Value-Oriented Legal Argumentation in Isabelle/HOL

Christoph Benzmüller
Freie Universität Berlin, Germany

David Fuenmayor
University of Luxembourg, Luxembourg
Freie Universität Berlin, Germany

Abstract

Literature in AI & Law contemplates argumentation in legal cases as an instance of theory construction. The task of a lawyer in a legal case is to construct a theory containing: (a) relevant generic facts about the world, (b) relevant legal rules such as precedents and statutes, and (c) contingent facts describing or interpreting the situation at hand. Lawyers then elaborate convincing arguments starting from these facts and rules, deriving into a positive decision in favour of their client, often employing sophisticated argumentation techniques involving such notions as burden of proof, *stare decisis*, legal balancing, etc. In this paper we exemplarily show how to harness Isabelle/HOL to model lawyer's argumentation using value-oriented legal balancing, while drawing upon shallow embeddings of combinations of expressive modal logics in HOL. We highlight the essential role of model finders (*Nitpick*) and "hammers" (*Sledgehammer*) in assisting the task of legal theory construction and share some thoughts on the practicability of extending the catalogue of ITP applications towards legal informatics.

2012 ACM Subject Classification Computing methodologies → Knowledge representation and reasoning

Keywords and phrases Higher order logic, preference logic, shallow embedding, legal reasoning

Digital Object Identifier 10.4230/LIPIcs.ITP.2021.7

Supplementary Material *Isabelle/HOL* formalisation sources are available online at GitHub [1].
Software (Isabelle/HOL formalisation sources): `https://github.com/cbenzmueller/LogiKEy` [1]
archived at `swh:1:dir:88ab2e30ba1f4807b3dad6abaf1fc1738e809706`

Acknowledgements We thank Bertram Lomfeld for encouraging us to take on this challenge; we also thank the anonymous reviewers for their valuable feedback.

1 Introduction

In this paper we explore (value-oriented) legal reasoning as a new application area for higher-order proof assistants. More specifically, we employ *Isabelle/HOL* [33] to formalise, verify, and enhance legal arguments as presented in the context of a legal case between two parties: a *plaintiff* and a *defendant*. In the spirit of previous work in the AI & Law tradition, we tackle the formal reconstruction of legal cases as a task of theory construction, namely, "building, evaluating and using theories" [5]. Thus, *"the task for a lawyer or a judge in a "hard case" is to construct a theory of the disputed rules that produces the desired legal result, and then to persuade the relevant audience that this theory is preferable to any theories offered by an opponent"* [32].

We utilise the framework of shallow semantical embeddings (SSE; cf. [7, 15]) of (combinations of) non-classical logics in classical higher-order logic (HOL). HOL, which is instantiated here as *Isabelle/HOL*, thereby serves as a *meta-logic*, rich enough to support the encoding of combinations of *object logics* (modal, conditional, deontic, etc. [6, 8, 9, 10]) allowing for the modelling of adaptable value systems. For this sake, we also integrate some basic notions from *formal concept analysis* (FCA) [22] to exemplarily illustrate the encoding of a theory of legal values as proposed by Lomfeld [30].

This paper improves an unpublished workshop paper [11]; *Paper structure:* In §2 we outline our object logic of choice, a modal logic of preferences [37], and we then present a SSE of this logic in the *Isabelle/HOL* proof assistant. Subsequently we depict in §3 the encoding of a logic of legal values by drawing upon FCA notions and Lomfeld's value theory. In §4 we demonstrate how the formalisation of relevant legal and world knowledge can be used for formally reconstructing value-oriented arguments for an exemplary property law case. We conclude in §5 with some comments on related work and further reflections and ideas for the prospective application of ITP in the legal domain.

2 Shallow Embedding of the Object Logic

2.1 Modal Preference Logic $\mathcal{PL}$

As will become evident later on, our object logic needs to provide the means for representing (conditional) preferences between propositions. For this sake we have chosen the modal logic of *ceteris paribus* preferences as introduced by van Benthem et al. [37], which we abbreviate by $\mathcal{PL}$ in the remainder. For the purpose of this present paper we will focus our discussion on $\mathcal{PL}$'s basic preference language, disregarding the mechanism of *ceteris paribus* clauses. Nevertheless, we have provided a complete encoding and assessment of $\mathcal{PL}$ in the associated *Isabelle/HOL* sources [1]. We will briefly outline below some relevant syntactic and semantic notions of $\mathcal{PL}$ and refer the reader to [37] for a complete exposition.

$\mathcal{PL}$ is composed of normal *S4* and *K4* modal operators, together with a *global* existential modality **E**. Combinations of these modalities enable us to capture a wide variety of propositional preference statements of the form $A \prec B$ (for different, indexed $\prec$-relations as shown below). The formulas of $\mathcal{PL}$ are inductively defined as follows (where $\mathtt{p}$ ranges over a set `Prop` of propositional constant symbols):

$$\varphi, \psi ::= \mathtt{p} \mid \varphi \wedge \psi \mid \neg\varphi \mid \Diamond^{\preceq}\varphi \mid \Diamond^{\prec}\varphi \mid \mathbf{E}\varphi$$

$\Diamond^{\preceq}\varphi$ is to be read as "φ is true in a state that is considered to be at least as good as the current state", $\Diamond^{\prec}\varphi$ as "φ is true in a state that is considered to be strictly better than the current state", and $\mathbf{E}\varphi$ as "there is a state where φ is true". $\Box^{\preceq}\varphi$, $\Box^{\prec}\varphi$ and $\mathbf{A}\varphi$ can be introduced to abbreviate $\neg\Diamond^{\preceq}\neg\varphi$, $\neg\Diamond^{\prec}\neg\varphi$ and $\neg\mathbf{E}\neg\varphi$, respectively. Further, standard logical connectives such as $\vee$, $\rightarrow$ and $\leftrightarrow$ can be defined as usual. We use **boldface** fonts to distinguish standard logical connectives of $\mathcal{PL}$ from their counterparts in HOL.

A preference model $\mathcal{M}$ is a triple $\mathcal{M} = \langle W, \preceq, V\rangle$ where: (i) W is a set of states; (ii) $\preceq$ is a so-called "betterness relation" that is reflexive and transitive (i.e. a preorder), where its strict subrelation $\prec$ is defined as: $w \prec v$ iff $w \preceq v \wedge v \npreceq w$ for all v and w (totality of $\preceq$, i.e. $v \preceq w$ or $w \preceq v$, is generally not assumed); (iii) V is a standard modal valuation. Below we show the truth conditions for $\mathcal{PL}$'s modal connectives (the rest are standard):

$\mathcal{M}, w \vDash \Diamond^{\preceq}\varphi$ iff $\exists v \in W$ such that $w \preceq v$ and $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \Diamond^{\prec}\varphi$ iff $\exists v \in W$ such that $w \prec v$ and $\mathcal{M}, v \vDash \varphi$
$\mathcal{M}, w \vDash \mathbf{E}\varphi$ iff $\exists v \in W$ such that $\mathcal{M}, v \vDash \varphi$

A formula φ is *true at* world $w \in W$ in model $\mathcal{M}$ if $\mathcal{M}, w \vDash \varphi$. φ is *globally true in* $\mathcal{M}$, denoted $\mathcal{M} \vDash \varphi$, if φ is *true at* every $w \in W$. Moreover, φ is *valid* (in a class of models $\mathcal{K}$) if *globally true in* every $\mathcal{M}$ ($\in \mathcal{K}$), denoted $\vDash_{\mathcal{PL}} \varphi$ ($\vDash_{\mathcal{K}} \varphi$).

Quite relevant to our purposes is the fact that $\mathcal{PL}$ introduces eight semantical definitions for binary preference operations on propositions ($\preceq_{EE}, \preceq_{AE}, \preceq_{EA}, \preceq_{AA}$, and their strict variants). They correspond, roughly speaking, to the four different ways of combining a pair of universal and existential quantifiers when "lifting" an ordering on worlds to an ordering

on sets of worlds (i.e. propositions). In this respect $\mathcal{PL}$ can be seen as a family of preference logics encompassing, in particular, the proposals by von Wright [38] and Halpern [24]. $\mathcal{PL}$ appears well suited for effective automation using the SSE approach, which has been an important selection criterion. This judgment is based on good prior experience with the SSE of related (monadic) modal logics [14, 15] whose semantics employs Kripke-style relational semantics.

2.2 Encoding $\mathcal{PL}$ in Meta-logic HOL

We employ the *shallow semantical embeddings* (SSE) technique [7, 15] to encode (a semantical characterisation of) the logical connectives of an *object logic* as λ-expressions in HOL. This essentially shows that the object logic can be unraveled as a fragment of HOL and hence automated as such. For (multi-)modal normal logics, like $\mathcal{PL}$, the relevant semantical structures are Kripke-style relational frames. $\mathcal{PL}$ formulas can thus be encoded as predicates in HOL taking worlds as arguments.[1]

As a result, we obtain a combined, interactive and automated, theorem prover and model finder for (an extended variant of) $\mathcal{PL}$ realised within *Isabelle/HOL*. This is a new contribution, since we are not aware of any other existing implementation and automation of such a logic. Moreover, the SSE technique supports the automated assessment of meta-logical properties of the embedded logic at a semantical level, which in turn provides practical evidence for the correctness of our encoding.

