify over an unbounded world domain.

7 TPTP implementation

The computational inputs, Python scripts, and solver logs are available in the project repository [16] at `https://github.com/pstanica/TFF-SUMO-ProjectsFiles`. The files `naval_binary.tff` and `naval_operations_corrected.tff` encode finite instances of the priority policy, while `naval_operations_scenario.tff` is the scenario input analyzed below. The repository also contains `verify_policy.py` for policy verification and `check_supplied_tff.py` for evaluating the scenario axioms in explicit finite interpretations. The

repository also includes `check_vampire_log.py` for checking the logged Vampire input and proof endpoint, `convert_original_to_smt2.py` for translating the scenario input to SMT-LIB, and `make_q4_witness.py` for generating the grounded countermodel check. A README gives the commands and explains the scope of each check.

7.1 Translation and the role of named propositions

TPTP provides typed first-order syntax for stating frame conditions and world-indexed predicates [17]. The encoding uses TFF without arithmetic. A world variable in the target encoding is an ordinary first-order variable introduced by the translation. It is distinct from adding world variables to the modal source language.

For a modal formula α, let $\mathrm{ST}_w(\alpha)$ denote its standard translation at w. Translate an atom p as $p(w)$, commute with Boolean connectives, and put

$$\begin{aligned}
\mathrm{ST}_w(K_a\alpha) &= \forall u\,(R_a^K(w,u) \to \mathrm{ST}_u(\alpha)),\\
\mathrm{ST}_w(B_a\alpha) &= \forall u\,(R_a^B(w,u) \to \mathrm{ST}_u(\alpha)),\\
\mathrm{ST}_w(D_h\alpha) &= \forall u\,(R_h^D(w,u) \to \mathrm{ST}_u(\alpha)),\\
\mathrm{ST}_w(\mathsf{Issued}_h(\alpha)) &= \mathsf{active}(h,w) \wedge \mathrm{ST}_w(D_h\alpha).
\end{aligned}$$

Translate applicability and priority atoms to $\mathsf{active}(h,w)$ and $\mathsf{less}(h,k,w)$. Definitions of selection predicates are then expanded or translated in the same way.

Proposition 24. *The standard translation preserves truth in a model and its first-order relational representation.*

Proof. Use structural induction on the modal formula. Atomic formulas have the same interpretation by construction, and the Boolean cases commute with translation. In the K_a case, the first-order universal quantifier ranges over exactly the worlds accessible through R_a^K for which the modal clause requires truth. The induction hypothesis identifies the truth of the inner translation with the truth of the inner modal formula at each such world. The B_a and D_h cases use the same argument with their respective relations. The issued-obligation case follows by conjunction with the applicability predicate. The selection definitions are finite Boolean formulas in these expressions, so preservation extends to them. □

A second representation, used for finite scenario requests, introduces a type `prop` with constants such as `issue_warning`. The predicate `holds(P,W)` supplies their interpretations. A variable of type `prop` ranges over proposition objects in that first-order domain, not over all formulas of the modal language. Without constructor clauses for negation, conjunction, and modal nesting, this representation does not interpret arbitrary formula syntax. Standard translation of a fixed formula avoids that difficulty by generating its Boolean and modal structure explicitly.

If the proposition sort is restricted to finitely many named constants by a domain-closure clause, that restriction is a modeling assumption. It supports finite enumeration of possible requests and can reduce parts of the search space. It also excludes unnamed requests. Neither the restriction nor a timing observation justifies a general complexity claim. Closure of the proposition sort alone does not close the world sort.

7.2 Accessibility clauses and semantic correspondence

The following clauses illustrate the frame conditions. The types are `world`, `agent`, and `authority`. The declarations of `k`, `b`, and `d` have result type `$o`, TPTP's Boolean type.

```
tff(world_t,type,world:$tType).
tff(agent_t,type,agent:$tType).
tff(authority_t,type,authority:$tType).
tff(k_t,type,k:(agent*world*world)>$o).
tff(b_t,type,b:(agent*world*world)>$o).
tff(d_t,type,d:(authority*world*world)>$o).
tff(k_reflexive,axiom, ![A:agent,W:world]: k(A,W,W)).
tff(k_symmetric,axiom, ![A:agent,W:world,U:world]:
  (k(A,W,U) => k(A,U,W))).
```

```
tff(k_transitive,axiom, ![A:agent,W:world,U:world,V:world]:
  ((k(A,W,U) & k(A,U,V)) => k(A,W,V))).
tff(b_serial,axiom, ![A:agent,W:world]:
  ?[U:world]: b(A,W,U)).
tff(b_transitive,axiom, ![A:agent,W:world,U:world,V:world]:
  ((b(A,W,U) & b(A,U,V)) => b(A,W,V))).
tff(b_euclidean,axiom, ![A:agent,W:world,U:world,V:world]:
  ((b(A,W,U) & b(A,W,V)) => b(A,U,V))).
tff(kb_inclusion,axiom, ![A:agent,W:world,U:world]:
  (b(A,W,U) => k(A,W,U))).
tff(d_serial,axiom, ![H:authority,W:world]:
  ?[U:world]: d(H,W,U)).
```

The inclusion clause is required for knowledge to imply belief. It is not a consequence of the separate S5 and KD45 frame conditions. Also, deontic seriality is unconditional for `d`. Inactivity is represented in the issued-obligation definition rather than by deleting successors.

