Journal of Computer and System Sciences 89 (2017) 209–245

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Axiomatizations for downward XPath on data trees

Sergio Abriola [a], María Emilia Descotte [b], Raul Fervari [c], Santiago Figueira [a,*]

[a] Departamento de Computación, FCEN, Universidad de Buenos Aires, and ICC-CONICET, Argentina
[b] Departamento de Computación, FCEN, Universidad de Buenos Aires, and IMAS-CONICET, Argentina
[c] FAMAF, Universidad Nacional de Córdoba, and CONICET, Argentina

ARTICLE INFO

Article history:
Received 9 May 2016
Received in revised form 7 March 2017
Accepted 18 May 2017
Available online 13 June 2017

Keywords:
XPath
XML
Data tree
Axiomatization
Normal form
Modal logic
Query language

ABSTRACT

We give sound and complete axiomatizations for XPath with data tests by 'equality' or 'inequality', and containing the single 'child' axis. This data-aware logic predicts over data trees, which are tree-like structures whose every node contains a label from a finite alphabet and a data value from an infinite domain. The language allows us to compare data values of two nodes but cannot access the data values themselves (i.e. there is no comparison by constants). Our axioms are in the style of equational logic, extending the axiomatization of data-oblivious XPath, by B. ten Cate, T. Litak and M. Marx. We axiomatize the full logic with tests by 'equality' and 'inequality', and also a simpler fragment with 'equality' tests only. Our axiomatizations apply both to node expressions and path expressions. The proof of completeness relies on a novel normal form theorem for XPath with data tests.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

XML (eXtensible Markup Language) is the most successful language for data exchange on the web. It meets the requirements of a flexible, generic and platform-independent language. An XML document is a hierarchical structure that can be abstracted as a data tree, where nodes have labels (such as *LastName*) from a finite domain, and data values (such as *Smith*) from an infinite domain. For some tasks, data values can be disregarded (for instance, checking whether a given XML document conforms to a schema specification). But many applications require data-aware query languages, that is, languages with the ability of comparing data values. Indeed, the possibility to perform *joins* in queries or comparing for equality of data values is a very common and necessary feature in database query languages.

XPath is the most widely used query language for XML documents; it is an open standard and constitutes a World Wide Web Consortium (W3C) Recommendation [10]. XPath has syntactic operators or 'axes' to navigate the tree using the 'child', 'parent', 'sibling', etc. accessibility relations, and can make tests on intermediate nodes. Core-XPath [16] is the fragment of XPath 1.0 containing only the navigational behavior of XPath, i.e. without any reference to the data in the queries.

Core-XPath can be seen as a *modal language*, such as those used in software verification, like Linear Temporal Logic (LTL) [6] or Propositional Dynamic Logic (PDL) [17]. XPath has been already investigated from a 'modal' point of view. In [21] this perspective is illustrated by showing how some results on Core-XPath fragments can be derived from classical results in modal logic. In particular, when the only accessibility relation is 'child', Core-XPath has many similarities with

* Corresponding author.
E-mail address: santiago@dc.uba.ar (S. Figueira).

http://dx.doi.org/10.1016/j.jcss.2017.05.008
0022-0000/© 2017 Elsevier Inc. All rights reserved.

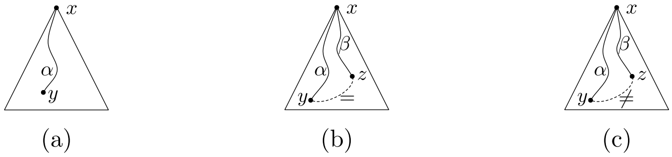

Fig. 1. In x it holds: (a) The *modal* diamond $\langle\alpha\rangle$; (b) the data-aware diamond $\langle\alpha=\beta\rangle$; (c) the data-aware diamond $\langle\alpha\neq\beta\rangle$.

basic modal logic (BML). First, in the absence of data, an XML document just becomes a tree whose every node has a label from a finite domain; this is a special kind of Kripke models, when labels are represented as propositional letters. Second, any property expressed in BML can be translated to Core-XPath and vice versa. There are also some differences: Core-XPath may express not only properties φ on nodes, called *node expressions*, but also on *paths*, called *path expressions*. When α is a path expression, its truth is evaluated on pairs of nodes instead of on individual nodes, as BML does. In a nutshell, α is true at (x, y) if the path from x to y (which is unique, since our models are trees) satisfies the condition expressed by α.

The formal syntax and semantics of Core-XPath will be given later in full detail, but let us now give a glimpse of it. If α is a path expression, in Core-XPath we can write $\langle\alpha\rangle$, a node expression saying that there is a descendant y of x such that (x, y) satisfies α (see Fig. 1(a)).

Imagine that α simply expresses "go to child; φ holds; end of path". Then the node expression $\langle\alpha\rangle$ is translated as $\Diamond\tilde{\varphi}$ in BML, where $\tilde{\varphi}$ is the recursive translation of φ to the BML language. For illustrating a more complex path expression, suppose that the path expression β expresses "go to child; φ holds; go to child; ψ holds; end of path". Then the node expression $\langle\beta\rangle$ is translated to the language of BML as $\Diamond(\tilde{\varphi}\wedge\Diamond\tilde{\psi})$. Core-XPath generalizes the 'diamond' $\Diamond$ operator of BML to complex diamonds $\langle\alpha\rangle$, where α describes a property on a path. Conversely, any formula $\Diamond\varphi$ of BML can be straightforwardly translated to Core-XPath as $\langle\alpha\rangle$, where α expresses "go to child; φ holds; end of path".

By the (finite) tree model property of BML, the validity of a formula with respect to the class of all Kripke models is equivalent to the validity in the class of (finite) tree-shaped Kripke models. Since there are truth-preserving translations to and from Core-XPath, it is not surprising that there exist axiomatizations of the node expressions fragment of Core-XPath with 'child' as the only accessibility operator. Interestingly, there are also axiomatizations of the path expressions fragment of it. Even more, there are also axiomatizations of all single axis fragments of Core-XPath (those where the only accessibility relation is the one of 'child', 'descendant', 'sibling', etc.), and also for the full Core-XPath language [22].

Core-Data-XPath [8] —here called XPath$_=$— is the extension of Core-XPath with (in)equality tests between attributes of elements in an XML document. The resemblance with modal languages is now more distant, since the models of XPath$_=$ cannot be represented by Kripke models. A first attempt to represent a data tree as a Kripke model, would be to let any data value v in the data tree correspond to a propositional letter p_v in the Kripke model [1]. However, this would be unfair: in BML, p_v is a licit formula expressing "the value is v" but this kind of construction is not permitted in XPath$_=$. Indeed, XPath$_=$ can only compare data values by equality or inequality at the end of paths, but it cannot compare the data value of a node with a constant. The rationale of this feature is twofold: on the one hand, it remains a finitary language; on the other, its semantics is invariant over renaming of data values. XPath$_=$ augments Core-XPath expressivity with 'data-aware diamonds' of the form $\langle\alpha=\beta\rangle$ and $\langle\alpha\neq\beta\rangle$. The former is true at x if there are descendants y and z of x such that α is true at (x, y) and β is true at (x, z), and y and z have the same data value (see Fig. 1(b)). The latter is true at x if there are y and z as before but such that y and z have distinct data values (see Fig. 1(c)). Observe that $\neg\langle\alpha=\beta\rangle$ expresses that all pairs of paths satisfying α and β respectively, starting in x, end up in nodes with different data values, while $\langle\alpha\neq\beta\rangle$ expresses that there is a pair of paths satisfying α and β respectively which end up in nodes with different data value. One can see that $\langle\alpha\neq\beta\rangle$ is not expressible in terms of Boolean combinations of expressions of the form $\langle\cdot=\cdot\rangle$.

Whilst the model theory of XPath$_=$ was recently investigated both for the node expressions fragment [12,13] and for the path expressions fragment, [3,2], the only other research into the proof theory of XPath$_=$ outside of this work is for a simple fragment [5].

Obtaining a complete axiomatization has applications in static analysis of queries, such as optimization through query rewriting. The idea here is to see equivalence axiom schemes as (undirected) rules for the rewriting of queries; in this context, the completeness of the axiomatic system means that a semantic equivalence between two node or path expressions must have a corresponding chain of rewriting rules that transform the first expression into the second one. Therefore, obtaining an axiomatization, along with all the proofs of the theorems involved in the demonstration of its completeness, can be used as a first step in finding effective strategies for rewriting queries into equivalent but less complex forms.

Studying complete axiomatizations can also give us an alternative method for solving the validity problem, which is undecidable for the full logic Core-Data-XPath [15], but it is decidable when the only axis present in the language is 'child', and in fact, also when adding 'descendant' [11] (and also for other fragments).

1.1. Contributions

We give sound and complete axiomatizations for XPath$_=$ with 'child' as the only axis. We extend the axiomatization of Core-XPath given in [22] with the needed axiom schemes to obtain all validities of Core-Data-XPath. Our axiomatizations will be equational: all axiom schemes are of the form $\varphi\equiv\psi$ for node expressions φ and ψ or of the form $\alpha\equiv\beta$ for path expressions α and β, and inference rules will be the standard ones of equational logic. We show that an equivalence $\varphi \equiv \psi$ is derivable in the axiomatic system if and only if for any data tree, and any node x in it, either φ and ψ are true at x or both are false at x. We also present a similar result for path expressions: an equivalence $\alpha \equiv \beta$ is derivable if and only if for any data tree, and any pair of nodes (x, y) in it, either α and β are true at (x, y) or both are false at (x, y). Our completeness proof relies on a normal form theorem for expressions of XPath$_=$ with 'child' axis, and a construction of a canonical model for any consistent formula in normal form inspired by [14].

We proceed gradually. To warm up, we first show an axiomatization for the fragment of XPath$_=$ with all Boolean operators, with data-aware diamonds of the form $\langle\alpha = \beta\rangle$, but keeping out those of the form $\langle\alpha \neq \beta\rangle$. This fragment is still interesting since it allows us to express the *join* query constructor. Then we give the axiomatization for the full XPath$_=$ with 'child' axis, whose proof is more involved but uses some ideas from the simpler case.

1.2. Related work

As we mentioned before, there exist axiomatizations for navigational fragments of XPath with different axes [22]. Axiomatizations of other fragments of Core-XPath have been investigated in [7], and extensions with XPath 2.0 features have been addressed in [23]. We found only a few attempts of axiomatizing modal logics with some notion of data value.

A logical framework to reason about data organization is investigated in [4]. They introduce *reference structures* as the model to represent data storage, and a *propositional labeled modal language* to talk about such structures. Both together model memory configurations, i.e., they allow storing data files, and retrieving information about other cells' content and location of files. A sentence $[\![m]\!]A$ is read as "memory cell m stores sentence A". Then, data is represented by mean of sentences: for instance, if data c_i represents a number N, c_i is the sentence "this is a number N" (same for other sorts of data). This representation is quite different from our approach. Nevertheless, according to our knowledge this is one of the first attempts on axiomatizing data-aware logics, by introducing a Hilbert-style axiomatization.

Tree Query Language (TQL) is a formalism based on *ambient logic*, designed as a query language for semi-structured data. It allows checking schema properties, extracting tags satisfying a property and also recursive queries. The TQL data model is *information trees*, and the notation to talk about information trees is called *info-terms*. In [9] an axiomatization for info-terms is given in terms of a minimal congruence. This axiomatization is sound and complete with respect to the information tree semantics. This is more related to our approach in the sense that we consider data values as an equivalence relation.

The most closely related work is [5], where an axiomatization for a very simple fragment of XPath, named *DataGL*, was given. Following our informal description of XPath$_=$, DataGL allows for constructions of the form $\langle\epsilon = \beta\rangle$ and $\langle\epsilon \neq \beta\rangle$, where ϵ represents the empty path and $\beta = \downarrow_*[\varphi]$ is a path of the form 'go to descendant; φ holds; end of path'. In particular, they introduce a sound and complete sequent calculus for this logic and derive PSPACE-completeness for the validity problem.

1.3. Organization

In §2 we give the formal syntax and semantics of XPath$_=$ with 'child' axis, called XPath$_=(\downarrow)$. As we already mentioned, we also study a special syntactical fragment, called XPath$_=(\downarrow)^-$, whose all data-aware diamonds are of the form $\langle\alpha = \beta\rangle$, and keeps out those of the form $\langle\alpha \neq \beta\rangle$. In §3 we give a sound and complete axiomatic system for XPath$_=(\downarrow)^-$: in §3.1 we state the needed axiom schemes, an extension of those introduced in [22]; in §3.2 we define the normal forms for XPath$_=(\downarrow)^-$ (these are not an extension of those defined in [22]) and state the corresponding normal form theorem; in §3.3 we show the completeness result, whose more complex part lies in proving that any node expression in normal form is satisfiable in a canonical model. In §4 we extend the previous axiom schemes to get a sound and complete axiomatic system for XPath$_=(\downarrow)$. We follow the same route as for XPath$_=(\downarrow)^-$: axiom schemes (§4.1), normal form (§4.2) and canonical model (§4.3). Those proofs requiring highly technical arguments were deferred to Appendix A. Finally, in §5 we close with some final remarks and future lines of research.

2. Preliminaries

Syntax of XPath$_=(\downarrow)$ We work with a simplification of XPath, stripped of its syntactic sugar and with the only axis being the 'child' relation, notated $\downarrow$. We consider fragments of XPath that correspond to the navigational part of XPath 1.0 with data equality and inequality. XPath$_=(\downarrow)$ is a two-sorted language, with *path expressions* (which we write α, β, γ) expressing properties of paths, and *node expressions* (which we write φ, ψ, ρ), expressing properties of nodes.

The language *Downward XPath*, notated XPath$_=(\downarrow)$ is defined by mutual recursion as follows:

$$\alpha, \beta ::= \epsilon \mid \downarrow \mid [\varphi] \mid \alpha\beta \mid \alpha \cup \beta$$

$$\varphi, \psi ::= a \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle\alpha\rangle \mid \langle\alpha = \beta\rangle \mid \langle\alpha \neq \beta\rangle, \qquad a \in \mathbb{A}$$

where $\mathbb{A}$ is a finite set of labels.

Other Boolean operators, such as $\vee$, $\rightarrow$, are defined as usual. We define the node expressions $\textsc{true}$ and $\textsc{false}$, and the path expression $\perp$, as follows:

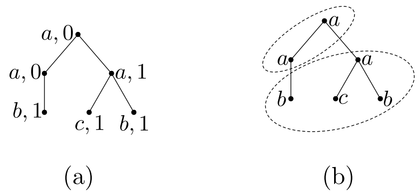

Fig. 2. (a) A data tree. Nodes are tagged with (ℓ, n) meaning that its label is ℓ and its data-value is n. (b) Our view of data tree: a node-labeled tree and a partition over its nodes.

$$\begin{aligned} \text{TRUE} &\stackrel{\text{def}}{=} \langle \epsilon \rangle \\ \text{FALSE} &\stackrel{\text{def}}{=} \neg\text{TRUE} \\ \bot &\stackrel{\text{def}}{=} [\neg\langle \epsilon \rangle] \end{aligned}$$

As we remark later, these expressions behave as expected in the axiomatic systems we design.

We notate $\text{XPath}_=(\downarrow)^-$ to the syntactic fragment which does not use the last rule $\langle \alpha \neq \beta \rangle$. An $\text{XPath}_=(\downarrow)$-*formula* [resp. $\text{XPath}_=(\downarrow)^-$-*formula*] is either a node expression or a path expression of $\text{XPath}_=(\downarrow)$ [resp. $\text{XPath}_=(\downarrow)^-$].

We define the *length* of an $\text{XPath}_=(\downarrow)$-path expression α, notated $\text{len}(\alpha)$, as follows:

$$\begin{aligned} \text{len}(\epsilon) &= 0 & \text{len}(\alpha\beta) &= \text{len}(\alpha) + \text{len}(\beta) \\ \text{len}(\downarrow) &= 1 & \text{len}(\alpha \cup \beta) &= \max\{\text{len}(\alpha), \text{len}(\beta)\} \\ \text{len}([\varphi]) &= 0 \end{aligned}$$

We write dd to denote the *downward depth* [13] of an $\text{XPath}_=(\downarrow)$-formula, which measures 'how deep' such formula can see, and is defined as follows:

$$\begin{aligned} \text{dd}(a) &= 0 & \text{dd}(\epsilon) &= 0 \\ \text{dd}(\neg\varphi) &= \text{dd}(\varphi) & \text{dd}(\downarrow) &= 1 \\ \text{dd}(\varphi \wedge \psi) &= \max\{\text{dd}(\varphi), \text{dd}(\psi)\} & \text{dd}([\varphi]) &= \text{dd}(\varphi) \\ \text{dd}(\langle\alpha\rangle) &= \text{dd}(\alpha) & \text{dd}(\alpha\beta) &= \max\{\text{dd}(\alpha), \text{len}(\alpha) + \text{dd}(\beta)\} \\ \text{dd}(\langle\alpha * \beta\rangle) &= \max\{\text{dd}(\alpha), \text{dd}(\beta)\} & \text{dd}(\alpha \cup \beta) &= \max\{\text{dd}(\alpha), \text{dd}(\beta)\}, \end{aligned}$$

where $a \in \mathbb{A}$, φ, ψ are node expressions of $\text{XPath}_=(\downarrow)$, α, β are path expressions of $\text{XPath}_=(\downarrow)$, and $* \in \{=, \neq\}$. Notice that the definition of $\text{dd}(\alpha\beta)$ is not symmetric because the concatenation is not so. Indeed, consider the path expressions $\alpha = \downarrow[\langle\downarrow\rangle]$ and $\beta = \downarrow\downarrow[\langle\downarrow\downarrow\downarrow\rangle]$. Then $\text{dd}(\alpha\beta) = \text{len}(\alpha) + \text{dd}(\beta) = 1 + 5 = 6$ but $\text{dd}(\beta\alpha) = \text{dd}(\beta) = 5$.

Data trees We introduce *data trees*, the structures in which we interpret $\text{XPath}_=(\downarrow)$-formulas. Usually, a data tree is defined as a tree whose every node contains a label from a finite alphabet $\mathbb{A}$ and a data value from an infinite domain. An example of a data tree is depicted in Fig. 2(a). Our logical language, whose formal semantics is defined below, will be able to compare the data value of two nodes by equality or inequality but it will not be able to compare against a concrete value. Hence we will work with an abstraction of the usual definition of data tree: instead of having data values in each node of the tree, we have an equivalence relation between the nodes or, equivalently, a partition. We identify two nodes with the same data value as being related by the equivalence relation, or belonging to the same equivalence class in the partition —see Fig. 2(b). While this is not the classical view of a data tree, it is more convenient for our purposes, and it is equivalent, as far as the semantics of our logical language is concerned.

Definition 1. Let $\mathbb{A}$ be a finite set of *labels*, a *data tree* $\mathcal{T}$ is a pair (T, π), where T is a tree (i.e. a connected acyclic graph such that every node has exactly one parent, except the root, which has no parent) whose nodes are labeled with elements from $\mathbb{A}$, and π is a partition over the nodes of T. We use T indistinctly to denote the set of nodes of $\mathcal{T}$ or the structure of the labeled tree. Given two nodes $x, y \in T$ we write $x \to y$ if y is a child of x and $x \xrightarrow{i} y$ (for $i \geq 1$) as a short for

$$(\exists z_0, \ldots, z_i \in T)\; x = z_0 \to z_1 \to \ldots \to z_i = y.$$

Observe that in particular $x \to y$ iff $x \xrightarrow{1} y$.

We denote with $[x]_\pi$ the class of x in the partition π, and with $label(x) \in \mathbb{A}$ the node's label. We say that $\mathcal{T}, x$ is a *pointed data tree*, and $\mathcal{T}, x, y$ is a *two-pointed data tree*.

Semantics of $\text{XPath}_=(\downarrow)$ Let us introduce the semantics of $\text{XPath}_=(\downarrow)$-formulas. Let $\mathcal{T} = (T, \pi)$ be a data tree. We define the semantics of $\text{XPath}_=(\downarrow)$ on $\mathcal{T}$ (notated as $[\![\cdot]\!]^{\mathcal{T}}$) in Table 1.

Let $\mathcal{T}, x$ be a pointed data tree and φ a node expression, we write $\mathcal{T}, x \models \varphi$ to denote $x \in [\![\varphi]\!]^{\mathcal{T}}$, and we say that $\mathcal{T}, x$ satisfies φ or that φ is true at $\mathcal{T}, x$. Let $\mathcal{T}, x, y$ be a two-pointed data tree and α a path expression, we write $\mathcal{T}, x, y \models \alpha$

Table 1
Semantics of XPath$_=(\downarrow)$.

$[\![\epsilon]\!]^{\mathcal{T}}$	$= \{(x,x) \mid x \in T\}$
$[\![\downarrow]\!]^{\mathcal{T}}$	$= \{(x,y) \mid x \to y\}$
$[\![\alpha\beta]\!]^{\mathcal{T}}$	$= \{(x,z) \mid \text{there exists } y \in T \text{ with } (x,y) \in [\![\alpha]\!]^{\mathcal{T}}, (y,z) \in [\![\beta]\!]^{\mathcal{T}}\}$
$[\![\alpha \cup \beta]\!]^{\mathcal{T}}$	$= [\![\alpha]\!]^{\mathcal{T}} \cup [\![\beta]\!]^{\mathcal{T}}$
$[\![[\varphi]]\!]^{\mathcal{T}}$	$= \{(x,x) \mid x \in [\![\varphi]\!]^{\mathcal{T}}\}$
$[\![a]\!]^{\mathcal{T}}$	$= \{x \in T \mid \mathit{label}(x) = a\}$
$[\![\neg\varphi]\!]^{\mathcal{T}}$	$= T \setminus [\![\varphi]\!]^{\mathcal{T}}$
$[\![\varphi \wedge \psi]\!]^{\mathcal{T}}$	$= [\![\varphi]\!]^{\mathcal{T}} \cap [\![\psi]\!]^{\mathcal{T}}$
$[\![\langle\alpha\rangle]\!]^{\mathcal{T}}$	$= \{x \in T \mid \text{there exists } y \in T \text{ with } (x,y) \in [\![\alpha]\!]^{\mathcal{T}}\}$
$[\![\langle\alpha = \beta\rangle]\!]^{\mathcal{T}}$	$= \{x \in T \mid \text{there exist } y,z \in T \text{ with } (x,y) \in [\![\alpha]\!]^{\mathcal{T}}, (x,z) \in [\![\beta]\!]^{\mathcal{T}}, [y]_\pi = [z]_\pi)\}$
$[\![\langle\alpha \neq \beta\rangle]\!]^{\mathcal{T}}$	$= \{x \in T \mid \text{there exist } y,z \in T \text{ with } (x,y) \in [\![\alpha]\!]^{\mathcal{T}}, (x,z) \in [\![\beta]\!]^{\mathcal{T}}, [y]_\pi \neq [z]_\pi)\}$

to denote $(x,y) \in [\![\alpha]\!]^{\mathcal{T}}$, and we say that $\mathcal{T}, x, y$ satisfies α or that α is true at $\mathcal{T}, x, y$. We say that a node expression φ is satisfiable in a data tree $\mathcal{T}$ if $\mathcal{T}, r \models \varphi$, where r is the root of $\mathcal{T}$. We say that φ is satisfiable if it is satisfiable in some data tree $\mathcal{T}$.

Example 2. Consider the data tree of Fig. 2 with root x.

1. $\langle \downarrow = \downarrow[a]\downarrow[b]\rangle$ is true at x because there is a path of length 1, and there is a path of length 2 (with labels a in the second node and b in the third one) ending in nodes with the same data value.
2. $\langle \epsilon = \downarrow\downarrow\rangle$ is false at x because there are no paths of length 2 ending in nodes with the same data value as x.
3. $\neg\langle \downarrow\downarrow \neq \downarrow\downarrow\rangle$ is true at x because all paths of length 2 end in nodes with the same data value.
4. $\langle \downarrow[a]\downarrow[b] = \epsilon\rangle$ is false at x because no path of length 2 with labels a in the second node and b in the third node, end in a node with the same data value as x.
5. $\langle \downarrow[a \wedge \langle\downarrow[b]\rangle] = \epsilon\rangle$ is true at x because x has a child with label a, satisfying $\langle\downarrow[b]\rangle$, and with the same data value as x.

We say that two node expressions φ, ψ of XPath$_=(\downarrow)$ are equivalent (notation: $\models \varphi \equiv \psi$) iff $[\![\varphi]\!]^{\mathcal{T}} = [\![\psi]\!]^{\mathcal{T}}$ for all data trees $\mathcal{T}$. Similarly, path expressions α, β of XPath$_=(\downarrow)$ are equivalent (notation: $\models \alpha \equiv \beta$) iff $[\![\alpha]\!]^{\mathcal{T}} = [\![\beta]\!]^{\mathcal{T}}$ for all data trees $\mathcal{T}$.

Let $\mathcal{T}, x, y$ and $\mathcal{T}', x', y'$ be two-pointed data trees, we say that $\mathcal{T}, x \equiv \mathcal{T}', x'$ [resp. $\mathcal{T}, x \equiv^- \mathcal{T}', x'$] iff for all node expressions φ of XPath$_=(\downarrow)$ [resp. XPath$_=(\downarrow)^-$] we have $\mathcal{T}, x \models \varphi$ iff $\mathcal{T}', x' \models \varphi$, and we say that $\mathcal{T}, x, y \equiv \mathcal{T}', x', y'$ [resp. $\mathcal{T}, x, y \equiv^- \mathcal{T}', x', y'$] iff for all path expressions α of XPath$_=(\downarrow)$ [resp. XPath$_=(\downarrow)^-$] $\mathcal{T}, x, y \models \alpha$ iff $\mathcal{T}', x', y' \models \alpha$.

Let $\mathcal{T} = (T, \pi)$ be a data tree. When T' is a subset of T, we write $\pi \restriction_{T'}$ to denote the restriction of the partition π to T'. Let $x \in T$, and let X be the set of x and all its descendants in T, i.e. $X = \{x\} \cup \{y \in T \mid (\exists i \geq 1)\ x \xrightarrow{i} y\}$. We define $\mathcal{T}\restriction_x = (T\restriction_x, \pi\restriction_x)$ as the data tree that consists of the subtree of T that is hanging from x, maintaining the partition of that portion.