We now give a succinct overview of the SSE of $\mathcal{PL}$ [1]. The embedding starts with declaring the HOL base type ι, corresponding to the domain of possible worlds/states in our formalisation. $\mathcal{PL}$ propositions are modelled as predicates on objects of type ι (i.e. as *truth-sets* of worlds) and, hence, they are given the type $(\iota \to o)$, which is abbreviated as σ in the remainder. The "betterness relation" $\preceq$ of $\mathcal{PL}$ is introduced as an uninterpreted constant symbol $\preceq_{(\iota\to\iota\to o)}$ in HOL, and its strict variant $\prec$ is introduced as an abbreviation $\prec_{(\iota\to\iota\to o)}$ standing for the HOL term $\lambda v\lambda w(v \leq w \wedge \neg(w \leq v))$; see Fig. 1. $\preceq$-accessible worlds are interpreted as those that are *at least as good* as the present one, and we hence postulate that $\preceq$ is a preorder, i.e. reflexive and transitive. In a next step the σ-type lifted logical connectives of $\mathcal{PL}$ are introduced as abbreviations for λ-terms in the meta-logic HOL. The conjunction operator $\wedge$ of $\mathcal{PL}$, for example, is introduced as an abbreviation $\wedge_{\sigma\to\sigma\to\sigma}$ which stands for the HOL term $\lambda\varphi_\sigma\lambda\psi_\sigma\lambda w_\iota(\varphi\ w\ \wedge\ \psi\ w)$, so that $\varphi_\sigma \wedge \psi_\sigma$ reduces to $\lambda w_\iota(\varphi\ w\ \wedge\ \psi\ w)$, denoting the set[2] of all worlds w in which both φ and ψ hold. Analogously, for negation, we introduce an abbreviation $\neg_{\sigma\to\sigma}$, which stands for $\lambda\varphi_\sigma\lambda w_\iota\neg(\varphi\ w)$.

The operators $\Diamond^{\preceq}$ and $\Diamond^{\prec}$ use $\preceq$ and $\prec$ as guards in their definitions. These operators are introduced as shorthands $\Diamond^{\preceq}_{\sigma\to\sigma}$ and $\Diamond^{\prec}_{\sigma\to\sigma}$ abbreviating the HOL terms $\lambda\varphi_\sigma\lambda w_\iota\exists v_\iota(w \preceq v \wedge \varphi\ v)$ and $\lambda\varphi_\sigma\lambda w_\iota\exists v_\iota(w \prec v \wedge \varphi\ v)$, respectively. $\Diamond^{\preceq}_{\sigma\to\sigma}\varphi_\sigma$ thus reduces to $\lambda w_\iota\exists v_\iota(w \preceq v \wedge \varphi\ v)$, denoting the set of all worlds w so that φ holds in some world v that is at least as good as w; analogous for $\Diamond^{\prec}_{\sigma\to\sigma}$. Additionally, the *global existential* modality $\mathbf{E}_{\sigma\to\sigma}$ is introduced as shorthand for the HOL term $\lambda\varphi_\sigma\lambda w_\iota\exists v_\iota(\varphi\ v)$. The duals $\Box^{\preceq}_{\sigma\to\sigma}\varphi_\sigma$, $\Box^{\prec}_{\sigma\to\sigma}\varphi_\sigma$ and $\mathbf{A}_{\sigma\to\sigma}\varphi$ can easily be added so that they are equivalent to $\neg\Diamond^{\preceq}_{\sigma\to\sigma}\neg\varphi_\sigma$, $\neg\Diamond^{\prec}_{\sigma\to\sigma}\neg\varphi_\sigma$ and $\neg\mathbf{E}_{\sigma\to\sigma}\neg\varphi$ respectively. A special predicate $\lfloor\varphi_\sigma\rfloor$ (read φ_σ is *valid*) for σ-type lifted $\mathcal{PL}$ formulas in HOL is defined as an abbreviation for $\forall w_\iota(\varphi_\sigma\ w)$.

[1] This corresponds to the well-known standard translation to first-order logic. Observe, however, that the additional expressivity of HOL allows us to also encode and flexibly combine non-normal modal logics (conditional, deontic, etc.) and to encode also different kinds of quantifiers; see e.g. [6, 8, 9, 10].

[2] In HOL (with Henkin semantics) sets are associated with their characteristic functions.

```
(*betterness relation ⪯ and strict betterness relation ≺*)
consts BR::γ ("_⪯_")
definition SBR::γ ("_≺_") where "v≺w ≡ (v⪯w) ∧ ¬(w⪯v)"
abbreviation "reflexive R ≡ ∀x. R x x"
abbreviation "transitive R ≡ ∀x y z. R x y ∧ R y z ⟶ R x z"
abbreviation "is_total R ≡ ∀x y. R x y ∨ R y x"
axiomatization where rBR: "reflexive BR" and tBR: "transitive BR"
lemma tSBR: "transitive SBR" using SBR_def tBR by blast (*derived from axioms*)
(*modal logic connectives (operating on truth-sets)*)
abbreviation c1::σ  ("⊥")   where "⊥ ≡ λw. False"
abbreviation c2::σ  ("⊤")   where "⊤ ≡ λw. True"
abbreviation c3::μ  ("¬_")  where "¬φ ≡ λw.¬(φ w)"
abbreviation c4::ν  (infixl"∧"85) where "φ∧ψ ≡ λw.(φ w)∧(ψ w)"
abbreviation c5::ν  (infixl"∨"83) where "φ∨ψ ≡ λw.(φ w)∨(ψ w)"
abbreviation c6::ν  (infixl"→"84) where "φ→ψ ≡ λw.(φ w)⟶(ψ w)"
abbreviation c7::ν  (infixl"↔"84) where "φ↔ψ ≡ λw.(φ w)⟷(ψ w)"
abbreviation c8::μ  ("□⪯_") where "□⪯φ ≡ λw.∀v.(w⪯v)⟶(φ v)"
abbreviation c9::μ  ("◇⪯_") where "◇⪯φ ≡ λw.∃v.(w⪯v)∧(φ v)"
abbreviation c10::μ ("□≺_") where "□≺φ ≡ λw.∀v.(w≺v)⟶(φ v)"
abbreviation c11::μ ("◇≺_") where "◇≺φ ≡ λw.∃v.(w≺v)∧(φ v)"
abbreviation c12::μ ("E_")  where "Eφ ≡ λw.∃v.(φ v)"
abbreviation c13::μ ("A_")  where "Aφ ≡ λw.∀v.(φ v)"
(*meta-logical predicate for global and validity*)
abbreviation g1::π ("⌊_⌋")  where "⌊ψ⌋ ≡  ∀w. ψ w"
(*some tests: dualities*)
lemma "⌊(◇⪯φ)↔(¬□⪯¬φ)⌋ ∧ ⌊(◇≺φ)↔(¬□≺¬φ)⌋ ∧ ⌊(Aφ)↔(¬E¬φ)⌋" by blast (*proof*)
```

Figure 1 SSE of basic $\mathcal{PL}$ in *Isabelle/HOL* (extract).

$\preceq$ is now "lifted" to a preference relation between $\mathcal{PL}$ propositions (sets of worlds).[3]

$$(\varphi_\sigma \preceq_{EE} \psi_\sigma)\ u_\iota \ \text{ iff } \ \exists s_\iota\ \varphi_\sigma s \wedge (\exists t_\iota\ \psi_\sigma t \wedge s \preceq t) \qquad (u_\iota \text{ arbitrary})$$
$$(\varphi_\sigma \preceq_{EA} \psi_\sigma)\ u_\iota \ \text{ iff } \ \exists t_\iota\ \psi_\sigma t \wedge (\forall s_\iota\ \varphi_\sigma s \rightarrow s \preceq t) \qquad (u_\iota \text{ arbitrary})$$
$$(\varphi_\sigma \preceq_{AE} \psi_\sigma)\ u_\iota \ \text{ iff } \ \forall s_\iota\ \varphi_\sigma s \rightarrow (\exists t_\iota\ \psi_\sigma t \wedge s \preceq t) \qquad (u_\iota \text{ arbitrary})$$
$$(\varphi_\sigma \preceq_{AA} \psi_\sigma)\ u_\iota \ \text{ iff } \ \forall s_\iota\ \varphi_\sigma s \rightarrow (\forall t_\iota\ \psi_\sigma t \rightarrow s \preceq t) \qquad (u_\iota \text{ arbitrary})$$

As an illustration, we can read $\varphi \prec_{AA} \psi$ as "every ψ-state being *better* than every φ-state", and read $\varphi \prec_{AE} \psi$ as "every φ-state having a *better* ψ-state" (similarly for others). Each of these non-trivial variants can be argued for [37, 27]. However, as we will reveal in §3, only the *EA*- and *AE*-variants satisfy the minimal conditions required for a logic of value aggregation. Moreover, they are the only ones that satisfy transitivity.

As shown in [37], the binary preference operators above are complemented by "syntactic" counterparts defined as derived operators using the language of $\mathcal{PL}$. In fact, both sets of definitions ("semantic" and "syntactic") coincide in general only for the *EE*- and *AE*-variants (other variants coincide only if $\preceq$ is a total/linear ordering). The "syntactic" variants are encoded below in HOL employing the σ-type lifted logic $\mathcal{PL}$ (using **boldface** to differentiate them).

$$(\varphi_\sigma \preceq_{EE} \psi_\sigma) := \boldsymbol{E}(\varphi_\sigma \wedge \Diamond^{\preceq}\psi_\sigma) \quad \text{and} \quad (\varphi_\sigma \prec_{EE} \psi_\sigma) := \boldsymbol{E}(\varphi_\sigma \wedge \Diamond^{\prec}\psi_\sigma)$$
$$(\varphi_\sigma \preceq_{EA} \psi_\sigma) := \boldsymbol{E}(\psi_\sigma \wedge \Box^{\prec}\neg\varphi_\sigma) \quad \text{and} \quad (\varphi_\sigma \prec_{EA} \psi_\sigma) := \boldsymbol{E}(\psi_\sigma \wedge \Box^{\preceq}\neg\varphi_\sigma)$$
$$(\varphi_\sigma \preceq_{AE} \psi_\sigma) := \boldsymbol{A}(\varphi_\sigma \rightarrow \Diamond^{\preceq}\psi_\sigma) \quad \text{and} \quad (\varphi_\sigma \prec_{AE} \psi_\sigma) := \boldsymbol{A}(\varphi_\sigma \rightarrow \Diamond^{\prec}\psi_\sigma)$$
$$(\varphi_\sigma \preceq_{AA} \psi_\sigma) := \boldsymbol{A}(\psi_\sigma \rightarrow \Box^{\prec}\neg\varphi_\sigma) \quad \text{and} \quad (\varphi_\sigma \prec_{AA} \psi_\sigma) := \boldsymbol{A}(\psi_\sigma \rightarrow \Box^{\preceq}\neg\varphi_\sigma)$$

[3] The variant $\preceq_{EA}$ as originally presented in [37] was in fact wrongly formulated. This mistake has been uncovered during the (iterative) formalisation process thanks to *Isabelle/HOL*.

