Modeling Intentional States with Subsystems of $\mathcal{ALC}$

Thomas Macaulay Ferguson[1,2][0000−0002−6494−1833]

[1] ILLC, University of Amsterdam, Amsterdam, The Netherlands
[2] Arché Research Centre, University of St. Andrews, St. Andrews, Scotland
tferguson@gradcenter.cuny.edu

Abstract. Although standard description logics are accurate tools to capture inferences from extensional knowledge bases, logics like $\mathcal{ALC}$ are unreasonably strong in *intentional contexts*. The successful modeling of cases in which, *e.g.*, a knowledge base is interpreted as describing a conversational agent's beliefs or goals thus requires a more constrained notion of closure. In this paper, we identify several inferences in $\mathcal{ALC}$ whose validity is questionable for such applications and describe two subsystems of $\mathcal{ALC}$ that better fit the use-case.

Keywords: Description logics · intentionality · hyperintensionality.

1 Introduction

Description logics like $\mathcal{ALC}$ and $\mathcal{SROIQ}$ are tools for knowledge representation, *i.e.*, the modeling of and reasoning about information. In their introduction to [2], Nardi and Brachman make a remark that reflects an important feature of standard description logics by counting them among a tradition of

> logic-based formalisms, which evolved out of the intuition that predicate calculus could be used unambiguously to capture facts about the world.[14, p. 2]

The phrase "predicate calculus" denotes the "classical" first-order logic of Frege and Russell. Although various description logics correspond to distinct *fragments* of classical first-order logic, the semantic foundation is almost universally a classical one. As a tool to "capture facts" about the world, this Boolean framework is well-suited to many use-cases. For example, a knowledge base representing metadata about an organization's assets will likely be interpreted *extensionally*, that is, the truth or falsity of an assertion make up the primary logical consideration. Thus, the classical framework is generally sound with respect to extensional contexts.

However, not all contexts about which one wishes to reason are extensional; many applications exist in which veridical considerations do not take center stage. If, *e.g.*, a conversational agent maintains a knowledge base representation of the beliefs of an interlocutor, it is implausible to count arbitrary tautologies as beliefs. As the expressivity increases to include, *e.g.*, number restrictions and role axioms, a logic like $\mathcal{SROIQ}$ will *grossly* overgenerate, closing knowledge bases under claims looking suspiciously like number theory and algebra, as well. In short, reasoning about *extensional facts about the world*—a task to which standard description logics are well-suited—is a distinct exercise from reasoning about *intentional states of cognitive agents*.

2 Challenges for $\mathcal{ALC}$ in Intentional Contexts

In this section, we review a number of ways in which $\mathcal{ALC}$—and thus all stronger description logics—is *deductively promiscuous* in intentional contexts. We briefly provide a précis of the description logic $\mathcal{ALC}$. Its language is built from pairwise disjoint classes of primitive *individual names*, *concept names*, and *role names*, natural interpretations of which are individuals (*i.e.*, constants in first-order logic terms), classes (*i.e.*, unary predicates), and relations (*i.e.*, binary predicates). These terms make up expressions when closed under syncategorematic terms allow us to represent many details of a domain.

The language includes primitive *top* and *bottom* concepts $\top$ and $\bot$, interpreted as a universal class and an empty class, respectively. The class of concepts is constructed by closing the primitive concepts under the following, where r is a role name:

$$C ::= \top | \bot | \neg C | C \sqcap C | C \sqcup C | \exists r.C | \forall r.C$$

$\neg C$ is the *complement* of C and $C \sqcap D$, $C \sqcup D$ are the *intersection* and *union* of C, D, respectively. We read $\exists r.C$ as the class of individuals a that bears the relation r to *some* C and $\forall r.C$ as the class of individuals a for which a bears r to *only* members of C. Then, where a, b are individual names, r is a role name, and C, D are concept names, *sentences* are formed as follows:

$$\varphi ::= a : C | \langle a, b \rangle : r | C \sqsubseteq D | C \equiv D$$

In the sequel, our choice of terms will be *self-documenting*, where the intended interpretations should be clear.

2.1 Examples

We identify cases in which $\mathcal{ALC}$ appears to overgenerate, licensing consequences too permissively for particular use-cases. That $\mathcal{ALC}$ is promiscuous should not be controversial; intentional interpretations of knowledge bases are *hyperintensional* in the sense of [7] and not closed under substitution of logical equivalents.

These concepts see implementation in OWL 2 [4] as the classes `owl:Thing` and `owl:Nothing`, respectively.

Case I: Assessing Mastery of a Domain Consider a conversational agent that is a component of a recommendation system, *e.g.*, an artificial assistant for an online storefront. The types of resources to which the agent directs users should be influenced by knowledge about those users. One particularly information-rich metric is the degree to which an interlocutor enjoys mastery over a particular domain. Popular, accessible examples of a particular genre of media, *e.g.*, are natural candidate recommendations for a user with no knowledge of that genre; those same examples would prove far less useful—quite likely, redundant—for a user with detailed command of the genre's esoterica.

Thus, the utility of a recommendation system's suggestions with respect to a domain will be enhanced for understanding the interlocutor's command of that domain. A clear indicator would be the breadth and degree of correctness of the interlocutor's *beliefs* about that domain, which could be measured from the number and accuracy of the interlocutor's assertions stored in an intentional knowledge base.