The logic XPath$_=(\downarrow)$ is local in the same way as the basic modal logic:

Proposition 3. *Let $(\mathcal{T}, \pi)$ be a data tree. Then*

- $\mathcal{T}, x \equiv \mathcal{T}\restriction_x, x$.
- *If y, z are descendants of x in $\mathcal{T}$, then $\mathcal{T}, y, z \equiv \mathcal{T}\restriction_x, y, z$.*

Inference rules An XPath$_=(\downarrow)$-*node equivalence* is an expression of the form $\varphi \equiv \psi$, where φ, ψ are node expressions of XPath$_=(\downarrow)$. An XPath$_=(\downarrow)$-*path equivalence* is an expression of the form $\alpha \equiv \beta$, where α, β are path expressions. An *axiom* is either a node equivalence or a path equivalence.

For P, Q both path expressions or both node expressions, we say that $P \equiv Q$ is *derivable* (or also that P is *provably equivalent* to Q) from a given set of axioms Σ (notation $\Sigma \vdash P \equiv Q$) if it can be obtained from them using the standard rules of equational logic:

1. $P \equiv P$.
2. If $P \equiv Q$, then $Q \equiv P$.
3. If $P \equiv Q$ and $Q \equiv R$, then $P \equiv R$.
4. If $P \equiv Q$ and R' is obtained from R by replacing some occurrences of P by Q, then $R \equiv R'$.

We write $\varphi \leq \psi$ when $\varphi \vee \psi \equiv \psi$, and we write $\alpha \leq \beta$ when $\alpha \cup \beta \equiv \beta$.

Table 2
Axiomatic system XP^- for $\mathrm{XPath}_=(\downarrow)^-$.

Axioms for labels			
LbAx1	TRUE	$\equiv$	$\bigvee_{a\in\mathbb{A}} a$
LbAx2	FALSE	$\equiv$	$a \wedge b$ (where $a \neq b$)
Path axiom schemes for predicates			
PrAx1	$(\alpha[\neg\langle\beta\rangle])\beta$	$\equiv$	$\perp$
PrAx2	[TRUE]	$\equiv$	ϵ
PrAx3	$[\varphi \vee \psi]$	$\equiv$	$[\varphi] \cup [\psi]$
Path axiom schemes for idempotent semirings			
IsAx1	$(\alpha \cup \beta) \cup \gamma$	$\equiv$	$\alpha \cup (\beta \cup \gamma)$
IsAx2	$\alpha \cup \beta$	$\equiv$	$\beta \cup \alpha$
IsAx3	$\alpha \cup \alpha$	$\equiv$	α
IsAx4	$\alpha(\beta\gamma)$	$\equiv$	$(\alpha\beta)\gamma$
IsAx5	$\epsilon\alpha$	$\equiv$	α
	$\alpha\epsilon$	$\equiv$	α
IsAx6	$\alpha(\beta \cup \gamma)$	$\equiv$	$(\alpha\beta) \cup (\alpha\gamma)$
	$(\alpha \cup \beta)\gamma$	$\equiv$	$(\alpha\gamma) \cup (\beta\gamma)$
IsAx7	$\perp \cup \alpha$	$\equiv$	α
Node axiom schemes			
NdAx1	φ	$\equiv$	$\neg(\neg\varphi \vee \psi) \vee \neg(\neg\varphi \vee \neg\psi)$
NdAx2	$\langle[\varphi]\rangle$	$\equiv$	φ
NdAx3	$\langle\alpha \cup \beta\rangle$	$\equiv$	$\langle\alpha\rangle \vee \langle\beta\rangle$
NdAx4	$\langle\alpha\beta\rangle$	$\equiv$	$\langle\alpha[\langle\beta\rangle]\rangle$
Node axiom schemes for equality			
EqAx1	$\langle\alpha = \beta\rangle$	$\equiv$	$\langle\beta = \alpha\rangle$
EqAx2	$\langle\alpha \cup \beta = \gamma\rangle$	$\equiv$	$\langle\alpha = \gamma\rangle \vee \langle\beta = \gamma\rangle$
EqAx3	$\varphi \wedge \langle\alpha = \beta\rangle$	$\equiv$	$\langle[\varphi]\alpha = \beta\rangle$
EqAx4	$\langle\alpha = \beta\rangle$	$\leq$	$\langle\alpha\rangle$
EqAx5	$\langle\gamma[\langle\alpha = \beta\rangle]\rangle$	$\leq$	$\langle\gamma\alpha = \gamma\beta\rangle$
EqAx6	$\langle\alpha = \alpha\rangle$	$\equiv$	$\langle\alpha\rangle$
EqAx7	$\langle\alpha = \epsilon\rangle \wedge \langle\beta = \epsilon\rangle$	$\leq$	$\langle\alpha = \beta\rangle$
EqAx8	$\langle\alpha = \beta[\langle\epsilon = \gamma\rangle]\rangle$	$\leq$	$\langle\alpha = \beta\gamma\rangle$

Definition 4 *(Consistent Node and Path Expressions).* Let Σ be a set of axioms. We say that a node expression [resp. path expression] P of $\mathrm{XPath}_=(\downarrow)$ is Σ-*consistent* if $\Sigma \nvdash P \equiv \text{FALSE}$ [resp. $\Sigma \nvdash P \equiv \perp$]. We define $\mathbf{Con}_\Sigma$ as the set of Σ-*consistent node expressions*.

3. Axiomatic system for $\mathrm{XPath}_=(\downarrow)^-$

3.1. Axiomatization

The main theorems of this article are the ones about the completeness of the proposed axiomatizations. These theorems have two main ingredients: one is a normal form theorem that allows to rewrite any consistent node or path expression in terms of normal forms. The other one is the construction of a canonical model for any consistent node expression in normal form. As it is usually the case, at the same time, we give (through the set of axioms) the definition of consistency. So, an axiom (or an axiom scheme) could have been added either because it was needed to prove the normal form theorem or because it was needed to guarantee that every unsatisfiable formula is inconsistent — the key fact is that we have a much better intuition of what should be satisfiable than of what should be consistent. Of course we should be careful that the added axioms are sound but that is quite intuitive.

In Table 2 we list the axiom schemes for the fragment $\mathrm{XPath}_=(\downarrow)^-$. This list includes all the axiom schemes from [22] for the logic Core-XPath with single axis 'child' (second, third and fourth blocks) and adds the new axiom schemes for data-aware diamonds of the form $\langle\alpha = \beta\rangle$ (last block). Also, remember that in our data trees each node satisfies exactly one label. We add two axiom schemes to handle this issue (first block). This is unessential for our development, and could be dropped to axiomatize $\mathrm{XPath}_=(\downarrow)$ over data trees whose every node is tagged with multiple labels, with minor changes to the definitions of normal forms.

Let XP^- be the set of all instantiations of the axiom schemes of Table 2 for a fixed alphabet $\mathbb{A}$. In the scope of this section we will often say that a node expression is *consistent* meaning that it is XP^--consistent (as in Definition 4).

Observe that PrAx4 from [22, Table 3], defined by

PrAx4	$(\alpha\beta)[\varphi]$	$\equiv$	$\alpha(\beta[\varphi])$

is not present in our axiomatization because, due to our language design, it is a particular case of IsAx4.

The following syntactic equivalences will be useful for the next sections:

Fact 5. As seen in [22], TRUE, FALSE, and $\perp$ behave as expected: $\mathsf{XP}^- \vdash \varphi \vee \text{TRUE} \equiv \text{TRUE}$, $\mathsf{XP}^- \vdash \alpha[\text{FALSE}] \equiv \perp$, et cetera. Furthermore, we have the following from [22, Table 6]:

Der1	$\mathsf{XP}^- \vdash \varphi \vee \psi \equiv \psi \vee \varphi$
Der2	$\mathsf{XP}^- \vdash \varphi \vee (\psi \vee \rho) \equiv (\varphi \vee \psi) \vee \rho$
Der12	$\mathsf{XP}^- \vdash \langle \alpha\beta \rangle \leq \langle \alpha \rangle$
Der13	$\mathsf{XP}^- \vdash \langle \alpha[\text{FALSE}] \rangle \equiv \text{FALSE}$
Der21	$\mathsf{XP}^- \vdash \alpha[\varphi][\psi] \equiv \alpha[\varphi \wedge \psi]$

We note that in order to prove the previous derivations one needs to use PrAx1, PrAx2, PrAx3, IsAx1, IsAx2 , IsAx4, IsAx5, IsAx6, IsAx7, NdAx1, NdAx2, NdAx3 and NdAx4.[1]

As a consequence of **Der1**, **Der2** and Huntington's equation NdAx1, we can derive all the axioms of Boolean algebras from the axioms in XP^- [19,18]. In what follows, we will often use the Boolean properties without explicitly referencing them. In particular, we use the fact that $\mathsf{XP}^- \vdash \psi \leq \text{FALSE}$ implies that ψ is an inconsistent node expression, and that $\mathsf{XP}^- \vdash \varphi \leq \psi$ implies that $\varphi \wedge \neg\psi$ is inconsistent.

Sometimes we use IsAx1, IsAx4, EqAx1, EqAx4, and EqAx6 without explicitly mentioning them. We omit such steps in order to make the proofs more readable.

It is not difficult to see that the axioms XP^- are sound for $\text{XPath}_=(\downarrow)^-$:

Proposition 6 *(Soundness of* $\text{XPath}_=(\downarrow)^-$*).*

1. *Let* φ *and* ψ *be node expressions of* $\text{XPath}_=(\downarrow)^-$. *Then* $\mathsf{XP}^- \vdash \varphi \equiv \psi$ *implies* $\models \varphi \equiv \psi$.
2. *Let* α *and* β *be path expressions of* $\text{XPath}_=(\downarrow)^-$. *Then* $\mathsf{XP}^- \vdash \alpha \equiv \beta$ *implies* $\models \alpha \equiv \beta$.

Proof. Equational rules are valid because we have compositional semantics, and the proof that all the axioms schemes from Table 2 are sound is straightforward. □

3.2. Normal forms

When working in Core-XPath, the only diamond in the language is the *modal* diamond of the form $\langle\alpha\rangle$, where α is a path expression. In the absence of data-aware diamonds any node expression $\langle[\varphi]\downarrow\beta\rangle$ is equivalent to $\varphi \wedge \langle\downarrow[\langle\beta\rangle]\rangle$. Hence when the only axis is 'child', the only path expressions that we need are of the form $\downarrow[\langle\cdot\rangle]$, of length 1, and therefore the only diamonds that we need are of the form $\langle\downarrow[\psi]\rangle$, which in the basic modal logic is written simply as $\Diamond\psi$. This rewriting of path expressions is carried out in [22], and so normal forms have somewhat the same flavor as in the basic modal logic.

When data shows up, this rewriting is no longer possible: the node expression $\langle\alpha = \beta\rangle$ checks if there are nodes with equal data value *at the end of paths* α *and* β. So these paths cannot be compressed as before. For an easy example, observe that the data-aware diamond $\langle\downarrow[a]\downarrow[b] = \epsilon\rangle$ is not equivalent to $\langle\downarrow[a \wedge \langle\downarrow[b]\rangle] = \epsilon\rangle$ (see items 4 and 5 of Example 2).

The normal forms we will introduce are inspired by the classic *Disjunctive Normal Form* (DNF) for propositional logic. Our normal forms will take into account path expressions of arbitrary length, and this makes our definition more involved than the one in [22]. We introduce them in this section for the language $\text{XPath}_=(\downarrow)^-$. This definition will be extended to the general logic $\text{XPath}_=(\downarrow)$ in §4.2.

We define the sets P_n^- and N_n^-, which contain the path and node expressions of $\text{XPath}_=(\downarrow)^-$, respectively, in normal form at level n:

Definition 7 *(Normal form for* $\text{XPath}_=(\downarrow)^-$*).*

$$
\begin{aligned}
P_0^- &= \{\epsilon\} \\
N_0^- &= \{a \wedge \langle\epsilon = \epsilon\rangle \mid a \in \mathbb{A}\} \\
P_{n+1}^- &= \{\epsilon\} \cup \left\{\downarrow[\psi]\beta \mid \psi \in N_n^-, \beta \in P_n^-\right\} \\
D_{n+1}^- &= \left\{\langle\alpha = \beta\rangle \mid \alpha, \beta \in P_{n+1}^-\right\}
\end{aligned}
$$

[1] **Der1** uses IsAx2, NdAx2, and NdAx3. **Der2** uses IsAx1 (and NdAx2 and NdAx3 again). We can now derive all the axioms of Boolean algebras by also using NdAx1. **Der12** also uses PrAx2, PrAx3, IsAx4, IsAx5, IsAx6, and NdAx4. **Der13** does not need further axioms. **Der21** also uses PrAx1 and IsAx7.

$$N^-_{n+1} = \left\{ a \wedge \bigwedge_{\varphi \in C} \varphi \wedge \bigwedge_{\varphi \in D^-_{n+1} \setminus C} \neg\varphi \mid C \subseteq D^-_{n+1}, a \in \mathbb{A} \right\} \cap \mathbf{Con}_{\mathsf{XP}^-}.$$

Observe that we define normal forms by mutual recursion among three kinds of sets: P^-_n, D^-_n and N^-_n (for some n), which are sets of path expressions, data-aware diamonds, and node expressions, respectively. They consist of expressions that can look forward up to a certain downward depth. The index n indicates which maximum downward depth we are exploring, both in path and node expressions. Base cases are the simplest expressions of each kind (depth 0). New path expressions are constructed by using node and path expressions ψ and β from a previous level of their respective type, and exploring one more step using $\downarrow$. Data-aware diamond expressions are auxiliary expressions consisting of equalities between two path expressions of the same level. Finally, node expressions in normal form at some level n are formed of consistent conjunctions of positive and negative data-aware diamond expressions of level n. Notice that at each level i, each conjunction in N^-_i has one conjunct of the form a with $a \in \mathbb{A}$ which provides a label for the current node. Finally, let us remark that it would suffice that N^-_0 contains formulas of the form a, for $a \in \mathbb{A}$. However, we include instead formulas of the form $a \wedge \langle \epsilon = \epsilon \rangle$ (containing the tautology $\langle \epsilon = \epsilon \rangle$) only for technical reasons.

Example 8. Let us see some examples of expressions in normal form. We consider only two labels a and b, and ignore redundancies (if we write $\langle \alpha = \beta \rangle$, we do not write $\langle \beta = \alpha \rangle$). The sets P^-_1 and D^-_1 are as follows:

$$P^-_1 = \{\downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon, \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon, \epsilon\}$$

$$\begin{aligned} D^-_1 = \{&\langle \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle, \langle \epsilon = \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle, \langle \epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle, \\ &\langle \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle, \langle \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle, \langle \epsilon = \epsilon \rangle\}. \end{aligned}$$

An example of a node expression in normal form at level 1, i.e. a node expression in N^-_1, is

$$\begin{aligned} \varphi = a \wedge \langle \epsilon = \epsilon \rangle \wedge \langle \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle \wedge \langle \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle \\ \wedge \langle \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle \wedge \neg\langle \epsilon = \downarrow[a \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle \wedge \neg\langle \epsilon = \downarrow[b \wedge \langle \epsilon = \epsilon \rangle]\epsilon \rangle. \end{aligned}$$

The following lemmas (9, 10 and 11) are very intuitive and their proofs are straightforward.

Lemma 9. *Let $\psi \in N^-_n$ and $\alpha, \beta \in P^-_n$. Let $\mathcal{T}, x$ be a pointed data tree, such that $\mathcal{T}, x \models \psi$ and $\mathcal{T}, x \models \langle \alpha = \beta \rangle$. Then $\langle \alpha = \beta \rangle$ is a conjunct of ψ.*

Proof. The case when $n = 0$ follows from the definitions of P^-_0 and N^-_0. If $n > 0$, since $\alpha, \beta \in P^-_n$, by definition of D^-_n, we have $\langle \alpha = \beta \rangle \in D^-_n$. Because $\psi \in N^-_n$, either $\langle \alpha = \beta \rangle$ or its negation is a conjunct of ψ. Suppose that the latter occurs, then $\mathcal{T}, x \models \neg\langle \alpha = \beta \rangle$, and, by hypothesis, $\mathcal{T}, x \models \langle \alpha = \beta \rangle$, which is a contradiction. $\square$

Lemma 10. *Let $\psi \in N^-_n$ and $\alpha \in P^-_n$. If $[\psi]\alpha$ is consistent then $\langle \alpha = \alpha \rangle$ is a conjunct of ψ. As an immediate consequence, if $\langle \downarrow[\psi]\alpha \rangle$ is consistent then $\langle \alpha = \alpha \rangle$ is a conjunct of ψ.*

Proof. Since $\alpha \in P^-_n$, then either $\alpha = \epsilon$ or α is of the form $\downarrow[\psi_1] \ldots \downarrow[\psi_k]\epsilon$ for some $1 \le k \le n$, and $\psi_i \in N^-_{n-i}$. If $\alpha = \epsilon$, we are done, as $\langle \epsilon = \epsilon \rangle$ is always a conjunct of ψ by consistency. Else, since $\langle \alpha = \alpha \rangle \in D^-_n$, $\langle \alpha = \alpha \rangle$ or its negation is a conjunct of ψ. By using **Der 21** from Fact 5, EqAx6 and PrAx1 consecutively, one can see that the latter case is not possible, because $[\psi]\alpha$ is consistent. Then $\langle \alpha = \alpha \rangle$ is a conjunct of ψ. $\square$

Lemma 11. *For every pair of distinct elements $\varphi, \psi \in N^-_n$, $\varphi \wedge \psi$ is inconsistent.*

Proof. If $n = 0$, then $\varphi = a \wedge \langle \epsilon = \epsilon \rangle$ and $\psi = b \wedge \langle \epsilon = \epsilon \rangle$, with $a, b \in \mathbb{A}$ and $a \neq b$. Then by LbAx2, we have $\mathsf{XP}^- \vdash \varphi \wedge \psi \equiv$ FALSE, i.e., $\varphi \wedge \psi$ is inconsistent.

Let φ and ψ be distinct normal forms of degree $n > 0$, then we have two possibilities:

- If φ and ψ differ in the conjunct of the form a with $a \in \mathbb{A}$, then we use an argument similar to the one used for the base case.
- If not, then there is $\sigma \in D^-_n$ such that, without loss of generality, σ is a conjunct of φ and $\neg\sigma$ is a conjunct of ψ. Therefore, because $\varphi \wedge \psi$ contains $\sigma \wedge \neg\sigma$ as a sub-expression, we have $\mathsf{XP}^- \vdash \varphi \wedge \psi \equiv$ FALSE, i.e., it is inconsistent.

This concludes the proof. $\square$

Lemma 12. *Let $\alpha, \beta \in P_n^-$. If there is a data tree $\mathcal{T}$ and nodes $x, y \in T$ such that $\mathcal{T}, x, y \models \alpha$ and $\mathcal{T}, x, y \models \beta$, then $\alpha = \beta$.*

Proof. Let $\alpha = \downarrow[\psi_1]...\downarrow[\psi_i]\epsilon$ and $\beta = \downarrow[\rho_1]...\downarrow[\rho_j]\epsilon$. By definition of the semantics of $\text{XPath}_=(\downarrow)^-$, $i = j$ since $\mathcal{T}, x, y \models \alpha$ and $\mathcal{T}, x, y \models \beta$. In particular, there are nodes $z_i \in T$, $1 \le k \le i$, such that $\mathcal{T}, z_i \models \psi_k, \mathcal{T}, z_i \models \rho_k$ for all $1 \le k \le i$. Using Proposition 6, we obtain that $\psi_k \wedge \rho_k$ is consistent for all $k = 1, \dots, i$, and thus by Lemma 11 we have that $\psi_k = \rho_k$ for all $k = 1, \dots, i$. Then we conclude that $\alpha = \beta$. $\square$

The following lemma is a normal form result for the special case of data-aware 'diamond' node expressions in D_n^-:

Lemma 13. *Let $n > 0$ and $a \in \mathbb{A}$. If $\varphi \in D_n^-$ is consistent then there are $\psi_1, \dots, \psi_k \in N_n^-$ such that $\mathsf{XP}^- \vdash a \wedge \varphi \equiv \bigvee_i \psi_i$*

Proof. Take

$$\psi = \bigvee \left(\left\{ a \wedge \bigwedge_{\psi \in C} \psi \wedge \bigwedge_{\psi \in D_n^- \setminus C} \neg\psi \mid C \subseteq D_n^-, \varphi \in C \right\} \cap \mathbf{Con}_{\mathsf{XP}^-} \right).$$

It can be seen that $\mathsf{XP}^- \vdash a \wedge \varphi \equiv \psi$. Notice that the above disjunction is not empty. Indeed, let $D_n^- \setminus \{\varphi\} = \{\psi_1, \dots, \psi_k\}$, and define $\varphi_0 = a \wedge \varphi$ and $\varphi_{i+1} = \varphi_i \wedge \psi_{i+1}$ if $\varphi_i \wedge \psi_{i+1}$ is consistent and $\varphi_{i+1} = \varphi_i \wedge \neg\psi_{i+1}$ otherwise. By NdAx1 either $\varphi_i \wedge \psi_{i+1}$ or $\varphi_i \wedge \neg\psi_{i+1}$ is consistent, and hence φ_i is consistent for all i. This means that φ_k is consistent and hence it is one of the disjuncts of the above formula. $\square$

The next lemma states that expressions in any P_n^- or N_n^- are provably equivalent to the union or disjunction, respectively, of expressions in higher levels of P_n^- or N_n^-.

Lemma 14. *Let $m > n$. If $\varphi \in N_n^-$ then there are $\varphi_1 \dots \varphi_k \in N_m^-$ such that $\mathsf{XP}^- \vdash \varphi \equiv \bigvee_i \varphi_i$. If $\alpha \in P_n^-$ then there are $\alpha_1 \dots \alpha_k \in P_m^-$ such that $\mathsf{XP}^- \vdash \alpha \equiv \bigcup_i \alpha_i$.*

Proof. Observe that it suffices to show this result for $m = n + 1$.

The basic idea is to proceed by induction over n, first proving the result for P_n^- and then using that for the case of N_n^-.

The base case for P_0^- is trivial, while the case for $\varphi \in N_0^-$ is easy by taking the disjunction of all node expressions in N_1^- which contain the same label as φ as a conjunct.

Now for the inductive case $\alpha = P_{n+1}^-$, if $\alpha = \epsilon$ then the result is trivial, and otherwise $\alpha = \downarrow[\psi]\beta$ with $\psi \in N_n^-$ and $\beta \in P_n^-$. We now use the inductive hypothesis on ψ and β and distribute into a union in P_{n+2}^- using PrAx3 and the path axiom schemes for idempotent semirings. The case $\varphi \in N_{n+1}^-$ is solved similarly, using that we know the result holds for path expressions in P_{n+1}^-. $\square$

It is easier to prove that every consistent formula is satisfiable over expressions in normal form than over the general case, as we can rely on the particular structure of those expressions. However, these proofs would be of little use if expressions in normal form only represented a small subset of all possible expressions. That is not really the case: Theorem 16 below will show that all node expressions (and also all path expressions) are provably equivalent to a disjunction of expressions in normal form.

Example 15. As a simple example of these equivalences, take the language with only three labels a, b, and c, and consider the node expression $\varphi = \neg a$. Then $\mathsf{XP}^- \vdash \varphi \equiv (b \wedge \langle\epsilon = \epsilon\rangle) \vee (c \wedge \langle\epsilon = \epsilon\rangle)$, where $b \wedge \langle\epsilon = \epsilon\rangle$ and $c \wedge \langle\epsilon = \epsilon\rangle$ are node expressions in N_0^-.

For a slightly more complex example, related with Example 8, take the language with only the labels a and b, and consider the node expression

$$\varphi = \langle[a]\downarrow[a] = \downarrow[b]\rangle \wedge \neg\langle\epsilon = \downarrow[a]\rangle.$$

Then $\mathsf{XP}^- \vdash \varphi \equiv \psi_1 \vee \psi_2$, where

$$\psi_1 = \psi \wedge \neg\langle\epsilon = \downarrow[b \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle$$

$$\psi_2 = \psi \wedge \langle\epsilon = \downarrow[b \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle$$

$$\begin{aligned}\psi = {} & a \wedge \langle\epsilon = \epsilon\rangle \wedge \langle\downarrow[a \wedge \langle\epsilon = \epsilon\rangle]\epsilon = \downarrow[b \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle \wedge \langle\downarrow[a \wedge \langle\epsilon = \epsilon\rangle]\epsilon = \downarrow[a \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle \wedge \\ & \wedge \langle\downarrow[b \wedge \langle\epsilon = \epsilon\rangle]\epsilon = \downarrow[b \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle \wedge \neg\langle\epsilon = \downarrow[a \wedge \langle\epsilon = \epsilon\rangle]\epsilon\rangle\end{aligned}$$

Theorem 16 *(Normal form for XPath$_=(\downarrow)^-$). Let φ be a consistent node expression of* XPath$_=(\downarrow)^-$ *such that* $\mathrm{dd}(\varphi) = n$*. Then* $\mathsf{XP}^- \vdash \varphi \equiv \bigvee_i \varphi_i$ *for some* $(\varphi_i)_{1\le i\le k} \in N_n^-$*. Let α be a consistent path expression of* XPath$_=(\downarrow)^-$ *such that* $\mathrm{dd}(\alpha) = n$*. Then* $\mathsf{XP}^- \vdash \alpha \equiv \bigcup_i [\varphi_i]\alpha_i$ *for some* $(\alpha_i)_{1\le i\le k} \in P_n^-$ *and* $(\varphi_i)_{1\le i\le k} \in N_n^-$*. Furthermore, if α is ϵ or starting with $\downarrow$ then* $\mathsf{XP}^- \vdash \alpha \equiv \bigcup_i \alpha_i$ *for some* $(\alpha_i)_{1\le i\le k} \in P_n^-$*.*

Proof. We show that if F is a consistent formula of XPath$_=(\downarrow)^-$ such that $\mathrm{dd}(F) = n$, then

a) if F is a node expression then $\mathsf{XP}^- \vdash F \equiv \bigvee_i \psi_i$ for some $(\psi_i)_{1\le i\le k} \in N_n^-$;
b) if F is a path expression then $\mathsf{XP}^- \vdash F \equiv \bigcup_i [\varphi_i]\alpha_i$ for some $(\alpha_i)_{1\le i\le k} \in P_n^-$ and $(\varphi_i)_{1\le i\le k} \in N_n^-$; furthermore, if F is ϵ or starts with $\downarrow$, then $\mathsf{XP}^- \vdash F \equiv \bigcup_i \alpha_i$ for some $(\alpha_i)_{1\le i\le k} \in P_n^-$.