Component	Encoding clauses	Semantic role
Knowledge	Reflexivity, symmetry, transitivity of `k`	Equivalence relation in Definition 2
Belief	Seriality, transitivity, Euclideanness of `b`	KD45 frame conditions
Knowledge and belief	Accessibility inclusion $R_a^B \subseteq R_a^K$	Ensures knowledge implies belief, equation (A5)
Normative standards	Seriality of `d`	Within-authority consistency
Applicable requests	Applicability guard and universal `d` clause	Definition 3
Priority	Irreflexivity and transitivity of `less`	World-relative strict partial order
Issued conflict	Two issued requests and incompatible contents	Definition 4 and background theory
Effective status	Relevance, maximality, agreement, disagreement	Axioms A12–A25

Table 1: Correspondence between model components and first-order clauses.

Axiom family	Clause or translated formula	Verification obligation
A1–A4	Knowledge frame clauses	Distribution, factivity, introspection
A5	`kb_inclusion`	Belief successors are knowledge successors
A6–A9	Belief frame clauses	Distribution, seriality, introspection
A10–A11	Normative frame clauses	Distribution and seriality
A12–A13	Guarded issued and weak-permission definitions	Applicability is conjoined explicitly
A14	Consequence of `maximal_def` and `defeated_def`	A higher relevant authority excludes maximality
A15	Consequence of normative seriality and A12	Opposing issued requirements from one authority are excluded
A16	`relevant_def`	Either sign makes an authority relevant
A17	`maximal_def`	No higher relevant authority exists
A18	`unresolved_def`	Both signs occur among maximal authorities
A19, A21	Consequences of the issued definition	Inactive authorities issue nothing
A20	`less_irreflexive`, `less_transitive`	Strict partial order at every world
A22	Issued-conflict definition	Opposing requests have distinct sources
A23	`defeated_def`	Relevant but nonmaximal for the topic
A24	`effective_pos_def`, `effective_neg_def`	Nonempty maximal set and agreement
A25	Mediation equivalent to `unresolved`	Mediation is exactly maximal disagreement

Table 2: Axiom-by-axiom correspondence for the full calculus.

Table 2 distinguishes defining clauses from consequences enforced by the frame and those definitions. There is no need to implement A14 as a clause negating an issued predicate. Its consequence is already present in the definition of defeat. Similarly, A15, A19, and A21 can be included as redundant checking clauses, but their semantic justification is the seriality and applicability construction. The proof of strong completeness supplies the syntactic correspondence, while Proposition 24 supplies preservation for the first-order translation.

7.3 A self-contained binary-topic encoding

Here is a compact encoding of one topic. The predicates `pos` and `neg` represent its two issued requirements. For example, `pos(H,W)` is defined by applicability and the standard translation of $D_h\varphi_t$, and `neg(H,W)` by applicability and the translation of $D_h\neg\varphi_t$. The following clauses assume those definitions and the frame clauses above. They do not assert both an obligation and its negation as a method of defeat.

```
tff(pos_t,type,pos:(authority*world)>$o).
tff(neg_t,type,neg:(authority*world)>$o).
tff(less_t,type,less:(authority*authority*world)>$o).
tff(relevant_t,type,relevant:(authority*world)>$o).
tff(maximal_t,type,maximal:(authority*world)>$o).
tff(defeated_t,type,defeated:(authority*world)>$o).
tff(ep_t,type,effective_pos:world>$o).
tff(en_t,type,effective_neg:world>$o).
tff(unresolved_t,type,unresolved:world>$o).
tff(less_irreflexive,axiom, ![H:authority,W:world]:
  ~less(H,H,W)).
tff(less_transitive,axiom,
  ![H:authority,J:authority,K:authority,W:world]:
  ((less(H,J,W) & less(J,K,W)) => less(H,K,W))).
tff(relevant_def,axiom, ![H:authority,W:world]:
  (relevant(H,W) <=> (pos(H,W) | neg(H,W)))).
tff(maximal_def,axiom, ![H:authority,W:world]:
  (maximal(H,W) <=> (relevant(H,W) &
    ~(?[J:authority]: (relevant(J,W) & less(H,J,W)))))).
tff(defeated_def,axiom, ![H:authority,W:world]:
  (defeated(H,W) <=> (relevant(H,W) & ~maximal(H,W)))).
tff(effective_pos_def,axiom, ![W:world]:
  (effective_pos(W) <=>
    ((?[H:authority]: maximal(H,W)) &
     (![J:authority]: (maximal(J,W) => pos(J,W)))))).
tff(effective_neg_def,axiom, ![W:world]:
  (effective_neg(W) <=>
    ((?[H:authority]: maximal(H,W)) &
     (![J:authority]: (maximal(J,W) => neg(J,W)))))).
tff(unresolved_def,axiom, ![W:world]:
  (unresolved(W) <=>
    ((?[H:authority]: (maximal(H,W) & pos(H,W))) &
     (?[J:authority]: (maximal(J,W) & neg(J,W)))))).
```