We further extend the lifted logic $\mathcal{PL}$ by adding quantifiers. This can be done by identifying $\forall x_\alpha s_\sigma$ with the HOL term $\lambda w_\iota \forall x_\alpha (s_\sigma w)$ and $\exists x_\alpha s_\sigma$ with $\lambda w_\iota \exists x_\alpha (s_\sigma w)$. This way quantified expressions can be seamlessly employed, e.g., for the representation of legal and world knowledge in §4.

A note on the heavy use of abbreviations as opposed to definitions is in order. One motivation is to show by the simplest possible means that the logic $\mathcal{PL}$ (and also our subsequent encodings in this paper) can be understood as a genuine fragment of HOL, and the introduction of the connectives of $\mathcal{PL}$ as syntactic sugar (abbreviations) for λ-terms in HOL does just that. No specific concepts, in particular no *Isabelle/HOL* specific ones, are needed to achieve our goals, making our work easily transferrable to other higher-order proof assistant systems. Another motivation is to show that proof automation with *Sledgehammer* already works very well using only abbreviations. Of course, by using definitions we could support, e.g., selective expansions of definitions, which could be a useful option for further proof optimisation. However, this was not yet necessary for the proof automation results obtained in this work. This, and related issues, are worth considering in further work.

2.3 Faithfulness of the SSE

The faithfulness (soundness & completeness) of the present SSE of $\mathcal{PL}$ in HOL follows from previous results for SSEs of propositional multi-modal logics [14] and their quantified extensions [15]. Soundness of the SSE states that our modelling does not give any "false positives", i.e., if $\vDash^{\mathrm{HOL}(\Gamma)} \lfloor\varphi_\sigma\rfloor$ then $\vDash_{\mathcal{PL}} \varphi$, and therefore $\vdash_{\mathcal{PL}} \varphi$ in the (complete) calculus axiomatised by [37]; here HOL(Γ) corresponds to HOL extended with the relevant types and constants plus a set Γ of axioms encoding $\mathcal{PL}$ semantic conditions, i.e., reflexivity and transitivity of $\preceq_{(\iota\to\iota\to o)}$. Completeness of the SSE means that our modelling does not give "false negatives", i.e., if $\vDash_{\mathcal{PL}} \varphi$ then $\vDash^{\mathrm{HOL}(\Gamma)} \lfloor\varphi_\sigma\rfloor$. Moreover, SSE completeness can be mechanically verified by deriving the σ-type lifted $\mathcal{PL}$ axioms and inference rules in HOL(Γ).[4]

3 A Logic for Value-oriented Legal Reasoning

On top of object logic $\mathcal{PL}$ we define a domain-specific logic for reasoning with values in the context of legal cases. We subsequently encode this logic of legal values in *Isabelle/HOL* and put it to the test.

Setting the Stage: Plaintiff vs. Defendant

In a preliminary step, the contending parties in a legal case, the "plaintiff" (p) and the "defendant" (d), are introduced as an (extensible) two-valued datatype c (for "contender") together with a function $(\cdot)^{-1}$ used to obtain for a given party the *other* one; i.e. $\mathtt{p}^{-1} = \mathtt{d}$ and $\mathtt{d}^{-1} = \mathtt{p}$. Moreover, we add a predicate For to model the ruling *for* a party and postulate: For $x \leftrightarrow \neg$For x^{-1}.

```
(*new datatype for parties/contenders (there could be more in principle)*)
datatype c = p | d (*plaintiff & defendant*)
fun other::"c⇒c" ("_⁻¹") where "p⁻¹ = d" | "d⁻¹ = p"
(*new constant symbol: finding/ruling for party*)
consts For::"c⇒σ"
axiomatization where ForAx: "⌊For x ↔ (¬For x⁻¹)⌋"
```

[4] See the corresponding sources in [1], where we conducted numerous experiments mechanically verifying meta-theoretical results on $\mathcal{PL}$.

Abstract Values and Value Principles

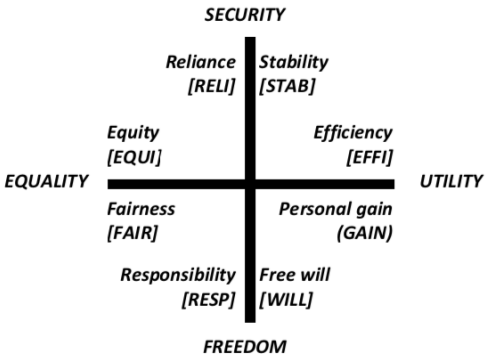

Figure 2 Value theory of Lomfeld [30].

Our approach to value-oriented legal reasoning draws upon recent work in legal theory by Lomfeld [30, 29] who considers a four-quadrant value space generated by two axes featuring antagonistic *abstract values* (FREEDOM vs. SECURITY & UTILITY vs. EQUALITY) at the extremes (Fig. 2).

A set of eight *value principles* are allocated to the four quadrants (two for each quadrant) as shown in Fig. 2. Additionally, Lomfeld's theory contemplates the encoding of legal rules as conditional preferences between conflicting value principles of the form: $R : (E_1 \wedge \cdots \wedge E_n) \Rightarrow A \prec B$. Hence, application of rule R involves *balancing* value principles A and B in context (i.e. under the conditions $E_1 \ldots E_n$).

To provide a concrete modelling of this theory in *Isabelle/HOL*, we have chosen to model value principles as sets of *abstract values*.[5] For the latter we have introduced a four-valued datatype ('t VAL). Observe that this datatype is parameterised with a type variable 't. In the remainder we take 't as being c. In doing this, we allow for the encoding of value principles w.r.t. a particular (favoured) legal party. In the remainder value principles are thus encoded as functions taking objects of type c (`p` or `d`) to sets of abstract values:

```
 9  (*new parameterized datatype for abstract values (wrt. a given party)*)
10 datatype 't VAL = FREEDOM 't | UTILITY 't | SECURITY 't | EQUALITY 't
11 type_synonym v = "(c)VAL⇒bool" (*principles: sets of (abstract) values*)
12 type_synonym cv = "c⇒v" (*principles are specified wrt. a given party*)
```

We have also introduced some convenient type-aliases; v for the type of sets of abstract values, and cv for its corresponding functional version (taking a legal party as parameter).

Instances of value principles (w.r.t. a legal party) are next introduced as sets of abstract values (w.r.t. a legal party), i.e., as objects of type cv. For this we introduce set-constructor operators for values (depicted as $\{\!|\ldots|\!\}$).

Recalling Fig. 2, we have, e.g., that the principle of STAB*ility* favouring the plaintiff (STAB^p) is encoded as a two-element set of abstract values (favouring the plaintiff), i.e., $\{\!|\text{SECURITY } p, \text{UTILITY } p|\!\}$. We do analogously for the other value principles.

[5] Here we suitably simplify Lomfeld's value theory to the effect that, e.g., STAB*ility* becomes identified with EFFI*ciency*. This is enough for our modelling work in §4. A more granular encoding of value principles is possible by adding a third axis to the value space in Fig. 2.

```
 (*notation for sets*)
abbreviation vset1 ("{_}") where "{φ} ≡ λx::(c)VAL. x=φ"
abbreviation vset2 ("{_,_}")   where "{α,β} ≡ λx::(c)VAL. x=α ∨ x=β"
 (*value principles*)
abbreviation stab::cv ("STAB-") where "STAB^x ≡ {SECURITY x, UTILITY  x}"
abbreviation effi::cv ("EFFI-") where "EFFI^x ≡ {UTILITY  x, SECURITY x}"
abbreviation gain::cv ("GAIN-") where "GAIN^x ≡ {UTILITY  x, FREEDOM  x}"
abbreviation will::cv ("WILL-") where "WILL^x ≡ {FREEDOM  x, UTILITY  x}"
abbreviation resp::cv ("RESP-") where "RESP^x ≡ {FREEDOM  x, EQUALITY x}"
abbreviation fair::cv ("FAIR-") where "FAIR^x ≡ {EQUALITY x, FREEDOM  x}"
abbreviation equi::cv ("EQUI-") where "EQUI^x ≡ {EQUALITY x, SECURITY x}"
abbreviation reli::cv ("RELI-") where "RELI^x ≡ {SECURITY x, EQUALITY x}"
```

From a modal logic point of view it is, alternatively, convenient to conceive value principles as truth-bearers, i.e., propositions (as sets of worlds or situations). To overcome this apparent dichotomy in the modelling of value principles (sets of abstract values vs. sets of worlds) we make use of the mathematical notion of a *Galois connection* as exemplified by the notion of *derivation operators* from the theory of *formal concept analysis* (FCA), a mathematical theory of concepts and concept hierarchies as formal ontologies. Below we succinctly discuss a couple of FCA notions relevant to our work. We refer the interested reader to [22] for an actual introduction to FCA.

Some FCA Notions

A *formal context* is a triple $K = \langle G, M, I\rangle$ where G is a set of *objects*, M is a set of *attributes*, and I is a relation between G and M (so-called *incidence relation*), i.e., $I \subseteq G \times M$. We read $\langle g, m\rangle \in I$ as "the object g has the attribute m". We define two so-called *derivation operators* $\uparrow$ and $\downarrow$ as follows:

$$A\uparrow := \{m \in M \mid \langle g, m\rangle \in I \text{ for all } g \in A\} \qquad \text{for } A \subseteq G$$
$$B\downarrow := \{g \in G \mid \langle g, m\rangle \in I \text{ for all } m \in B\} \qquad \text{for } B \subseteq M$$

$A\uparrow$ is the set of all attributes shared by all objects from A, called the *intent* of A. Dually, $B\downarrow$ is the set of all objects sharing all attributes from B, called the *extent* of B. This pair of derivation operators thus forms an antitone *Galois connection* between (the powersets of) G and M, i.e. we always have that $B \subseteq A\uparrow$ iff $A \subseteq B\downarrow$.