Because of EqAx6, it is enough to prove the lemma for the fragment of XPath$_=(\downarrow)^-$ without diamonds of the form $\langle\alpha\rangle$. We proceed by induction on the complexity of F, denoted by c, and defined for the specific purpose of this proof as follows:

$$\begin{array}{rclrcl} \mathsf{c}(a) & = & 1 & \mathsf{c}(\epsilon) & = & 0 \\ \mathsf{c}(\neg\varphi) & = & 1+\mathsf{c}(\varphi) & \mathsf{c}(\downarrow) & = & 1 \\ \mathsf{c}(\varphi\wedge\psi) & = & 1+\mathsf{c}(\varphi)+\mathsf{c}(\psi) & \mathsf{c}(\alpha\beta) & = & \mathsf{c}(\alpha)+\mathsf{c}(\beta) \\ \mathsf{c}(\langle\alpha=\beta\rangle) & = & 1+\mathsf{c}(\alpha)+\mathsf{c}(\beta) & \mathsf{c}(\alpha\cup\beta) & = & 1+\mathsf{c}(\alpha)+\mathsf{c}(\beta) \\ & & & \mathsf{c}([\varphi]) & = & 2+\mathsf{c}(\varphi) \end{array}$$

Observe that the only node expressions of least complexity (namely, 1) are the labels a or $\langle\epsilon=\epsilon\rangle$, that the only path expressions of least complexity (namely, 0) are those of the form $\epsilon\ldots\epsilon$, and that the only path expressions of complexity 1 consist of one $\downarrow$ symbol concatenated with any number of ϵ symbols at both sides (that number might be 0, leaving the path expression $\downarrow$). Observe also that $\mathsf{c}(\varphi\wedge\langle\alpha=\beta\rangle) < \mathsf{c}(\langle[\varphi]\alpha=\beta\rangle)$.

Base case If the complexity of F is 0 then it is the path expression $\epsilon\ldots\epsilon$, which is provably equivalent to ϵ by IsAx5. Since $\epsilon \in P_0^-$, *b*) is immediate. If the complexity of F is 1 then F is either a node expression which consists of a single label or $\langle\epsilon=\epsilon\rangle$, or F is the path expression $\downarrow$ (eventually concatenated with ϵ but those path expressions are all provably equivalent to $\downarrow$ by EqAx6). If $F = a$ ($a \in \mathbb{A}$), then *a*) is immediate, since using EqAx6 and Boolean reasoning we have $\mathsf{XP}^- \vdash F \equiv a \wedge \langle\epsilon=\epsilon\rangle$, so $a\wedge\langle\epsilon=\epsilon\rangle \in N_0^-$, and we finish by applying Lemma 14. If $F = \langle\epsilon=\epsilon\rangle$, then *a*) follows from EqAx6, LbAx1, Boolean reasoning and Lemma 14. If $F = \downarrow$, by IsAx5, LbAx1 and PrAx2 we have $\mathsf{XP}^- \vdash F \equiv \downarrow[\bigvee_{a\in\mathbb{A}} a]\epsilon$, and by PrAx3 and IsAx6, we conclude $\mathsf{XP}^- \vdash F \equiv \bigcup_{a\in\mathbb{A}} \downarrow[a]\epsilon$ (observe that $\downarrow[a]\epsilon \equiv \downarrow[a\wedge\langle\epsilon=\epsilon\rangle]\epsilon \in P_1^-$).

Induction If the complexity of F is greater than 1, then F involves some of the operators $\neg, \wedge, \langle\ \rangle, \cup, [\ ]$ or a concatenation different from the ones of complexity 1 mentioned above. We will perform the inductive step for each of these operators.

If $F = \varphi\wedge\psi$ or $\neg\varphi$, we reason as in the propositional case. If $F = \varphi\wedge\psi$, we use the inductive hypothesis on φ and ψ to obtain that $\mathsf{XP}^- \vdash F \equiv \bigvee_i \varphi_i \wedge \bigvee_j \psi_j$, where $\varphi_i \in N^-_{\mathrm{dd}(\varphi)}$ for all i and $\psi_j \in N^-_{\mathrm{dd}(\psi)}$ for all j. Actually, we can assume that $\varphi_i, \psi_j \in N_n^-$ for all i, j by Lemma 14. We now use Boolean distributive laws to prove that F is equivalent to $\bigvee_{i,j}(\varphi_i\wedge\psi_j)$. We then use Lemma 11 plus the consistency of F to remove from that expression redundant conjunctions (if $\varphi_i = \psi_j$, from $\varphi_i\wedge\psi_j$ we just keep φ_i) and inconsistent conjunctions (cases where $\varphi_i \ne \psi_j$).

If $F = \neg\varphi$, we have by inductive hypothesis that $\mathsf{XP}^- \vdash \neg\varphi \equiv \neg\bigvee_{1\le i\le m}\varphi_i$, and we can again assume by Lemma 14 that $\varphi_i \in N_n^-$ for all i. Expanding each φ_i into $a_i \wedge \bigwedge_{\rho\in C_i}\rho \wedge \bigwedge_{\rho\in D_n^-\setminus C_i}\neg\rho$ (where $C_i \subseteq D_n^-$) and then using Boolean algebra, we have $\mathsf{XP}^- \vdash \neg\varphi \equiv \bigwedge_{1\le i\le m}(\neg a_i \vee \bigvee_{\rho\in C_i}\neg\rho \vee \bigvee_{\rho\in D_n^-\setminus C_i}\rho)$. We now use Boolean distributive laws to get $\mathsf{XP}^- \vdash \neg\varphi \equiv \bigvee_{\omega\in\Omega}\bigwedge_{1\le i\le m}\omega(i)$, where each $\omega(i)$ is either $\neg a_i$, some $\neg\rho$ for $\rho\in C_i$, or some $\rho \in D_n^-\setminus C_i$, and where Ω contains all possible strings ω of length m formed in that way. We now use LbAx1 to get $\mathsf{XP}^- \vdash \neg\varphi \equiv \bigvee_{\omega\in\Omega}\bigvee_{a\in\mathbb{A}} a \wedge \bigwedge_{1\le i\le m}\omega(i)$. Then, we eliminate repetitions in conjunctions of node expressions, and use properties of Boolean algebra to eliminate inconsistencies; also, as each disjunct has some positive occurrences of some $a\in\mathbb{A}$, we can use LbAx2 and eliminate the (redundant) occurrences of negation of labels. So now we have that $\mathsf{XP}^- \vdash \neg\varphi \equiv \bigvee_{\omega\in\Omega}\bigvee_{a\in\mathbb{A}}\psi_{\omega,a}$, where each $\psi_{\omega,a}$ is of the form $a\wedge\bigwedge_{\rho\in C}\rho\wedge\bigwedge_{\rho\in D}\neg\rho$, with $C, D \subseteq D_n^-$ and $C\cap D = \emptyset$. However we do not necessarily have $D = D_n^-\setminus C$, so these conjunctions may not add up to be of the form of a node expression in N_n^-: to add the conjunctions needed in order to get normal forms, we proceed as in the proof of Lemma 13 to complete each $a\wedge\bigwedge_{\rho\in C}\rho\wedge\bigwedge_{\rho\in D}\neg\rho$ into $\bigvee_j(a\wedge\bigwedge_{\rho\in C_j}\wedge\bigwedge_{\rho\in D_n^-\setminus C_j}\neg\rho)$, where $C\subseteq C_j$ for all j. Finally, we have obtained a set $(C_k)_{k\in K}$ of subsets of D_n^- such that $\mathsf{XP}^- \vdash \neg\varphi \equiv \bigvee_{k\in K}(a_k\wedge\bigwedge_{\rho\in C_k}\rho\wedge\bigwedge_{\rho\in D_n^-\setminus C_k}\neg\rho)$.

If F is of the form $\langle\alpha=\beta\rangle$, we reason as follows. Since $c(\alpha) < c(\langle\alpha=\beta\rangle)$, by inductive hypothesis, we have $\mathsf{XP}^- \vdash \alpha \equiv \bigcup_i[\varphi_i]\alpha_i$ for some $\alpha_i \in P_n^-$ and $\varphi_i \in N_n^-$ (We may have to use also Lemma 14, PrAx3, IsAx6, and **Der21** of Fact 5 if $\mathrm{dd}(\alpha) < \mathrm{dd}(\langle\alpha=\beta\rangle)$). Similarly, we can turn β into $\bigcup_j[\psi_j]\beta_j$. Using EqAx2, EqAx3, and EqAx1, we obtain $\mathsf{XP}^- \vdash F \equiv \bigvee_{i,j}\varphi_i\wedge\psi_j\wedge\langle\alpha_i=\beta_j\rangle$. We then use LbAx1 and Boolean reasoning to get $\mathsf{XP}^- \vdash F \equiv \bigvee_{i,j}\varphi_i\wedge\psi_j\wedge(\bigvee_{a\in\mathbb{A}} a\wedge\langle\alpha_i=\beta_j\rangle)$, and then distribute the $\wedge$, use Lemma 13 over each $a\wedge\langle\alpha_i=\beta_i\rangle$, and eliminate inconsistencies using Lemma 11 to obtain our desired result.

Suppose that F is a path expression. Without loss of generality, we can assume that $F \neq [\varphi]$ or $F \neq \alpha \cup \beta$ because in those cases, there exist an equivalent expression of the same complexity that is a concatenation ($[\varphi]\epsilon$ or $(\alpha \cup \beta)\epsilon$ respectively). Then we only need to prove the result for the concatenation in order to conclude the proof. Also without loss of generality we may assume that F does not start with ϵ, since in that case there exist an equivalent expression of the same complexity that doesn't start with ϵ. In case F is a concatenation $F = \alpha\beta$ that doesn't start with ϵ, we split the proof in three different cases according to the form of α (note that, by IsAx4, we can assume that α is not a concatenation itself).

If F is of the form $[\varphi]\beta$ then by IsAx5 we may suppose that β ends in ϵ and by **Der21** of Fact 5 we may suppose that β is either ϵ or starts with $\downarrow$ (notice that ϵ does not count in the complexity of a formula and that the expression in the left hand side of **Der21** of Fact 5 has a complexity greater than the one in the right hand side). By inductive hypothesis, φ is provably equivalent to $\bigvee_i \varphi_i$ for some $(\varphi_i)_i \in N_n^-$ (We may have to use Lemma 14 to increase the degree). Therefore, by PrAx3, $[\varphi]$ is provably equivalent to $\bigcup_i [\varphi_i]$. By inductive hypothesis, β is provably equivalent to $\bigcup_j \beta_j$ for some $(\beta_j)_j \in P_n^-$ (again, we may have to use Lemma 14). Hence F is provably equivalent to $(\bigcup_i [\varphi_i])(\bigcup_j \beta_j)$, and by IsAx6 we conclude that F is provably equivalent to $\bigcup_{i,j} [\varphi_i]\beta_j$ as we wanted to show.

If F is of the form $\downarrow\beta$, we use inductive hypothesis to show that β is provably equivalent to $\bigcup_i [\varphi_i]\beta_i$ for some $(\beta_i)_i \in P_{n-1}^-$ and $(\varphi_i)_i \in N_{n-1}^-$. By IsAx6, we conclude that F is provably equivalent to $\bigcup_i \downarrow[\varphi_i]\beta_i$, and $\downarrow[\varphi_i]\beta_i \in P_n^-$ as we wanted to show.

Finally, if F is of the form $(\gamma \cup \delta)\beta$, then, by IsAx6, $F \equiv (\gamma\beta) \cup (\delta\beta)$. The result follows from inductive hypothesis for $\gamma\beta$ and $\delta\beta$ (as usual, we may have to use Lemma 14, PrAx3, IsAx6, and **Der21** of Fact 5 to increase the degree). $\square$

3.3. Completeness for node and path expressions

In this section we show that for node expressions φ and ψ of $\mathrm{XPath}_=(\downarrow)^-$, the equivalence $\varphi \equiv \psi$ is derivable from the axiom schemes of Table 2 if and only if φ is $\mathrm{XPath}_=(\downarrow)^-$-semantically equivalent to ψ. We also show the same result for path expressions of $\mathrm{XPath}_=(\downarrow)^-$.

We first introduce the main lemma of this section, and then continue to its consequences; as the proof of this lemma is very extensive, we postpone it to Section 3.3.1.

Lemma 17. *Any node expression $\varphi \in N_n^-$ is satisfiable.*

Based on the above lemma, we arrive to the next theorem, which is the main result of this section:

Theorem 18 *(Completeness of $XPath_=(\downarrow)^-$).*

1. *Let φ and ψ be node expressions of* $\mathrm{XPath}_=(\downarrow)^-$. *Then* $\mathsf{XP}^- \vdash \varphi \equiv \psi$ *iff* $\models \varphi \equiv \psi$.
2. *Let α and β be path expressions of* $\mathrm{XPath}_=(\downarrow)^-$. *Then* $\mathsf{XP}^- \vdash \alpha \equiv \beta$ *iff* $\models \alpha \equiv \beta$.

Proof. Let us show item 1. Soundness follows from Proposition 6.

For completeness, suppose $\models \varphi \equiv \psi$. Now assume that φ is consistent and ψ is not. On the one hand, by Theorem 16, there is n such that φ is provably equivalent to $\bigvee_{1\le i\le k} \varphi_i$, for $\varphi_i \in N_n^-$. By Lemma 17, we have that in particular φ_1 (and hence φ) is satisfiable. On the other hand, by Proposition 6, ψ is unsatisfiable, and this contradicts the fact that $\models \varphi \equiv \psi$. This shows that if φ is consistent then so is ψ. Symmetrically, one can show that if ψ is consistent, then so is φ. Therefore, either φ and ψ are consistent or φ and ψ are inconsistent. In the latter case, we trivially have $\mathsf{XP}^- \vdash \varphi \equiv \psi$.

In case φ and ψ are consistent, by Theorem 16 and Lemma 14, there is n and node expressions φ' and ψ' which are disjunctions of node expressions in N_n^- such that $\mathsf{XP}^- \vdash \varphi \equiv \varphi'$ and $\mathsf{XP}^- \vdash \psi \equiv \psi'$.

Suppose that φ' contains a disjunct φ'' which is not a disjunct of ψ'. By Lemma 17, φ'' is satisfiable in some data tree $\mathcal{T}$. By Lemma 11, for any disjunct ψ'' of ψ' we have that $\varphi'' \wedge \psi''$ is inconsistent, and by Proposition 6, unsatisfiable. Hence ψ' is not satisfiable in $\mathcal{T}$, and so $\not\models \varphi \equiv \psi$, a contradiction. The case when ψ' contains a disjunct which is not a disjunct of φ' is analogous.

Then φ' and ψ' are identical, modulo reordering of disjunctions, and so $\mathsf{XP}^- \vdash \varphi' \equiv \psi'$ which implies $\mathsf{XP}^- \vdash \varphi \equiv \psi$.

For item 2, soundness follows from Proposition 6. For completeness, suppose $\models \alpha \equiv \beta$.

Suppose that α is consistent and β is not. On the one hand, by Theorem 16, there is n such that α is provably equivalent to $\bigcup_{1\le i\le k} [\varphi_i]\alpha_i$, with $\alpha_i \in P_n^-$ and $\varphi_i \in N_n^-$. Furthermore, we can assume that $[\varphi_1]\alpha_1$ is consistent (if it is not, we simply remove it from the disjunction) and so $\langle \alpha_1 = \alpha_1 \rangle$ is a conjunct of φ_1 by Lemma 10. By Lemma 17, the node expression φ_1 is satisfiable. Then, since $\langle \alpha_1 = \alpha_1 \rangle$ is a conjunct of φ_1, the path expression $[\varphi_1]\alpha_1$ is satisfiable, and so α is satisfiable. On the other hand, by Proposition 6, β is unsatisfiable, and this contradicts the fact that $\models \alpha \equiv \beta$. This shows that if α is consistent then so is β. Symmetrically, one can show that if β is consistent, then so is α. Therefore, either α and β are consistent or α and β are inconsistent. In the latter case, we trivially have $\mathsf{XP}^- \vdash \alpha \equiv \beta$.

Suppose both α and β are consistent. By Theorem 16 plus Lemma 14 we have that there is n and path expressions $\alpha_1 \ldots \alpha_k, \beta_1 \ldots \beta_\ell$ in P_n^- and node expressions $\varphi_1 \ldots \varphi_k, \psi_1 \ldots \psi_\ell \in N_n^-$ such that $\mathsf{XP}^- \vdash \alpha \equiv \bigcup_{1\le i\le k} [\varphi_i]\alpha_i$ and $\mathsf{XP}^- \vdash \beta \equiv$

$\bigcup_{1\leq j\leq \ell}[\psi_j]\beta_j$. Furthermore, we can assume that $\langle\alpha_i = \alpha_i\rangle$ is a conjunct of φ_i for $i = 1\ldots k$ and $\langle\beta_j = \beta_j\rangle$ is a conjunct of ψ_j for $j = 1\ldots\ell$.

Now, suppose that

$$[\varphi_i]\alpha_i \notin \{[\psi_1]\beta_1, \ldots, [\psi_\ell]\beta_\ell\} \tag{1}$$

for some i. Since $\varphi_i \in N_n^-$, by Lemma 17, there is a data tree $\mathcal{T} = (T, \pi)$ with root r such that $\mathcal{T}, r \models \varphi_i$. Since $\langle\alpha_i = \alpha_i\rangle$ is a conjunct of φ_i, we have that there is $y \in T$ such that $\mathcal{T}, r, y \models \alpha_i$.

Let us show that $\mathcal{T}, r, y \not\models [\psi_j]\beta_j$ for any $j \leq \ell$. Fix any j. By (1), we have that $\varphi_i \neq \psi_j$ or $\alpha_i \neq \beta_j$. In the first case, $\mathcal{T}, r, y \not\models [\psi_j]\beta_j$ follows from Lemma 11 and Proposition 6 (in particular $\mathcal{T}, r \not\models \psi_j$). If $\varphi_i = \psi_j$, we have $\alpha_i \neq \beta_j$ and $\mathcal{T}, r, y \not\models [\psi_j]\beta_j$ follows from Lemma 12.

So we have that $\mathcal{T}, r, y \models \alpha$ but $\mathcal{T}, r, y \not\models \beta$, a contradiction with our hypothesis that $\models \alpha \equiv \beta$. Hence for any i there is j such that $[\varphi_i]\alpha_i = [\psi_j]\beta_j$. Analogously one can show that for any j there is i such that $[\psi_j]\beta_j = [\varphi_i]\alpha_i$. Then $\bigcup_{1\leq i\leq k}[\varphi_i]\alpha_i$ and $\bigcup_{1\leq j\leq \ell}[\psi_j]\beta_j$ are identical, modulo reordering of unions, and so $\mathsf{XP}^- \vdash \alpha \equiv \beta$. □

All we need to complete the argument is to prove Lemma 17. Doing this involves the rest of the section.

3.3.1. Canonical model

In order to prove Lemma 17, we construct, recursively in n and for every $\varphi \in N_n^-$, a data tree $\mathcal{T}^\varphi = (T^\varphi, \pi^\varphi)$ such that φ is satisfiable in $\mathcal{T}^\varphi$.

For the base case, if $\varphi \in N_0^-$ and $\varphi = a \wedge \langle\epsilon = \epsilon\rangle$ with $a \in \mathbb{A}$, simply define the data tree $\mathcal{T}^\varphi = (T^\varphi, \pi^\varphi)$ where T^φ is a tree which consists of the single node x with label a, and $\pi^\varphi = \{\{x\}\}$.

Now let $\varphi \in N_{n+1}^-$. Since φ is a conjunction as in Definition 7, it is enough to guarantee that the following conditions hold (we now observe that these conditions are enough because of EqAx1, but we usually avoid these observations of symmetry):

(C1) If $a \in \mathbb{A}$ is a conjunct of φ, then the root r^φ of $\mathcal{T}^\varphi$ has label a.
(C2) If $\langle\epsilon = \downarrow[\psi]\alpha\rangle$ is a conjunct of φ, then there is a child $r^{\mathbf{v}}$ (where $\mathbf{v} = (\psi, \alpha)$; we will introduce this notation in time to formalize the construction) of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied and a node $x^{\mathbf{v}}$ with the same data value as r^φ such that $\mathcal{T}^\varphi, r^{\mathbf{v}}, x^{\mathbf{v}} \models \alpha$.
(C3) If $\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ is a conjunct of φ, then there are two children $r_1^{\mathbf{u}}$, $r_2^{\mathbf{u}}$ of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, and there are nodes $x^{\mathbf{u}}$ and $y^{\mathbf{u}}$ with the same data value such that $\mathcal{T}^\varphi, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$ and $\mathcal{T}^\varphi, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$.
(C4) If $\neg\langle\epsilon = \downarrow[\psi]\alpha\rangle$ is a conjunct of φ, then for each child z of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied, if x is a node such that $\mathcal{T}^\varphi, z, x \models \alpha$, then the data value of x is different than the one of r^φ.
(C5) If $\neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ is a conjunct of φ, then for each children z_1, z_2 of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, if w_1, w_2 are nodes such that $\mathcal{T}^\varphi, z_1, w_1 \models \alpha$ and $\mathcal{T}^\varphi, z_2, w_2 \models \beta$, then the data values of w_1 and w_2 are different.

Since the construction of the canonical model requires some technical notation that might hinder the understanding of the ideas behind it, we will begin with an intuitive description of the construction.

Insight into the construction

The idea to achieve all the previous conditions is to incrementally build a tree such that it satisfies at its root conditions (C1), (C4), and (C5), then also (C2) (without spoiling any previous conditions), and finally also (C3).

First we start with a root r^φ labeled a, where a is the label present in φ (so that condition (C1) is satisfied). At this point in the construction, as we only have one node, conditions (C4) and (C5) are trivially satisfied. On the contrary, (C2) and (C3) might not be satisfied, and require a positive action (i.e. changing the current model) to make them true. We want to add witnesses that guarantee the satisfaction of (C2) and (C3), and we will achieve this by the use of the inductive hypothesis to construct new trees that we will hang as children of r^φ. However, adding witnesses jeopardizes the satisfaction of (C4) and (C5), so we need to do it carefully enough.

First we add witnesses in order to satisfy condition (C2). If $\psi \in N_n^-$, by inductive hypothesis, there exists a tree $\mathcal{T}^\psi$ such that ψ is satisfied at $\mathcal{T}^\psi$. Also, if $\langle\epsilon = \downarrow[\psi]\alpha\rangle$ is a conjunct of φ, by the consistency of φ, Lemma 10 and the inductive hypothesis, there is a pair of nodes satisfying α in $\mathcal{T}^\psi$ and starting at its root. In this case, we will hang a copy of $\mathcal{T}^\psi$ (or perhaps a slightly modified copy of it constructed in order not to spoil condition (C4)) as a child of r^φ and merge the equivalence class of r^φ to the equivalence class of the endpoint $x^{\mathbf{v}}$ of a specially chosen pair of nodes satisfying α and beginning at the root of $\mathcal{T}^\psi$ (see Fig. 3(a)). This is the only merging required; other classes in $\mathcal{T}^\psi$ remain disjoint from the previous constructed part of $\mathcal{T}^\varphi$. In this way, we will guarantee condition (C2) (see Fig. 3(b)). Since the other equivalence classes of $\mathcal{T}^\psi$ will remain disjoint from the rest of the tree $\mathcal{T}^\varphi$ all along the construction and since two different normal forms cannot be satisfied at the same point (see Lemma 11 plus Proposition 6), the only way in which this process could spoil condition (C4) is that there is $\beta \in P_n^-$ such that $\neg\langle\epsilon = \downarrow[\psi]\beta\rangle$ is a conjunct of φ and a pair of nodes satisfying β

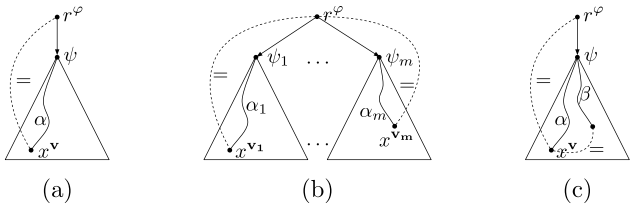

Fig. 3. (a) A witness for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$; (b) we repeat the process of (a) for each conjunct $\langle \epsilon = \downarrow[\psi_1]\alpha_1 \rangle, \ldots, \langle \epsilon = \downarrow[\psi_m]\alpha_m \rangle$ of φ; (c) by adding a witness for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$, we may be creating an unwanted witness for $\langle \epsilon = \downarrow[\psi]\beta \rangle$.

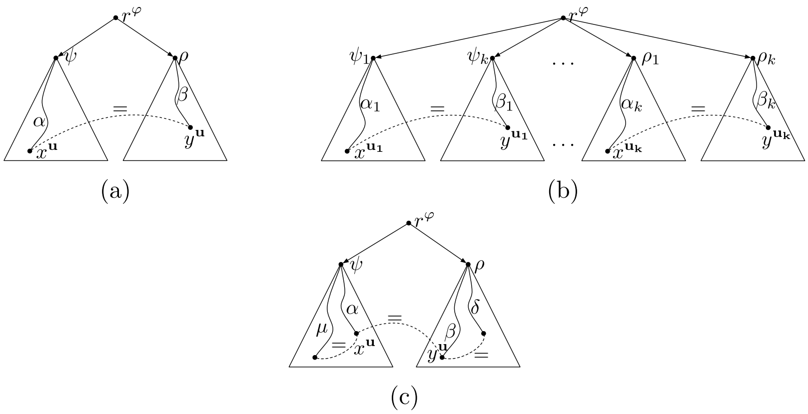

Fig. 4. (a) A witness for $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$; (b) we repeat the process of (a) for each conjunct $\langle \downarrow[\psi_1]\alpha_1 = \downarrow[\rho_1]\beta_1 \rangle, \ldots, \langle \downarrow[\psi_k]\alpha_k = \downarrow[\rho_k]\beta_k \rangle$ of φ; (c) by adding a witness for $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$, we may be creating an unwanted witness for $\langle \downarrow[\psi]\mu = \downarrow[\rho]\delta \rangle$.

in $\mathcal{T}^\psi$, starting at its root and ending in a point with the same data value as $x^{\mathbf{v}}$ (see Fig. 3(c)). But Lemma 19 ensures that (maybe with changes to $\mathcal{T}^\psi$) we can choose $x^{\mathbf{v}}$ to avoid this situation. Then, since we only add nodes to the equivalence class of the root r^φ by this process, the only way in which we could spoil condition (C5) is if φ has conjuncts $\langle \epsilon = \downarrow[\psi]\mu \rangle$, $\langle \epsilon = \downarrow[\rho]\delta \rangle$ and $\neg\langle \downarrow[\psi]\mu = \downarrow[\rho]\delta \rangle$ for some $\psi, \rho \in N_n^-$, $\mu, \delta \in P_n^-$. But this is clearly unsatisfiable and so our axioms should tell us that it is inconsistent (see EqAx7).