The formulas quantify over the authority sort. To implement the finite-authority model exactly, add distinct constants for the members of $\mathcal{AU}$ and a domain-closure axiom saying that every authority equals one of them. Alternatively, expand each authority quantifier into a finite conjunction or disjunction. Without closure, a first-order model may contain additional authorities, and the existence of a maximal relevant member does not follow from strict partial order alone. For example, an infinite ascending chain has no maximal member.

For three authorities the closure clause is

```
tff(authority_closure,axiom, ![H:authority]:
  (H=roe | H=unclos | H=nsd)).
```

The three constants require type declarations and pairwise disequalities. Priority information must specify absent comparisons when they matter. Under open-world reasoning, failure to supply `less(roe,nsd,w0)` does not entail its negation. To assert incomparability, supply both negative comparisons. Similarly, ruling out unnamed or additional requests requires complete request information or appropriate closure clauses.

For atomic named contents, the guarded normative definition has the form

```
tff(issued_def,axiom,
  ![H:authority,P:prop,W:world]:
  (issued(H,P,W) <=> (active(H,W) &
    (![U:world]: (d(H,W,U) => holds(P,U))))).
```

For the same named atomic contents, A13 is translated as

```
tff(weak_permitted_t,type,weak_permitted:(authority*prop*world)>$o).
tff(weak_permission_def,axiom,
  ![H:authority,P:prop,W:world]:
  (weak_permitted(H,P,W) <=> (active(H,W) &
    ~(![U:world]: (d(H,W,U) => ~holds(P,U)))))).
```

Here `prop`, `active`, `holds`, and `issued` require their corresponding type declarations. A negative alternative can be translated directly using $\tilde{\texttt{holds(P,U)}}$. A separate constant called `not_p` represents negation only if its interpretation is constrained to be the complement of that of `p`.

Issued conflict between different named contents can be defined using an explicit incompatibility table. If `incompatible(P,Q)` is supplied, conflict holds when distinct authorities issue P and Q. This table must represent the background-theory consequence test rather than serve as a substitute for it. For the binary topic, maximal agreement is computed using the positive and negative classification above. For a larger request vocabulary, a pairwise blocker rule is a different policy and should not be identified with maximal agreement without a proof of equivalence.

7.4 The naval instance and epistemic uncertainty

At the actual world `w0`, the scenario stipulates a detected threat, a nearby strategic asset, and territorial location. Applicability is given by

$$\mathsf{Active}_{\mathrm{UNCLOS}} \leftrightarrow \mathsf{territorial}, \qquad \mathsf{Active}_{\mathrm{ROE}}, \qquad \mathsf{Active}_{\mathrm{NSD}} \leftrightarrow \mathsf{assetNearby}.$$

Scenario rules derive the warning-first requirement from territorial location, the immediate response from the detected threat, and the protection requirement from asset proximity. They must constrain the respective normative alternatives rather than infer a normal obligation solely from the truth of its content at the actual world.

The actual truth of territorial location does not imply that the commander knows it. To represent uncertainty, an epistemic successor of `w0` must make `territorial` false. Then the universal knowledge clause gives $\neg K_a\mathsf{territorial}$. A successor with territorial location true can additionally witness $\neg K_a\neg\mathsf{territorial}$. If both possibilities are to remain compatible with belief, provide corresponding belief successors as well. Such uncertainty is not the conjunction of believing territorial location and believing its negation, which serial belief relations exclude.

For the binary intervention topic, UNCLOS issues the negative alternative and ROE the positive alternative at `w0`. NSD's protection requirement is on a different topic and so does not enter this maximality computation. The supplied priority UNCLOS $\prec_{w_0}$ ROE therefore selects immediate action. The issued response conflict remains true. If protection is changed to an opposing response requirement and NSD and ROE remain incomparable and maximal, the intervention topic becomes unresolved.