A *formal concept* (in a context K) is defined as a pair $\langle A, B\rangle$ such that $A \subseteq G$, $B \subseteq M$, $A\uparrow = B$, and $B\downarrow = A$. We call A and B the *extent* and the *intent* of the concept $\langle A, B\rangle$, respectively. Indeed $\langle A\uparrow\downarrow, A\uparrow\rangle$ and $\langle B\downarrow, B\downarrow\uparrow\rangle$ are always concepts.

The set of concepts in a formal context is partially ordered by set inclusion of their extents, or, dually, by the (reversing) inclusion of their intents. In fact, for a given formal context this ordering forms a complete lattice: its *concept lattice*. Conversely, it can be shown that every complete lattice is isomorphic to the concept lattice of some formal context. We can thus define lattice-theoretical meet and join operations on FCA concepts in order to obtain an algebra of concepts:[6]

$$\langle A_1, B_1\rangle \wedge \langle A_2, B_2\rangle := \langle (A_1 \cap A_2)\ , (B_1 \cup B_2)\downarrow\uparrow\rangle$$
$$\langle A_1, B_1\rangle \vee \langle A_2, B_2\rangle := \langle (A_1 \cup A_2)\uparrow\downarrow\ , (B_1 \cap B_2)\rangle$$

[6] This result can be seamlessly stated for infinite meets and joins (infima and suprema) in the usual way. It corresponds to the first part of the so-called *basic theorem on concept lattices* [22].

Value Principles

We now extend the encoding (SSE) of our object logic $\mathcal{PL}$, exploiting the high expressivity of our meta-logic HOL. We define two FCA derivation operators $\uparrow$ and $\downarrow$ employing the corresponding definitions from above. For this we take G as the domain set of worlds corresponding to the type ι and M as a domain set of abstract values, corresponding in the current modelling approach to the type VAL. In doing this, each value principle (set of abstract values) becomes associated with a proposition (set of worlds) by means of the operator $\downarrow$ (conversely for $\uparrow$). We encode this by defining a binary *incidence* relation $\mathcal{I}$ between worlds/states (type ι) and abstract values (type VAL). We define $\downarrow$ so that $V\downarrow$ denotes the set of all worlds that are $\mathcal{I}$-related to every value in V (analogously for $V\uparrow$).

We introduce an alternative notation: $[V] := V\downarrow$ which may enhance readability in some cases.

```
(**Value Theory*)
consts Irel::"ι⇒v" ("ℐ") (*incidence relation worlds-values*)
 (*derivation operators (cf. theory of "formal concept analysis") *)
abbreviation intent::"σ⇒v" ("_↑") where "W↑ ≡ λv. ∀x. W x ⟶ ℐ x v"
abbreviation extent::"v⇒σ" ("_↓") where "V↓ ≡ λw. ∀x. V x ⟶ ℐ w x"
abbreviation extent_brkt ("[_]") where "[V] ≡ V↓" (*alternative notation*)
```

Recalling the semantics of the object logic $\mathcal{PL}$ from our discussion in §2.1, we can give an intuitive reading for truth at a world in a preference model to terms of the form $P\downarrow$; namely, we can read $\mathcal{M}, w \vDash P\downarrow$ as "principle P provides a reason for (state of affairs) w to obtain". In the same vein, we can read $\mathcal{M} \vDash A \to P\downarrow$ as "principle P provides a reason for proposition A being the case".

Transferring these insights to our current modelling in *Isabelle/HOL*, we can intuitively read, e.g., the formula $\text{STAB}^d\downarrow w$ (of type *bool*) as: "the legal principle of stability is *justifiably* promoted in favour of the defendant (in situation w)". In a similar vein, we can read $\lfloor \text{For } d \to \text{STAB}^d\downarrow \rfloor$ as "promoting (legal) stability in favour of the defendant justifies deciding for him/her (in any situation)".

Value Aggregation and Preference

As discussed above, our logic of legal values must provide means for expressing conditional preferences between principles of the form: $(E_1 \wedge \cdots \wedge E_n) \Rightarrow A \prec B$. The conditional $\Rightarrow$ is modelled in this work using $\mathcal{PL}$'s material conditional $\to$, while noting that a defeasible conditional operator can indeed be defined and added by employing $\mathcal{PL}$'s modal operators [20, 28]. We can also define a binary preference connective $\prec$ for propositions by reusing any of the eight preference "lifting" variants in $\mathcal{PL}$ as discussed in §2. However, this choice cannot be arbitrary, since it needs to interact with value aggregation in an appropriate way.

Lomfeld's theory also contemplates a mechanism for expressing aggregation of value principles (as reasons). We thus define a binary value aggregation connective $\oplus$, observing that it should satisfy particular logical constraints in interaction with a (suitably selected) value preference relation $\prec$:

$$(A \prec B) \to (A \prec B \oplus C) \quad \text{but not} \quad (A \prec B \oplus C) \to (A \prec B) \qquad \text{aggregation on the right}$$
$$(A \oplus C \prec B) \to (A \prec B) \quad \text{but not} \quad (A \prec B) \to (A \oplus C \prec B) \qquad \text{aggregation on the left}$$
$$(B \prec A) \wedge (C \prec A) \to (B \oplus C \prec A) \qquad \text{union property (optional)}$$

The aggregation connectives are most conveniently defined using join (resp. set union) operators, which gives us commutativity. As it happens, only the $\prec_{AE}/\preceq_{AE}$ and $\prec_{EA}/\preceq_{EA}$ variants from §2 satisfy the first two conditions. They are also the only variants satisfying

transitivity. Moreover, if we choose to enforce the third aggregation principle (union property), then we are left with only one variant to consider, namely $\prec_{AE}/\preceq_{AE}$. This variant also offers several benefits for our current modelling purposes: it can be faithfully encoded in the language of $\mathcal{PL}$ [37] and its behaviour is well documented in the literature [24] [27, Ch. 4].

After extensive computer-supported experiments in *Isabelle/HOL* (see [1]) we have identified the following candidate definitions satisfying all desiderata. First, for value aggregation $\oplus$:[7]

$$A \oplus_{(1)} B := (A \cap B)\downarrow \quad \text{and} \quad A \oplus_{(2)} B := (A\downarrow \vee B\downarrow)$$

Then, for a binary preference connective $\prec$ between propositions we have:

$$\varphi \prec_{(1)} \psi := \varphi \preceq_{AE} \psi \quad \text{and} \quad \varphi \prec_{(2)} \psi := \varphi \prec_{AE} \psi$$

For the rest of this work we will illustratively employ the second set of definitions indexed by (2).

Promoting Values

We still need to consider the mechanism by which we can link legal decisions, together with other legally relevant facts, to legal values. We conceive of such a mechanism as a sentence schema, which reads intuitively as: "Taking decision D in the presence of facts F promotes/advances legal (value) principle V". The formalisation of this schema corresponds to a new predicate *Promotes(F,D,V)*, where F is a conjunction of facts relevant to the case (a proposition), D is the legal decision, and V is the value principle thereby promoted.[8]

$$\mathit{Promotes}(F, D, V) := \quad F \rightarrow \Box^{\prec}(D \leftrightarrow \Diamond^{\prec} V\downarrow)$$

Promotes(F,D,V) can be given an intuitive reading: "in every F-situation we have that, in all better states, the admissibility of promoting value V both entails and justifies (as a reason) taking decision D".

```
35 (*connective for aggregating value principles*)
36 abbreviation aggr ("[_⊕_]") where "[V₁⊕V₂] ≡ (V₁↓) ∨ (V₂↓)"
37 (*chosen variant for preference relation (cf. Halpern (1997)*)
38 abbreviation pref::"σ⇒σ⇒σ"  ("_≺_") where "φ ≺ ψ ≡ φ ≺ᴬᴱ ψ"
39 (*schema for value principle promotion*)
40 abbreviation "Promotes F D V ≡ ⌊F → □≺(D ↔ ◇≺(V↓))⌋"
```

Value Conflict

Another important idea inspired from Lomfeld's value theory [29, 30] is the notion of value *conflict*. Recalling Fig. 2, values are disposed around two axis of value coordinates, with values lying at contrary poles playing antagonistic roles. For our modelling purposes it makes thus sense to consider a predicate *Conflict* on worlds (i.e. a proposition) signalling situations where value conflicts appear.

```
41 (*proposition for testing for value conflict*)
42 abbreviation conflict ("Conflict_") where (*conflict for value support*)
43 "Conflictˣ ≡ [SECURITYˣ] ∧ [EQUALITYˣ] ∧ [FREEDOMˣ] ∧ [UTILITYˣ]"
```

[7] Observe that $\oplus_1$ is based upon the join operation on the corresponding FCA formal concepts. $\oplus_2$ is a strengthening of the first, since $(A \oplus_2 B) \subseteq (A \oplus_1 B)$.