We now proceed to add witnesses in order to satisfy condition (C3). By an argument similar to the one given for condition (C2), if $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ is a conjunct of φ, there are, by inductive hypothesis, trees $\mathcal{T}^\psi$ and $\mathcal{T}^\rho$ at which ψ and ρ are satisfied and pairs of nodes satisfying α and β starting at their respective roots. We will hang a copy of each of those trees (or perhaps slightly modified copies of them) as children of r^φ and we will merge the equivalence classes (in $\mathcal{T}^\psi$ and $\mathcal{T}^\rho$) of the ending points $x^{\mathbf{u}}, y^{\mathbf{u}}$ of a specially chosen pair of nodes satisfying α (and starting at the root of $\mathcal{T}^\psi$) and a specially chosen pair of nodes satisfying β (and starting at the root of $\mathcal{T}^\rho$) as mentioned before (see Fig. 4(a)). Note that all the classes in $\mathcal{T}^\psi$ and $\mathcal{T}^\rho$ remain disjoint from the previous constructed part of $\mathcal{T}^\varphi$. In this way, we guarantee condition (C3) (see Fig. 4(b)). Since we are not adding any nodes to the class of r^φ, it is clear that we cannot spoil condition (C2) by performing this procedure. With a similar argument as the one given before, the only way in which we can spoil condition (C5) is that there are $\psi, \rho \in N_n^-$, $\alpha, \beta, \mu, \delta \in P_n^-$ such that $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\neg\langle \downarrow[\psi]\mu = \downarrow[\rho]\delta \rangle$ are conjuncts of φ, a pair of nodes satisfying μ beginning at the root of $\mathcal{T}^\psi$ and ending in a point with the same data value as $x^{\mathbf{u}}$, and a pair of nodes satisfying δ beginning at the root of $\mathcal{T}^\rho$ and ending in a point with the same data value as $y^{\mathbf{u}}$ (see Fig. 4(c)). But Lemma 19 ensures that we can choose $x^{\mathbf{u}}$ and $y^{\mathbf{u}}$ to avoid this situation.

Formalization

In order to formalize the construction described above, we introduce the following key lemma:

Lemma 19. *Let $\psi_0 \in N_n^-$, $\alpha, \beta_1, \ldots, \beta_m \in P_n^-$. Suppose that there exists a tree $\mathcal{T}^{\psi_0} = (T^{\psi_0}, \pi^{\psi_0})$ with root r^{ψ_0} such that $\mathcal{T}^{\psi_0}, r^{\psi_0} \models \psi_0$ and for all $i = 1, \ldots, m$ there exists $\gamma_i \in P_{n+1}^-$ such that $\langle \gamma_i = \downarrow[\psi_0]\alpha \rangle \wedge \neg\langle \gamma_i = \downarrow[\psi_0]\beta_i \rangle$ is consistent. Then there exists a tree $\widetilde{\mathcal{T}^{\psi_0}} = (\widetilde{T^{\psi_0}}, \widetilde{\pi^{\psi_0}})$ with root $\widetilde{r^{\psi_0}}$ and a node x such that:*

- $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}} \models \psi_0$,
- $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$, *and*
- $[x]_{\widetilde{\pi^{\psi_0}}} \neq [y]_{\widetilde{\pi^{\psi_0}}}$ *for all y such that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, y \models \beta_i$ for some $i = 1, \ldots, m$.*

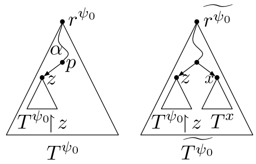

Fig. 5. $T^x = T^{\psi_0}\restriction z$ is a new subtree with disjoint data values to the rest of $\widetilde{\mathcal{T}^{\psi_0}}$. The new node x satisfies $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$.

Proof. Suppose that $\alpha = \downarrow[\psi_1]\dots\downarrow[\psi_{j_0}]\epsilon$, where $\psi_k \in N^-_{n-k}$ for all $k = 1,\dots, j_0$. If $j_0 = 0$ (that is, $\alpha = \epsilon$), then it suffices to take $\widetilde{\mathcal{T}^{\psi_0}} = \mathcal{T}^{\psi_0}$ and $x = r^{\psi_0}$. We only need to show that then $\neg\langle\epsilon = \beta_i\rangle$ is a conjunct of ψ_0 for all i. Indeed, assuming instead that $\langle\epsilon = \beta_i\rangle$ is a conjunct of ψ_0 for some i, we have

$$\begin{aligned}
\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle &\equiv \langle\gamma_i = \downarrow[\psi_0 \wedge \langle\epsilon = \beta_i\rangle]\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle \\
&\qquad\qquad (\text{Hypothesis } \langle\epsilon = \beta_i\rangle \text{ is a conjunct of } \psi_0) \\
&\le \langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle \qquad (\textbf{Der21} \text{ (Fact 5) \& EqAx8}) \\
&\equiv \text{FALSE} \qquad (\text{Boolean})
\end{aligned}$$

which is a contradiction with our assumption that $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ is consistent, by standard propositional reasoning.

If $j_0 > 0$, to define $\widetilde{\mathcal{T}^{\psi_0}} = (\widetilde{T^{\psi_0}}, \widetilde{\pi^{\psi_0}})$ we modify the tree $\mathcal{T}^{\psi_0} = (T^{\psi_0}, \pi^{\psi_0})$. From the consistency of $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle$ for some i, by Lemma 10, we conclude that $\langle\alpha = \alpha\rangle$ is a conjunct of ψ_0. Hence there is $z \in T^{\psi_0}$, $z \neq r^{\psi_0}$, such that $\mathcal{T}^{\psi_0}, r^{\psi_0}, z \models \alpha$.

Before proceeding to complete the proof of this case, we give a sketch of it. We prove that we cannot have a witness for β_i with the same data value as z in the subtree $T^{\psi_0}\restriction z$. Intuitively this is because, in that case, α would be a prefix of β_i, say $\beta_i = \alpha\delta$, and $\langle\epsilon = \delta\rangle$ would be a conjunct of ψ_{j_0}. Then $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ would be unsatisfiable (and thus it should be inconsistent) for any choice of γ_i which is a contradiction. But our hypotheses are not enough to avoid having a witness for β_i in the class of z outside $T^{\psi_0}\restriction z$; thus we need to change the tree in order to achieve the desired properties. We replicate the subtree $T^{\psi_0}\restriction z$ but using fresh data values (different from any other data value already present in $\mathcal{T}^{\psi_0}$), see Fig. 5. It is clear that in this way, the second and the third conditions will be satisfied by the root of this new subtree. The first condition will also remain true because the positive conjuncts will remain valid since we are not suppressing any nodes, and the negative ones will not be affected either because every node we add has a fresh data value.

Now we formalize the previous intuition. Call p the parent node of the aforementioned $z \in T^{\psi_0}$ and define $\widetilde{T^{\psi_0}}$ as $T^{\psi_0} \sqcup T^x$, where we define T^x as $T^{\psi_0}\restriction z$, and in $\widetilde{T^{\psi_0}}$ the root of T^x is a child x of p. Define $\widetilde{\pi^{\psi_0}}$ as $\pi^{\psi_0} \sqcup \pi^z$; observe that the data values of T^x differ from all those of the rest of $\widetilde{T^{\psi_0}}$ (see Fig. 5).

We now check that this new tree $\widetilde{T^{\psi_0}}$ satisfies ψ_0 at its root $\widetilde{r^{\psi_0}}$. We prove by induction that x_j, the j-th ancestor of x (namely $x_j \xrightarrow{j} x$, and we let $x_0 := x$), satisfies $\widetilde{\mathcal{T}^{\psi_0}}, x_j \models \psi_{j_0 - j}$. This proves both that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}} \models \psi_0$ and that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$. For the base case $j = 0$, the result is straightforward from Proposition 3: T^x is a copy of $T^{\psi_0}\restriction z$, with z satisfying ψ_{j_0}. For the inductive case, assume that the result holds for $x_0, \dots, x_j$. We want to see that it holds for x_{j+1}. To do this, we verify that every conjunct of ψ_{j_0-j-1} is satisfied at x_{j+1}:

- If the conjunct is a label, it is clear that x_{j+1} still has that label in $\widetilde{\mathcal{T}^{\psi_0}}$, as it has not been changed by the construction.
- If the conjunct is of the form $\langle\mu_1 = \mu_2\rangle$, then it must still hold in $\widetilde{\mathcal{T}^{\psi_0}}$ by inductive hypothesis plus the fact that our construction did not remove nodes.
- If the conjunct is of the form $\neg\langle\mu_1 = \mu_2\rangle$, we observe that, by inductive hypothesis plus the fact that the data classes of nodes in T^x are disjoint with those of the rest of $\widetilde{\mathcal{T}^{\psi_0}}$, then $\langle\mu_1 = \mu_2\rangle$ can only be true in x_{j+1} if there are witnesses y_1, y_2 in the same equivalence class in the new subtree T^x such that $\widetilde{\mathcal{T}^{\psi_0}}, x_{j+1}, y_1 \models \mu_1$ and $\widetilde{\mathcal{T}^{\psi_0}}, x_{j+1}, y_2 \models \mu_2$. In that case, we have that

  $$\mu_1 = \downarrow[\psi_{j_0-j}]\downarrow\dots\downarrow[\psi_{j_0}]\hat{\mu_1} \quad \text{and} \quad \mu_2 = \downarrow[\psi_{j_0-j}]\downarrow\dots\downarrow[\psi_{j_0}]\hat{\mu_2}$$

  for some $\hat{\mu_1}, \hat{\mu_2}$, and that $\widetilde{\mathcal{T}^{\psi_0}}, x_0, y_1 \models \hat{\mu_1}$, $\widetilde{\mathcal{T}^{\psi_0}}, x_0, y_2 \models \hat{\mu_2}$. Therefore, by Lemma 9, $\langle\hat{\mu_1} = \hat{\mu_2}\rangle$ is a conjunct of ψ_{j_0}, and then $\mathcal{T}^{\psi_0}, z \models \langle\hat{\mu_1} = \hat{\mu_2}\rangle$, a contradiction with our assumption that $\neg\langle\downarrow[\psi_{j_0-j}]\downarrow\dots\downarrow[\psi_{j_0}]\hat{\mu_1} = \downarrow[\psi_{j_0-j}]\downarrow\dots\downarrow[\psi_{j_0}]\hat{\mu_2}\rangle$ is a conjunct of $\psi_{j_0-(j+1)}$ and $\mathcal{T}^{\psi_0}, x_{j+1} \models \psi_{j_0-(j+1)}$.

To conclude the proof, we only need to check that $[x]_{\widetilde{\pi^{\psi_0}}} \neq [y]_{\widetilde{\pi^{\psi_0}}}$ for all y such that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, y \models \beta_i$ for some $i = 1, \dots, m$. Suppose that $\beta_i = \downarrow[\rho_1]\dots\downarrow[\rho_{l_0}]\epsilon$. If $l_0 < j_0$ or $\rho_l \neq \psi_l$ for some $l = 1, \dots, j_0$, then the result follows immediately from the construction. Otherwise, $l_0 \geq j_0$ and $\rho_l = \psi_l$ for all $l = 1, \dots, j_0$ and so, by hypothesis, there exists $\gamma_i \in P^-_{n+1}$ such that

$$\langle\gamma_i = \downarrow[\psi_0]\downarrow[\psi_1]\downarrow\dots\downarrow[\psi_{j_0}]\epsilon\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\dots\downarrow[\psi_{j_0}]\downarrow[\rho_{j_0+1}]\dots\downarrow[\rho_{l_0}]\epsilon\rangle$$

is consistent. We prove that $\neg\langle\epsilon = \downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle$ is a conjunct of ψ_{j_0}, from which our desired property follows immediately since we have proved that $\widetilde{\mathcal{T}^{\psi_0}}, x \models \psi_{j_0}$. Aiming for a contradiction, suppose instead that $\langle\epsilon = \downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle$ is a conjunct of ψ_{j_0}. Then, as $\alpha = \downarrow[\psi_1]\ldots\downarrow[\psi_{j_0}]\epsilon$, we can derive that $\alpha \equiv \alpha[\langle\epsilon = \downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle]$ (**Der21** (Fact 5)). Also observe that $\mathsf{XP}^- \vdash \beta_i \equiv \alpha\downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon$, and then we have

$$\begin{aligned}
\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle &\equiv \langle\gamma_i = \downarrow[\psi_0]\alpha[\langle\epsilon = \downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle]\rangle && (\textbf{Der21}\ (\text{Fact 5}))\\
&\leq \langle\gamma_i = \downarrow[\psi_0]\alpha\downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle && (\text{EqAx8})\\
&\equiv \langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle
\end{aligned}$$

But using simple propositional reasoning, we have a contradiction with our hypothesis that $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ was consistent, a contradiction that came from assuming that $\langle\epsilon = \downarrow[\rho_{j_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle$ was a conjunct of ψ_{j_0}. $\square$

Now that we have proved this lemma, we proceed to the formal construction of $\mathcal{T}^{\varphi}$, for $\varphi \in N^-_{n+1}$ (recall the base case at the beginning of §3.3.1).

Consider the following sets:

$$\mathbf{V} = \{(\psi, \alpha) \mid \langle\epsilon = \downarrow[\psi]\alpha\rangle \text{ is a conjunct of } \varphi\}$$

$$\mathbf{U} = \{(\psi, \alpha, \rho, \beta) \mid \langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle \text{ is a conjunct of } \varphi\}$$

Rule 1. Witnesses for $\mathbf{v} = (\boldsymbol{\psi}, \boldsymbol{\alpha}) \in \boldsymbol{V}$ We define a data tree $\mathcal{T}^{\mathbf{v}} = (T^{\mathbf{v}}, \pi^{\mathbf{v}})$ with root $r^{\mathbf{v}}$. By inductive hypothesis, there exists a tree $\mathcal{T}^{\psi}$ such that ψ is satisfiable in that tree. In Lemma 19, consider

$$\begin{aligned}
\psi_0 &:= \psi\\
\mathcal{T}^{\psi_0} &:= \mathcal{T}^{\psi}\\
\alpha &:= \alpha\\
\{\beta_1, \ldots, \beta_m\} &:= \{\beta \in P^-_n \mid \neg\langle\epsilon = \downarrow[\psi]\beta\rangle \text{ is a conjunct of } \varphi\}\\
\gamma_i &:= \epsilon \text{ for all } i = 1, \ldots, m
\end{aligned}$$

Then there exists $\widetilde{\mathcal{T}^{\psi}} = (\widetilde{T^{\psi}}, \widetilde{\pi^{\psi}})$ with root $\widetilde{r^{\psi}}$ and a node x such that

- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}} \models \psi$,
- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, x \models \alpha$,
- $[x]_{\widetilde{\pi^{\psi}}} \neq [y]_{\widetilde{\pi^{\psi}}}$ for all y such that there is $\beta \in P^-_n$ with $(\psi, \beta) \notin \mathbf{V}$, and $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, y \models \beta$

Define $\mathcal{T}^{\mathbf{v}}$ as $\widetilde{\mathcal{T}^{\psi}}$, and $x^{\mathbf{v}}$ as x. The root $r^{\mathbf{v}}$ and the partition $\pi^{\mathbf{v}}$ are the ones of $\widetilde{\mathcal{T}^{\psi}}$.

Rule 2. Witnesses for $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in U$ We define data trees $\mathcal{T}^{\mathbf{u}}_1 = (T^{\mathbf{u}}_1, \pi^{\mathbf{u}}_1)$ and $\mathcal{T}^{\mathbf{u}}_2 = (T^{\mathbf{u}}_2, \pi^{\mathbf{u}}_2)$ with roots $r^{\mathbf{u}}_1$, $r^{\mathbf{u}}_2$ respectively. By inductive hypothesis, there exist trees $\mathcal{T}^{\psi} = (T^{\psi}, \pi^{\psi})$ (with root r^{ψ}) and $\mathcal{T}^{\rho} = (T^{\rho}, \pi^{\rho})$ (with root r^{ρ}) such that ψ is satisfiable in $\mathcal{T}^{\psi}$ and ρ is satisfiable in $\mathcal{T}^{\rho}$. In Lemma 19, consider

$$\begin{aligned}
\psi_0 &:= \psi\\
\mathcal{T}^{\psi_0} &:= \mathcal{T}^{\psi}\\
\alpha &:= \alpha\\
\{\beta_1, \ldots, \beta_m\} &:= \{\gamma \in P^-_n \mid \neg\langle\downarrow[\rho]\beta = \downarrow[\psi]\gamma\rangle \text{ is a conjunct of } \varphi\}\\
\gamma_i &:= \downarrow[\rho]\beta \text{ for all } i = 1, \ldots, m
\end{aligned}$$

Then there exist $\widetilde{\mathcal{T}^{\psi}} = (\widetilde{T^{\psi}}, \widetilde{\pi^{\psi}})$ with root $\widetilde{r^{\psi}}$ and a node x such that:

- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}} \models \psi$,
- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, x \models \alpha$,
- $[x]_{\widetilde{\pi^{\psi}}} \neq [y]_{\widetilde{\pi^{\psi}}}$ for all y such that there is $\gamma \in P^-_n$ with $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, y \models \gamma$ and $\neg\langle\downarrow[\rho]\beta = \downarrow[\psi]\gamma\rangle$ is a conjunct of φ.

Define $T^{\mathbf{u}}_1$ as $\widetilde{T^{\psi}}$, $\pi^{\mathbf{u}}_1$ as $\widetilde{\pi^{\psi}}$, $r^{\mathbf{u}}_1$ as $\widetilde{r^{\psi}}$ and $x^{\mathbf{u}} = x \in T^{\mathbf{u}}_1$. Now let

$$\{\mu_1, \ldots, \mu_r\} = \left\{\mu \in P^-_n \mid \text{there exists } y \in T^{\mathbf{u}}_1 \text{ such that } \mathcal{T}^{\mathbf{u}}_1, r^{\mathbf{u}}_1, y \models \mu \text{ and } [y]_{\pi^{\mathbf{u}}_1} = [x^{\mathbf{u}}]_{\pi^{\mathbf{u}}_1}\right\}.$$

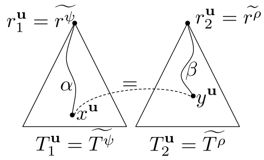

Fig. 6. The data trees $\mathcal{T}_1^{\mathbf{u}} = (T_1^{\mathbf{u}}, \pi_1^{\mathbf{u}})$ and $\mathcal{T}_2^{\mathbf{u}} = (T_2^{\mathbf{u}}, \pi_2^{\mathbf{u}})$ for some $\mathbf{u} \in \mathbf{U}$. $\pi^{\mathbf{u}_1}$ and $\pi^{\mathbf{u}_2}$ are disjoint except that the equivalence class of $x^{\mathbf{u}}$ is merged with the equivalence class of $y^{\mathbf{u}}$.

Then it follows that $\langle \downarrow[\rho]\beta = \downarrow[\psi]\mu_j \rangle$ is a conjunct of φ for all $j = 1, \ldots, r$. In Lemma 19, consider

$$\begin{aligned}
\psi_0 &:= \rho \\
\mathcal{T}^{\psi_0} &:= \mathcal{T}^{\rho} \\
\alpha &:= \beta \\
\{\beta_1, \ldots, \beta_m\} &:= \{\delta \in P_n^- \mid \exists j = 1, \ldots, r \text{ with } \neg\langle \downarrow[\rho]\delta = \downarrow[\psi]\mu_j \rangle \text{ is a conjunct of } \varphi\} \\
\gamma_i &:= \downarrow[\psi]\mu_j \text{ for } j = 1, \ldots r \text{ such that } \langle \downarrow[\rho]\beta_i = \downarrow[\psi]\mu_j \rangle \text{ is a conjunct of } \varphi.
\end{aligned}$$

Then there exist a tree $\widetilde{\mathcal{T}^{\rho}} = (\widetilde{T^{\rho}}, \widetilde{\pi^{\rho}})$ with root $\widetilde{r^{\rho}}$ and a node y such that

- $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}} \models \rho$,
- $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}}, y \models \beta$,
- $[y]_{\widetilde{\pi^{\rho}}} \neq [z]_{\widetilde{\pi^{\rho}}}$ for all z such that there is $\delta \in P_n^-$ and $j = 1, \ldots, r$ with $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}}, z \models \delta$ and $\neg\langle \downarrow[\rho]\delta = \downarrow[\psi]\mu_j \rangle$ is a conjunct of φ.

Define $T_2^{\mathbf{u}}$ as $\widetilde{T^{\rho}}$, $\pi_2^{\mathbf{u}}$ as $\widetilde{\pi^{\rho}}$, $r_2^{\mathbf{u}}$ as $\widetilde{r^{\rho}}$ and $y^{\mathbf{u}} = y$. Without loss of generality, we assume that $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ are disjoint.

Now we define a partition $\pi^{\mathbf{u}}$ over $T_1^{\mathbf{u}} \cup T_2^{\mathbf{u}}$ as

$$\pi^{\mathbf{u}} = \pi_1^{\mathbf{u}} \cup \pi_2^{\mathbf{u}} \cup \{[x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}} \cup [y^{\mathbf{u}}]_{\pi_2^{\mathbf{u}}}\} \setminus \{[x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}}, [y^{\mathbf{u}}]_{\pi_2^{\mathbf{u}}}\}.$$

In other words, the rooted data tree $(T_1^{\mathbf{u}}, \pi^{\mathbf{u}}\restriction_{T_1^{\mathbf{u}}}, r_1^{\mathbf{u}})$ is just a copy of $(\widetilde{T^{\psi}}, \widetilde{\pi^{\psi}}, \widetilde{r^{\psi}})$, with a special node named $x^{\mathbf{u}}$ which satisfies $T_1^{\mathbf{u}}, \pi^{\mathbf{u}}, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$. Analogously, the pointed data tree $(T_2^{\mathbf{u}}, \pi^{\mathbf{u}}\restriction_{T_2^{\mathbf{u}}}, r_2^{\mathbf{u}})$ is a copy of $(\widetilde{T^{\rho}}, \widetilde{\pi^{\rho}}, \widetilde{r^{\rho}})$, with a special node named $y^{\mathbf{u}}$ which satisfies $T_2^{\mathbf{u}}, \pi^{\mathbf{u}}, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$. Notice that the equivalence class $\sim$ induced by $\pi^{\mathbf{u}}$ (defined over the disjoint sets $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$) is defined as $z \sim w$ iff $w \in [z]_{\widetilde{\pi^{\psi}}}$ or $w \in [z]_{\widetilde{\pi^{\rho}}}$, or both $w \in [x^{\mathbf{u}}]_{\widetilde{\pi^{\psi}}}$ and $z \in [y^{\mathbf{u}}]_{\widetilde{\pi^{\rho}}}$ or both $w \in [y^{\mathbf{u}}]_{\widetilde{\pi^{\rho}}}$ and $z \in [x^{\mathbf{u}}]_{\widetilde{\pi^{\psi}}}$. See Fig. 6.

The following remark will be used later to prove that Rule 2 does not spoil condition (C5) (cf. Fig. 4(c)):

Remark 20. Let $(\psi, \alpha, \rho, \beta) \in \mathbf{U}$. If $\neg\langle \downarrow[\psi]\mu = \downarrow[\rho]\delta \rangle$ is a conjunct of φ, then $[y^{\mathbf{u}}]_{\pi_2^{\mathbf{u}}} \neq [y]_{\pi_2^{\mathbf{u}}}$ for all y such that $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y \models \delta$ or $[x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}} \neq [x]_{\pi_1^{\mathbf{u}}}$ for all x such that $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x \models \mu$.

Proof. The result is immediate from Rule 2: If neither of the conditions is satisfied, then $\mu = \mu_j$ for some $j = 1, \ldots, r$ and so $\langle \downarrow[\rho]\delta = \downarrow[\psi]\mu \rangle$ is a conjunct of φ which is a contradiction. □

The rooted data tree $(\boldsymbol{T}^{\boldsymbol{\varphi}}, \boldsymbol{\pi}^{\boldsymbol{\varphi}}, \boldsymbol{r}^{\boldsymbol{\varphi}})$ Now, using Rule 1 and Rule 2, we define T^{φ} as the tree which consists of a root r^{φ} with label $a \in \mathbb{A}$ if a is a conjunct of φ, and with children

$$(T^{\mathbf{v}})_{\mathbf{v} \in \mathbf{V}}\ ,\ (T_1^{\mathbf{u}})_{\mathbf{u} \in \mathbf{U}}\ ,\ (T_2^{\mathbf{u}})_{\mathbf{u} \in \mathbf{U}}.$$

We assume that the nodes of all such trees are pairwise disjoint. Define π^{φ} over T^{φ} by

$$\pi^{\varphi} = \left(\bigcup_{\mathbf{v} \in \mathbf{V}} \pi^{\mathbf{v}} \setminus \{[x^{\mathbf{v}}]_{\pi^{\mathbf{v}}} \mid \mathbf{v} \in \mathbf{V}\} \right) \cup \left\{ \{r^{\varphi}\} \cup \bigcup_{\mathbf{v} \in \mathbf{V}} [x^{\mathbf{v}}]_{\pi^{\mathbf{v}}} \right\} \cup \bigcup_{\mathbf{u} \in \mathbf{U}} \pi^{\mathbf{u}}.$$

In other words, T^{φ} has a root, named r^{φ}, and children $(r^{\mathbf{v}})_{\mathbf{v} \in \mathbf{V}}$, $(r_1^{\mathbf{u}})_{\mathbf{u} \in \mathbf{U}}$, $(r_2^{\mathbf{u}})_{\mathbf{u} \in \mathbf{U}}$. Each of these children is the root of its corresponding tree inside T^{φ} as defined above, i.e. for each $\mathbf{v} \in \mathbf{V}$, $r^{\mathbf{v}}$ is the root of $T^{\mathbf{v}}$, and for each $\mathbf{u} \in \mathbf{U}$, $r_i^{\mathbf{u}}$ is the root of $T_i^{\mathbf{u}}$ $(i = 1, 2)$. All these subtrees are disjoint, and π^{φ} is defined as the disjoint union of partitions $\pi^{\mathbf{v}}$ for $\mathbf{v} \in \mathbf{V}$, and all $\pi^{\mathbf{u}}$ for $\mathbf{u} \in \mathbf{U}$, *with the exception* that we put into the same class the nodes r^{φ} and $(x^{\mathbf{v}})_{\mathbf{v} \in \mathbf{V}}$. See Fig. 7.