7.5 An explicit finite model of the naval topic

The following finite model exhibits the issued conflict, its effective outcome, and the commander's uncertainty about applicability. Let the worlds be w_0, w_1, w_2, the agent be the commander a, and the authorities be ROE, UNCLOS, and NSD. The positive content p is immediate response, and the negative content $\neg p$ is the warning-first choice in this binary topic. Set both R_a^K and R_a^B to $W \times W$. The agent therefore considers all three worlds possible. Let p be true only at w_2. Make threat detection and asset proximity true at all three worlds, so both are known by the commander.

Activate ROE and NSD at every world, and activate UNCLOS at w_0 and w_2 but not at w_1. This matches a territorial-location atom that is true at w_0, w_2 and false at w_1. Use the following normative successor sets at every source world w,

$$R^D_{\mathrm{ROE}}(w) = \{w_2\}, \qquad R^D_{\mathrm{UNCLOS}}(w) = \{w_1\}, \qquad R^D_{\mathrm{NSD}}(w) = W.$$

Every successor set is nonempty. ROE's alternatives satisfy p, UNCLOS's alternatives satisfy $\neg p$, and NSD's alternatives include both signs. In this topic NSD therefore issues neither alternative. If an asset-protection atom is added and made true at all worlds, NSD can require protection while still not taking a side on the intervention topic.

The strict comparisons at every world are UNCLOS below ROE and UNCLOS below NSD, with no others. This is a strict partial order. ROE and NSD are incomparable, but NSD's presence above UNCLOS does not enter maximality for this topic because NSD has no issued response requirement. Table 3 lists the resulting truth values.

Formula	w_0	w_1	w_2
p	false	false	true
$\mathsf{Active}_{\mathrm{UNCLOS}}$	true	false	true
$\mathsf{Issued}_{\mathrm{ROE}}(p)$	true	true	true
$\mathsf{Issued}_{\mathrm{UNCLOS}}(\neg p)$	true	false	true
$\mathsf{Issued}_{\mathrm{NSD}}(p)$	false	false	false
$\mathsf{Issued}_{\mathrm{NSD}}(\neg p)$	false	false	false
C_t	true	false	true
E_t^+	true	true	true
E_t^-	false	false	false
U_t	false	false	false

Table 3: Issued and effective requirements in the three-world model.

At w_0, A22 records the issued conflict, A14 and A23 mark UNCLOS's request as defeated, and A24$^+$ makes the immediate response effective. The issued UNCLOS request remains true there. At w_1, the authority is inactive, so A12 makes its issued request false even though its normative standard still satisfies $D_{\mathrm{UNCLOS}}\neg p$. This is the point at which the applicability guard prevents an inactive source from contributing an obligation.

Since w_1 is a knowledge successor of w_0 and UNCLOS is inactive there, $K_a\mathsf{Issued}_{\mathrm{UNCLOS}}(\neg p)$ fails at w_0. Separately, $K_a\neg p$ fails because the successor w_2 satisfies p. These are different counterexamples to different knowledge assertions. The effective positive status holds at all three worlds, so $K_aE_t^+$ holds at w_0. Thus an agent can lack knowledge that a particular source applies while knowing the priority policy's effective conclusion.

The file `naval_binary.tff` specifies this finite interpretation with domain closure for worlds, agents, and authorities. Its conjecture includes persistence of the issued conflict, the effective positive outcome, absence of unresolved status, and the two negative knowledge assertions. The repository script `verify_policy.py` provides the accompanying policy checks. This model gives a consistency witness for that finite instance. The named-proposition input is analyzed separately in Section 7.6.

7.6 Benchmark queries and reported results

The named-proposition input `naval_operations_scenario.tff` in the project repository [16] contains 46 active axioms and one active conjecture. The other three benchmark conjectures are present as commented alternatives. The analysis below concerns this input and its active clauses.

Table 4 records the reported Vampire 5.0 observations and identifies the conjectures in the scenario input. The per-query proof outputs and exact execution commands are not included in the computational package. These observations concern separate runs from the logged executions described in Section 7.7. Associating each timing with its exact configuration requires the corresponding proof output and execution details. Each conjecture must also be examined against its premises to determine what a successful proof establishes.

The first two conjectures are

```
conflict(immediate_action,issue_warning,w0)
effective_obligation(immediate_action,w0)
```

Q1 follows from the two response obligations, their mutual exclusion, and the conflict definition. Although authority constants are not explicitly declared pairwise distinct in this input, ROE and UNCLOS must be distinct in every model. If they were identical, their active normative relation at `w0` would have a successor

Query	Conjecture	Reported status	Reported time
Q1	Issued response conflict	Theorem	< 1 s
Q2	Effective immediate response	Theorem	< 1 s
Q3	No knowledge of warning proposition	Theorem	< 1 s
Q4	No unresolved conflict	Timeout	300 s

Table 4: Reported benchmark observations and the conjectures identified in the scenario input.

satisfying both response contents, contradicting mutual exclusion. This supplies the distinct-source witness required by `conflict_def`.