Example 1 *A conversational agent assists the interlocutor as she navigates an online storefront devoted to classical music. After the initial interactions, the agent has recorded the following ABox* $\mathcal{A}_0$ *as reflecting the agent's explicit beliefs about classical music:*

– $\mathsf{Mahler} : \forall \mathsf{composerOf}.\mathsf{Symphony}$
– $\langle \mathsf{Mahler}, \mathsf{Kindertotenlieder} \rangle : \mathsf{composerOf}$
– $\mathsf{Kindertotenlieder} : \neg \mathsf{Symphony}$
– $\mathsf{MahlerSymphony5} : \mathsf{Symphony} \sqcap \mathsf{Modernist}$

Suppose that the agent takes the percentage of correct *classifications* of musical pieces within the concept $\mathsf{Symphony}$ as a coarse metric for the agent's command of classical music. Given the intended interpretation of the terms, it appears that the interlocutor has correctly classified each of the two compositions mentioned in $\mathcal{A}_0$; intuitively, she has evidenced some expertise with respect to the domain.

Clearly, this task requires *some* deductive machinery; to infer the correct classification of $\mathsf{MahlerSymphony5}$ requires an inference from $\mathsf{MahlerSymphony5} : \mathsf{Symphony} \sqcap \mathsf{Modernist}$ to $\mathsf{MahlerSymphony5} : \mathsf{Symphony}$. But the deductive machinery of $\mathcal{ALC}$ leads to a conflict with this intuition. The interlocutor has evidently made a *mistake* over the course of the interaction by asserting that Mahler *only* composed symphonies. Thus, $\mathcal{A}_0$ is inconsistent, whence, *e.g.*, $\mathcal{A}_0 \vDash_{\mathcal{ALC}} \mathsf{MahlerSymphony5} : \neg \mathsf{Symphony}$, despite the interlocutor's *correct* classification of $\mathsf{MahlerSymphony5}$. Any simple mistake that triggers a similar inconsistency in an intentional knowledge base will, therefore, entirely warp the picture of the interlocutor's domain knowledge.

Case II: Assessing an Interlocutor's Goals Consider a further case in which an artificial conversational agent is tasked with directing interlocutors to resources. The appropriateness of resources will undoubtedly be measured in terms of how well those resources fit the interlocutor's *goals*. Such cases might see the initial interactions leveraged to populate an intentional knowledge base representing the interlocutor's *goals* or *requirements*.

Example 2 *A conversational agent assists an interlocutor* Self *in an online technology storefront. The interlocutor's goals are represented by the following ABox:*

$$\mathcal{A}_1 = \{\mathsf{Self} : \exists \mathsf{purchases}.\mathsf{GamingComputer} \sqcap \mathsf{DesktopComputer}\}$$

As in Example 1, there is a need for *some* deductions under which $\mathcal{A}_1$ is closed. Suppose that the artificial agent has a directive such that whenever an interlocutor counts $\mathsf{Self} : \exists \mathsf{purchases}.\mathsf{DesktopComputer} \sqcap \mathsf{GamingComputer}$ as a goal, the interlocutor should be directed to a particular resource (*e.g.*, a webpage). However, this trigger condition is not explicitly included in $\mathcal{A}_1$; it must be *inferred*.

However, given $\mathcal{A}_1$, $\mathcal{ALC}$ delivers additional inferences under which a collection of goals clearly should not be closed. For example, we should hesitate to license an inference to the further goal of

$$\mathsf{Self} : \exists \mathsf{purchases}.(\mathsf{TernaryComputer} \sqcup (\mathsf{GamingComputer} \sqcap \mathsf{DesktopComputer})) \sqcap (\mathsf{GamingComputer} \sqcap \mathsf{DesktopComputer})$$

Despite equivalence with the member of the interlocutor's explicit goal from $\mathcal{A}_1$, to count the above as a *goal* has several shortcomings. For one, the sentence is a bit of a *red herring*—one can imagine that, if read lazily, resources may be wasted on trying to connect the user with an instance of $\mathsf{TernaryComputer}$.

This candidate becomes more problematic when one considers the matter of *awareness*. The notion, identified in [9] as a key cognitive feature that can constrain agents' capacity for reasoning, is defined so that "an agent is aware of a formula if he can compute whether or not it is true in a given situation within a certain time or space bound."[9, p. 41] Simply put:

> How can someone say that he knows or doesn't know about p if p is a concept he is completely unaware of?[9, p. 40]

This notion of awareness, in the present context, can be recast as a matter of *conceptual proficiency*. If a knowledge base $\langle \mathcal{T}, \mathcal{A} \rangle$ is, *e.g.*, a representation of an agent's goals, it is implausible to say that $a : C$ is among those goals in cases in which the agent *fails to grasp* the use of C, *i.e.*, is *unaware* of how to distinguish cases in which a falls under C from cases in which a does not fall under C.

Given the intended interpretation, the concept $\mathsf{TernaryComputer}$ is a highly *esoteric* one. As the typical consumer of computer products—*e.g.* the interlocutor of Example 2—is unlikely to count the above as a goal, to assume its equivalence would be *counterproductive* to the aims of the conversational agent. Thus, pragmatic considerations demonstrate the problematic overgeneration following from the application of $\mathcal{ALC}$ to this domain.

Note that the suggestion that $\mathcal{ALC}$ is too strong is *not* to say that $\mathcal{ALC}$ fails as a characterization of *normative* dimensions of logic. Indeed, it remains plausible that *e.g.* α *ought to accept* φ given other beliefs. What is *implausible*, however, is that $\mathcal{ALC}$—and *a fortiori* more expressive description logics like $\mathcal{SROIQ}$—captures the *descriptive* dimension, that is, a characterization of the closure of the *occurrent beliefs* of an agent α.