The following fact follows easily by construction:

Fact 21. The partition restricted to the trees $T^{\mathbf{v}}$ for $\mathbf{v} \in \mathbf{V}$ and the partition restricted to the trees $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ for $\mathbf{u} \in \mathbf{U}$ remains unchanged. More formally:

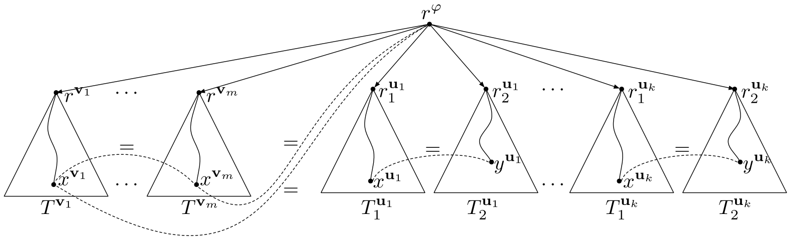

Fig. 7. The data tree $\mathcal{T}^\varphi$, with root r^φ, when $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ and $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. Nodes $r^\varphi, x^{\mathbf{v}_1}, \dots, x^{\mathbf{v}_m}$ are in the same equivalence class, and for each i nodes $x^{\mathbf{u}_i}$ and $y^{\mathbf{u}_i}$ are in the same equivalence class.

- For each $\mathbf{v} = (\psi, \alpha) \in \mathbf{V}$, we have $\pi^\varphi \restriction_{T^{\mathbf{v}}} = \pi^{\mathbf{v}}$.
- For each $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U}$, we have $\pi^\varphi \restriction_{T_1^{\mathbf{u}}} = \pi_1^{\mathbf{u}}$, and $\pi^\varphi \restriction_{T_2^{\mathbf{u}}} = \pi_2^{\mathbf{u}}$.

We conclude from Proposition 3 and the construction that:

Fact 22. The validity of a formula in a child of r^φ is preserved in $\mathcal{T}^\varphi$. More formally:

- For each $\mathbf{v} \in \mathbf{V}$ and $x, y \in T^{\mathbf{v}}$ we have $\mathcal{T}^\varphi, x \equiv^- \mathcal{T}^{\mathbf{v}}, x$ and $\mathcal{T}^\varphi, x, y \equiv^- \mathcal{T}^{\mathbf{v}}, x, y$.
- For each $\mathbf{u} \in \mathbf{U}$, $i \in \{1, 2\}$ and $x, y \in T_i^{\mathbf{u}}$ we have $\mathcal{T}^\varphi, x \equiv^- \mathcal{T}_i^{\mathbf{u}}, x$ and $\mathcal{T}^\varphi, x, y \equiv^- \mathcal{T}_i^{\mathbf{u}}, x, y$.

It only remains to check that conditions (C1)–(C5) at the beginning of §3.3 are satisfied:

Verification of (C1) This condition is trivially satisfied.

Verification of (C2) Suppose $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ is a conjunct of φ. Then, by Rule 1, there is $x^{\mathbf{v}} \in T^\varphi$ such that $[r^\varphi]_{\pi^\varphi} = [x^{\mathbf{v}}]_{\pi^\varphi}$, with $\mathbf{v} = (\psi, \alpha)$. We also know by construction that $\mathcal{T}^{\mathbf{v}}, r^{\mathbf{v}} \models \psi$ and $\mathcal{T}^{\mathbf{v}}, r^{\mathbf{v}}, x^{\mathbf{v}} \models \alpha$. By Fact 22 we conclude $\mathcal{T}^\varphi, r^\varphi \models \langle \epsilon = \downarrow[\psi]\alpha \rangle$.

Verification of (C3) Suppose $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ is a conjunct of φ. Then, by Rule 2, there are $x^{\mathbf{u}}, y^{\mathbf{u}} \in \mathcal{T}^\varphi$ such that $[x^{\mathbf{u}}]_{\pi^\varphi} = [y^{\mathbf{u}}]_{\pi^\varphi}$, with $\mathbf{u} = (\psi, \alpha, \rho, \beta)$. We also know on the one hand that $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}} \models \psi$ and $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}} \models \rho$, and on the other hand that $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$ and $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$. By Fact 22 we conclude $\mathcal{T}^\varphi, r^\varphi \models \langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$.

Verification of (C4) Suppose $\neg\langle \epsilon = \downarrow[\psi]\alpha \rangle$ is a conjunct of φ. Aiming for a contradiction, suppose that $\mathcal{T}^\varphi, r^\varphi \models \langle \epsilon = \downarrow[\psi]\alpha \rangle$. Then there is a successor z of r^φ in which ψ holds, and by construction plus Lemma 11, z is the root of some copy of a data tree $\widetilde{\mathcal{T}^\psi}$. Moreover, there is $x \in \widetilde{T^\psi}$ such that $\mathcal{T}^\varphi, z, x \models \alpha$, with $[x]_{\pi^\varphi} = [r^\varphi]_{\pi^\varphi}$. In addition to this, $(\psi, \alpha) \notin \mathbf{V}$ and so, by Rule 1, $[x]_{\pi^\varphi} \neq [x^{\mathbf{v}}]_{\pi^\varphi}$ for all $\mathbf{v} \in \mathbf{V}$. Then, by construction, $[x]_{\pi^\varphi} \neq [r^\varphi]_{\pi^\varphi}$ which is a contradiction.

Verification of (C5) Suppose $\neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ is a conjunct of φ. Aiming for a contradiction, suppose that $\mathcal{T}^\varphi, r^\varphi \models \langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$. Then there are successors z_1 and z_2 of r^φ in which ψ and ρ holds, respectively. Also, by construction and Lemma 11, z_1 and z_2 are the roots of some copies of data trees $\widetilde{\mathcal{T}^\psi}$ and $\widetilde{\mathcal{T}^\rho}$ (note that we are using the notation $\widetilde{\mathcal{T}^\psi}$ and $\widetilde{\mathcal{T}^\rho}$ either if the tree is the one obtained by inductive hypothesis or a modified version of it). Moreover, there are descendants w_1 and w_2 such that $\mathcal{T}^\varphi, z_1, w_1 \models \alpha$, $\mathcal{T}^\varphi, z_2, w_2 \models \beta$ and $[w_1]_{\pi^\varphi} = [w_2]_{\pi^\varphi}$. We now have two cases to analyze:

- $\widetilde{\mathcal{T}^\psi} = \widetilde{\mathcal{T}^\rho}$: In this case, because of Lemma 11, $\psi = \rho$. And we have $\widetilde{\mathcal{T}^\psi}, \widetilde{r^\psi} \models \langle \alpha = \beta \rangle$, and as a consequence $\langle \alpha = \beta \rangle$ has to be a conjunct of ψ (Lemma 9). We prove that in this case $\langle \downarrow[\psi]\alpha' = \downarrow[\psi]\alpha' \rangle$ can not be a conjunct of φ for any $\alpha' \in P_n^-$: If this were the case, $\langle \downarrow[\psi]\alpha' = \downarrow[\psi]\alpha' \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ would be consistent, but:

$$
\begin{aligned}
&\langle \downarrow[\psi]\alpha' = \downarrow[\psi]\alpha' \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle \\
&\quad \leq \langle \downarrow[\psi] \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle && (\textbf{Der12}\ (\text{Fact 5})) \\
&\quad \equiv \langle \downarrow[\psi \wedge \langle \alpha = \beta \rangle] \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle && (\langle \alpha = \beta \rangle \text{ is a conjunct of } \psi) \\
&\quad \equiv \langle \downarrow[\psi][\langle \alpha = \beta \rangle] \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle && (\textbf{Der21}\ (\text{Fact 5})) \\
&\quad \leq \langle \downarrow[\psi]\alpha = \downarrow[\psi]\beta \rangle \wedge \neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle && (\text{EqAx5}) \\
&\quad \equiv \text{FALSE} && (\text{Boolean})
\end{aligned}
$$

which is a contradiction.

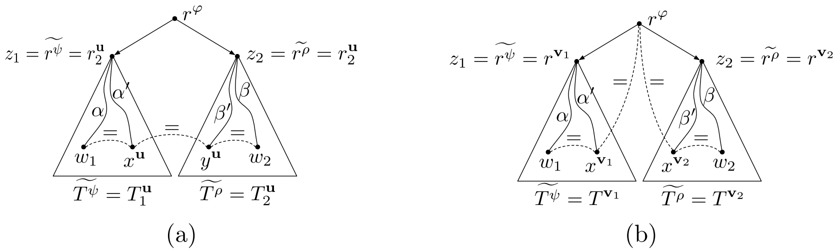

Fig. 8. Nodes w_1 and w_2 are in the same equivalence class because (a) Rule 2 was applied via $\mathbf{u} = (\psi, \alpha', \rho, \beta') \in \mathbf{U}$, or (b) Rule 1 was applied twice via $\mathbf{v}_1 = (\psi, \alpha')$, $\mathbf{v}_2 = (\rho, \beta') \in \mathbf{V}$.

Table 3
Additional axiom schemes to allow for data inequality tests. The axiomatic system XP consists of all the instantiations of this table, plus the ones of Table 2.

Node axiom schemes for inequality				
NeqAx1	$\langle \alpha \neq \beta \rangle$	$\equiv$	$\langle \beta \neq \alpha \rangle$	Analogous to EqAx1 – EqAx5 but with symbol $\neq$ instead of $=$
NeqAx2	$\langle \alpha \cup \beta \neq \gamma \rangle$	$\equiv$	$\langle \alpha \neq \gamma \rangle \vee \langle \beta \neq \gamma \rangle$	
NeqAx3	$\varphi \wedge \langle \alpha \neq \beta \rangle$	$\equiv$	$\langle [\varphi]\alpha \neq \beta \rangle$	
NeqAx4	$\langle \alpha \neq \beta \rangle$	$\leq$	$\langle \alpha \rangle$	
NeqAx5	$\langle \gamma[\langle \alpha \neq \beta \rangle] \rangle$	$\leq$	$\langle \gamma\alpha \neq \gamma\beta \rangle$	
NeqAx6	$\langle \alpha = \gamma \rangle \wedge \langle \beta = \eta \rangle$	$\leq$	$\langle \alpha = \beta \rangle \vee \langle \gamma \neq \eta \rangle$	
NeqAx7	$\langle \alpha \neq \gamma \rangle \wedge \langle \beta = \eta \rangle$	$\leq$	$\langle \alpha \neq \beta \rangle \vee \langle \gamma \neq \eta \rangle$	
NeqAx8	$\langle \gamma = \eta[\neg\langle \alpha = \beta \rangle \wedge \langle \alpha \rangle]\beta \rangle$	$\leq$	$\langle \gamma \neq \eta\alpha \rangle$	
NeqAx9	$\langle \gamma \neq \eta[\neg\langle \alpha \neq \beta \rangle \wedge \langle \alpha \rangle]\beta \rangle$	$\leq$	$\langle \gamma \neq \eta\alpha \rangle$	
NeqAx10	$\langle \gamma = \eta[\neg\langle \alpha \neq \alpha \rangle \wedge \langle \alpha = \beta \rangle]\alpha \rangle$	$\leq$	$\langle \gamma = \eta\beta \rangle$	

Then, $\langle{\downarrow}[\psi]\alpha' = {\downarrow}[\psi]\alpha'\rangle$ is not a conjunct of φ and so, it follows easily from the consistency of φ that $(\psi, \alpha') \notin \mathbf{V}$ for all $\alpha' \in P_n^-$. And also $(\psi, \alpha', \rho', \beta') \notin \mathbf{U}$ for all $\alpha', \beta' \in P_n^-$, $\rho' \in N_n^-$. This gives a contradiction by construction because in this case it would not be a copy of a tree $\widetilde{\mathcal{T}^\psi}$.

- $\widetilde{\mathcal{T}^\psi} \neq \widetilde{\mathcal{T}^\rho}$: In this case, there are two possibilities to consider:
  - One possibility is that $[w_1]_{\pi^\varphi} = [w_2]_{\pi^\varphi}$ because Rule 2 was applied (see Fig. 8 (a)). Then there is $\mathbf{u} = (\psi, \alpha', \rho, \beta') \in \mathbf{U}$ (the symmetric case is analogous). In this case we have $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha'$ and $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta'$. Furthermore, since $[w_1]_{\pi^\varphi} = [w_2]_{\pi^\varphi}$, we have that $[w_1]_{\pi_1^{\mathbf{u}}} = [x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}}$ and $[w_2]_{\pi_2^{\mathbf{u}}} = [y^{\mathbf{u}}]_{\pi_2^{\mathbf{u}}}$ which is a contradiction by Remark 20.
  - The other possibility is that $[w_1]_{\pi^\varphi} = [w_2]_{\pi^\varphi}$ because Rule 1 was applied twice (see Fig. 8(b)). Then there exist $\mathbf{v}_1 = (\psi, \alpha')$, $\mathbf{v}_2 = (\rho, \beta') \in \mathbf{V}$. In this case we have $\mathcal{T}^{\mathbf{v}_1}, r^{\mathbf{v}_1}, x^{\mathbf{v}_1} \models \alpha'$ and $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2}, x^{\mathbf{v}_2} \models \beta'$. Furthermore, since $[w_1]_{\pi^\varphi} = [w_2]_{\pi^\varphi}$, we have that $[w_1]_{\pi^{\mathbf{v}_1}} = [x^{\mathbf{v}_1}]_{\pi^{\mathbf{v}_1}}$ and $[w_2]_{\pi^{\mathbf{v}_2}} = [y^{\mathbf{v}_2}]_{\pi^{\mathbf{v}_2}}$. Then, by Rule 1, (ψ, α) and (ρ, β) belong to $\mathbf{V}$ which gives a contradiction because of the consistency of φ plus EqAx7.

4. Axiomatic system for XPath$_=(\downarrow)$

4.1. Axiomatization

In this section we introduce additional axiom schemes to handle inequalities. Axioms schemes in Table 3 extend those from Table 2 to form a complete axiomatic system for the full logic XPath$_=(\downarrow)$. Observe that NeqAx1 – NeqAx5 are analogous to EqAx1 – EqAx5.

Let XP be the set of all instantiations of the axiom schemes from Table 2 plus the ones from Table 3. In the scope of this section we will often say that a node expression is *consistent* meaning that it is XP-consistent (as in Definition 4).

Sometimes we use NeqAx1 and NeqAx4 without explicitly mentioning them. We omit such steps in order to make the proofs more readable. We also note that NeqAx2 and NeqAx3 are necessary for the proof of Theorem 28, which is omitted; they have to be used in the same way as EqAx2 and EqAx3 in the proof of Theorem 16.

It is not difficult to see that the axioms XP are sound for XPath$_=(\downarrow)$:

Proposition 23 *(Soundness of XPath$_=(\downarrow)$).*

1. *Let φ and ψ be node expressions of* XPath$_=(\downarrow)$*. Then* XP $\vdash \varphi \equiv \psi$ *implies* $\models \varphi \equiv \psi$.
2. *Let α and β be path expressions of* XPath$_=(\downarrow)$*. Then* XP $\vdash \alpha \equiv \beta$ *implies* $\models \alpha \equiv \beta$.

4.2. Normal forms

We define the sets P_n and N_n, that contain the path and node expressions of XPath$_=(\downarrow)$, respectively, in normal form at level n:

Definition 24 *(Normal form for XPath$_=(\downarrow)$).*

$$\begin{aligned}
P_0 &= \{\epsilon\}\\
N_0 &= \{a \wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle \mid a \in \mathbb{A}\}\\
P_{n+1} &= \{\epsilon\} \cup \{\downarrow[\psi]\beta \mid \psi \in N_n, \beta \in P_n\}\\
D_{n+1} &= \{\langle \alpha = \beta \rangle \mid \alpha, \beta \in P_{n+1}\} \cup \{\langle \alpha \neq \beta \rangle \mid \alpha, \beta \in P_{n+1}\}\\
N_{n+1} &= \left\{ a \wedge \bigwedge_{\varphi \in C} \varphi \wedge \bigwedge_{\varphi \in D_{n+1} \setminus C} \neg\varphi \mid C \subseteq D_{n+1}, a \in \mathbb{A} \right\} \cap \mathbf{Con}_{\mathsf{XP}}.
\end{aligned}$$

Normal forms are built using the same idea from §3.2, but considering also data-aware diamonds with inequalities. Again, let us remark that it would suffice that N_0 contains formulas of the form a, for $a \in \mathbb{A}$, but we include instead formulas of the form $a \wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle$ (containing the tautologies $\langle \epsilon = \epsilon \rangle$ and $\neg\langle \epsilon \neq \epsilon \rangle$) for technical reasons. For instance, considering again two labels a and b, the node expressions of N_0 are

$$\psi = a \wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle \quad \text{and} \quad \theta = b \wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle.$$

The sets P_1 and D_1 are as follows:

$$\begin{aligned}
P_1 &= \{\downarrow[\psi]\epsilon, \downarrow[\theta]\epsilon, \epsilon\}\\
D_1 &= \{\langle \epsilon = \epsilon \rangle, \langle \downarrow[\psi]\epsilon = \downarrow[\theta]\epsilon \rangle, \langle \epsilon = \downarrow[\psi]\epsilon \rangle, \langle \epsilon = \downarrow[\theta]\epsilon \rangle, \langle \downarrow[\psi]\epsilon = \downarrow[\psi]\epsilon \rangle, \langle \downarrow[\theta]\epsilon = \downarrow[\theta]\epsilon \rangle,\\
&\quad \langle \epsilon \neq \epsilon \rangle, \langle \downarrow[\psi]\epsilon \neq \downarrow[\theta]\epsilon \rangle, \langle \epsilon \neq \downarrow[\psi]\epsilon \rangle, \langle \epsilon \neq \downarrow[\theta]\epsilon \rangle, \langle \downarrow[\psi]\epsilon \neq \downarrow[\psi]\epsilon \rangle, \langle \downarrow[\theta]\epsilon \neq \downarrow[\theta]\epsilon \rangle\}.
\end{aligned}$$

An example of a node expression in normal form at level 1, i.e. a node expression in N_1, is

$$\begin{aligned}
\varphi = a &\wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle \wedge \langle \downarrow[\psi]\epsilon = \downarrow[\theta]\epsilon \rangle \wedge \langle \downarrow[\psi]\epsilon = \downarrow[\psi]\epsilon \rangle \wedge \langle \downarrow[\theta]\epsilon = \downarrow[\theta]\epsilon \rangle \wedge\\
&\wedge \langle \epsilon \neq \downarrow[\psi]\epsilon \rangle \wedge \langle \epsilon \neq \downarrow[\theta]\epsilon \rangle \wedge \langle \downarrow[\psi]\epsilon \neq \downarrow[\theta]\epsilon \rangle \wedge \neg\langle \epsilon = \downarrow[\psi]\epsilon \rangle \wedge \neg\langle \epsilon = \downarrow[\theta]\epsilon \rangle \wedge\\
&\wedge \langle \downarrow[\theta]\epsilon \neq \downarrow[\theta]\epsilon \rangle \wedge \langle \downarrow[\psi]\epsilon \neq \downarrow[\psi]\epsilon \rangle.
\end{aligned}$$

Analogs of Lemmas 9, 10 and 11 hold in this case, with the same proofs as those given for the case of XPath$_=(\downarrow)^-$:

Lemma 25. *Let $* \in \{=, \neq\}$, $\psi \in N_n$ and $\alpha, \alpha' \in P_n$. Let $\mathcal{T}, u$ be a pointed data tree, such that $\mathcal{T}, u \models \psi$ and $\mathcal{T}, u \models \langle \alpha * \alpha' \rangle$. Then $\langle \alpha * \alpha' \rangle$ is a conjunct of ψ.*

Lemma 26. *Let $\psi \in N_n$ and $\alpha \in P_n$. If $[\psi]\alpha$ is consistent then $\langle \alpha = \alpha \rangle$ is a conjunct of ψ. As an immediate consequence, if $\langle \downarrow[\psi]\alpha \rangle$ is consistent then $\langle \alpha = \alpha \rangle$ is a conjunct of ψ.*

Lemma 27. *For every pair of distinct elements $\varphi, \psi \in N_n$, $\varphi \wedge \psi$ is inconsistent.*

We omit the proof of the following theorem, since it is analogous to the one for XP$^-$ (Theorem 16):

Theorem 28 *(Normal form for XPath$_=(\downarrow)$). Let φ be a consistent node expression of XPath$_=(\downarrow)$ such that $\mathrm{dd}(\varphi) = n$. Then $\mathsf{XP} \vdash \varphi \equiv \bigvee_i \varphi_i$ for some $(\varphi_i)_{1 \le i \le k} \in N_n$. Let α be a consistent path expression of XPath$_=(\downarrow)$ such that $\mathrm{dd}(\alpha) = n$. Then $\mathsf{XP} \vdash \alpha \equiv \bigcup_i [\varphi_i]\alpha_i$ for some $(\alpha_i)_{1 \le i \le k} \in P_n$ and $(\varphi_i)_{1 \le i \le k} \in N_n$. Furthermore, if α is ϵ or starting with $\downarrow$ then $\mathsf{XP} \vdash \alpha \equiv \bigcup_i \alpha_i$ for some $(\alpha_i)_{1 \le i \le k} \in P_n$.*

The following two technical lemmas, whose proofs are deferred to Appendix A, will be needed for the construction of the canonical model:

Lemma 29. *Let $* \in \{=, \neq\}$, $\gamma \in P_n$, $\psi_i \in N_{n-i}$ for $i = 1, \ldots, i_0$, $\alpha, \beta \in P_{n-i_0}$ such that*

$$\langle \gamma * \downarrow[\psi_1] \ldots \downarrow[\psi_{i_0}]\alpha \rangle \wedge \neg\langle \gamma * \downarrow[\psi_1] \ldots \downarrow[\psi_{i_0}]\beta \rangle$$

is consistent and $\neg\langle \alpha \neq \alpha \rangle$ is a conjunct of ψ_{i_0}. Then $\neg\langle \alpha = \beta \rangle$ is a conjunct of ψ_{i_0}.

Lemma 30. *Let $\psi \in N_n$, $\alpha, \beta \in P_n$ such that $\langle \downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha \rangle \wedge \neg\langle \downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma \rangle$ is consistent and $\neg\langle \alpha \neq \alpha \rangle$ is a conjunct of ψ. Then $\neg\langle \alpha = \gamma \rangle$ is a conjunct of ψ.*

4.3. Completeness for node and path expressions

In this section we show that for node expressions φ and ψ of XPath$_=(\downarrow)$, the equivalence $\varphi \equiv \psi$ is derivable from the axiom schemes of Table 2 plus Table 3 if and only if φ is XPath$_=(\downarrow)$-semantically equivalent to ψ. We also show the corresponding result for path expressions of XPath$_=(\downarrow)$.

Theorem 31 *(Completeness of XPath$_=(\downarrow)$).*

1. *Let φ and ψ be node expressions of* XPath$_=(\downarrow)$. *Then* $\mathsf{XP} \vdash \varphi \equiv \psi$ *iff* $\models \varphi \equiv \psi$.
2. *Let α and β be path expressions of* XPath$_=(\downarrow)$. *Then* $\mathsf{XP} \vdash \alpha \equiv \beta$ *iff* $\models \alpha \equiv \beta$.

The proof of the above theorem is analogous to that of Theorem 18. The critical part of the argumentation is the analog of Lemma 17 for the more expressive logic XPath$_=(\downarrow)$:

Lemma 32. *Any node expression $\varphi \in N_n$ is satisfiable.*

The rest of this section, namely §4.3.1, is devoted to the proof of Lemma 32.

4.3.1. Canonical model

We construct, recursively in n and for every $\varphi \in N_n$, a data tree $\mathcal{T}^\varphi = (T^\varphi, \pi^\varphi)$ such that φ is satisfiable in $\mathcal{T}^\varphi$.

For the base case, if $\varphi \in N_0$ and $\varphi = a \wedge \langle \epsilon = \epsilon \rangle \wedge \neg\langle \epsilon \neq \epsilon \rangle$ with $a \in \mathbb{A}$, we define the data tree $\mathcal{T}^\varphi = (T^\varphi, \pi^\varphi)$ where T^φ is a tree which consists of the single node x with label a, and $\pi^\varphi = \{\{x\}\}$.