Q2's conclusion is also an axiom named `effective_obligation_immediate_action`. Thus a proof of this conjecture can use the conclusion directly and does not independently validate priority resolution. The input also has a forward priority rule that derives an effective requirement from a lower-priority conflicting requirement. That rule does not require maximality of the winning source. In a larger hierarchy it can therefore select a source that is itself overridden. The maximality definitions in A17 and A24 supply the additional condition needed for the policy proved in this paper. The finite examples accompanying that policy do not assert the effective response as an input axiom.

Q3 is `~knows(commander,issue_warning,w0)`. The input supplies a belief-accessible world outside territorial waters, an inclusion clause making that world knowledge-accessible from `w0`, and a global implication from a warning to territorial location. At the witness world that implication excludes the warning proposition. The universal knowledge definition therefore proves Q3, whose subject is the action proposition. Knowledge that UNCLOS applies and knowledge that its requirement is effective are separate assertions. Also, the source's knowledge-to-belief inclusion clause applies only to the commander at `w0`. A5 in the full calculus requires the corresponding inclusion for every agent and source world.

The active Q4 conjecture is

```
tff(query_no_unresolvable_conflict,conjecture,
    ~unresolvable_conflict(w0)).
```

Although the active input contains no axiom negating Q4, the timeout does not establish Q4 or its negation. The next proposition shows that the 46 axioms admit finite models both satisfying and falsifying Q4. This non-entailment result follows from the two models independently of the proof-search outcome.

Proposition 25. *The active axioms of `naval_operations_scenario.tff` are satisfiable. They entail neither* $\neg\mathsf{unresolvable_conflict}(w_0)$ *nor its negation. This remains the case if the world, agent, authority, and proposition domains are restricted to the displayed constants and those constants are required to be distinct within each sort.*

Proof. Construct two structures on three worlds w_0, w_1, w_2, one commander, the three distinct named authorities, and the six distinct named propositions. Knowledge and belief accessibility are universal. Make threat detection, asset proximity, and asset protection true at all worlds. Make territorial location true at w_0, w_2 and false at w_1. Make the warning proposition true only at w_0 and immediate action true only at w_2.

Activate ROE and NSD at all worlds and UNCLOS at w_0, w_2. Give UNCLOS the singleton normative successor set $\{w_0\}$ from every world and ROE the singleton set $\{w_2\}$. In the first structure give NSD the singleton set $\{w_2\}$, and in the second give it $\{w_0\}$. All these relations are serial. Interpret each modal predicate by its defining universal-successor clause in the input.

At w_0, let the only priority comparisons be UNCLOS below ROE and UNCLOS below NSD. At other worlds let there be no priority comparisons. Interpret `effective_obligation` as true exactly for immediate action at w_0. Interpret `conflict` and `unresolvable_conflict` by their defining clauses.

Both structures satisfy the actual-world facts, the known-threat and known-asset clauses, the positive and negative territorial belief witnesses, and the local belief-to-knowledge inclusion. The warning-to-territorial implication holds because the only warning world is w_0. The response contents are mutually exclusive, and w_2 witnesses compatibility of immediate action with protection. Applicability implies each source's stated requirement, including NSD's protection requirement in both structures. All active sources have normative successors, and the authority-consistency clause follows from the obligation definition. The supplied priority facts and negative comparisons hold.

The only incompatible pair of distinct named contents is warning and immediate action. At w_0, the forward priority-resolution clause therefore requires only immediate action to be effective, as assigned. Its explicit effective-response axiom is also satisfied. The proposition-exhaustion clause holds by the chosen domain, and the defining clauses hold by interpretation. Thus every active axiom is satisfied in both structures.

In the first structure, ROE and NSD require immediate action, while UNCLOS requires warning. Every opposing pair is comparable, so no unresolved conflict exists at w_0. In the second structure, NSD additionally requires warning, since its normative successor is w_0, where both warning and protection hold. ROE still requires immediate action. These two sources are incomparable, so the unresolved-conflict definition holds at w_0. NSD's additional warning requirement violates no axiom. The existence of an action-and-protection world does not constrain which protection worlds are NSD's normative alternatives. The two structures therefore disagree on Q4 while satisfying the same active axioms, proving both non-entailment assertions and satisfiability. □

The supplied `prop_exhaustive` clause already closes the proposition sort to six names. It does not close the authority or world sorts, but Proposition 25 shows that closing those sorts alone would not settle Q4. The additional issue is that the premises require certain obligations without excluding other obligations. In a normal modal interpretation, a source can require protection and also require a warning unless its normative alternatives or the permitted request vocabulary exclude that combination.