[8] We adopt the terminology of *advancing* or *promoting* a value from the literature [16, 34, 5] understanding it in a teleological sense: a decision promoting a value principle means taking that decision *for the sake* of honouring the principle; thus seeing the value principle as a reason for taking that decision.

```
1  theory ValueOntologyTestLong imports ValueOntology (** Benzmüller, Fuenmayor & Lomfeld, 2021 **)
2  begin
3  lemma "True" nitpick[satisfy,show_all,card ι=10] oops
4  lemma "⌊Conflictᵖ⌋" nitpick[satisfy,card ι=4] nitpick oops (*contingent*)
5   (*derivation operators satisfy main properties of Galois connections*)
6  lemma G:      "B ⊑ A↑ ⟷ A ⊑ B↓" by blast
7  lemma G1:     "A ⊑ A↑↓" by simp
8  lemma G2:     "B ⊑ B↓↑" by simp
9  lemma G3:     "A₁ ⊑ A₂ ⟶ A₂↑ ⊑ A₁↑" by simp
10 lemma G4:     "B₁ ⊑ B₂ ⟶ B₂↓ ⊑ B₁↓" by simp
11 lemma cl1:    "A↑ = A↑↓↑" by blast
12 lemma cl2:    "B↓ = B↓↑↓" by blast
13 lemma dual1a: "(A₁ ⊔ A₂)↑ = (A₁↑ ⊓ A₂↑)" by blast
14 lemma dual1b: "(B₁ ⊔ B₂)↓ = (B₁↓ ⊓ B₂↓)" by blast
15 lemma         "(A₁ ⊓ A₂)↑ ⊑ (A₁↑ ⊔ A₂↑)" nitpick oops (*countermodel*)
16 lemma         "(B₁ ⊓ B₂)↓ ⊑ (B₁↓ ⊔ B₂↓)" nitpick oops (*countermodel*)
17 lemma dual2a: "(A₁↑ ⊔ A₂↑) ⊑  (A₁ ⊓ A₂)↑" by blast
18 lemma dual2b: "(B₁↓ ⊔ B₂↓) ⊑  (B₁ ⊓ B₂)↓" by blast
19  (*value conflict tests*)
20 lemma "⌊[RELIᵖ] ∧ [WILLᵖ] ⟶ Conflictᵖ⌋" by simp
21 lemma "⌊Conflictᵖ ⟶ [RELIᵖ] ∧ [WILLᵖ]⌋" by simp
22 lemma "⌊[RELIᵖ] ∧ [WILLᵖ]⌋" nitpick[satisfy] nitpick oops (*contingent*)
23 lemma "⌊[FAIRᵈ] ∧ [EFFIᵈ]⌋" nitpick[satisfy] nitpick oops (*contingent*)
24 lemma "⌊(¬Conflictᵖ) ∧ [FAIRᵈ] ∧ [EFFIᵈ]⌋"
25  nitpick[satisfy,show_all] nitpick oops (*contingent: p & d independent*)
26 lemma "⌊(¬Conflictᵈ) ∧ (¬Conflictᵖ) ∧ [RELIᵈ] ∧ [WILLᵖ]⌋"
27  nitpick[satisfy,show_all] nitpick oops (*contingent: p & d independent*)
28  (*values in two non-opposed quadrants: no conflict*)
29 lemma "⌊[WILLˣ] ∧ [STABˣ] ⟶ Conflictˣ⌋" nitpick oops (*countermodel found*)
30 lemma "⌊[WILLˣ] ∧ [GAINˣ] ∧ [EFFIˣ] ∧ [STABˣ] ⟶ Conflictˣ⌋" nitpick oops
31  (*values in two opposed quadrants: conflict*)
32 lemma "⌊[RESPˣ] ∧ [STABˣ] ⟶ Conflictˣ⌋" by simp
33  (*values in three quadrants: conflict*)
34 lemma "⌊[WILLˣ] ∧ [EFFIˣ] ∧ [RELIˣ] ⟶ Conflictˣ⌋" by simp
35  (*values in opposed quadrants for different parties: no conflict*)
36 lemma "⌊[EQUIˣ] ∧ [GAINʸ] ⟶ (Conflictˣ ∨ Conflictʸ)⌋" nitpick oops (*cntmdl*)
37 lemma "⌊[RESPˣ] ∧ [STABʸ] ⟶ (Conflictˣ ∨ Conflictʸ)⌋" nitpick oops (*cntmdl*)
38  (*value preferences tests*)
39 lemma "⌊[WILLˣ]≺[WILLˣ⊕STABˣ]⌋" nitpick nitpick[satisfy] oops (*contingent*)
40 lemma "⌊[WILLˣ]≺[STABˣ]⌋ ⟶ ⌊[WILLˣ]≺[WILLˣ⊕STABˣ]⌋" by blast
41 lemma "⌊[WILLˣ]≺[STABˣ]⌋ ⟶ ⌊[WILLˣ]≺[RELIˣ⊕STABˣ]⌋" by blast
42 lemma "⌊[WILLˣ]≺[WILLˣ⊕STABˣ]⌋ ⟶ ⌊[WILLˣ]≺[STABˣ]⌋" (*nitpick*) nitpick[satisfy] oops (*ctgnt?*)
43 lemma "⌊[WILLˣ]≺[RELIˣ⊕STABˣ]⌋ ⟶ ⌊[WILLˣ]≺[STABˣ]⌋" nitpick nitpick[satisfy] oops (*contingent*)
44 lemma "⌊[WILLˣ⊕STABˣ]≺[WILLˣ]⌋" nitpick nitpick[satisfy] oops (*contingent*)
45 lemma "⌊[WILLˣ⊕STABˣ]≺[WILLˣ]⌋ ⟶ ⌊[STABˣ]≺[WILLˣ]⌋" by metis
46 lemma "⌊[RELIˣ⊕STABˣ]≺[WILLˣ]⌋ ⟶ ⌊[STABˣ]≺[WILLˣ]⌋" by metis
47 lemma "⌊[STABˣ]≺[WILLˣ]⌋ ⟶ ⌊[WILLˣ⊕STABˣ]≺[WILLˣ]⌋" nitpick nitpick[satisfy] oops (*contingent*)
48 lemma "⌊[STABˣ]≺[WILLˣ]⌋ ⟶ ⌊[RELIˣ⊕STABˣ]≺[WILLˣ]⌋" nitpick nitpick[satisfy] oops (*contingent*)
49  (*basic properties*)
50 lemma "⌊[X]≺[X]⌋" nitpick nitpick[satisfy] oops (*contingent*)
51 lemma "⌊(([X]≺[Y]) ∧ ([Y]≺[Z])) ⟶ ([X]≺[Z])⌋" using tSBR by blast (*transitive*)
52 lemma "⌊([X]≺[Y]) ∧ ([Y]≺[X])⌋ ⟶ X = Y" nitpick oops (*not antisymmetric*)
53 end
```

Figure 3 Testing the logic of legal values.

Testing the Encoding

In order to test the adequacy of our modelling, some implied and non-implied knowledge is studied. We briefly discuss some of the conducted tests as shown in Fig. 3.

Among others, we verify that the pair of operators for *extension* (↓) and *intension* (↑), cf. *formal concept analysis* [22], constitute indeed a Galois connection (Lines 6–18), and we carry out some further tests on the value theory concerning value aggregation and consistency (Lines 20ff.).

In our modelling of the notion of *value conflict*, promoting values (for the same party) from two opposing value quadrants, say RELI & WILL, should entail a value conflict; theorem provers quickly confirm this as shown in Fig. 3 (Line 20). However, promoting values from two non-opposed quadrants, such as WILL & STAB (Line 29) should not imply conflict:

```
Nitpick found a model for card ι = 1:

  Types:
    c = {d, p}
    c VAL = {FREEDOM d, FREEDOM p, UTILITY d, UTILITY p, EQUALITY d, EQUALITY p, SECURITY d, SECURITY p}
  Constants:
    BR = (λx. _)((ι1, ι1) := True)
    For = (λx. _)((d, ι1) := False, (p, ι1) := True)
    I = (λx. _)
        ((ι1, FREEDOM d) := False, (ι1, FREEDOM p) := True, (ι1, UTILITY d) := False, (ι1, UTILITY p) := True,
         (ι1, EQUALITY d) := False, (ι1, EQUALITY p) := True, (ι1, SECURITY d) := False, (ι1, SECURITY p) := True)
    other = (λx. _)(d := p, p := d)
```

Figure 4 Satisfying model for the statement in Line 22 of Fig. 3.

the model finder *Nitpick*[9] computes and reports a countermodel (not shown here) to the stated conjecture. A value conflict is also not implied if values from opposing quadrants are promoted for different parties (Lines 36-37).

Note that the notion of value conflict has deliberately not been aligned with inconsistency in meta-logic HOL. This way we can represent conflict situations in which, for instance, RELI and WILL (being conflicting values, see Line 20 in Fig. 3) are promoted for the plaintiff (p), without leading to a logical inconsistency in *Isabelle/HOL* (thus avoiding "explosion"). In Line 22 of Fig. 3, for example, *Nitpick* is called simultaneously in both modes in order to confirm the contingency of the statement; as expected both a model (cf. Fig. 4) and countermodel (not displayed here) for the statement are returned. This value conflict (w.r.t. p) can also be spotted by inspecting the satisfying models generated by *Nitpick*. One of such models is depicted in Fig. 4, where it is shown that (in the given possible world ι_1) all of the abstract values (EQUALITY, SECURITY, UTILITY, and FREEDOM) are simultaneously promoted for p, which implies a value conflict according to our definition.

Analysing the model structures returned by *Nitpick* has indeed been very helpful to gain a deeper insight into $\mathcal{PL}$ semantic structures. This becomes particularly relevant for complex modelling tasks where a clear understanding is often initially lacking.

Further tests in Fig. 3 (Lines 39-48) assess the behaviour of the aggregation operator $\oplus$ in combination with value preferences. We test for a correct behaviour when "strengthening", resp. "weakening", the right-hand side (Lines 39-43). As an illustration, in line 41, if STAB is preferred over WILL, then STAB combined with, say, RELI is also preferred over WILL alone. Similar test are conducted for "strengthening", resp. "weakening", the left-hand side (Lines 44-48).

Finally, we verify (lines 50–52) basic properties of the preference relation.

4 A Case Study in Property Law

To illustrate our approach, we formalise and assess, employing *Isabelle/HOL*, a well-known benchmark case in AI & Law involving the appropriation of wild animals: *Pierson vs. Post*. In a nutshell: *Pierson killed and carried off a fox which Post already was hunting with hounds on public land. The Court found for Pierson* (cf. [2, 34, 16], and also [23] for the significance of this case as a benchmark).

We start with some words on the modelling of background (legal & world) knowledge.