Now, let $\varphi \in N_{n+1}$. Since φ is a conjunction as in Definition 24, it is enough to guarantee that the following conditions hold (observe that we are using EqAx1 and NeqAx1 but we usually avoid these observations of symmetry):

(C1) If $a \in \mathbb{A}$ is a conjunct of φ, then the root r^φ of $\mathcal{T}^\varphi$ has label a.
(C2) If $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ is a conjunct of φ, then there is a child $r^{\mathbf{v}}$ of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied, and a node $x^{\mathbf{v}}$ with the same data value as r^φ such that $\mathcal{T}^\varphi, r^{\mathbf{v}}, x^{\mathbf{v}} \models \alpha$.
(C3) If $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ is a conjunct of φ, then there is a child $r^{\mathbf{v}}$ of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied, and a node $x^{\mathbf{v}}$ with different data value than r^φ such that $\mathcal{T}^\varphi, r^{\mathbf{v}}, x^{\mathbf{v}} \models \alpha$.
(C4) If $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ is a conjunct of φ, then there are two children $r_1^{\mathbf{u}}$, $r_2^{\mathbf{u}}$ of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, and there are nodes $x^{\mathbf{u}}$ and $y^{\mathbf{u}}$ with the same data value such that $\mathcal{T}^\varphi, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$ and $\mathcal{T}^\varphi, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$.
(C5) If $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ is a conjunct of φ, then there are two children $r_1^{\mathbf{u}}$, $r_2^{\mathbf{u}}$ of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, and there are nodes $x^{\mathbf{u}}$ and $y^{\mathbf{u}}$ with different data value such that $\mathcal{T}^\varphi, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$ and $\mathcal{T}^\varphi, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$.
(C6) If $\neg\langle \epsilon = \downarrow[\psi]\alpha \rangle$ is a conjunct of φ, then for each child z of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied, if x is a node such that $\mathcal{T}^\varphi, z, x \models \alpha$, then the data value of x is different than the one of r^φ.
(C7) If $\neg\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ is a conjunct of φ, then for each child z of the root r^φ of $\mathcal{T}^\varphi$ at which ψ is satisfied, if x is a node such that $\mathcal{T}^\varphi, z, x \models \alpha$, then the data value of x is the same as the one of r^φ.
(C8) If $\neg\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ is a conjunct of φ, then for each children z_1, z_2 of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, if w_1, w_2 are nodes such that $\mathcal{T}^\varphi, z_1, w_1 \models \alpha$ and $\mathcal{T}^\varphi, z_2, w_2 \models \beta$, then the data values of w_1 and w_2 are different.
(C9) If $\neg\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ is a conjunct of φ, then for each children z_1, z_2 of the root r^φ of $\mathcal{T}^\varphi$ at which ψ and ρ are satisfied respectively, if w_1, w_2 are nodes such that $\mathcal{T}^\varphi, z_1, w_1 \models \alpha$ and $\mathcal{T}^\varphi, z_2, w_2 \models \beta$, then w_1 and w_2 have the same data value.

As in §3.3.1, we first give an intuitive description of the construction of the model, and then proceed to formalize it:

4.3.2. Insight into the construction

The construction given in §3.3.1 has some similarities with the one we are about to present. As before, we will hang, from a common root, copies of trees given by inductive hypothesis to guarantee the satisfaction of some conjuncts of φ. Like in the previous case, we may need to introduce some changes on those trees in order to avoid spoiling the satisfaction of other conjuncts.

However, this construction is far more complex than the one for XPath$_=(\downarrow)^-$. In the previous case, when adding new witnesses with fresh data values, one only needed to be careful enough to avoid putting in the same class nodes that should be in different classes. Now, in addition to that (which is also harder to achieve, as witnessed by the differences between Lemmas 19 and 41 explained at the end of the latter), one also needs to guarantee conditions of the form $\neg\langle \mu \neq \delta \rangle$ with $\mu, \delta \in P_{n+1}$ which force the merging of classes of *every* witness of the kind of paths involved that could appear along the construction.

Unlike the case of XPath$_=(\downarrow)^-$, each pair of path expressions μ, δ in P_{n+1} will occur in two conjuncts of φ instead of one (we do not care about symmetric repetitions). Indeed, in the case of XPath$_=(\downarrow)^-$, for μ, δ in P_{n+1}^-, we either have

$\langle \mu = \delta \rangle$ or $\neg\langle \mu = \delta \rangle$ as a conjunct of a node expression in N^{-}_{n+1}. Now we have four choices because we also have either $\langle \mu \neq \delta \rangle$ or $\neg\langle \mu \neq \delta \rangle$, and hence *two* conjuncts containing μ and δ will occur in node expressions of N_{n+1}. We cannot treat as separate from each other those two conjuncts in which the same pair μ, δ appear, so we first split P_{n+1} into four subsets to deal with diamonds that compare against the constant empty path:

$$\mathbf{V}_{=,\neq} = \{(\psi, \alpha) \mid \psi \in N_n, \alpha \in P_n, \langle \epsilon = \downarrow[\psi]\alpha \rangle \text{ and } \langle \epsilon \neq \downarrow[\psi]\alpha \rangle \text{ are conjuncts of } \varphi\}$$

$$\mathbf{V}_{=,\neg\neq} = \{(\psi, \alpha) \mid \psi \in N_n, \alpha \in P_n, \langle \epsilon = \downarrow[\psi]\alpha \rangle \text{ and } \neg\langle \epsilon \neq \downarrow[\psi]\alpha \rangle \text{ are conjuncts of } \varphi\}$$

$$\mathbf{V}_{\neg=,\neq} = \{(\psi, \alpha) \mid \psi \in N_n, \alpha \in P_n, \neg\langle \epsilon = \downarrow[\psi]\alpha \rangle \text{ and } \langle \epsilon \neq \downarrow[\psi]\alpha \rangle \text{ are conjuncts of } \varphi\}$$

$$\mathbf{V}_{\neg=,\neg\neq} = \{(\psi, \alpha) \mid \psi \in N_n, \alpha \in P_n, \neg\langle \epsilon = \downarrow[\psi]\alpha \rangle \text{ and } \neg\langle \epsilon \neq \downarrow[\psi]\alpha \rangle \text{ are conjuncts of } \varphi\}$$

We make the following observations regarding the above definitions:

Observation 33. For $(\psi, \alpha) \in \mathbf{V}_{\neg=,\neg\neq}$, our axioms should tell us that either $\langle \alpha \rangle$ is not a conjunct of ψ or $\downarrow[\psi]\beta$ does not appear in any other positive conjunct of φ. If this is not the case, then φ would be clearly unsatisfiable and thus our axiomatic system would not be complete. This assertion is a consequence of the following lemma plus **Der12** of Fact 5. It is important to remark that the axioms required for the proof can be easily proven sound.

Lemma 34. *Let $\psi \in N_n$, $\alpha \in P_n$, $\gamma \in P_{n+1}$. If $\neg\langle \gamma = \downarrow[\psi]\alpha \rangle \wedge \neg\langle \gamma \neq \downarrow[\psi]\alpha \rangle \wedge \langle \gamma \rangle \wedge \langle \downarrow[\psi] \rangle$ is consistent, then $\neg\langle \alpha = \alpha \rangle$ is a conjunct of ψ.*

Proof. See Appendix A. □

Then, by Lemma 27 plus the fact that we will construct our model by hanging from the root the trees given by inductive hypothesis, we should not be worried about the satisfaction of either $\neg\langle \epsilon = \downarrow[\psi]\alpha \rangle$ nor $\neg\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ because we will never create a pair of nodes witnessing the path $\downarrow[\psi]\alpha$.

Observation 35. For $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$, our axioms should tell us that in a tree $\mathcal{T}^{\psi}$, any pair of nodes satisfying α ends in a node in the same equivalence class, since we want to put any such node in the class of the root r^{φ}. The following lemma has this property as an immediate consequence.

Lemma 36. *Let $* \in \{=, \neq\}$, $\psi \in N_n$, $\alpha, \beta \in P_n$, $\gamma \in P_{n+1}$. If $\langle \gamma = \downarrow[\psi]\alpha \rangle \wedge \neg\langle \gamma \neq \downarrow[\psi]\alpha \rangle \wedge \neg\langle \gamma * \downarrow[\psi]\beta \rangle$ is consistent, then $\neg\langle \alpha * \beta \rangle$ is a conjunct of ψ.*

Proof. See Appendix A. □

Observation 37. For $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$ and $(\psi, \beta) \in \mathbf{V}_{\neg=,\neq}$, Lemma 36 also tells us that in a tree $\mathcal{T}^{\psi}$, any pairs of nodes satisfying α and β end in points in different equivalence classes; which is also necessary to be able to satisfy φ.

Observation 38. For $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$ and $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$, in order to obtain a witness for $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$, our axioms should tell us that in a tree $\mathcal{T}^{\psi}$ we can find a pair of nodes satisfying α starting from the root, and such that its ending node is in a different class from that of the ending node of any pair of nodes satisfying β and beginning at the root of that tree. The following lemma combined with Observation 35 has this as an immediate consequence.

Lemma 39. *Let $* \in \{=, \neq\}$, $\psi \in N_n$, $\alpha, \beta \in P_n$, $\gamma \in P_{n+1}$. If $\langle \gamma = \downarrow[\psi]\alpha \rangle \wedge \neg\langle \gamma \neq \downarrow[\psi]\alpha \rangle \wedge \langle \gamma * \downarrow[\psi]\beta \rangle$ is consistent, then $\langle \alpha * \beta \rangle$ is a conjunct of ψ.*

Proof. See Appendix A. □

Observation 40. For $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$, in order to obtain a witness for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$, we need a tree in which ψ is satisfied and a pair of nodes (beginning at the root of that tree) satisfying α and ending in a node such that: it is in the class of the ending nodes of pairs of nodes satisfying β for $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$, but it is not in the class of any ending node of a pair of nodes satisfying γ for $(\psi, \gamma) \in \mathbf{V}_{\neg=,\neq}$. In case there exists $\beta \in P_n$ such that $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$, any tree at which ψ is satisfied will work by the previous observations and lemmas. But in case $(\psi, \beta) \notin \mathbf{V}_{=,\neg\neq}$ for all $\beta \in P_n$, we will have to make use of Lemma 41 (the analogous of Lemma 19 for this case).

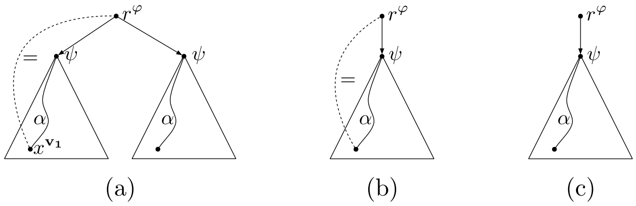

Fig. 9. (a) Witnesses for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ and $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ for $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$; (b) A witness for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ for $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$; (c) A witness for $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ for $(\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$.

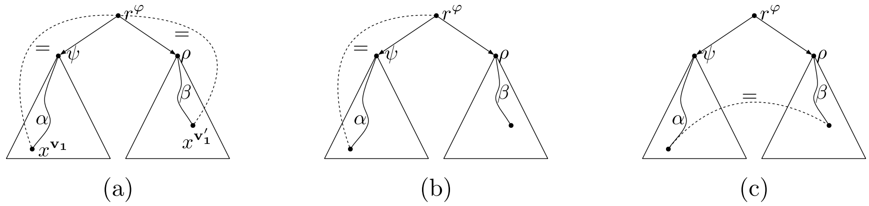

Fig. 10. For $(\psi, \alpha), (\rho, \beta) \in \mathbf{V}_{=,\neq}$, (a) Witnesses for $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$; (b) Witnesses for $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$; For $(\psi, \alpha) \in \mathbf{V}_{=,\neq}, (\rho, \beta) \in \mathbf{V}_{\neg=,\neq}$ or $(\psi, \alpha), (\rho, \beta) \in \mathbf{V}_{\neg=,\neq}$ (c) Witnesses for $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$.

Processing data-aware diamonds of the form $(\neg)\langle \epsilon * \downarrow[\psi]\alpha \rangle$ Having all these observations at hand, we begin by analyzing the following (non-disjoint) cases to construct our tree $\mathcal{T}^{\varphi}$:

(Case 1) For $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$, we add two witnesses. One for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ from which we merge the class of the ending point $x^{\mathbf{v_1}}$ of a pair of nodes satisfying α as in Observation 40 with the class of r^{φ}. We add another witness for $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ (remember Observation 38). See Fig. 9(a).

(Case 2) For $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$, we add one witness for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ (see Fig. 9(b)) and, at the end of the construction, we will merge the class of any node x such that $r^{\varphi}, x \models \downarrow[\psi]\alpha$ with the class of r^{φ} (remember Observation 35).

(Case 3) For $(\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$, we add one witness for $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ (See Fig. 9(c)). Note that $\langle \epsilon \neq \downarrow[\psi]\alpha \rangle \wedge \neg\langle \epsilon = \downarrow[\psi]\alpha \rangle$ will be satisfied by Observations 37 and 40.

Processing data-aware diamonds of the form $(\neg)\langle \downarrow[\psi]\alpha * \downarrow[\rho]\beta \rangle$ For conjuncts of φ the form $(\neg)\langle \downarrow[\psi]\alpha * \downarrow[\rho]\beta \rangle$ that do not involve comparison with the constant path ϵ, we have that, depending on which of the sets $\mathbf{V}_{=,\neq}, \mathbf{V}_{=,\neg\neq}, \mathbf{V}_{\neg=,\neq}, \mathbf{V}_{\neg=,\neg\neq}$ do (ψ, α) and (ρ, β) belong to, many of the four possible combinations ($\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$, $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\neg\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$, etc.) are not possible as conjuncts for a consistent φ. More specifically:

(Case 4) If we have $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\neg\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ as conjuncts of φ, then all the following cases should be impossible since, in that case, φ would be clearly unsatisfiable and thus it should be inconsistent: (ψ, α) or (ρ, β) in $\mathbf{V}_{=,\neq}$, (ψ, α) or (ρ, β) in $\mathbf{V}_{\neg=,\neg\neq}$, one in $\mathbf{V}_{=,\neg\neq}$ and the other in $\mathbf{V}_{\neg=,\neq}$. Besides, if both belong to $\mathbf{V}_{=,\neg\neq}$, since we merge the class of any node x such that $r^{\varphi}, x \models \downarrow[\psi]\alpha$ or $r^{\varphi}, x \models \downarrow[\rho]\beta$, those conjuncts $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\neg\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ will be satisfied. If both belong to $\mathbf{V}_{\neg=,\neq}$, we need to force these conjuncts by merging the class of any node x such that $r^{\varphi}, x \models \downarrow[\psi]\alpha$ or $r^{\varphi}, x \models \downarrow[\rho]\beta$ (note that we have such nodes by (Case 3)). It is important to notice that this process does not add nodes to the class of the root since such nodes x are never in the same equivalence class than any $x^{\mathbf{v_1}}$ from (Case 1) nor in the same equivalence class of a witness of $\langle \downarrow[\mu]\delta \rangle$ for $(\mu, \delta) \in \mathbf{V}_{=,\neg\neq}$.

(Case 5) If we have $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ as conjuncts of φ, then it cannot be the case that (ψ, α) or (ρ, β) belong to $\mathbf{V}_{\neg=,\neg\neq}$. It is also not possible that both belong to $\mathbf{V}_{=,\neg\neq}$ or one to $\mathbf{V}_{=,\neg\neq}$ and the other to $\mathbf{V}_{\neg=,\neq}$. Besides, if $(\psi, \alpha), (\rho, \beta)$ belong to $\mathbf{V}_{=,\neq}$, then $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ and $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ are already satisfied: $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ by the witnesses for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ and $\langle \epsilon = \downarrow[\rho]\beta \rangle$ (see Fig. 10(a)), $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ by the witnesses for $\langle \epsilon = \downarrow[\psi]\alpha \rangle$ and $\langle \epsilon \neq \downarrow[\rho]\beta \rangle$ (see Fig. 10(b) and remember Observation 38). In case one belongs to $\mathbf{V}_{=,\neq}$ and the other to $\mathbf{V}_{=,\neg\neq}$, the argument is similar. If one belongs to $\mathbf{V}_{=,\neq}$ and the other to $\mathbf{V}_{\neg=,\neq}$ or both to $\mathbf{V}_{\neg=,\neq}$, $\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ will be satisfied using arguments similar to the previous ones; but we need to add witnesses to guarantee the satisfaction of $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ (see Fig. 10(c)). In some cases, the merging performed in (Case 4), would have already merged the classes of a witness for $\langle \downarrow[\psi]\alpha \rangle$ and a witness for $\langle \downarrow[\rho]\beta \rangle$, in the remaining cases, we will need to force that merging carefully enough not to spoil conditions (C6) and (C8) (we will use Lemma 41 to achieve that).

Finally, these last two cases are satisfied automatically:

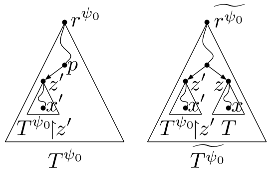

Fig. 11. $T = T^{\psi_0}\restriction z$ is a new subtree with a special node x such that its class of data values is disjoint to the rest of $\widetilde{\mathcal{T}^{\psi_0}}$ and $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$.

(Case 6) If we have $\neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ and $\langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$ as conjuncts of φ, then all the following cases should be impossible: (ψ, α) or (ρ, β) in $\mathbf{V}_{\neg=,\neg\neq}$, both in $\mathbf{V}_{=,\neq}$ or $\mathbf{V}_{=,\neg\neq}$. Besides, if one belongs to $\mathbf{V}_{=,\neg\neq}$ and the other to $\mathbf{V}_{\neg=,\neq}$ or if one belongs to $\mathbf{V}_{=,\neq}$ and the other to $\mathbf{V}_{\neg=,\neq}$ or if they both belong to $\mathbf{V}_{\neg=,\neq}$, $\neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ and $\langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$ will be satisfied automatically —the last two cases may not be as intuitive as others but are also true and we will give a detailed proof in time.

(Case 7) If we have $\neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ and $\neg\langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$ as conjuncts of φ, then the only case that should not lead to an inconsistency is when at least one of (ψ, α) and (ρ, β) is in $\mathbf{V}_{\neg=,\neg\neq}$ and, in this case, $\neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ and $\neg\langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$ will be satisfied automatically.

Formalization

In order to formalize the construction described above, we introduce the following lemma, which is key to guarantee conditions (C2) and (C4) without spoiling conditions (C6) and (C8):

Lemma 41. *Let $\psi_0 \in N_n$, $\alpha, \beta_1, \ldots, \beta_m \in P_n$. Suppose that there exists a tree $\mathcal{T}^{\psi_0} = (T^{\psi_0}, \pi^{\psi_0})$ with root r^{ψ_0} such that $\mathcal{T}^{\psi_0}, r^{\psi_0} \models \psi_0$ and for all $i = 1, \ldots, m$ there exists $\gamma_i \in P_{n+1}$ such that $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ is consistent. Then there exists a tree $\widetilde{\mathcal{T}^{\psi_0}} = (\widetilde{T^{\psi_0}}, \widetilde{\pi^{\psi_0}})$ with root $\widetilde{r^{\psi_0}}$ and a node x such that:*

- $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}} \models \psi_0$,
- $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$, *and*
- $[x]_{\widetilde{\pi^{\psi_0}}} \neq [y]_{\widetilde{\pi^{\psi_0}}}$ *for all y such that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, y \models \beta_i$ for some $i = 1, \ldots, m$.*

Proof. Suppose $\alpha = \downarrow[\psi_1]\ldots\downarrow[\psi_{j_0}]\epsilon$ where $\psi_k \in N_{n-k}$ for all $k = 1, \ldots, j_0$ and let

$$k_0 = \min_{0 \le k \le j_0} \left\{k \mid \neg\langle\downarrow[\psi_{k+1}]\ldots\downarrow[\psi_{j_0}]\epsilon \neq \downarrow[\psi_{k+1}]\ldots\downarrow[\psi_{j_0}]\epsilon\rangle \text{ is a conjunct of } \psi_k\right\}.$$

In case $k_0 = 0$ (i.e. $\neg\langle\alpha \neq \alpha\rangle$), by Lemma 29, $\neg\langle\alpha = \beta_i\rangle$ is a conjunct of ψ_0 for all $i = 1, \ldots, m$. Then $\widetilde{\mathcal{T}^{\psi_0}} = (T^{\psi_0}, \pi^{\psi_0})$ satisfies the desired properties. The intuitive idea behind this application of Lemma 29 is that in case every ending point of a pair of nodes satisfying α is in the same equivalence class, then there cannot be pairs of nodes satisfying α and β_i ending in points with the same data value, because in that case $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ would be unsatisfiable and thus inconsistent, which is a contradiction with our hypothesis.

In case $k_0 \neq 0$, by consistency, there are $z', x' \in T^{\psi_0}$ such that $\mathcal{T}^{\psi_0}, r^{\psi_0}, z' \models \downarrow[\psi_1]\ldots\downarrow[\psi_{k_0}]$ and $\mathcal{T}^{\psi_0}, z', x' \models \downarrow[\psi_{k_0+1}]\ldots\downarrow[\psi_{j_0}]$. Before proceeding to complete the proof of this case, we give an intuitive idea. We prove that we cannot have a witness for β_i with the same data value than x' in the subtree $T^{\psi_0}\restriction z'$. Intuitively this is because, in that case, α and β_i would have a common prefix. Let us say that

$$\beta = \downarrow[\psi_1]\ldots\downarrow[\psi_{k_0}]\downarrow[\rho_{k_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon \quad \text{and} \quad \langle\downarrow[\psi_{k_0+1}]\ldots\downarrow[\psi_{j_0}]\epsilon = \downarrow[\rho_{k_0+1}]\ldots\downarrow[\rho_{l_0}]\epsilon\rangle$$

is a conjunct of ψ_{k_0}. Then, since $\neg\langle\downarrow[\psi_{k_0+1}]\ldots\downarrow[\psi_{j_0}]\epsilon \neq \downarrow[\psi_{k_0+1}]\ldots\downarrow[\psi_{j_0}]\epsilon\rangle$ is also a conjunct of ψ_{k_0}, $\langle\gamma_i = \downarrow[\psi_0]\alpha\rangle \wedge \neg\langle\gamma_i = \downarrow[\psi_0]\beta_i\rangle$ would be unsatisfiable (and thus inconsistent) for any choice of γ_i, which is a contradiction. But our hypotheses do not guarantee that we would not have a witness for β_i in the class of x' outside $T^{\psi_0}\restriction z'$, and therefore we need to change the tree in order to achieve the desired properties. We replicate the subtree $T^{\psi_0}\restriction z'$ but using a fresh data value (different from any other data value already present in $\mathcal{T}^{\psi_0}$) for the class of the companion of x' that we call x; see Fig. 11. It is clear that in this way, the second and the third conditions will be satisfied by x. The first condition will also remain true because, intuitively, the positive conjuncts will remain valid since we are not suppressing any nodes, and the negative ones that compare by equality will not be affected because every new node has either the same data value than its companion or a fresh data value. The argument for negative conjuncts that compare by inequality is based on the way in which we have chosen k_0 (see a detailed proof below).

Now we formalize the previous intuition. Let p be the parent of z' ($k_0 > 0$). As we did in the proof of Lemma 19, we define $\widetilde{T^{\psi_0}}$ by adding a new child z of p and a data tree $\mathcal{T} = (T, \pi)$ hanging from z. This tree T is a copy of $T^{\psi_0}\restriction z$, and we call x to the companion of x'. $\widetilde{\pi^{\psi_0}}$ is defined as π^{ψ_0} with the exception that the class of x is new (the classes of the other nodes of T are merged with the classes of their companions) (see Fig. 11).

We first prove by induction that z_j, the j-th ancestor of z (namely $z_j \xrightarrow{j} z$, and we let $z_0 := z$), satisfies $\widetilde{\mathcal{T}^{\psi_0}}, z_j \models \psi_{k_0-j}$. This will prove both that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}} \models \psi_0$ and that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, x \models \alpha$. By Proposition 3, it is straightforward from the construction that ψ_{k_0} is satisfied at z (the companion of z') which proves the base case. For the inductive case, assume the result holds for $z_0, \ldots, z_j$. We want to see that it holds for z_{j+1}. To do this, we check that every conjunct of ψ_{k_0-j-1} is satisfied at z_{j+1}:

- If the conjunct is a label, it is clear that z_{j+1} has that label in $\widetilde{\mathcal{T}^{\psi_0}}$, as it has not been changed by the construction.
- If the conjunct is of the form $\langle \mu_1 = \mu_2 \rangle$ or $\langle \mu_1 \neq \mu_2 \rangle$, then it must still hold in $\widetilde{\mathcal{T}^{\psi_0}}$ by inductive hypothesis and the fact that our construction did not remove nodes.
- If the conjunct is of the form $\neg\langle \mu_1 = \mu_2 \rangle$, we observe that, by inductive hypothesis plus the way in which we have constructed $\widetilde{\mathcal{T}^{\psi_0}}$, we have that: If $\widetilde{\mathcal{T}^{\psi_0}}, z_{j+1} \models \langle \mu_1 = \mu_2 \rangle$ then $\mathcal{T}^{\psi_0}, z_{j+1} \models \langle \mu_1 = \mu_2 \rangle$ (for a complete proof of this assertion, one can use arguments similar to the ones used in Lemma 19) which is a contradiction with the fact that $\mathcal{T}^{\psi_0}, z_{j+1} \models \psi_{k_0-(j+1)}$. Then $\widetilde{\mathcal{T}^{\psi_0}}, z_{j+1} \models \neg\langle \mu_1 = \mu_2 \rangle$.
- If the conjunct is of the form $\neg\langle \mu_1 \neq \mu_2 \rangle$, by inductive hypothesis plus the way in which we have constructed $\widetilde{\mathcal{T}^{\psi_0}}$, $\langle \mu_1 \neq \mu_2 \rangle$ can only be true in z_{j+1} if there are witnesses y_1, y_2 in distinct equivalence classes such that $\widetilde{\mathcal{T}^{\psi_0}}, z_{j+1}, y_1 \models \mu_1$, $\widetilde{\mathcal{T}^{\psi_0}}, z_{j+1}, y_2 \models \mu_2$ and at least one of them is in the new subtree T. In that case, without loss of generality, we have that $\mu_1 = {\downarrow}[\psi_{k_0-j}] \ldots {\downarrow}[\psi_{k_0}]\hat{\mu}_1$. Then, by definition of k_0, $\langle {\downarrow}[\psi_{k_0-j}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \neq {\downarrow}[\psi_{k_0-j}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \rangle$ is a conjunct of ψ_{k_0-j-1}. Therefore, by consistency and NeqAx7, $\langle {\downarrow}[\psi_{k_0-j}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \neq \mu_2 \rangle$ or $\neg\langle \mu_2 = \mu_2 \rangle$ is also a conjunct of ψ_{k_0-j-1}. If the latter occurs, we have a contradiction by the previous item. If $\langle {\downarrow}[\psi_{k_0-j}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \neq \mu_2 \rangle$ is a conjunct of ψ_{k_0-j-1}, by Lemma 29 $\neg\langle {\downarrow}[\psi_{k_0+1}] \ldots {\downarrow}[\psi_{j_0}]\epsilon = \hat{\mu}_1 \rangle$ is a conjunct of ψ_{k_0}. Then, by construction, the class of y_1 in $\widetilde{\mathcal{T}^{\psi_0}}$ is equal to the class of its companion and so we can assume that $y_1 \notin T$. Analogously we can assume that $y_2 \notin T$ but, as we have already said, by inductive hypothesis plus the way in which we have constructed $\widetilde{\mathcal{T}^{\psi_0}}$, $\langle \mu_1 \neq \mu_2 \rangle$ cannot be satisfied at z_{j+1} by witnesses y_1, y_2 if neither of them is in the new subtree T.