3 $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$

We now introduce two subsystems of $\mathcal{ALC}$ (in the same language) that are arguably more appropriate to the closure of intentional knowledge bases than stock $\mathcal{ALC}$. The propositional deductive system that serves as the basis for both proposals is the logic of *analytic containment* AC introduced by Richard Angell in [1]. Angell's case that AC captures a semantic notion of *synonymy* has been reinforced in a number of works—*e.g.*, [6], [12], [15].

The interpretation of AC offered in [10] is based on Kleene's three truth values $\mathcal{V}_3$. In [13], Kleene introduces three-valued matrices for connectives to account for cases in which a recursive procedure calculating truth values fails to converge:

> In this section, we shall introduce new senses of the propositional connectives, in which, *e.g.*, $Q(x) \vee R(x)$ will be defined in some cases when $Q(x)$ or $R(x)$ is undefined. It will be convenient to use truth tables, with three "truth values" $\mathfrak{t}$ ('true'), $\mathfrak{f}$ ('false') and $\mathfrak{e}$ ('undefined'), in describing the senses which the connectives shall now have.[13, p. 332]

Kleene considers that for a predicate there is a "range of definition" over which it is defined. A fruitful and relevant interpretation of these ranges is that $\varphi(c)$ evaluates to $\mathfrak{e}$ when an agent lacks *proficiency* with the concept $\varphi(x)$ and is unable to determine a truth value. If an agent is not familiar with the use of a predicate—or does not have a clear grasp of how a predicate may apply to certain objects—an atomic formula may be viewed as not truth-evaluable.

[10] assigns statements φ *pairs* of truth values—*i.e.* values from the set $\mathcal{V}_3^2$—where the first coordinate indicates the degree of *support* for φ and the second coordinate represents the degree of *refutation* or *evidence against* φ. This yields nine available values for $\mathcal{AC}$:

$\langle \mathfrak{t}, - \rangle$: α has evidence supporting φ	$\langle -, \mathfrak{t} \rangle$: α has evidence refuting φ
$\langle \mathfrak{e}, - \rangle$: α unaware of φ's truth conditions	$\langle -, \mathfrak{e} \rangle$: α unaware of φ's falsity conditions
$\langle \mathfrak{f}, - \rangle$: α lacks evidence supporting φ	$\langle -, \mathfrak{f} \rangle$: α lacks evidence refuting φ

E.g., that $\mathsf{Brouwer} : \mathsf{Topologist}$ is assigned the value $\langle \mathfrak{t}, \mathfrak{f} \rangle$ might correspond to a case in which an agent—who is aware of how to classify objects as elements of either $\mathsf{Topologist}$ or $\neg\mathsf{Topologist}$—has unequivocal support for $\mathsf{Brouwer}$'s falling under that concept. That $\mathsf{Caesar}:\mathsf{PrimeNumber}$ is assigned $\langle \mathfrak{e}, \mathfrak{e} \rangle$ may be understood as indicating that an agent is unaware of how Caesar's being a $\mathsf{PrimeNumber}$ could either be demonstrated or refuted.

The present interpretation of the truth values of [10] is relatively conservative; the account essentially enriches Belnap's concern for support and refutation of [3] with a device representing *awareness* in the sense of [9]. The interpretation diverges from the standard as concerns the availability of a truth value like $\langle \mathfrak{f}, \mathfrak{e} \rangle$, as it countenances cases in which an agent can be aware of a statement's *truth conditions* while lacking awareness of its *falsity conditions*. If one supposes that awareness of a concept must be *in toto*, then eliminating the incompatible values (*i.e.*, $\langle \mathfrak{t}, \mathfrak{e} \rangle$, $\langle \mathfrak{f}, \mathfrak{e} \rangle$, $\langle \mathfrak{e}, \mathfrak{t} \rangle$, $\langle \mathfrak{e}, \mathfrak{f} \rangle$) yields a set of values $\mathcal{V}_3^{2\star}$ inducing five-valued logic $\mathsf{S}^\star_{\mathtt{fde}}$, offered by Daniels in [8] as the logic of the intentional contexts of *fictions*.

3.1 Semantics

As the truth values of AC and $\mathsf{S}^{\star}_{\mathtt{fde}}$ are based on weak Kleene logic wK, the model theory for the description logics $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^{\star}}$ will reference wK as well. The weak truth functions $\dot{\sim}$, $\dot{\wedge}$, and $\dot{\vee}$ are defined as follows:

Definition 1 *The* weak Kleene truth tables *are:*

$\dot{\sim}$	
$\mathfrak{t}$	$\mathfrak{f}$
$\mathfrak{e}$	$\mathfrak{e}$
$\mathfrak{f}$	$\mathfrak{t}$

$\dot{\wedge}$	$\mathfrak{t}$	$\mathfrak{e}$	$\mathfrak{f}$
$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{e}$	$\mathfrak{f}$
$\mathfrak{e}$	$\mathfrak{e}$	$\mathfrak{e}$	$\mathfrak{e}$
$\mathfrak{f}$	$\mathfrak{f}$	$\mathfrak{e}$	$\mathfrak{f}$

$\dot{\vee}$	$\mathfrak{t}$	$\mathfrak{e}$	$\mathfrak{f}$
$\mathfrak{t}$	$\mathfrak{t}$	$\mathfrak{e}$	$\mathfrak{t}$
$\mathfrak{e}$	$\mathfrak{e}$	$\mathfrak{e}$	$\mathfrak{e}$
$\mathfrak{f}$	$\mathfrak{t}$	$\mathfrak{e}$	$\mathfrak{f}$

As observed in [11], the available accounts of quantification for Kleene logics are not sufficient to support the class theory necessary for a description logic. Instead, the following quantifiers on $\mathcal{V}_3^2$ are offered:

Definition 2 *The* restricted Kleene quantifiers *are functions* $\dot{\exists}$ *and* $\dot{\forall}$ *mapping a nonempty sets* $X \subseteq \mathcal{V}_3^2$ *to truth values from* $\mathcal{V}_3$ *as follows:*

$$\dot{\exists}(X) = \begin{cases} \mathfrak{t} & \text{if } \langle \mathfrak{t}, \mathfrak{t} \rangle \in X \\ \mathfrak{e} & \text{if for all } \langle u, v \rangle \in X, \text{ either } u = \mathfrak{e} \text{ or } v = \mathfrak{e} \\ \mathfrak{f} & \text{if } \langle \mathfrak{t}, \mathfrak{t} \rangle \notin X \text{ and for some } \langle u, v \rangle \in X,\ u \neq \mathfrak{e} \text{ and } v \neq \mathfrak{e} \end{cases}$$

$$\dot{\forall}(X) = \begin{cases} \mathfrak{t} & \text{if } \langle \mathfrak{t}, \mathfrak{f} \rangle, \langle \mathfrak{t}, \mathfrak{e} \rangle \notin X \text{ and for some } \langle u, v \rangle \in X,\ u \neq \mathfrak{e} \text{ and } v \neq \mathfrak{e} \\ \mathfrak{e} & \text{if for all } \langle u, v \rangle \in X, \text{ either } u = \mathfrak{e} \text{ or } v = \mathfrak{e} \\ \mathfrak{f} & \text{if } \{\langle \mathfrak{t}, \mathfrak{f} \rangle, \langle \mathfrak{t}, \mathfrak{e} \rangle\} X \neq \varnothing \text{ and for some } \langle u, v \rangle \in X,\ u \neq \mathfrak{e} \text{ and } v \neq \mathfrak{e} \end{cases}$$

The guiding principle here is that a statement with restricted quantification like $[\forall x \varphi(x)]\psi(x)$—read: "all φ's are ψ's"—is truth-evaluable only in case the two domains can be meaningfully compared. Evaluating such a sentence presupposes a domain of comparison; if a user is *unaware* of a range over which $\varphi(x)$ and $\psi(x)$ could be compared, it is impossible to demonstrate *or refute* the statement.

We've noted that the presentation of AC is *bilateral*—one coordinate tracking truth, the other falsity. The truth function for negation $\ddot{\sim}$, then, merely *toggles* truth and falsity, *i.e.*, exchanges the coordinates of a truth value. DeMorgan's laws suggest that extending this bilateralism to the truth functions requires that one weak Kleene truth function operates on the arguments' *first coordinates* while *its dual truth function* operates on their *second coordinates.*

Definition 3 *The* AC *truth functions* $\ddot{\sim}$, $\ddot{\wedge}$, *and* $\ddot{\vee}$ *are defined so that:*

- $\ddot{\sim}(\langle u, v \rangle) = \langle v, u \rangle$
- $\langle u_0, v_0 \rangle \mathbin{\ddot{\wedge}} \langle u_1, v_1 \rangle = \langle u_0 \mathbin{\dot{\wedge}} u_1, v_0 \mathbin{\dot{\vee}} v_1 \rangle$
- $\langle u_0, v_0 \rangle \mathbin{\ddot{\vee}} \langle u_1, v_1 \rangle = \langle u_0 \mathbin{\dot{\vee}} u_1, v_0 \mathbin{\dot{\wedge}} v_1 \rangle$

Quantifiers will be defined over sets X of pairs of values from $\mathcal{V}_3^2$—*i.e.*, pairs of pairs of Kleene truth values. For convenience, let $\pi^{i,j}$ map values $\langle \langle u_0, v_0 \rangle, \langle u_1, v_1 \rangle \rangle$ to the pair of the ith coordinate of $\langle u_0, v_0 \rangle$ and the jth cordinate of $\langle u_1, v_1 \rangle$ and let $\pi^{i,j}[X]$ be the image of X under $\pi^{i,j}$. Then:

Definition 4 *The* AC *restricted quantifiers* $\ddot{\exists}$ *and* $\ddot{\forall}$ *are defined so that:*

- $\ddot{\exists}(X) = \langle \dot{\exists}(\pi^{0,0}[X]), \dot{\forall}(\pi^{0,1}[X]) \rangle$
- $\ddot{\forall}(X) = \langle \dot{\forall}(\pi^{0,0}[X]), \dot{\exists}(\pi^{0,1}[X]) \rangle$

Most of the interpretative work is being done by the weak Kleene interpretation, with the AC functions merely imposing a bilateral reading. The presentation of the proper AC model theory is necessarily brief; for detailed discussion of the connectives or quantifiers, see [10] or [11], respectively.

We now can define the model theory for $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$, starting with an *interpretation*:

Definition 5 *An interpretation over* $\mathcal{V}_3^2$ *is a function* $\mathcal{I}$ *paired with domain* $\Delta^{\mathcal{I}}$, *described by the following clauses:*