[9] *Nitpick* [19] searches for, respectively enumerates, finite models or countermodels to a conjectured statement/lemma. By default *Nitpick* searches for countermodels, and model finding is enforced by stating the parameter keyword "satisfy". These models are given as concrete interpretations of relevant terms in the given context so that the conjectured statement is satisfied or falsified.

4.1 Legal & World Knowledge

The realistic modelling of concrete legal cases requires further legal & world knowledge (LWK) to be taken into account. For the sake of illustration, we introduce here only a small and monolithic *Isabelle/HOL* theory[10] called "GeneralKnowledge". This includes a small excerpt of a much simplified "animal appropriation taxonomy", where we associate "animal appropriation" kinds of situations with the value preferences they imply (as conditional preference relations).

In a realistic setting this knowledge base would be further split and structured similarly to other legal or general ontologies, e.g., in the *Semantic Web*. Note, however, that the expressiveness in our approach, unlike in many other legal ontologies or taxonomies, is by no means limited to definite underlying (but fixed) logical language foundations. We could thus easily decide for a more realistic modelling, e.g., avoiding simplifying propositional abstractions. For instance, the proposition "appWildAnimal", representing the appropriation of one or more wild animals, can anytime be replaced by a more complex formula (featuring, e.g., quantifiers, modalities or defeasible conditionals).

We now briefly outline the encoding of our example LWK (see [1] for the full details).

First, some non-logical constants that stand for kinds of legally relevant situations (here: of appropriation) are introduced, and their meaning is constrained by some postulates:

```
(*LWK: kinds of situations addressed*)
consts appObject::σ  appAnimal::σ (*appropriation of objects/animals in general*)
       appWildAnimal::σ  appDomAnimal::σ (*appropriation of wild/domestic animals*)
(*LWK: postulates for kinds of situations*)
axiomatization where
 W1: "⌊appAnimal ⟶ appObject⌋" and
 W2: "⌊¬(appWildAnimal ∧ appDomAnimal)⌋" and
 W3: "⌊appWildAnimal ⟶ appAnimal⌋"  and
 W4: "⌊appDomAnimal ⟶ appAnimal⌋"
```

Then the "default" legal rules for several situations (here: appropriation of animals) are formulated as conditional preference relations:

```
(*LWK: (prima facie) value preferences for kinds of situations*)
axiomatization where
 R1: "⌊appAnimal ⟶ ([STAB^p] ≺ [STAB^d])⌋" and
 R2: "⌊appWildAnimal ⟶ ([WILL^x⁻¹] ≺ [STAB^x])⌋" and
 R3: "⌊appDomAnimal  ⟶ ([STAB^x⁻¹] ≺ [RELI^x⊕RESP^x])⌋"
```

For example, rule R2 could be read as: "In a wild-animals-appropriation kind of situation, promoting STAB*ility* in favour of a party (say, the plaintiff) is preferred over promoting WILL in favour of the other party (defendant)". If there is no more specific legal rule from a precedent or a codified statute then these "default'[11] preference relations determine the result. Moreover, we can have rules conditioned on more concrete legal *factors*.[12] As a

[10] Isabelle documents are suggestively called "theories". They correspond to top-level modules bundling together related definitions, theories, proofs, etc.

[11] We use of the term "default" in the colloquial sense, noting however, that there exist in fact several (non-monotonic) logical systems aimed at modelling such a kind of *defeasible* behaviour for rules/conditionals (i.e., meaning that they can be "overruled"). One of them has been suggestively called "default logic". We refer to [25] for a discussion.

[12] The introduction of legal *factors* is an established practice in the implementation of case-based legal systems (cf. [3] for an overview). They can be conceived –as we do– as propositions abstracted from the facts of a case by the analyst/modeller in order to allow for assessing and comparing cases at a higher level of abstraction. Factors are typically either pro-plaintiff or pro-defendant, and their being true or false (resp. present or absent) in a concrete case can serve to invoke relevant precedents or statutes.

didactic example, the legal rule R4 states that the Own*ership* (say, the plaintiff's) of the land on which the appropriation took place, together with the fact that the opposing party (defendant) acted out of Mal*ice* implies a value preference of RELI*ance* and RESP*onsibility* over STAB*ility*. This last rule has indeed been chosen to reflect the famous common law precedent of Keeble vs. Hickeringill [16, 2].

```
(*LWK: conditional value preferences, e.g. from precedents*)
axiomatization where
 R4: "⌊(Mal x⁻¹ ∧ Own x)  ⟶ ([STABˣ⁻¹] ≺ [RESPˣ⊕RELIˣ])⌋"
```

As already discussed, for ease of illustration, terms like "appWildAnimal" are modelled here as simple propositional constants. In practice, however, they may later be replaced, or logically implied, by a more realistic modelling of the relevant situational facts, utilising suitably complex (even higher-order, if needed) formulas depicting states of affairs to some desired level of granularity.

For the sake of modelling the appropriation of objects, we have introduced an additional type e (for "entities") that can be employed for terms denoting individuals (things, animals, etc.) when modelling legally relevant situations. Some simple vocabulary and taxonomic relationships (here for wild and domestic animals) are specified to illustrate this.

```
(*LWK: domain vocabulary*)
typedecl e    (*declares new type for 'entities'*)
consts
  Animal::"e⇒σ" Domestic::"e⇒σ" Fox::"e⇒σ" Parrot::"e⇒σ" Pet::"e⇒σ" FreeRoaming::"e⇒σ"
(*LWK: domain knowledge (about animals)*)
axiomatization where
 W5: "⌊∀a. Fox a ⟶ Animal a⌋"  and
 W6: "⌊∀a. Parrot a ⟶ Animal a⌋" and
 W7: "⌊∀a. (Animal a ∧ FreeRoaming a ∧ ¬Pet a) ⟶ ¬Domestic a⌋" and
 W8: "⌊∀a. Animal a ∧ Pet a ⟶ Domestic a⌋"
```

As mentioned before, we have introduced some convenient legal *factors* into our example LWK to allow for the encoding of legal knowledge originating from precedents or statutes at a more abstract level. In our approach these factors are to be logically implied (as deductive arguments) from the concrete facts of the case (as exemplified in §4 below). Observe that our framework also allows us to introduce definitions for those factors for which clear legal specifications exist. At the present stage, we will provide some simple postulates constraining factors' interpretation.

```
(*LWK: legally-relevant, situational 'factors'*)
consts Own::"c⇒σ"    (*object is owned by party c*)
       Poss::"c⇒σ"   (*party c has actual possession of object*)
       Intent::"c⇒σ" (*party c has intention to possess object*)
       Mal::"c⇒σ" (*party c acts out of malice*)
       Mtn::"c⇒σ" (*party c respons. for maintenance of object*)
(*LWK: meaning postulates for general notions*)
axiomatization where
 W9: "⌊Poss x ⟶ (¬Poss x⁻¹)⌋" and
 W10: "⌊Own x ⟶ (¬Own x⁻¹)⌋"
```

Recalling §3 we relate the introduced factors to value principles and outcomes by means of the *Promotes* predicate. Finally, the consistency of all axioms and rules provided is confirmed by *Nitpick*.

```
(*LWK: relate values, outcomes and situational 'factors'*)
axiomatization where
 F1: "Promotes (Intent x) (For x) WILL^x" and
 F2: "Promotes (Mal x) (For x⁻¹) RESP^x" and
 F3: "Promotes (Poss x) (For x) STAB^x" and
 F4: "Promotes (Mtn x) (For x) RESP^x" and
 F5: "Promotes (Own x) (For x) RELI^x"
(*Theory is consistent, (non-trivial) model found*)
lemma True nitpick[satisfy,card ι=4] oops
```

4.2 Pierson vs. Post

We illustrate our reasoning framework by encoding the classic property law case Pierson vs. Post.

Ruling for Pierson

The formal modelling of an argument in favour of Pierson is outlined next (the entire formalisation of this argument is presented in the sources [1]).

First we introduce some minimal vocabulary: a constant α of type e (denoting the appropriated animal), and the relations *pursue* and *capture* between the animal and one of the parties (of type c). A background (generic) theory as well as the (contingent) case facts as suitably interpreted by Pierson's party are then stipulated:

```
 (*case-specific 'world-vocabulary'*)
consts α::"e" (*appropriated animal (fox in this case) *)
consts Pursue::"c⇒e⇒σ" Capture::"c⇒e⇒σ"
(************** pro-defendant (Pierson) argument **************)
 (*defendant's theory*)
abbreviation "dT1 ≡ ⌊(∃c. Capture c α ∧ ¬Domestic α) → appWildAnimal⌋"
abbreviation "dT2 ≡ ⌊∀c. Pursue c α → Intent c⌋"
abbreviation "dT3 ≡ ⌊∀c. Capture c α → Poss c⌋"
abbreviation "d_theory ≡ dT1 ∧ dT2 ∧ dT3"
 (*defendant's facts*)
abbreviation "dF1 w ≡ Fox α w"
abbreviation "dF2 w ≡ FreeRoaming α w"
abbreviation "dF3 w ≡ ¬Pet α w"
abbreviation "dF4 w ≡ Pursue p α w"
abbreviation "dF5 w ≡ Capture d α w"
abbreviation "d_facts ≡ dF1 ∧ dF2 ∧ dF3 ∧ dF4 ∧ dF5"
```

The aforementioned decision of the court for Pierson was justified by the majority opinion. The essential preference relation in the case is implied in the idea that appropriation of (free-roaming) wild animals requires actual corporal possession. The manifest corporal link to the possessor creates legal certainty, which is represented by the value *stability* (STAB) and outweighs the mere *will* to possess (WILL) by the plaintiff; cf. the arguments of classic lawyers cited by the majority opinion [23]: "pursuit alone vests no property" (Justinian institutes), and "corporal possession creates legal certainty" (Pufendorf). Recalling Fig. 2 in §3, this corresponds to a preference for the abstract value SECURITY over FREEDOM.