There is a further difference at inactive sources. The input defines `obligatory` as an unguarded box and imposes seriality only on active authorities. An extra inactive authority with no normative successors then vacuously requires every named proposition. Moreover, the source variables in `unresolvable_def` need not be distinct and are not bound to the particular source witnesses in `conflict_def`. Such an inactive source can therefore supply both requirements in the unresolved test. A12's applicability guard, unconditional normative seriality, and A18's maximal signed disagreement avoid that behavior in the formal construction.

The file `check_supplied_tff.py` parses the fragment used by this input and evaluates every active axiom in the two finite structures from Proposition 25. It also checks the extra-inactive-source example. The checker is a finite-interpretation evaluator, not a general TFF implementation or a theorem prover. Evaluating all active axioms checks that both witnesses satisfy the theory used in the non-entailment proof.

The companion `naval_operations_corrected.tff` retains the six named proposition objects while implementing guarded issued obligations, weak permission, complete finite-domain information, and maximal agreement for the intervention topic. Its effective outcome is defined from the requests and priority rather than asserted as a scenario fact. It is a finite instance of the stated policy, with a separate checked interpretation. The scenario input supports the analysis of Q1–Q4, whereas the finite policy input tests the guarded maximal-agreement definitions.

7.7 Logged execution of the finite policy instance

A run of `naval_operations_corrected.tff` with Vampire 5.0.0 returned SZS status `Theorem`. The recorded executable is a release build at commit `b56c5534b`, dated 24 October 2025. The output contains a TPTP proof ending in ⊥, reports 175.412 seconds elapsed, and records peak memory usage of 217 MB. Table 5 gives these measurements together with the supplied scenario logs for Vampire and Z3.

Input identified by log	Outcome	Elapsed time (s)	Peak memory (MB)
Vampire finite policy instance	Theorem	175.412	217
Vampire scenario log	Time limit	1,000.000	1,422
Z3 scenario translation	Unknown	600.550	993.48
Z3 finite countermodel check	Sat	0.010	18.51

Table 5: Execution measurements recorded in the supplied solver logs.

The finite instance uses one combined conjecture with twelve conjuncts. It checks the three issued requirements, the persistent issued conflict, effectiveness of immediate action, exclusion of an effective warning, and absence of an unresolved intervention topic. It also checks the two distinct negative knowledge assertions, knowledge of the effective positive status, and weak permission for both response alternatives under NSD.

Proving the conjunction establishes all twelve properties from the same input theory. It does not provide twelve separate timing measurements.

The proof embeds the combined conjecture and 26 input axioms used in its derivation. These formulas match the corresponding formulas in the supplied finite-instance file up to bound-variable renaming and Boolean grouping. The repository script `check_vampire_log.py` checks the correspondence of the input formulas and the completed proof endpoint against `vampire_corrected.log`. The individual inference steps have not been replayed in an independent proof checker. Separately, the explicit interpretation checked in Section 7.5 and its named-proposition counterpart satisfy the finite input axioms. This consistency witness prevents the theorem result from being interpreted solely as a consequence of inconsistent premises.

The repository log `vampire_scenario_original.log` records termination at a time limit of 1,000 seconds during saturation, with 1,422 MB peak memory. It contains no embedded conjecture or input axioms, so its exact input cannot be reconstructed from the log alone. Its filename identifies it as the scenario run, and the supplied scenario file has Q4 active. The timeout itself establishes neither Q4 nor its negation. Proposition 25 supplies the separate model-theoretic analysis of that input.

The repository log `z3_original.log` for the SMT-LIB translation of the scenario returns `unknown`, reports 600.55 seconds of solver time, and records maximum memory of 993.48 MB. It records 2,305,966 quantifier instantiations, but supplies neither a model nor a proof. The explicit reason for `unknown` is not recorded. The translated input `naval_operations_original.smt2`, generated by `convert_original_to_smt2.py`, contains the 46 active scenario axioms and the negation of the active Q4 conjecture. Accordingly, `sat` would exhibit a countermodel to Q4, while `unsat` would establish Q4 from the axioms. The observed `unknown` establishes neither conclusion and does not alter Proposition 25.

For a separate verification task, `make_q4_witness.py` generates `naval_q4_countermodel_check.smt2` by expanding all quantifiers over the finite domains of the second structure in Proposition 25 and fixing its complete predicate interpretation. It checks the 46 scenario axioms together with the negation of Q4 as a quantifier-free Boolean problem. This checks the exhibited witness rather than replacing unrestricted satisfiability by an equivalent finite search. The generated assertions have been evaluated independently against the 318 Boolean values in that witness. The repository log `z3_countermodel.log` for this separate check returns `sat` in 0.01 seconds, with maximum memory of 18.51 MB. It reports elimination of all 318 Boolean variables by equation solving. The result verifies the specified finite countermodel against the grounded scenario axioms and negated Q4. It is a witness-verification measurement, not a timing for unrestricted quantified satisfiability.