- $\mathcal{I}(a : \top) = \langle \mathfrak{e}, v \rangle$ *and* $\mathcal{I}(a : \bot) = \langle v, \mathfrak{e} \rangle$ *for some* $v \in \mathcal{V}_3$
- $\mathcal{I}(a : C) \in \mathcal{V}_3^2$ *and* $\mathcal{I}(\langle a, b \rangle : r) \in \mathcal{V}_3^2$
- $\mathcal{I}(a : \neg C) = \ddot{\sim} \mathcal{I}(a : C)$
- $\mathcal{I}(a : C \sqcap D) = \mathcal{I}(a : C) \mathbin{\ddot{\wedge}} \mathcal{I}(a : D)$
- $\mathcal{I}(a : C \sqcup D) = \mathcal{I}(a : C) \mathbin{\ddot{\vee}} \mathcal{I}(a : D)$
- $\mathcal{I}(a : \exists r.C) = \ddot{\exists}(\{\langle \mathcal{I}(\langle a, b \rangle : r), \mathcal{I}(b : C) \rangle \mid b \in \Delta\})$
- $\mathcal{I}(a : \forall r.C) = \ddot{\forall}(\{\langle \mathcal{I}(\langle a, b \rangle : r), \mathcal{I}(b : C) \rangle \mid b \in \Delta\})$
- $\mathcal{I}(C \sqsubseteq D) = \ddot{\forall}(\{\langle \mathcal{I}(a : C), \mathcal{I}(a : D) \rangle \mid a \in \Delta\})$
- $\mathcal{I}(C \equiv D) = \mathcal{I}(C \sqsubseteq D) \mathbin{\ddot{\wedge}} \mathcal{I}(D \sqsubseteq C)$

We call an interpretation an AC interpretation if it ranges over $\mathcal{V}_3^2$ and an $\mathsf{S}^\star_{\mathtt{fde}}$ interpretation if it ranges over $\mathcal{V}_3^{2\star}$. In either case, we say that a sentence φ is *true* in $\mathcal{I}$ if the first coordinate of $\mathcal{I}(\varphi) = \mathfrak{t}$; a set of sentences is true if each sentence is true.

These definitions provide very natural accounts of semantic consequence for the two systems as *preservation of truth*. For knowledge bases $\langle \mathcal{T}, \mathcal{A} \rangle$, we define:

Definition 6 $\langle \mathcal{T}, \mathcal{A} \rangle \vDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ *if for all* AC *models,* φ *is true if* $\mathcal{T} \cup \mathcal{A}$ *is true*

Definition 7 $\langle \mathcal{T}, \mathcal{A} \rangle \vDash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$ *if for all* $\mathsf{S}^\star_{\mathtt{fde}}$ *models,* φ *is true if* $\mathcal{T} \cup \mathcal{A}$ *is true*

3.2 Proof Theory

The logics $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$ can be provided sound and complete tableau calculi over a common set of rules, differing only on closure conditions for branches. The calculi are *signed*, where each node is of the form $[v]\ \varphi$, for a sentence φ and Kleene truth value v. For elegance, we also include two additional signs: $\mathfrak{m}$ (understood as not-$\mathfrak{e}$) and $\mathfrak{n}$ (understood as not-$\mathfrak{t}$). Following the notation of [5], we let $+$ indicate the introduction of a branch on a tableau and $\circ$ indicate that two signed formulae are to appear on the same branch.

Definition 8 *The collection of rules* **ALC-AC** *is defined by the following rules:*

$$\frac{[\mathfrak{m}]\ \ \varphi}{[\mathfrak{t}]\ \ \varphi + [\mathfrak{f}]\ \ \varphi} \quad \frac{[\mathfrak{n}]\ \ \varphi}{[\mathfrak{f}]\ \ \varphi + [\mathfrak{e}]\ \ \varphi} \quad \frac{}{[\mathfrak{t}]\ \ a : \top} \quad \frac{}{[\mathfrak{t}]\ \ a : \neg\bot} \quad \frac{[v]\ \ a : \neg\neg C}{[v]\ \ a : C}$$

$$\frac{[v]\ \ a : \neg(C \sqcap D)}{[v]\ \ a : (\neg C \sqcup \neg D)} \quad \frac{[v]\ \ a : \neg(C \sqcup D)}{[v]\ \ a : (\neg C \sqcap \neg D)} \quad \frac{[v]\ \ a : \neg(\exists r.C)}{[v]\ \ a : \forall r.\neg C} \quad \frac{[v]\ \ a : \neg(\forall r.C)}{[v]\ \ a : \exists r.\neg C}$$

$$\frac{[v]\ \ a : C \sqcap D}{+_{v_0 \dot{\wedge} v_1 = v}\{[v_0]\ \ a : C \circ [v_1]\ \ a : D\}} \quad \frac{[v]\ \ a : C \sqcup D}{+_{v_0 \dot{\vee} v_1 = v}\{[v_0]\ \ a : C \circ [v_1]\ \ a : D\}}$$

$$\frac{[\mathfrak{t}]\ \ a : \exists r.C}{[\mathfrak{t}]\ \ \langle a, c\rangle : r \circ [\mathfrak{t}]\ \ c : C} \quad \frac{[\mathfrak{f}]\ \ a : \exists r.C}{[\mathfrak{m}]\ \ \langle a, c\rangle : r \circ [\mathfrak{m}]\ \ c : C \circ ([\mathfrak{n}]\ \ \langle a, d\rangle : r + [\mathfrak{f}]\ \ d : C)}$$

$$\frac{[\mathfrak{e}]\ \ a : \exists r.C}{[\mathfrak{e}]\ \ \langle a, d\rangle : r + [\mathfrak{e}]\ \ d : C}$$

$$\frac{[\mathfrak{t}]\ \ a : \forall r.C}{[\mathfrak{m}]\ \ \langle a, c\rangle : r \circ [\mathfrak{m}]\ \ c : C \circ ([\mathfrak{n}]\ \ \langle a, d\rangle : r + [\mathfrak{t}]\ \ d : C)} \quad \frac{[\mathfrak{e}]\ \ a : \forall r.C}{[\mathfrak{e}]\ \ \langle a, d\rangle : r + [\mathfrak{e}] d : C}$$