We can see that this legal rule R2, as introduced in the previous section (§4.1) is indeed employed by *Isabelle/HOL*'s automated tools to prove that, given a suitable defendant's theory, the (contingent) facts imply a decision in favour of Pierson in all "better' worlds (which we could read deontically as a sort of obligation):

```
 (*decision for defendant (Pierson) can be proven automatically*)
theorem Pierson: "d_theory ⟶ ⌊d_facts → □≺For d⌋"
  by (smt F1 F3 ForAx R2 W5 W7 other.simps tSBR)
```

The previous "one-liner" proof has indeed been suggested by *Sledgehammer* [17, 18] which we credit, together with *Nitpick* [19], for doing the heavy lifting in our work. A proof argument in favour of Pierson that uses the same dependencies can also be constructed interactively using *Isabelle*'s human-readable proof language *Isar* [39]. The individual steps of the proof are this time formulated with respect to an explicit world/situation parameter w. The argument goes roughly as follows:

1. From Pierson's facts and theory we infer that in the disputed situation w a wild animal has been appropriated: appWildAnimal w.
2. In this context, by applying the value preference rule R2, we get that promoting STAB in favour of Pierson is preferred over promoting WILL in favour of Post: $\lfloor[\text{WILL}^p] \prec [\text{STAB}^d]\rfloor$.
3. The admissibility of promoting WILL in favour of Post thus entails the admissibility of promoting STAB in favour of Pierson: $\lfloor\Diamond^{\prec}[\text{WILL}^p] \rightarrow \Diamond^{\prec}[\text{STAB}^d]\rfloor$.
4. Moreover, after instantiating the *value promotion* schema F1 (§4.1) for Post (p), and acknowledging that his pursuing of the animal (Pursue $p\ \alpha$) entails his intention to possess (Intent p), we obtain (for the given situation w) an obligation/recommendation to "align" any ruling for Post with the admissibility of promoting WILL in his favour: $\Box^{\prec}(\text{For } p \leftrightarrow \Diamond^{\prec}[\text{WILL}^p])\ w$.
5. Analogously, in view of Pierson's (d) capture of the animal (Capture $d\ \alpha$), thus having taken possession of it (Poss d), we infer from the instantiation of *value promotion* schema F3 (for Pierson) an obligation/recommendation to align a ruling for Pierson with the admissibility of promoting the value principle STAB (in his favour): $\Box^{\prec}(\text{For } d \leftrightarrow \Diamond^{\prec}[\text{STAB}^d])\ w$.
6. From (4) and (5) in combination with the courts duty to find a ruling for one of both parties (ForAx) we infer, for the given situation w, that either the admissibility of promoting WILL in favour of Post or the admissibility of promoting STAB in favour of Pierson (or both) hold in every "better" world/situation (thus becoming a recommended/obligatory condition): $\Box^{\prec}(\Diamond^{\prec}[\text{WILL}^p] \vee \Diamond^{\prec}[\text{STAB}^d])\ w$.
7. From this and (3) we thus get that the admissibility of promoting STAB in favour of Pierson is recommended/obligatory in the given context w: $\Box^{\prec}(\Diamond^{\prec}[\text{STAB}^d])\ w$.
8. And this together with (5) finally implies the recomendation/obligation to rule in favour of Pierson in the given context w: $\Box^{\prec}(\text{For } d\ v)$.

```
23  (*we reconstruct the reasoning process leading to the decision for the defendant*)
24 theorem Pierson': assumes d_theory and "d_facts w" shows "□≺For d w"
25 proof -
26   have 1: "appWildAnimal w"     using W5 W7 assms by blast
27   have 2: "⌊[WILLᵖ]≺[STABᵈ]⌋"  using 1 R2 assms by fastforce
28   have 3: "⌊(◇≺[WILLᵖ]) → ◇≺[STABᵈ]⌋" using 2 tSBR by smt
29   have 4: "□≺(For p ↔ ◇≺[WILLᵖ]) w" using F1 assms by meson
30   have 5: "□≺(For d ↔ ◇≺[STABᵈ]) w" using F3 assms by meson
31   have 6: "□≺((◇≺[WILLᵖ]) ∨ (◇≺[STABᵈ])) w" using 4 5 ForAx by (smt other.simps)
32   have 7: "□≺(◇≺[STABᵈ]) w" using 3 6 by blast
33   have 8: "□≺(For d) w" using 5 7 by simp
34   then show ?thesis by simp
35 qed
```

The consistency of Pierson's assumptions (theory and facts) together with the other postulates from the previously introduced Isabelle theories "GeneralKnowledge" and "Value-Ontology" is verified by generating a (non-trivial) model using *Nitpick* (Line 38). Further tests confirm that the decision for Pierson (and analogously for Post) is compatible with the premises and, moreover, that for neither party value conflicts are implied.

```
(************** Further checks (using model finder) ****************)
 (*defendant's theory and facts are logically consistent*)
lemma "d_theory ∧ ⌊d_facts⌋" nitpick[satisfy,card ι=3] oops (*(non-trivial) model found*)
 (*decision for defendant is compatible with premises and lacks value conflicts*)
lemma "⌊¬Conflict^p⌋ ∧ ⌊¬Conflict^d⌋ ∧ d_theory ∧ ⌊d_facts ∧ For d⌋"
  nitpick[satisfy,card ι=3] oops (* (non-trivial) model found*)
 (*situations where decision holds for plaintiff are compatible too*)
lemma "⌊¬Conflict^p⌋ ∧ ⌊¬Conflict^d⌋ ∧ d_theory ∧ ⌊d_facts ∧ For p⌋"
  nitpick[satisfy,card ι=3] oops (* (non-trivial) model found*)
```

Finally, observe that an analogous (deductively valid) argument for Post cannot follow from the given theory and situational facts. This is not surprising given that they have been deliberately chosen to suit Pierson's case. We show next, how it is indeed possible to construct a case (theory) suiting Post using our approach.

Ruling for Post

We model a possible counterargument by Post claiming an interpretation (i.e. a *distinction* in case law methodology) in that the animal, being vigorously pursued (with large dogs and hounds) by a professional hunter, is not "free-roaming". In doing this, the value preference $\lfloor[\text{WILL}^p] \prec [\text{STAB}^d]\rfloor$ (for appropriation of wild animals) as in the previous Pierson's argument does not obtain. Furthermore, Post's party postulates an alternative (suitable) value preference for hunting situations.

```
 (*case-specific 'world-vocabulary'*)
consts α::"e" (*appropriated animal (fox in this case) *)
consts Pursue::"c⇒e⇒σ" Capture::"c⇒e⇒σ"
(****************** pro-plaintiff (Post) argument ****************)
 (*acknowledges from defendant's theory*)
abbreviation "dT2 ≡ ⌊∀c. Pursue c α ⟶ Intent c⌋"
abbreviation "dT3 ≡ ⌊∀c. Capture c α ⟶ Poss c⌋"
 (*theory amendment: the animal was chased by a professional hunter (Post); protecting
   hunters' labor, thus fostering economic efficiency, prevails over legal certainty.*)
consts Hunter::"c⇒σ" hunting::"σ" (*new kind of situation: hunting*)
 (*plaintiff's theory*)
abbreviation "pT1 ≡ ⌊(∃c. Hunter c ∧ Pursue c α) ⟶ hunting⌋"
abbreviation "pT2 ≡ ∀x. ⌊hunting ⟶ ([STAB^x⁻¹] ≺ [EFFI^x⊕WILL^x])⌋" (*case-specific rule*)
abbreviation "pT3 ≡ ∀x. Promotes (hunting ∧ Hunter x) (For x) EFFI^x"
abbreviation "p_theory ≡ pT1 ∧ pT2 ∧ pT3 ∧ dT2 ∧ dT3"
 (*plaintiff's facts*)
abbreviation "pF1 w ≡ Fox α w"
abbreviation "pF2 w ≡ Hunter p w"
abbreviation "pF3 w ≡ Pursue p α w"
abbreviation "pF4 w ≡ Capture d α w"
abbreviation "p_facts ≡ pF1 ∧ pF2 ∧ pF3 ∧ pF4"
```

Note that an alternative legal rule (i.e. a possible argument for *overruling* in case law methodology) is presented in Line 16 above, entailing a value preference of the value combination *efficiency* (EFFI) and *will* (WILL) over *stability* (STAB): $\lfloor[\text{STAB}^d] \prec [\text{EFFI}^p \oplus \text{WILL}^p]\rfloor$. Following the argument put forward by the dissenting opinion in the original case, we might justify this new rule (inverting the initial value preference in the presence of EFFI) by pointing to the alleged public benefit of hunters getting rid of foxes, since the latter cause depredations in farms.

Accepting these modified assumptions the deductive validity of a decision for Post can in fact be proved and confirmed automatically, again, thanks to *Sledgehammer*:

```
(*decision for plaintiff (Post) can be proven automatically (needs approx. 20s)*)
theorem Post: "p_theory ⟶ ⌊p_facts → □⁻¹For p⌋"
  by (smt F1 F3 ForAx tBR SBR_def other.simps)
```

Similar to above, a detailed, interactive proof for the argument in favour of Post has been encoded and verified in *Isabelle/Isar*. We have also conducted further tests confirming the consistency of the assumptions and the absence of value conflicts (see sources in [1]).

5 Conclusion

Supporting interactive and automated value-oriented legal argumentation on the computer is a non-trivial challenge which we address, for reasons as defended e.g. by Bench-Capon [4], with symbolic AI techniques and formal methods. Motivated by recent pleas for *explainable and trustworthy AI*, our primary goal is to work towards the development of ethico-legal governors for future generations of intelligent system, or more generally, towards some form of (legally and ethically) *reasonable machines* [12], capable of exchanging rational justifications for the actions they take. While building up a capacity to engage in value-oriented legal argumentation is just one of a multitude of challenges this vision is faced with, it would clearly constitute an important stepping stone.