To conclude the proof, we only need to check that $[x]_{\widetilde{\pi^{\psi_0}}} \neq [y]_{\widetilde{\pi^{\psi_0}}}$ for all y such that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, y \models \beta_i$ for some $i = 1, \ldots, m$. Suppose that $\beta_i = {\downarrow}[\rho_1] \ldots {\downarrow}[\rho_{l_0}]\epsilon$. If $l_0 < k_0$ or $\rho_l \neq \psi_l$ for some $l = 1, \ldots, k_0$, then the result follows immediately from construction. If not, by hypothesis, there exists $\gamma_i \in P_{n+1}$ such that $\langle \gamma_i = {\downarrow}[\psi_0] \ldots {\downarrow}[\psi_{j_0}]\epsilon \rangle \wedge \neg\langle \gamma_i = {\downarrow}[\psi_0] \ldots {\downarrow}[\psi_{k_0}]{\downarrow}[\rho_{k_0+1}] \ldots {\downarrow}[\rho_{l_0}]\epsilon \rangle$ is consistent and $\neg\langle {\downarrow}[\psi_{k_0+1}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \neq {\downarrow}[\psi_{k_0+1}] \ldots {\downarrow}[\psi_{j_0}]\epsilon \rangle$ is a conjunct of ψ_{k_0}. Then, by Lemma 29, $\neg\langle {\downarrow}[\psi_{k_0+1}] \ldots {\downarrow}[\psi_{j_0}]\epsilon = {\downarrow}[\rho_{k_0+1}] \ldots {\downarrow}[\rho_{l_0}]\ \epsilon \rangle$ is a conjunct of ψ_{k_0}. This together with the fact that the class of x is disjoint with the part of $\widetilde{T^{\psi_0}}$ outside of T, shows that $[x]_{\widetilde{\pi^{\psi_0}}} \neq [y]_{\widetilde{\pi^{\psi_0}}}$ if y is such that $\widetilde{\mathcal{T}^{\psi_0}}, \widetilde{r^{\psi_0}}, y \models \beta_i$, which concludes the proof. $\square$

It might be useful for the reader to note the differences between Lemmas 41 and 19, since this is one of the reasons why the completeness result for $\mathrm{XPath}_=(\downarrow)$ is more complicated than for $\mathrm{XPath}_=(\downarrow)^-$. The main differences between those two lemmas are:

- In Lemma 41, if we would replicate the subtree hanging from a witness of $\langle \alpha \rangle$ then, due to the fact that we are working with the complete fragment (with inequality tests also), we would not be able to prove that each ancestor of that node satisfies the desired formulas. So we are forced to find that minimum k_0 that tells us which subtree we should replicate.
- In Lemma 19, we can use new data for every new node since, again, we are not working with inequality tests. But when it comes to the complete fragment, we need to be more careful in the way we define the partition in $\widetilde{\mathcal{T}^{\psi_0}}$ changing only the class of the new witness of $\langle \alpha \rangle$.

Now that we have this key lemma, we proceed to the formal construction of $\mathcal{T}^{\varphi}$. We define some special sets of quadruples $(\psi, \alpha, \rho, \beta)$ with $\psi, \rho \in N_n$, $\alpha, \beta \in P_n$:

- **U** is the set of quadruples $(\psi, \alpha, \rho, \beta)$ such that one of the following holds:
  - $(\psi, \alpha), (\rho, \beta) \in \mathbf{V}_{\neg=,\neq}$, and $\langle {\downarrow}[\psi]\alpha = {\downarrow}[\rho]\beta \rangle$, $\langle {\downarrow}[\psi]\alpha \neq {\downarrow}[\rho]\beta \rangle$ are conjuncts of φ, or
  - $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$, $(\rho, \beta) \in \mathbf{V}_{\neg=,\neq}$, and $\langle {\downarrow}[\psi]\alpha = {\downarrow}[\rho]\beta \rangle$, $\langle {\downarrow}[\psi]\alpha \neq {\downarrow}[\rho]\beta \rangle$ are conjuncts of φ.

  Cf. (Case 5).
- **Z** is the set of all quadruples $(\psi, \alpha, \rho, \beta)$ such that $(\psi, \alpha), (\rho, \beta) \in \mathbf{V}_{\neg=,\neq}$, and $\langle {\downarrow}[\psi]\alpha = {\downarrow}[\rho]\beta \rangle$, $\neg\langle {\downarrow}[\psi]\alpha \neq {\downarrow}[\rho]\beta \rangle$ are conjuncts of φ.
  Cf. (Case 4).

The following lemma states that the relation between the elements of $\mathbf{V}_{\neg=,\neq}$ defined by the set **Z** is transitive, a fact which will be needed to prove that φ is indeed satisfied in the constructed tree:

Lemma 42. *If $(\psi, \alpha, \rho, \beta), (\rho, \beta, \theta, \gamma) \in \mathbf{Z}$, then $(\psi, \alpha, \theta, \gamma) \in \mathbf{Z}$.*

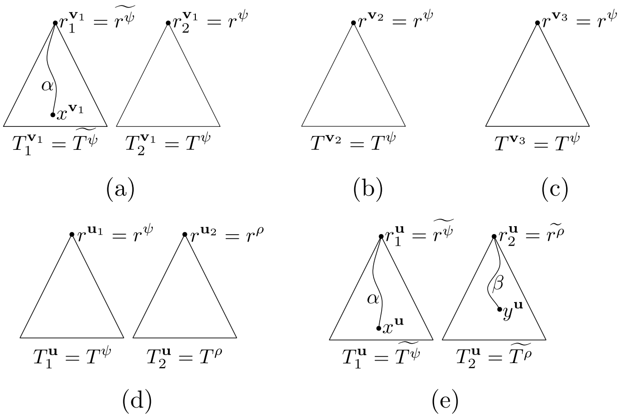

Fig. 12. Witnesses for (a) $\mathbf{v_1} = (\psi, \alpha) \in \mathbf{V}_{=,\neq}$; (b) $\mathbf{v_2} = (\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$; (c) $\mathbf{v_3} = (\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$; (d) $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U_1}$; (e) $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U_2}$.

Proof. By NeqAx7 and the consistency of φ, $\neg\langle\downarrow[\psi]\alpha \neq \downarrow[\theta]\gamma\rangle$ is a conjunct of φ. Then, by consistency of φ plus NeqAx6, $\langle\downarrow[\psi]\alpha = \downarrow[\theta]\gamma\rangle$ is also a conjunct of φ which concludes the proof. $\square$

Now that we have these lemmas, we proceed to construct $\mathcal{T}^\varphi$ as follows:

Rule 1. Witnesses for $\mathbf{v}_1 = (\psi, \alpha) \in \mathbf{V}_{=,\neq}$ (cf. (Case 1)) We define data trees $\mathcal{T}_1^{\mathbf{v_1}} = (T_1^{\mathbf{v_1}}, \pi_1^{\mathbf{v_1}})$ and $\mathcal{T}_2^{\mathbf{v_1}} = (T_2^{\mathbf{v_1}}, \pi_2^{\mathbf{v_1}})$ with roots $r_1^{\mathbf{v_1}}$ and $r_2^{\mathbf{v_1}}$ respectively. In order to choose appropriate witnesses for $\langle\epsilon = \downarrow[\psi]\alpha\rangle$ and $\langle\epsilon \neq \downarrow[\psi]\alpha\rangle$, we need the following lemma:

Lemma 43. *Let* $\mathbf{v_1} = (\psi, \alpha) \in \mathbf{V}_{=,\neq}$. *Then there exist* $\widetilde{\mathcal{T}^\psi} = (\widetilde{T^\psi}, \widetilde{\pi^\psi})$ *with root* $\widetilde{r^\psi}$ *and a node x such that:*

- $\widetilde{\mathcal{T}^\psi}, \widetilde{r^\psi} \models \psi$,
- $\widetilde{\mathcal{T}^\psi}, \widetilde{r^\psi}, x \models \alpha$,
- $[x]_{\widetilde{\pi^\psi}} = [y]_{\widetilde{\pi^\psi}}$ *for all y such that there is* $\beta \in P_n$ *with* $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$ *and* $\widetilde{T^\psi}, \widetilde{r^\psi}, y \models \beta$,
- $[x]_{\widetilde{\pi^\psi}} \neq [z]_{\widetilde{\pi^\psi}}$ *for all z such that there is* $\gamma \in P_n$ *with* $(\psi, \gamma) \in \mathbf{V}_{\neg=,\neq}$ *and* $\widetilde{T^\psi}, \widetilde{r^\psi}, z \models \gamma$.

Proof. We first analyze the case that there exists $\beta \in P_n$ such that $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$. Then, by Lemmas 36 and 39, the result is immediate from the fact that we are assuming there is a tree $\mathcal{T}^\psi$ satisfying ψ at its root. The idea is that by inductive hypothesis, there exists $\mathcal{T}^\psi = (T^\psi, \pi^\psi)$ satisfying ψ at is root. Then, Lemma 36 guarantees that every witness of some β as described before belongs to the same class in π^ψ and that every witness of some γ as described before does not belong to this class. Finally, Lemma 39 shows the existence of the desired node x.

To conclude the proof, suppose that $(\psi, \beta) \notin \mathbf{V}_{=,\neg\neq}$ for all $\beta \in P_n$. Then the result follows from Lemma 41. $\square$

Using Lemma 43, define $T_1^{\mathbf{v_1}}$ as $\widetilde{T^\psi}$, $\pi_1^{\mathbf{v_1}}$ as $\widetilde{\pi^\psi}$, $r_1^{\mathbf{v_1}}$ as $\widetilde{r^\psi}$ and $x^{\mathbf{v_1}} = x \in T_1^{\mathbf{v_1}}$. Also, by inductive hypothesis, there exists a tree $\mathcal{T}^\psi = (T^\psi, \pi^\psi)$ with root r^ψ such that $\mathcal{T}^\psi, r^\psi \models \psi$. Define $T_2^{\mathbf{v_1}}$ as T^ψ, $\pi_2^{\mathbf{v_1}}$ as π^ψ and $r_2^{\mathbf{v_1}}$ as r^ψ. Without loss of generality, we assume that $T_1^{\mathbf{v_1}}$ and $T_2^{\mathbf{v_1}}$ are disjoint. In other words, the rooted data tree $(T_1^{\mathbf{v_1}}, \pi_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}})$ is just a copy of $(\widetilde{T^\psi}, \widetilde{\pi^\psi}, \widetilde{r^\psi})$ with a special node named $x^{\mathbf{v_1}}$ and $(T_2^{\mathbf{v_1}}, \pi_2^{\mathbf{v_1}}, r_2^{\mathbf{v_1}})$ is just a copy of (T^ψ, π^ψ) disjoint with $(T_1^{\mathbf{v_1}}, \pi_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}})$. See Fig. 12(a).

Rule 2. Witnesses for $\mathbf{v}_2 = (\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$ (c.f (Case 2)) We define a data tree $\mathcal{T}^{\mathbf{v_2}} = (T^{\mathbf{v_2}}, \pi^{\mathbf{v_2}})$ with root $r^{\mathbf{v_2}}$. By inductive hypothesis, there exists $\mathcal{T}^\psi = (T^\psi, \pi^\psi)$, with root r^ψ such that $\mathcal{T}^\psi, r^\psi \models \psi$. Define $T^{\mathbf{v_2}}$ as T^ψ, $\pi^{\mathbf{v_2}}$ as π^ψ, and $r^{\mathbf{v_2}}$ as r^ψ. In other words, the rooted data tree $(T^{\mathbf{v_2}}, \pi^{\mathbf{v_2}}, r^{\mathbf{v_2}})$ is just a copy of $(T^\psi, \pi^\psi, r^\psi)$. See Fig. 12(b).

Rule 3. Witnesses for $\mathbf{v}_3 = (\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$ (cf. (Case 3)) We define a data tree $\mathcal{T}^{\mathbf{v_3}} = (T^{\mathbf{v_3}}, \pi^{\mathbf{v_3}})$ with root $r^{\mathbf{v_3}}$. By inductive hypothesis, there exists $\mathcal{T}^\psi = (T^\psi, \pi^\psi)$, with root r^ψ such that $\mathcal{T}^\psi, r^\psi \models \psi$. Define $T^{\mathbf{v_3}}$ as T^ψ, $\pi^{\mathbf{v_3}}$ as π^ψ, and $r^{\mathbf{v_3}}$ as r^ψ. In other words, the rooted data tree $(T^{\mathbf{v_3}}, \pi^{\mathbf{v_3}}, r^{\mathbf{v_3}})$ is just a copy of $(T^\psi, \pi^\psi, r^\psi)$. See Fig. 12(c).

Rule 4. Witnesses for $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U}$ (cf. (Case 5)) We define data trees $\mathcal{T}_1^{\mathbf{u}} = (T_1^{\mathbf{u}}, \pi_1^{\mathbf{u}})$ and $\mathcal{T}_2^{\mathbf{u}} = (T_2^{\mathbf{u}}, \pi_2^{\mathbf{u}})$ with roots $r_1^{\mathbf{u}}$, $r_2^{\mathbf{u}}$ respectively.

By inductive hypothesis, there exist trees $\mathcal{T}^\psi = (T^\psi, \pi^\psi)$ (with root r^ψ) and $\mathcal{T}^\rho = (T^\rho, \pi^\rho)$ (with root r^ρ) such that $\mathcal{T}^\psi, r^\psi \models \psi$ and $\mathcal{T}^\rho, r^\rho \models \rho$.

Now, in order to consider the information given by $\mathbf{U}$ and its interaction with $\mathbf{Z}$, we split $\mathbf{U}$ into two different subsets:

- $\mathbf{U_1}$ is the set of $(\psi, \alpha, \rho, \beta) \in \mathbf{U}$ for which there are $\gamma, \delta \in P_n$ such that:
  - $(\psi, \gamma, \rho, \delta) \in \mathbf{Z}$,
  - $\langle \gamma = \alpha \rangle$ is a conjunct of ψ,
  - $\langle \delta = \beta \rangle$ is a conjunct of ρ.
- $\mathbf{U_2} = \mathbf{U} \setminus \mathbf{U_1}$.

For $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U_1}$, define $T_1^{\mathbf{u}}$ as T^{ψ}, $\pi_1^{\mathbf{u}}$ as π^{ψ}, $r_1^{\mathbf{u}}$ as r^{ψ} and define $T_2^{\mathbf{u}}$ as T^{ρ}, $\pi_2^{\mathbf{u}}$ as π^{ρ}, $r_2^{\mathbf{u}}$ as r^{ρ}. Without loss of generality, we assume that $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ are disjoint.

In other words, the rooted data tree $(T_1^{\mathbf{u}}, \pi_1^{\mathbf{u}}, r_1^{\mathbf{u}})$ is just a copy of $(T^{\psi}, \pi^{\psi}, r^{\psi})$ and the pointed data tree $(T_2^{\mathbf{u}}, \pi_2^{\mathbf{u}}, r_2^{\mathbf{u}})$ is a copy of $(T^{\rho}, \pi^{\rho}, r^{\rho})$. See Fig. 12(d). Note that these are the cases in which the satisfaction of $\langle \downarrow[\psi]\alpha = \downarrow[\rho]\beta \rangle$ will be guaranteed by the merging described in (Case 4).

For $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U_2}$, in Lemma 41 consider

$$
\begin{aligned}
\psi_0 &:= \psi \\
\mathcal{T}^{\psi_0} &:= \mathcal{T}^{\psi} \\
\alpha &:= \alpha \\
\{\beta_1, \dots, \beta_m\} &:= \{\gamma \in P_n \mid \neg\langle \downarrow[\rho]\beta = \downarrow[\psi]\gamma \rangle \text{ is a conjunct of } \varphi\} \\
\gamma_i &:= \downarrow[\rho]\beta \text{ for all } i = 1, \dots, m
\end{aligned}
$$

Then there exist $\widetilde{\mathcal{T}^{\psi}} = (\widetilde{T^{\psi}}, \widetilde{\pi^{\psi}})$ with root $\widetilde{r^{\psi}}$ and a node x such that:

- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}} \models \psi$,
- $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, x \models \alpha$,
- $[x]_{\widetilde{\pi^{\psi}}} \neq [y]_{\widetilde{\pi^{\psi}}}$ for all y such that there is $\gamma \in P_n$ with $\widetilde{\mathcal{T}^{\psi}}, \widetilde{r^{\psi}}, y \models \gamma$ and $\neg\langle \downarrow[\rho]\beta = \downarrow[\psi]\gamma \rangle$ is a conjunct of φ.

Define $T_1^{\mathbf{u}}$ as $\widetilde{T^{\psi}}$, $\pi_1^{\mathbf{u}}$ as $\widetilde{\pi^{\psi}}$, $r_1^{\mathbf{u}}$ as $\widetilde{r^{\psi}}$ and $x^{\mathbf{u}} = x \in T_1^{\mathbf{u}}$. Now let

$$
\{\mu_1, \dots, \mu_r\} = \left\{ \mu \in P_n \mid \text{ there exists } \; y \in T_1^{\mathbf{u}} \text{ such that } \mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, y \models \mu \text{ and } [y]_{\pi_1^{\mathbf{u}}} = [x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}} \right\}.
$$

Then it follows that $\langle \downarrow[\rho]\beta = \downarrow[\psi]\mu_j \rangle$ is a conjunct of φ for all $j = 1, \dots, r$.

In Lemma 41, consider

$$
\begin{aligned}
\psi_0 &:= \rho \\
\mathcal{T}^{\psi_0} &:= \mathcal{T}^{\rho} \\
\alpha &:= \beta \\
\{\beta_1, \dots, \beta_m\} &:= \{\delta \in P_n \mid \exists j = 1, \dots, r \text{ with } \neg\langle \downarrow[\rho]\delta = \downarrow[\psi]\mu_j \rangle \text{ is a conjunct of } \varphi\} \\
\gamma_i &:= \downarrow[\psi]\mu_j \text{ for } j = 1, \dots r \text{ such that } \langle \downarrow[\rho]\beta_i = \downarrow[\psi]\mu_j \rangle \text{ is a conjunct of } \varphi
\end{aligned}
$$

Then there exist a tree $\widetilde{\mathcal{T}^{\rho}} = (\widetilde{T^{\rho}}, \widetilde{\pi^{\rho}})$ with root $\widetilde{r^{\rho}}$ and a node y such that

- $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}} \models \rho$,
- $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}}, y \models \beta$,
- $[y]_{\widetilde{\pi^{\rho}}} \neq [z]_{\widetilde{\pi^{\rho}}}$ for all z such that there is $\delta \in P_n$ and $j = 1, \dots, r$ with $\widetilde{\mathcal{T}^{\rho}}, \widetilde{r^{\rho}}, z \models \delta$ and $\neg\langle \downarrow[\rho]\delta = \downarrow[\psi]\mu_j \rangle$ is a conjunct of φ.

Define $T_2^{\mathbf{u}}$ as $\widetilde{T^{\rho}}$, $\pi_2^{\mathbf{u}}$ as $\widetilde{\pi^{\rho}}$, $r_2^{\mathbf{u}}$ as $\widetilde{r^{\rho}}$ and $y^{\mathbf{u}} = y$. Without loss of generality, we assume that $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ are disjoint.

In other words, the rooted data tree $(T_1^{\mathbf{u}}, \pi^{\mathbf{u}}\restriction_{T_1^{\mathbf{u}}}, r_1^{\mathbf{u}})$ is just a copy of $(\widetilde{T^{\psi}}, \widetilde{\pi^{\psi}}, \widetilde{r^{\psi}})$, with a special node named $x^{\mathbf{u}}$ which satisfies $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x^{\mathbf{u}} \models \alpha$. Analogously, the pointed data tree $(T_2^{\mathbf{u}}, \pi^{\mathbf{u}}\restriction_{T_2^{\mathbf{u}}}, r_2^{\mathbf{u}})$ is a copy of $(\widetilde{T^{\rho}}, \widetilde{\pi^{\rho}}, \widetilde{r^{\rho}})$, with a special node named $y^{\mathbf{u}}$ which satisfies $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y^{\mathbf{u}} \models \beta$. See Fig. 12(e).

Notice that this rule differs from Rule 2 of §3.3.1 in the fact that we do not merge the classes of $x^{\mathbf{u}}$ and $y^{\mathbf{u}}$ yet. We will perform that merging only at the end of the construction. This is not really important and we could have merged the classes at this step; the reason for doing it at the end is only a technical issue. The proof of Fact 49 will be easier to understand this way.

The following remark will be used later to prove that φ is indeed satisfied in the constructed tree. Its proof is omitted since it is analogous to the proof of Remark 20:

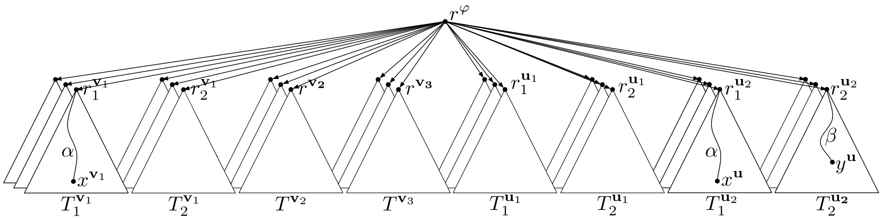

Fig. 13. The tree T^φ (without any partition yet).

Remark 44. Let $(\psi, \alpha, \rho, \beta) \in \mathbf{U_2}$. If $\neg\langle\downarrow[\psi]\mu = \downarrow[\rho]\delta\rangle$ is a conjunct of φ, then $[y^{\mathbf{u}}]_{\pi_2^{\mathbf{u}}} \neq [y]_{\pi_2^{\mathbf{u}}}$ for all y such that $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y \models \delta$ or $[x^{\mathbf{u}}]_{\pi_1^{\mathbf{u}}} \neq [x]_{\pi_1^{\mathbf{u}}}$ for all x such that $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x \models \mu$. $\square$

The rooted data tree $(T^\varphi, \pi^\varphi, r^\varphi)$ As shown in Fig. 13, now we define T^φ, using our Rules, as the tree which consists of a root r^φ with label $a \in \mathbb{A}$ if a is a conjunct of φ, and with children

$$(T_1^{\mathbf{v_1}})_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}}, (T_2^{\mathbf{v_1}})_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}}, (T^{\mathbf{v_2}})_{\mathbf{v_2}\in\mathbf{V}_{=,\neg\neq}}, (T^{\mathbf{v_3}})_{\mathbf{v_3}\in\mathbf{V}_{\neg=,\neq}}, (T_1^{\mathbf{u}})_{\mathbf{u}\in\mathbf{U}}, (T_2^{\mathbf{u}})_{\mathbf{u}\in\mathbf{U}}.$$

As a first step we provisionally define $\widetilde{\pi^\varphi}$ over T^φ by

$$\widetilde{\pi^\varphi} = \{\{r^\varphi\}\} \cup \bigcup_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}} (\pi_1^{\mathbf{v_1}} \cup \pi_2^{\mathbf{v_1}}) \cup \bigcup_{\mathbf{v_2}\in\mathbf{V}_{=,\neg\neq}} \pi^{\mathbf{v_2}} \cup \bigcup_{\mathbf{v_3}\in\mathbf{V}_{\neg=,\neq}} \pi^{\mathbf{v_3}} \cup \bigcup_{\mathbf{u}\in\mathbf{U}} (\pi_1^{\mathbf{u}} \cup \pi_2^{\mathbf{u}}).$$

It is important to notice that, up to this point in the construction, the tree hanging from each child of the root preserves its original partition.

In order to consider the information given by $\mathbf{Z}$ (cf. (Case 4)), we split $\mathbf{V}_{\neg=,\neq}$ into two subsets:

$$\mathbf{V}'_{\neg=,\neq} = \left\{(\psi,\alpha) \in \mathbf{V}_{\neg=,\neq} \mid \text{ for all } (\rho,\beta) \in \mathbf{V}_{\neg=,\neq},\ (\psi,\alpha,\rho,\beta) \notin \mathbf{Z}\right\},$$
$$\mathbf{V}''_{\neg=,\neq} = \mathbf{V}_{\neg=,\neq} \setminus \mathbf{V}'_{\neg=,\neq}.$$

The following property of the set $\mathbf{V}''_{\neg=,\neq}$ will be used to prove that φ is indeed satisfied at the constructed tree:

Lemma 45. *Let* $(\theta,\delta), (\theta,\delta') \in \mathbf{V}_{\neg=,\neq}$. *Suppose that* $(\theta,\delta) \in \mathbf{V}''_{\neg=,\neq}$ *and* $\neg\langle\delta' \neq \delta'\rangle$ *and* $\langle\delta = \delta'\rangle$ *are conjuncts of* θ. *Then* $(\theta,\delta,\theta,\delta') \in \mathbf{Z}$.

Proof. By NeqAx7, $\neg\langle\downarrow[\theta]\delta \neq \downarrow[\theta]\delta\rangle$ is a conjunct of φ. By EqAx5 plus **Der21** of Fact 5, $\langle\downarrow[\theta]\delta = \downarrow[\theta]\delta'\rangle$ is a conjunct of φ. If we suppose that $\langle\downarrow[\theta]\delta' \neq \downarrow[\theta]\delta'\rangle$ is a conjunct of φ, by Lemma 30, we have that $\neg\langle\delta = \delta'\rangle$ is a conjunct of θ which is a contradiction. Then we can assume that $\neg\langle\downarrow[\theta]\delta' \neq \downarrow[\theta]\delta'\rangle$ is a conjunct of φ and so we can conclude from NeqAx7 that $\neg\langle\downarrow[\theta]\delta \neq \downarrow[\theta]\delta'\rangle$ is a conjunct of φ. Then we have that $(\theta,\delta,\theta,\delta') \in \mathbf{Z}$. $\square$

As a particular case of Lemma 45, we have:

Remark 46. Let $(\theta,\delta), (\theta,\delta') \in \mathbf{V}_{\neg=,\neq}$. Suppose that $(\theta,\delta), (\theta,\delta') \in \mathbf{V}''_{\neg=,\neq}$ and $\langle\delta = \delta'\rangle$ is a conjunct of θ. Then $(\theta,\delta,\theta,\delta') \in \mathbf{Z}$.