$$\frac{[\mathfrak{f}]\ \ a : \forall r.C}{[\mathfrak{m}]\ \ \langle a, c\rangle : r \circ [\mathfrak{m}]\ \ c : C \circ [\mathfrak{t}]\ \ \langle a, c'\rangle : r \circ [\mathfrak{n}]\ \ c' : C}$$

$$\frac{[\mathfrak{t}]\ \ C \sqsubseteq D}{[\mathfrak{m}]\ \ c : C \circ [\mathfrak{m}]\ \ c : D \circ ([\mathfrak{n}]\ \ d : C + [\mathfrak{t}]\ \ d : D)} \quad \frac{[\mathfrak{e}]\ \ C \sqsubseteq D}{[\mathfrak{e}]\ \ d : C + [\mathfrak{e}]\ \ d : D}$$

$$\frac{[\mathfrak{f}]\ \ C \sqsubseteq D}{[\mathfrak{m}]\ \ c : C \circ [\mathfrak{m}]\ \ c : D \circ [\mathfrak{t}]\ \ c' : C \circ [\mathfrak{n}]\ \ c' :} \quad \frac{[v]\ \ C \equiv D}{+_{v_0 \dot{\wedge} v_1 = v}\{[v_0]\ \ C \sqsubseteq D \circ [v_1]\ \ D \sqsubseteq C\}}$$

where v is any element of $\mathcal{V}_3$, c or c' are new to a branch, and d is arbitrary.

To introduce the calculi, we need a further definition:

Definition 9 *A tableau with rules from* **ALC-AC** fits $\langle \mathcal{T}, \mathcal{A}\rangle \vdash \varphi$ *if its initial segment is some permutation of the signed sentences* $[\mathfrak{t}]\ \ \tau$ *for each* $\tau \in \mathcal{T}$, $[\mathfrak{t}]\ \ \alpha$ *for each* $\alpha \in \mathcal{A}$, *and* $[\mathfrak{n}]\ \ \varphi$.

We need two notions of closure of a branch in order to define tableau calculi for the two systems.

Definition 10 *A branch of a tableau with rules from* **ALC-AC** *is* AC-*closed if there is a sentence* φ *and distinct* $v, v' \in \mathcal{V}$ *such that* $[v]\ \ \varphi$ *and* $[v']\ \ \varphi$ *are on the branch.*

Definition 11 *A branch of a tableau with rules from* **ALC-AC** *is* S-*closed if it is either* AC-*closed* or *there is a sentence* $a : C$ *and distinct* $v, v' \in \mathcal{V}$ *such that* $[v]\ \ a : C$ *and* $[v']\ \ a : \neg C$ *are on the branch where either* $v = \mathfrak{e}$ *or* $v' = \mathfrak{e}$.

If every branch of a tableau is AC-closed (respectively, S-closed), we simply say that the tableau itself is AC-closed (respectively, S-closed). Deducibility for the two systems is naturally defined as the appropriate type of closure. For a knowledge base $\langle \mathcal{T}, \mathcal{A}\rangle$,

Definition 12 $\langle \mathcal{T}, \mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ *if every fitting tableau is* AC-*closed*

Definition 13 $\langle \mathcal{T}, \mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$ *if every fitting tableau is* S-*closed*

3.3 Formal Results on $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$

The systems $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$ arise from—and are thus embeddable in—the systems with restricted quantification introduced in [11]. Although the latter logics lack logical constants corresponding to $\top$ and $\bot$, they can be enriched with privileged unary predicates $\mathbf{T}$ and $\mathbf{F}$ so that:

- For all $\mathcal{I}$ and t, $\mathcal{I}(\mathbf{T}(t)) = \langle \mathfrak{t}, v\rangle$ and $\mathcal{I}(\mathbf{F}(t)) = \langle v, \mathfrak{t}\rangle$ for some $v \in \mathcal{V}_3$.
- $[\mathfrak{t}]$ $\mathbf{T}(t)$ and $[\mathfrak{t}]$ $\neg\mathbf{T}(t)$ are added as axioms to the tableau calculi

Such an expansion clearly preserves soundness and completeness. We will thus assume in the sequel that these predicates are part of the first-order language; references to the systems of [11] include the additional features without loss of generality.

We can recover soundness and completeness by taking advantage of this embedding with the following translation:

Definition 14 *The translation $\cdot^\tau$ from $\mathcal{L}_{\mathcal{ALC}}$ to the first-order language $\mathcal{L}_{\mathsf{ACrQ}}$*

- $(a : C)^\tau = C(a)$ *and* $(\langle a, b\rangle : r)^\tau = r(a, b)$ *for primitive* C *and* r
- $(a : \neg C)^\tau = {\sim}(a : C)^\tau$
- $(a : C \sqcap D)^\tau = (a : C)^\tau \wedge (a : D)^\tau$ *and* $(a : C \sqcup D)^\tau = (a : C)^\tau \vee (a : D)^\tau$
- $(a : \forall r.C)^\tau = [\forall x\ \ r(a, x)](x : C)^\tau$ *and* $(a : \exists r.C)^\tau = [\exists x\ \ r(a, x)](x : C)^\tau$
- $(C \sqsubseteq D)^\tau = [\forall x\ \ (x : C)^\tau](x : D)^\tau$ *and* $(C \equiv D)^\tau = (C \sqsubseteq D)^\tau \wedge (D \sqsubseteq C)^\tau$

The fidelity of $\cdot^\tau$ is witnessed by four lemmas. The first two ensure *model-theoretic* agreement:

Lemma 1 $\langle \mathcal{T}, \mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ *iff* $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vDash_{\mathsf{AC}} \varphi^\tau$

Proof. Suppose that for $\mathcal{ALC}_{\mathsf{AC}}$ and AC valuations $\mathcal{I}$ and $\mathcal{I}'$, $\mathcal{I}(\varphi) = \mathcal{I}'(\varphi^\tau)$ for all atoms φ, *i.e.*, whenever φ is of the form $a : C$ or $\langle a, b\rangle : r$ for primitive C and r. Then a simple induction on complexity of formulae establishes that the agreement lifts to complex sentences as well. Thus, every countermodel to $\langle \mathcal{T}, \mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ is isomorphic to a countermodel to $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vDash_{\mathsf{AC}} \varphi^\tau$ and *vice versa*, whence $\langle \mathcal{T}, \mathcal{A}\rangle \nvDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ iff $\mathcal{T}^\tau \cup \mathcal{A}^\tau \nvDash_{\mathsf{AC}} \varphi^\tau$.