Custom software systems for legal case-based reasoning have been developed in the AI & law community, beginning with the influential HYPO system in the 1980s [36] (cf. also the review paper [3]). In later years, there was a gradual shift of interest from rule-based non-monotonic reasoning (e.g., logic programming) to argumentation-based approaches (see [35] for an overview); however, we are not aware of any other work that uses higher-order theorem proving and proof assistants (the argumentation logic of [26] is an early related effort that is worth mentioning). Another important aspect of our work concerns value-based legal reasoning and deliberation, where a considerable amount of work has been presented in response to the challenge posed by Berman and Hafner [16]. Our approach, based mainly on Lomfeld's theory [30, 29], has also been influenced by some of this work, in particular [34, 2, 5]. We think that some of the recent work that uses expressive deontic logics for value balancing (cf. [31] and the references therein) can be integrated into our approach.

The approach presented and illustrated in this work adapts and implements the multi-layered LogiKEy knowledge engineering methodology [13] to enable the application of off-the-shelf interactive and automated theorem proving technology for classical higher-order logic in ethical-legal reasoning. LogiKEy has been extended in this work to include an additional modeling layer, the value ontology. The value ontology forms a bridge between the legal and general world knowledge layer and the object logic layer in LogiKEy. Isabelle/HOL has proven to be an excellent base technology to support the presented formalization work and the conducted experiments. We are particularly pleased with the good performance of the *Nitpick* model finder and the integrated automated theorem provers (provided by *Sledgehammer*), which provided very useful feedback at all modeling layers, including fully automated proofs for formal justification of the discussed judgments.

Further work includes refining the modelling of Lomfeld's value theory in combination with providing more expressive (combinations of) object logics. With respect to the latter, the use of material implication to model defeasible or "default" rules (among others) has proven sufficient for the illustrative purposes of this paper, but it is important to note that more realistic modeling of legal cases must also provide mechanisms to deal with the inevitable emergence of conflicts and contradictions in normative reasoning (overruling, conflict resolution, etc.). In line with the LogiKEy approach, we can indeed introduce a defeasible conditional operator

by reusing the modal operators of $\mathcal{PL}$(as discussed, e.g., in [20, 28]), or alternatively by the SSE of a suitable conditional logic in HOL [6]. Various kinds of paraconsistent negations could also be considered for the non-explosive representation of (and recovery from) contradictions by purely object-logical means (cf. [21] for an appropriate SSE). It is the pluralistic nature of our approach, realised within a dynamic modelling framework, that enables and supports such improvements without requiring technical adjustments to the underlying base reasoning technology.

References

1 Isabelle/HOL sources for this formalisation work. `http://logikey.org`, 2021. Subfolder: Preference-Logics/EncodingLegalBalancing.

2 Trevor J. M. Bench-Capon. The missing link revisited: The role of teleology in representing legal argument. *Artificial Intelligence and Law*, 10(1-3):79–94, 2002.

3 Trevor J. M. Bench-Capon. HYPO's legacy: introduction to the virtual special issue. *Artificial Intelligence and Law*, 25(2):205–250, 2017.

4 Trevor J. M. Bench-Capon. The need for good old-fashioned AI and Law. In W. Hötzendorfer, C. Tschol, and F. Kummer, editors, *In International Trends in Legal Informatics: A Festschrift for Erich Schweighofer*. Weblaw AG, 2020.

5 Trevor J. M. Bench-Capon and Giovanni Sartor. A model of legal reasoning with cases incorporating theories and value. *Artificial Intelligence*, 150:97–143, 2003.

6 Christoph Benzmüller. Cut-elimination for quantified conditional logic. *Journal of Philosophical Logic*, 46(3):333–353, 2017. `doi:10.1007/s10992-016-9403-0`.

7 Christoph Benzmüller. Universal (meta-)logical reasoning: Recent successes. *Science of Computer Programming*, 172:48–62, 2019. `doi:10.1016/j.scico.2018.10.008`.

8 Christoph Benzmüller, Ali Farjami, Paul Meder, and Xavier Parent. I/O logic in HOL. *Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications (Special Issue: Reasoning for Legal AI)*, 6(5):715–732, 2019.

9 Christoph Benzmüller, Ali Farjami, and Xavier Parent. Åqvist's dyadic deontic logic E in HOL. *Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications (Special Issue: Reasoning for Legal AI)*, 6(5):733–755, 2019.

10 Christoph Benzmüller, Ali Farjami, and Xavier Parent. Dyadic deontic logic in hol: Faithful embedding and meta-theoretical experiments. In Matthias Armgardt, Hans Christian Nordtveit Kvernenes, and Shahid Rahman, editors, *New Developments in Legal Reasoning and Logic: From Ancient Law to Modern Legal Systems*, volume 23 of *Logic, Argumentation & Reasoning*. Springer Nature Switzerland AG, 2021. `doi:10.1007/978-3-030-70084-3`.

11 Christoph Benzmüller, David Fuenmayor, and Bertram Lomfeld. Encoding legal balancing: Automating an abstract ethico-legal value ontology in preference logic, 2020. Workshop on Models of Legal Reasoning (MLR 2020), hosted by 17th Conference on Principles of Knowledge Representation and Reasoning (KR 2020). Unpublished paper available at: `https://www.researchgate.net/publication/342380027`.

12 Christoph Benzmüller and Bertram Lomfeld. Reasonable machines: A research manifesto. In Ute Schmid, Franziska Klügl, and Diedrich Wolter, editors, *KI 2020: Advances in Artificial Intelligence – 43rd German Conference on Artificial Intelligence, Bamberg, Germany, September 21–25, 2020, Proceedings*, volume 12352 of *Lecture Notes in Artificial Intelligence*, pages 251–258. Springer, Cham, 2020. `doi:10.1007/978-3-030-58285-2_20`.

13 Christoph Benzmüller, Xavier Parent, and Leendert van der Torre. Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support. *Artificial Intelligence*, 287:103348, 2020. `doi:10.1016/j.artint.2020.103348`.

14 Christoph Benzmüller and Lawrence C. Paulson. Multimodal and intuitionistic logics in simple type theory. *The Logic Journal of the IGPL*, 18(6):881–892, 2010. `doi:10.1093/jigpal/jzp080`.

15 Christoph Benzmüller and Lawrence C. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013. doi:10.1007/s11787-012-0052-y.

16 Donald Berman and Carole Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings 4th ICAIL*, pages 50–59. New York: ACM Press, 1993.

17 Jasmin C. Blanchette, Sascha Böhme, and Lawrence C. Paulson. Extending Sledgehammer with SMT solvers. *Journal of Automated Reasoning*, 51(1):109–128, 2013.

18 Jasmin C. Blanchette, Cezary Kaliszyk, Lawrence C. Paulson, and Josef Urban. Hammering towards QED. *Journal of Formalized Reasoning*, 9(1):101–148, 2016.

19 Jasmin C. Blanchette and Tobias Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In Matt Kaufmann and Lawrence C. Paulson, editors, *ITP 2010*, volume 6172 of *LNCS*, pages 131–146. Springer, 2010.

20 Craig Boutilier. Toward a logic for qualitative decision theory. In *Principles of knowledge representation and reasoning*, pages 75–86. Elsevier, 1994. doi:10.1016/B978-1-4832-1452-8.50104-4.

21 David Fuenmayor. Topological semantics for paraconsistent and paracomplete logics. *Archive of Formal Proofs*, 2020. , Formal proof development. URL: https://isa-afp.org/entries/Topological_Semantics.html.

22 Bernhard Ganter and Rudolf Wille. *Formal concept analysis: mathematical foundations*. Springer Berlin, 2012.

23 Thomas F. Gordon and Douglas Walton. Pierson vs. Post revisited. *Frontiers in Artificial Intelligence and Applications*, 144:208, 2006.

24 Joseph Y. Halpern. Defining relative likelihood in partially-ordered preferential structures. *Journal of Artificial Intelligence Research*, 7:1–24, 1997.

25 Robert Koons. Defeasible Reasoning. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

26 P. Krause, S. Ambler, Elvang-Goransson, M., and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 1995. doi:10.1111/j.1467-8640.1995.tb00025.x.

27 Fenrong Liu. *Changing for the better: Preference dynamics and agent diversity*. PhD thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam, 2008.

28 Fenrong Liu. *Reasoning about Preference Dynamics*. Springer Netherlands, 2011. doi:10.1007/978-94-007-1344-4.

29 Bertram Lomfeld. *Die Gründe des Vertrages: Eine Diskurstheorie der Vertragsrechte*. Mohr Siebeck, Tübingen, 2015.

30 Bertram Lomfeld. Grammatik der Rechtfertigung: Eine kritische Rekonstruktion der Rechts(fort)bildung. *Kritische Justiz*, 52(4), 2019.

31 Juliano Maranhão and Giovanni Sartor. Value assessment and revision in legal interpretation. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL 2019, Montreal, QC, Canada, June 17-21, 2019*, pages 219–223, 2019. doi:10.1145/3322640.3326709.

32 L. Thorne McCarty. An implementation of Eisner v. Macomber. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 276–286, 1995.

33 Tobias Nipkow, Lawrence C. Paulson, and Markus Wenzel. *Isabelle/HOL: A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.

34 Henry Prakken. An exercise in formalising teleological case-based reasoning. *Artificial Intelligence and Law*, 10(1-3):113–133, 2002.

35 Henry Prakken and Giovanni Sartor. Law and logic: A review from an argumentation perspective. *Artificial Intelligence*, 227:214–225, 2015.

36 Edwina L. Rissland and Kevin D. Ashley. A case-based system for trade secrets law. In *Proceedings of the 1st international conference on Artificial Intelligence and Law*, pages 60–66, 1987.

37 Johan van Benthem, Patrick Girard, and Olivier Roy. Everything else being equal: A modal logic for *ceteris paribus* preferences. *J. Philos. Log.*, 38(1):83–125, 2009. doi:10.1007/s10992-008-9085-3.

38 Georg Henrik von Wright. *The logic of preference*. Edinburgh University Press, 1963.

39 Makarius Wenzel. Isabelle/Isar—a generic framework for human-readable proof documents. *From Insight to Proof-Festschrift in Honour of Andrzej Trybulec*, 10(23):277–298, 2007.