Both Vampire logs identify the same Vampire build. They do not record the hardware configuration or complete command line. Moreover, the input theories and conjectures differ. The measurements therefore document these executions and do not establish a general speedup or a scaling law. In particular, the finite conjunction's 175.412-second proof should not be reported as a sub-second proof of each benchmark query. Reproduction requires the exact repository revision, the proof and solver outputs, and the checking inputs described above.

7.8 Scope of validation

Frame conditions can be tested independently by translating instances of factivity, distribution, positive and negative introspection, and belief consistency. Each test should state its modal formula and include exactly the required accessibility clauses. Common knowledge requires an additional operator and cannot be inferred from the individual knowledge clauses alone. Likewise, a knowledge-closure test must identify the modality used on its implication premise.

Table 6 reports the recorded measurements for six frame tests with formulas expressible in the present language. Each formula displays its agent index. The corresponding problem files and proof outputs are not included in the computational package, so the recorded frame-test timings cannot be reproduced from that package. A common-knowledge-to-individual-knowledge test belongs to an extension with common knowledge, and a closure test using an unspecified box cannot be assigned to a frame until that box is identified. Neither is used as evidence for this calculus.

Such checks test the translation and frame assumptions. They do not establish a complete axiomatization, a general runtime bound, or correctness of every scenario rule. The standard-translation proof gives a semantic justification, and test cases can check a concrete implementation against that justification. The

Test	Formula	Reported time (s)
Factivity	$K_a p \to p$	0.017
Positive introspection	$K_a p \to K_a K_a p$	0.023
Negative introspection	$\neg K_a p \to K_a \neg K_a p$	0.013
Belief consistency	$\neg(B_a p \wedge B_a \neg p)$	0.013
Belief introspection	$B_a p \to B_a B_a p$	0.062
Knowledge distribution	$K_a(p \to q) \wedge K_a p \to K_a q$	0.008

Table 6: Reported Vampire 5.0 frame tests for formulas in the stated language.

quadratic bound concerns selection from supplied data, while the finite-model bound in Theorem 22 concerns evaluation in an explicit model. Neither predicts saturation-prover timings for an open first-order problem.

Larger authority sets and richer content theories are needed to measure how the implementation scales beyond this three-authority example. The mathematical policy remains limited to binary topics, with joint feasibility checked separately. Priorities can depend on world conditions, but there is no temporal transition structure. The model also idealizes agents as closed under the chosen modal principles and excludes internally inconsistent normative standards for a single authority.

8 Future research directions

A first extension is to admit internally inconsistent norm sets while preserving the distinction between source requirements and effective outcomes. The serial normative relation currently excludes contradictory contents from one active authority. Explicit norm identifiers and a separate consequence relation could represent inconsistent source material without treating it as a single normal modal theory.

Permissions require a distinction between weak permission and permission explicitly granted by a norm. An explicit permission may oppose a prohibition even when it does not require the permitted action. Extending the selection policy to such cases needs a definition of conflict between norm types and an account of whether permissive rules can defeat obligations. It would be inappropriate to obtain this behavior merely by treating permission as another serial box operator.

For nonbinary decisions, the system should select a jointly feasible set of requirements relative to a background theory. Pairwise compatibility is insufficient, as the example following Theorem 18 shows. One possible policy would compare feasible request sets using priority information. Such a policy would need a defined selection criterion, an explanation of each selected request, and proofs of correctness and computational cost.

A temporal extension could represent changes in applicability, priority, and deadlines. The present model already permits different priority orders at different worlds, but it does not specify how the actual situation changes over time. Adding a transition relation would allow those changes to be distinguished from epistemic alternatives. Quantitative uncertainty is another possible extension, provided that probability assignments are kept distinct from normative priority.

Our TPTP encoding uses modal accessibility relations (Definition 2) for compatibility with first-order ATP tools. This encoding directly implements the Kripke semantics presented in Section 4. We have used modal operators from the Suggested Upper Merged Ontology (SUMO) [12, 14][1], and the basic axiomatizations described in [13, 8]. However we extend those operators and axioms with our system of authorities under uncertainty. We are also, for the moment, encoding our extensions directly in TFF rather than coding them in SUMO and relying on its automated translation to Typed Higher-order Form (THF) [2]. An alternative encoding using SUMO's higher-order approach is possible, where knowledge becomes a relation over formula objects: `knows(Agent, Formula)` rather than implicit in accessibility relations.

Finally, reproducible benchmark material should include a fixed repository revision, proof outputs, consistency checks, and the complete execution environment. It should separately test applicability guards, persistent issued conflicts, agreement among incomparable maximal authorities, and disagreement requiring

[1] `https://www.ontologyportal.org`

mediation. These cases distinguish the policy defined here from a rule that only looks for pairwise higher-priority blockers.