Lemma 2 $\langle \mathcal{T}, \mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$ *iff* $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vDash_{\mathsf{S}^\star_{\mathsf{fde}}} \varphi^\tau$

Proof. As $\mathcal{ALC}_{\mathsf{S}^\star}$ models are *a fortiori* $\mathcal{ALC}_{\mathsf{AC}}$ models (*mutatis mutandis* for $\mathsf{ALC}_{\mathsf{S}^\star}$), this lemma follows immediately from Lemma 2.

And two further lemmas ensure *proof-theoretic* agreement:

Lemma 3 $\langle \mathcal{T}, \mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ *iff* $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vdash_{\mathbf{ACrQ}} \varphi^\tau$

Proof. The rules **ALC-AC** are selected to agree with those of the tableaux calculus **ACrQ** of [11]. By comparing the two calculi, whenever a rule is applied to a signed sentence $[v]\ \varphi$, yielding further branches with signed sentences $[v_i]\ \varphi_i$, an analogous rule applied to $[v]\ \varphi^\tau$ yields signed sentences $[v_i]\ \varphi_i^\tau$. Thus, applying $\cdot^\tau$ to every sentence converts any $\mathcal{ALC}_{\mathsf{AC}}$ tableau to a **ACrQ** tableau. Conversely, any **ACrQ** tableau in which all initial formulae are in the range of $\cdot^\tau$ will include *only* sentences in the range of $\cdot^\tau$; moreover, translating each node with $\cdot^{\tau^{-1}}$ will translate the tableau to an isomorphic $\mathcal{ALC}_{\mathsf{AC}}$ tableau. It follows that $\langle\mathcal{T},\mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ will hold precisely when $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vdash_{\mathbf{ACrQ}} \varphi^\tau$ holds.

Lemma 4 $\langle\mathcal{T},\mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$ *iff* $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vdash_{\mathbf{SrQ}} \varphi^\tau$

Proof. In [11], the tableau calculus **SrQ** is by adding to **ACrQ** rules that, in the present terminology, ensure AC-closure of a branch whenever $[\mathfrak{e}]\ \varphi$ and $[\mathfrak{e}]\ \neg\varphi$ appear on that branch. *I.e.*, a branch AC-closes with the rules if and only if it S-closes without the rules. Thus, Lemma 3 applies to establish the same correspondence between $\mathcal{ALC}_{\mathsf{S}^\star}$ and **SrQ**.

Lemmas 1–4 suffice to show the adequacy of the proof theory with respect to the semantics. We conclude the section with the following theorems:

Theorem 1 $\langle\mathcal{T},\mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$ *iff* $\langle\mathcal{T},\mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$

Proof. By Lemma 3, the left-hand side is equivalent to $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vdash_{\mathbf{ACrQ}} \varphi^\tau$. By the results of [11], this is equivalent to $\mathcal{T}^\tau \cup \mathcal{A}^\tau \vDash_{\mathsf{AC}} \varphi^\tau$. Finally, appeal to Lemma 1 ensures equivalence with the right-hand side $\langle\mathcal{T},\mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{AC}}} \varphi$.

Theorem 2 $\langle\mathcal{T},\mathcal{A}\rangle \vdash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$ *iff* $\langle\mathcal{T},\mathcal{A}\rangle \vDash_{\mathcal{ALC}_{\mathsf{S}^\star}} \varphi$

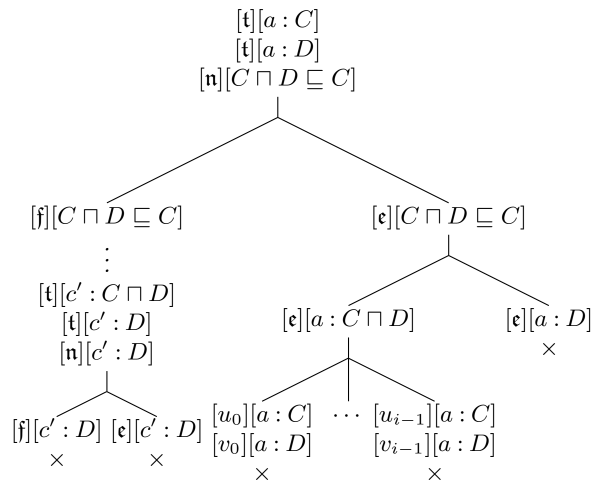
Proof. Follow the form of Theorem 1, including appeals to Lemmas 2 and 4.

4 Concluding Remarks

As knowledge-driven, conversational artificial intelligence is introduced to more domains, the necessity of knowledge representation over intentional contexts becomes increasingly pressing. The above discussion is intended as an opening move to address this problem by proffering of two natural candidate solutions.

Much remains to be done, *e.g.*, determination of complexity of the calculi, producing implementations of the systems, and empirical verification—or refutation—that $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$ are found satisfactory by users. We conclude, though, by providing some examples that bear on the issues described in Examples 1 and 2, showing that the above systems are sufficiently strong for the cases but weak enough to avoid the overgeneration of stock $\mathcal{ALC}$.

Example 3 *Let* $\mathcal{A} = \{a : C, a : D\}$. *Then* $\mathcal{A} \vDash_{\mathcal{ALC}_{\mathsf{AC}}} C \sqcap D \sqsubseteq C$. *We provide the following tableau demonstrating this fact:*



N.b. *that because for each* $j < i$*, either* $u_j = \mathfrak{e}$ *or* $v_j = \mathfrak{e}$*, each branch beneath* $[\mathfrak{e}][a : C \sqcap D]$ *will close.*

A similar tableau proof can be easily produced (and is thus left to the reader) demonstrating that if $\mathcal{A} = \{a : C \sqcap D\}$, then $\mathcal{A} \vDash_{\mathcal{ALC}_{\mathsf{AC}}} a : D \sqcap C$, as Example 2 required.

However, the forms of problematic inferences identified in Examples 1 and 2 can be shown to *fail* in $\mathcal{ALC}_{\mathsf{AC}}$ and $\mathcal{ALC}_{\mathsf{S}^\star}$. For example, Example 1 identified that in $\mathcal{ALC}$, one inconsistent classification of a term corrupts all of an agent's *correct* classifications in an ABox. Such mistakes are not catastrophic in $\mathcal{ALC}_{\mathsf{AC}}$:

Example 4 *Let* $\mathcal{A} = \{a : C, b : \neg C, \langle m, b\rangle : r, m : \forall r.C\}$*. Then* $\mathcal{A} \nvDash_{\mathcal{ALC}_{\mathsf{AC}}} a : \neg C$*. A model in which* $\mathcal{I}(a : C) = \langle \mathfrak{t}, \mathfrak{f}\rangle$*,* $\mathcal{I}(b : C) = \langle \mathfrak{t}, \mathfrak{t}\rangle$*, and* $\mathcal{I}(\langle m, b\rangle : r) = \langle \mathfrak{t}, \mathfrak{f}\rangle$ *can be recognized as a counterexample.*

Example 2 identified how the closure of an ABox under $\mathcal{ALC}$ may introduce subject-matter with which an agent is unaware or inproficient. The following shows how $\mathcal{ALC}_{\mathsf{AC}}$ prevents this:

Example 5 *Let* $\mathcal{A} = \{a : C\}$*. Then* $\mathcal{A} \nvDash_{\mathcal{ALC}_{\mathsf{AC}}} a : C \sqcap (C \sqcup D)$*. A model in which* $\mathcal{I}(a : C) = \langle \mathfrak{t}, \mathfrak{f}\rangle$ *and* $\mathcal{I}(a : D) = \langle \mathfrak{e}, \mathfrak{e}\rangle$ *can be confirmed to be a counterexample.*

Of course, the *adequacy* of the systems introduced here will be determined by user's experiences with *implementations*. Prior to such surveys, however, they at least can be seen to enjoy a *superior fit* to the use-cases than $\mathcal{ALC}$ and can therefore be recognized as a step in the right direction.

A reviewer has remarked that Example 3 reflects an inability to reason from an *empty ABox*. In the systems described in this paper, one is not *prevented* from reasoning from an empty TBox and ABox. But an empty TBox and ABox pair is agnostic about an agent's *awareness*, constituting insufficient grounds to infer anything about the concepts with which an agent is familiar or proficient.

References

1. Angell, R.B.: Three systems of first degree entailment. Journal of Symbolic Logic **42**(1), 147–148 (1977)
2. Baader, F., Calvanese, D., McGuinnes, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook. Cambridge University Press, Cambridge, 1 edn. (2003)
3. Belnap, Jr., N.D.: How a computer should think. In: Ryle, G. (ed.) Contemporary Aspects of Philosophy, pp. 30–56. Oriel Press, Stockfield (1977)
4. Bock, C., Fokoue, A., Haase, P., Hoekstra, R., Horrocks, I., Ruttenberg, A., Sattler, U., Smith, M.: OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax. W3C, Second edn. (2012), http://www.w3.org/TR/owl2-syntax/
5. Carnielli, W.A.: Systematization of finite many-valued logics through the method of tableaux. Journal of Symbolic Logic **52**(2), 473–493 (1987)
6. Correia, F.: Grounding and truth functions. Logique et Analyse **53**(211), 251–279 (2010)
7. Cresswell, M.J.: Hyperintensional logic. Studia Logica **34**, 25–38 (1975)
8. Daniels, C.: A note on negation. Erkenntnis **32**(3), 423–429 (1990)
9. Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence **34**(1), 39–76 (1988)
10. Ferguson, T.M.: Faulty Belnap computers and subsystems of FDE. Journal of Logic and Computation **26**(5), 1617–1636 (2016)
11. Ferguson, T.M.: Tableaux and restricted quantification for systems related to weak Kleene logic. In: Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX 2021). Springer, Switzerland (2021), to appear
12. Fine, K.: Angellic content. Journal of Philosophical Logic **45**(2), 199–226 (2016)
13. Kleene, S.C.: Introduction to Metamathematics. North-Holland Publishing Company, Amsterdam (1952)
14. Nardi, D., Brachman, R.J.: An introduction to description logics. In: Baader, F., Calvanese, D., McGuinnes, D., Nardi, D., Patel-Schneider, P. (eds.) The Description Logic Handbook, pp. 1–40. Cambridge University Press, Cambridge, 1 edn. (2003)
15. Yablo, S.: Aboutness. Princeton University Press, Princeton (2014)