9 Conclusion

We have defined a model for reasoning about uncertain authority applicability and conflicting obligations, with a finite policy that separates issued requirements from effective ones. Applicability guards prevent inactive authorities from issuing obligations, and priority selection adds outcome predicates without negating the issued requirements that explain the conflict. Weak permission is defined through the dual of the normative modality and can permit both response alternatives without requiring either. For each binary topic, agreement among maximal relevant authorities selects a requirement, while disagreement is recorded as requiring mediation.

The axiomatization is sound and strongly complete for the stated models. Its canonical construction verifies modal successors, the authority-order conditions, and the truth lemma. The resolution definitions preserve the underlying model and its satisfiable theories. They give exactly one status per binary topic and can be evaluated in quadratic time when obligation truth values and priority comparisons are supplied. Logical incompatibility and modal evaluation have separate costs, and joint feasibility across topics requires an additional check. The typed first-order translation makes these distinctions explicit and provides a basis for implementation.

The naval input illustrates the distinction between issued requirements and effective outcomes. Its reported timings require a qualified interpretation because the effective-response query is also an input axiom. Two checked finite models show that the premises do not determine its negative unresolved-conflict query. The companion finite encoding gives an explicit instance of the guarded maximal-agreement policy. A logged Vampire run proves its twelve-part conjecture in 175.412 seconds, with a separate finite-model check establishing consistency of the input. Z3 also verifies the grounded countermodel to the scenario query in 0.01 seconds. The Vampire scenario timeout and the Z3 unknown result for unrestricted quantified search remain inconclusive. Extending conflict selection to explicit permissive norms, inconsistent source norms, and jointly feasible nonbinary choices remains a further task.

Acknowledgements. The authors would like to thank the editors for the prompt handling of our paper and the reviewers for the very constructive comments which improved the editorial and technical content of our manuscript.

References

[1] C. Benzmüller and L. C. Paulson (2013). Quantified multimodal logics in simple type theory. *Logica Universalis* 7(1), 7–20.

[2] Benzmüller C., and Pease A. (2010). Reasoning with Embedded Formulas and Modalities in SUMO, in *The ECAI-10 Workshop on Automated Reasoning about Context and Ontology Evolution, ARCOE-10*, August 16-17, 2010, Lisbon, Portugal.

[3] C. Benzmüller and A. Pease (2012). Higher-order aspects and context in SUMO. *Journal of Web Semantics* 12–13, 104–117.

[4] P. S. Berman (2012). *Global Legal Pluralism: A Jurisprudence of Law Beyond Borders*. Cambridge University Press.

[5] P. Blackburn, M. de Rijke, and Y. Venema (2001). *Modal Logic*. Cambridge University Press.

[6] A. Church (1936). A note on the Entscheidungsproblem. *Journal of Symbolic Logic* 1(1), 40–41.

[7] S. A. Cook (1971). The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, 151–158.

[8] Feener, C., Pease, A., and Thompson, R., (2026). Embedding Deontic Logic Operators from a Large Ontology in Higher-order Logic, presented at the *FLoC26 workshop on Automated Reasoning in Quantified Non-Classical Logics*.

[9] G. Governatori, F. Olivieri, A. Rotolo, and S. Scannapieco (2013). Computing strong and weak permissions in defeasible logic. *Journal of Philosophical Logic* 42(6), 799–829.

[10] J. F. Horty (2001). *Agency and Deontic Logic*. Oxford University Press.

[11] L. Kovács and A. Voronkov (2013). First-order theorem proving and Vampire. In *Computer Aided Verification*, LNCS 8044, 1–35. Springer.

[12] I. Niles and A. Pease (2001). Toward a standard upper ontology. In *Proceedings of the Second International Conference on Formal Ontology in Information Systems*, 2–9.

[13] Pease, A., (2025). Modal and Higher Order Logical Reasoning with SUMO. in Aikaterini-Lida Kalouli, Tracy Holloway King, Stephen Pulman and Annie Zaenen (eds.), *Semantics at the Crossroads: From Theoretical Explorations to Implementations*, 165-186. Konstanz: PubliKon.

[14] A. Pease (2011). *Ontology: A Practical Guide*. Articulate Software Press.

[15] G. Sartor (2005). *Legal Reasoning: A Cognitive Approach to the Law*. Springer.

[16] P. Stănică. TFF-SUMO project files. `https://github.com/pstanica/TFF-SUMO-ProjectsFiles/tree/main/ModalNavalOperationsProject`.

[17] G. Sutcliffe (2017). The TPTP problem library and associated infrastructure: From CNF to TH0, TPTP v6.4.0. *Journal of Automated Reasoning* 59(4), 483–502.