Proof. Use NeqAx5 plus **Der21** of Fact 5 and NeqAx7. $\square$

We classify the elements of $\mathbf{V}''_{\neg=,\neq}$ according to the following equivalence relation:

$$[(\psi,\alpha)] = [(\rho,\beta)] \quad \text{iff} \quad (\psi,\alpha,\rho,\beta) \in \mathbf{Z}.$$

Observe that this relation is reflexive by NeqAx7, it is clearly symmetric and it is transitive by Lemma 42. We name the equivalence classes $A_1,\dots,A_m$. We define $\widehat{\pi^\varphi}$ over T^φ taking into account the information given by $\mathbf{V}_{=,\neq}, \mathbf{V}_{=,\neg\neq}$ and $\mathbf{Z}$. $\widehat{\pi^\varphi}$ is the smallest equivalence relation containing $\widetilde{\pi^\varphi}$ such that:

- $[x^{\mathbf{v_1}}]_{\widehat{\pi^\varphi}} = [r^\varphi]_{\widehat{\pi^\varphi}}$ for all $\mathbf{v_1} \in \mathbf{V}_{=,\neq}$,
- $[x]_{\widehat{\pi^\varphi}} = [r^\varphi]_{\widehat{\pi^\varphi}}$ for all $x \in M$,
- For all $i = 1,\dots,m$ $[x]_{\widehat{\pi^\varphi}} = [y]_{\widehat{\pi^\varphi}}$ for all $x, y \in L_i$

where

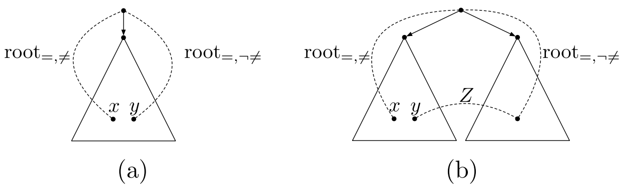

Fig. 14. Examples of (hypothetical) "gluings".

$$M = \{x \mid \text{ there exists } (\psi, \alpha) \in \mathbf{V}_{=,\neg\neq} \text{ and a child } z \text{ of } r^{\varphi} \text{ such that}$$
$$T^{\varphi}, \widetilde{\pi^{\varphi}}, z \models \psi \text{ and } T^{\varphi}, \widetilde{\pi^{\varphi}}, z, x \models \alpha\}$$
$$L_i = \{x \mid \text{ there exists } (\psi, \alpha) \in A_i \text{ and a child } z \text{ of } r^{\varphi} \text{ such that}$$
$$T^{\varphi}, \widetilde{\pi^{\varphi}}, z \models \psi \text{ and } T^{\varphi}, \widetilde{\pi^{\varphi}}, z, x \models \alpha\}$$

for all $i = 1, \dots, m$.

In the previous "gluing", we forced our model to satisfy all diamonds of the form $\langle \epsilon = \downarrow[\psi]\alpha \rangle$, $\neg\langle \epsilon \neq \downarrow[\psi]\alpha \rangle$ and $\neg\langle \downarrow[\psi]\alpha \neq \downarrow[\rho]\beta \rangle$ that need to be forced.

It is important to notice that, up to here, the tree hanging from each child of the root still preserves its partition:

Fact 47. The partition restricted to the trees $T_1^{\mathbf{v_1}}$, $T_2^{\mathbf{v_1}}$ for $\mathbf{v_1} \in \mathbf{V}_{=,\neq}$, the partition restricted to the trees $T^{\mathbf{v_2}}$ for $\mathbf{v_2} \in \mathbf{V}_{=,\neg\neq}$, the partition restricted to the trees $T^{\mathbf{v_3}}$ for $\mathbf{v_3} \in \mathbf{V}_{\neg=,\neq}$ and the partition restricted to the trees $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ for $\mathbf{u} \in \mathbf{U}$ remain unchanged. More formally:

- For each $\mathbf{v_1} = (\psi, \alpha) \in \mathbf{V}_{=,\neq}$ and $i \in \{1, 2\}$, we have $\widehat{\pi^{\varphi}} \upharpoonright_{T_i^{\mathbf{v_1}}} = \pi_i^{\mathbf{v_1}}$.
- For each $\mathbf{v_2} = (\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$, we have $\widehat{\pi^{\varphi}} \upharpoonright_{T^{\mathbf{v_2}}} = \pi^{\mathbf{v_2}}$.
- For each $\mathbf{v_3} = (\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$, we have $\widehat{\pi^{\varphi}} \upharpoonright_{T^{\mathbf{v_3}}} = \pi^{\mathbf{v_3}}$.
- For each $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U}$ and $i \in \{1, 2\}$, we have $\widehat{\pi^{\varphi}} \upharpoonright_{T_i^{\mathbf{u}}} = \pi_i^{\mathbf{u}}$.

Proof. We give a sketch of the proof and leave the details to the reader. If we think we have three kinds of "gluings", root$_{=,\neq}$-kind, root$_{=,\neg\neq}$-kind and Z-kind, then the way in which two equivalence classes in the same subtree can (hypothetically) be glued together is by a sequence of these gluings. The examples displayed in Fig. 14 shows that in (a), the classes of nodes x and y were glued together by a sequence of the form root$_{=,\neq}$-root$_{=,\neg\neq}$; in (b), the classes of nodes x and y were glued together by a sequence of the form root$_{=,\neq}$-root$_{=,\neg\neq}$-Z.

We give a list of the ingredients for the complete proof.

- By Rule 1, every witness for $\langle \downarrow[\psi]\alpha \rangle$ with $(\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$ in $\mathcal{T}_1^{\mathbf{v_1}}$ is in a different class (according to $\widetilde{\pi^{\varphi}}$) than $x^{\mathbf{v_1}}$ for all $\mathbf{v_1} \in \mathbf{V}_{=,\neq}$. Thus we do not have sequences containing root$_{=,\neq}$-Z or Z-root$_{=,\neq}$.
- Lemma 36 implies that every witness for $\langle \downarrow[\psi]\alpha \rangle$ with $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$ and every witness for $\langle \downarrow[\psi]\beta \rangle$ with $(\psi, \beta) \in \mathbf{V}_{\neg=,\neq}$ in the same subtree belong to different classes in that subtree. As a particular case, every $x \in M$ and $y \in L_i$ in the same subtree belong to different classes. Thus we do not have sequences containing root$_{=,\neg\neq}$-Z or Z-root$_{=,\neg\neq}$.
- Since we use a different copy at each application of Rule 1, we do not have sequences starting and ending with root$_{=,\neq}$.
- Lemma 36 implies that every witness for $\langle \downarrow[\psi]\alpha \rangle$ with $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$ in the same subtree belong to the same equivalence class in that subtree. Thus we do not have to worry about sequences starting and ending with root$_{=,\neg\neq}$ because this kind of sequences do not glue different classes.
- By Rule 1, every $x^{\mathbf{v_1}}$ is in the same class that every witness in the same subtree of $\langle \downarrow[\psi]\alpha \rangle$ with $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$. Thus we do not have to worry about sequences starting with root$_{=,\neq}$ and ending with root$_{=,\neg\neq}$ (or vice versa) because this kind of sequences do not glue different classes.
  Combining the previous items, it only remains to consider sequences of only Z-kind gluings.
- If $x, x' \in L_i$ in the same subtree, a very simple derivation involving NeqAx5 , shows that $[x]_{\widetilde{\pi^{\varphi}}} = [x']_{\widetilde{\pi^{\varphi}}}$. Thus we do not have to worry about sequences of the form Z (just one Z-kind gluing).
- By Lemma 48 below, we do not have to worry about longer sequences of all Z-kind gluings.

This concludes the proof of the Fact. $\square$

Lemma 48. *Let* $\psi, \theta_0, \dots, \theta_m \in N_n$, $\alpha, \beta, \delta_0, \delta_0', \dots, \delta_m, \delta_m' \in P_n$, $x, x', y, y' \in T^{\psi}$, $x_0, y_0 \in T^{\theta_0}, \dots,$ $x_m, y_m \in T^{\theta_m}$ *such that (see Fig. 15):*

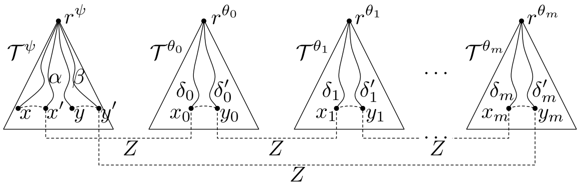

Fig. 15. The hypothesis of Lemma 48.

- $[x]_{\pi^\psi} = [x']_{\pi^\psi}$, $[y]_{\pi^\psi} = [y']_{\pi^\psi}$, $T^\psi, r^\psi, x' \models \alpha$, $T^\psi, r^\psi, y' \models \beta$,
- $[x_i]_{\pi^{\theta_i}} = [y_i]_{\pi^{\theta_i}}$, $T^{\theta_i}, r^{\theta_i}, x_i \models \delta_i$, $T^{\theta_i}, r^{\theta_i}, y_i \models \delta_i'$ *for* $i = 0 \ldots m$, *and*
- $(\theta_0, \delta_0, \psi, \alpha) \in \mathbf{Z}$, $(\theta_i, \delta_i, \theta_{i-1}, \delta_{i-1}') \in \mathbf{Z}$ *for* $i = 1 \ldots m$, $(\theta_m, \delta_m', \psi, \beta) \in \mathbf{Z}$.

Then $[x]_{\pi^\psi} = [y]_{\pi^\psi}$.

(Notation: For $\rho \in N_n$, *we use* $\mathcal{T}^\rho = (T^\rho, \pi^\rho)$ *with root* r^ρ *to denote* any *tree in which* ρ *is satisfiable, namely the one given by inductive hypothesis, or the modified one* $\widetilde{\mathcal{T}^\rho}$*.)*

Proof. Observe that, by Lemma 45 plus Lemma 42, $(\psi, \alpha, \psi, \beta) \in \mathbf{Z}$. Then, by NeqAx7 plus Lemma 36, $\neg\langle \alpha \neq \beta \rangle$ is a conjunct of ψ and so $[x]_{\pi^\psi} = [y]_{\pi^\psi}$. □

Finally, define π^φ over T^φ by

$$\pi^\varphi = \left(\widehat{\pi^\varphi} \setminus \left(\{[x^{\mathbf{u}}]_{\widehat{\pi^\varphi}}\}_{\mathbf{u}\in\mathbf{U_2}} \cup \{[y^{\mathbf{u}}]_{\widehat{\pi^\varphi}}\}_{\mathbf{u}\in\mathbf{U_2}}\right)\right) \cup \bigcup_{\mathbf{u}\in\mathbf{U_2}} \{[x^{\mathbf{u}}]_{\widehat{\pi^\varphi}} \cup [y^{\mathbf{u}}]_{\widehat{\pi^\varphi}}\}.$$

In other words, T^φ has a root, named r^φ, and children

$$(T_1^{\mathbf{v_1}})_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}}, (T_2^{\mathbf{v_1}})_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}}, (T^{\mathbf{v_2}})_{\mathbf{v_2}\in\mathbf{V}_{=,\neg\neq}}, (T^{\mathbf{v_3}})_{\mathbf{v_3}\in\mathbf{V}_{\neg=,\neq}}, (T_1^{\mathbf{u}})_{\mathbf{u}\in\mathbf{U}}, (T_2^{\mathbf{u}})_{\mathbf{u}\in\mathbf{U}}.$$

Each of these children is the root of its corresponding tree inside T^φ as defined above. All these subtrees are disjoint, and π^φ is defined as the disjoint union of the partitions *with the exception* that we put into the same class:

- the nodes r^φ, $(x^{\mathbf{v_1}})_{\mathbf{v_1}\in\mathbf{V}_{=,\neq}}$ and every witness of $\langle\downarrow[\psi]\alpha\rangle$ with $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$,
- a witness for $\langle\downarrow[\psi]\alpha\rangle$ and a witness for $\langle\downarrow[\rho]\beta\rangle$ if $(\psi, \alpha, \rho, \beta) \in \mathbf{U_2}$,
- every pair of witnesses of $\langle\downarrow[\psi]\alpha\rangle$ and $\langle\downarrow[\rho]\beta\rangle$ respectively with $(\psi, \alpha, \rho, \beta) \in \mathbf{Z}$.

In the previous gluing, we forced our model to satisfy all diamonds of the form $\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ that need to be forced.

The following Fact is key to prove that φ is satisfied in $\mathcal{T}^\varphi$:

Fact 49. The partition restricted to the trees $T_1^{\mathbf{v_1}}$, $T_2^{\mathbf{v_1}}$ for $\mathbf{v_1} \in \mathbf{V}_{=,\neq}$, the partition restricted to the trees $T^{\mathbf{v_2}}$ for $\mathbf{v_2} \in \mathbf{V}_{=,\neg\neq}$, the partition restricted to the trees $T^{\mathbf{v_3}}$ for $\mathbf{v_3} \in \mathbf{V}_{\neg=,\neq}$ and the partition restricted to the trees $T_1^{\mathbf{u}}$ and $T_2^{\mathbf{u}}$ for $\mathbf{u} \in \mathbf{U}$ remain unchanged. More formally:

- For each $\mathbf{v_1} = (\psi, \alpha) \in \mathbf{V}_{=,\neq}$ and $i \in \{1, 2\}$, we have $\pi^\varphi\restriction_{T_i^{\mathbf{v_1}}} = \pi_i^{\mathbf{v_1}}$.
- For each $\mathbf{v_2} = (\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$, we have $\pi^\varphi\restriction_{T^{\mathbf{v_2}}} = \pi^{\mathbf{v_2}}$.
- For each $\mathbf{v_3} = (\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$, we have $\pi^\varphi\restriction_{T^{\mathbf{v_3}}} = \pi^{\mathbf{v_3}}$.
- For each $\mathbf{u} = (\psi, \alpha, \rho, \beta) \in \mathbf{U}$ and $i \in \{1, 2\}$, we have $\pi^\varphi\restriction_{T_i^{\mathbf{u}}} = \pi_i^{\mathbf{u}}$.

Proof. We give a guide for the proof and we leave the details to the reader.

Now think that we have four kinds of "gluings", root$_{=,\neq}$-kind, root$_{=,\neg\neq}$-kind, Z-kind and U_2-kind, then the way in which two equivalence classes in the same subtree can (hypothetically) be glued together is by a sequence of these gluings. In the example displayed in Fig. 16 the classes of nodes x and y were glued together by a sequence of the form Z-Z-U_2.

We give a list of the ingredients for the complete proof.

- We have already observed that the same assertions hold if we change π^φ for $\widehat{\pi^\varphi}$ so we are only interested in sequences that involve some gluing of kind U_2. Moreover, we can assume all the observations made in the proof of Fact 47.
- The fact that $x^{\mathbf{v_1}}$ and $x^{\mathbf{u}}$ (or $y^{\mathbf{u}}$) are always in different subtrees tells us that we do not have sequences containing root$_{=,\neq}$-U_2 or U_2-root$_{=,\neq}$.

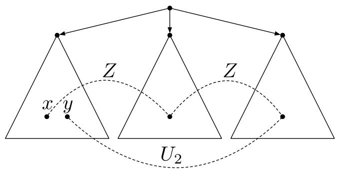

Fig. 16. Example of (hypothetical) "gluing".

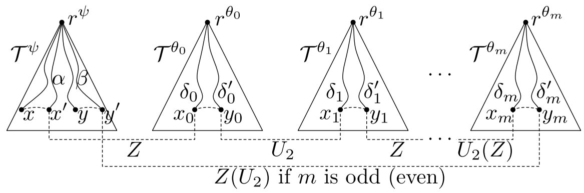

Fig. 17. The hypothesis of Lemma 50.

- Lemma 36 implies that every witness for $\langle\downarrow[\psi]\alpha\rangle$ with $(\psi,\alpha)\in\mathbf{V}_{=,\neg\neq}$ and every witness for $\langle\downarrow[\psi]\beta\rangle$ with $(\psi,\beta)\in\mathbf{V}_{\neg=,\neq}$ in the same subtree belong to different classes in that subtree. Thus we do not have sequences containing root$_{=,\neg\neq}$-U_2 or U_2-root$_{=,\neg\neq}$ coming from $\mathbf{u}=(\psi,\alpha,\rho,\beta)\in U_2$ with $(\psi,\alpha),(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$. Besides, suppose that we have one of those sequences coming from $\mathbf{u}=(\psi,\alpha,\rho,\beta)\in U_2$ with $(\psi,\alpha)\in\mathbf{V}_{=,\neq},(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$ (the symmetric case is analogous) and $(\psi,\mu)\in\mathbf{V}_{=,\neg\neq}$. Then, by the consistency of φ plus NeqAx6, we can conclude that $\neg\langle\downarrow[\psi]\mu=\downarrow[\rho]\beta\rangle$ is a conjunct of φ. This gives us a contradiction by Remark 44. Thus we do not have sequences containing root$_{=,\neg\neq}$-U_2 or U_2- emphroot$_{=,\neg\neq}$ at all.
- By Lemma 45 plus Lemma 42, we can reduce sequences with two consecutive Z-kind gluings to sequences not having two consecutive Z-kind gluings.
- Since we use new subtrees for each $\mathbf{u}\in\mathbf{U_2}$, we cannot have sequences containing U_2-U_2 neither sequences starting and ending with U_2.
- By Lemma 50 below, we cannot have sequences that alternate Z-kind gluings with U_2-kind gluings.
- One can think that the gluing of the classes $[x^{\mathbf{u}}]_{\widehat{\pi^\varphi}}$ and $[y^{\mathbf{u}}]_{\widehat{\pi^\varphi}}$ is made one at a time since they are finite.

This concludes the proof of the Fact. $\square$

Lemma 50. *Let $\psi,\theta_0,\ldots,\theta_m\in N_n$, $\alpha,\beta,\delta_0,\delta'_0,\ldots,\delta_m,\delta'_m\in P_n$, $x,x',y,y'\in T^\psi$, $x_0,y_0\in T^{\theta_0},\ldots,$ $x_m,y_m\in T^{\theta_m}$. The following conditions (see Fig. 17) cannot be satisfied all at the same time:*

- $[x]_{\pi^\psi}=[x']_{\pi^\psi}$, $[y]_{\pi^\psi}=[y']_{\pi^\psi}$, $T^\psi,r^\psi,x'\models\alpha$, $T^\psi,r^\psi,y'\models\beta$,
- $[x_i]_{\pi^{\theta_i}}=[y_i]_{\pi^{\theta_i}}$, $T^{\theta_i},r^{\theta_i},x_i\models\delta_i$, $T^{\theta_i},r^{\theta_i},y_i\models\delta'_i$ *for* $i=0\ldots m$,
- $(\theta_0,\delta_0,\psi,\alpha)\in\mathbf{Z}$,
- *for* $i=1\ldots m$, $(\theta_i,\delta_i,\theta_{i-1},\delta'_{i-1})\in\begin{cases}\mathbf{U_2} & \text{if } i \text{ is odd,}\\ \mathbf{Z} & \text{otherwise,}\end{cases}$
- $(\theta_m,\delta'_m,\psi,\beta)\in\begin{cases}\mathbf{Z} & \text{if } m \text{ is odd,}\\ \mathbf{U}_2 & \text{otherwise.}\end{cases}$

(Notation: For $\rho\in N_n$, we use $\mathcal{T}^\rho=(T^\rho,\pi^\rho)$ with root r^ρ to denote any *tree in which ρ is satisfiable, namely the one given by inductive hypothesis, or the modified one $\widetilde{\mathcal{T}^\rho}$.)*

Proof. We proceed by induction on m:

- Case $m=0$ (see Fig. 18(a)):
  Since $(\psi,\beta,\theta_0,\delta'_0)\in\mathbf{U_2}$, $(\psi,\alpha,\theta_0,\delta_0)\in\mathbf{Z}$ and $\langle\delta_0=\delta'_0\rangle$ is a conjunct of θ_0, we have that $\neg\langle\alpha=\beta\rangle$ is a conjunct of ψ. But, on the other hand, by Remark 44, we know that $\langle\downarrow[\psi]\beta=\downarrow[\theta_0]\delta_0\rangle$ is a conjunct of φ which implies, by Lemma 39, that $\langle\alpha=\beta\rangle$ is a conjunct of ψ, a contradiction.
- If $m=1$ (see Fig. 18(b)):
  By Remark 44, $\langle\downarrow[\theta_0]\delta_0=\downarrow[\theta_1]\delta'_1\rangle$ is a conjunct of φ and then, by NeqAx7, $(\theta_0,\delta_0,\theta_1,\delta'_1)\in\mathbf{Z}$. This gives a contradiction with the fact that $(\theta_0,\delta'_0,\theta_1,\delta_1)\in\mathbf{U_2}$ plus the fact that $\langle\delta_0=\delta'_0\rangle$ is a conjunct of θ_0 and $\langle\delta_1=\delta'_1\rangle$ is a conjunct of θ_1.
- For the induction, suppose $m\geq 2$:

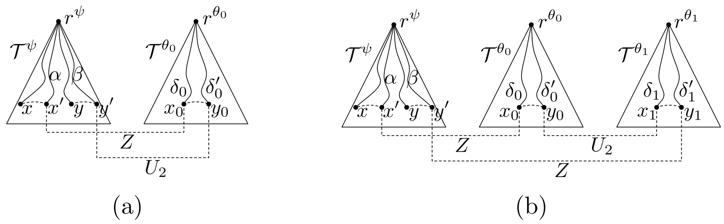

Fig. 18. Proof of Lemma 50. (a) case $m = 0$. (b) case $m = 1$.

In case m is odd, by Remark 44, $\langle\downarrow[\theta_{m-1}]\delta_{m-1} = \downarrow[\theta_m]\delta'_m\rangle$ is a conjunct of φ and then, by NeqAx7, $(\theta_{m-1}, \delta_{m-1}, \theta_m, \delta'_m) \in \mathbf{Z}$. By Lemma 42, $(\psi, \beta, \theta_{m-2}, \delta'_{m-2}) \in \mathbf{Z}$ and the result follows from inductive hypothesis for $m-2$.

In case m is even, by Remark 44, $\langle\downarrow[\theta_0]\delta_0 = \downarrow[\theta_1]\delta'_1\rangle$ is a conjunct of φ and then, by NeqAx7 plus NeqAx7, $(\theta_0, \delta_0, \theta_1, \delta'_1) \in \mathbf{Z}$. By Lemma 42, $(\psi, \alpha, \theta_2, \delta_2) \in \mathbf{Z}$ and the result follows from inductive hypothesis for $m-2$.

This concludes the proof. □

We conclude from Proposition 3 and the construction that:

Fact 51. The validity of a formula in a child of r^φ is preserved in $\mathcal{T}^\varphi$. More formally:

- For each $\mathbf{v_1} \in \mathbf{V}_{=,\neq}$, $i \in \{1,2\}$ and $x, y \in T_i^{\mathbf{v_1}}$ we have $\mathcal{T}^\varphi, x \equiv \mathcal{T}_i^{\mathbf{v_1}}, x$ and $\mathcal{T}^\varphi, x, y \equiv \mathcal{T}_i^{\mathbf{v_1}}, x, y$.
- For each $\mathbf{v_2} \in \mathbf{V}_{=,\neg\neq}$ and $x, y \in T^{\mathbf{v_2}}$ we have $\mathcal{T}^\varphi, x \equiv \mathcal{T}^{\mathbf{v_2}}, x$ and $\mathcal{T}^\varphi, x, y \equiv \mathcal{T}^{\mathbf{v_2}}, x, y$.
- For each $\mathbf{v_3} \in \mathbf{V}_{\neg=,\neq}$ and $x, y \in T^{\mathbf{v_3}}$ we have $\mathcal{T}^\varphi, x \equiv \mathcal{T}^{\mathbf{v_3}}, x$ and $\mathcal{T}^\varphi, x, y \equiv \mathcal{T}^{\mathbf{v_3}}, x, y$.
- For each $\mathbf{u} \in \mathbf{U}$, $i \in \{1,2\}$ and $x, y \in T_i^{\mathbf{u}}$ we have $\mathcal{T}^\varphi, x \equiv \mathcal{T}_i^{\mathbf{u}}, x$ and $\mathcal{T}^\varphi, x, y \equiv \mathcal{T}_i^{\mathbf{u}}, x, y$.

It only remains to prove that the conditions (C1) – (C9) from the beginning of §4.3.1 are satisfied in the tree we have constructed:

Verification of (C1) This condition is trivially satisfied.

Verification of (C2) Suppose $\langle\epsilon = \downarrow[\psi]\alpha\rangle$ is a conjunct of φ. Then there are two possibilities, $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$ or $(\psi, \alpha) \in \mathbf{V}_{=,\neg\neq}$.

- In the first case, by Rule 1 and construction, there exists $x^{\mathbf{v}_1} \in T^\varphi$ such that $[r^\varphi]_{\pi^\varphi} = [x^{\mathbf{v}_1}]_{\pi^\varphi}$ with $\mathbf{v_1} = (\psi, \alpha)$. By construction, we also know $\mathcal{T}_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}} \models \psi$ and $\mathcal{T}_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}}, x^{\mathbf{v}_1} \models \alpha$. Then, by Fact 51, $\mathcal{T}^\varphi, r^\varphi \models \langle\epsilon = \downarrow[\psi]\alpha\rangle$.
- In the second case, $\langle\downarrow[\psi]\alpha\rangle$ is consistent. Then, by construction plus Lemma 26, there is $x \in T^\varphi$ such that $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2} \models \psi$, $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2}, x \models \alpha$ and $[r^\varphi]_{\pi^\varphi} = [x]_{\pi^\varphi}$ with $\mathbf{v_2} = (\psi, \alpha)$. Then, by Fact 51, $\mathcal{T}^\varphi, r^\varphi \models \langle\epsilon = \downarrow[\psi]\alpha\rangle$.

Verification of (C3) Suppose $\langle\epsilon \neq \downarrow[\psi]\alpha\rangle$ is a conjunct of φ. Then there are two possibilities, $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$ or $(\psi, \alpha) \in \mathbf{V}_{\neg=,\neq}$.

- In the first case, by Rule 1 plus Lemmas 26, 36 and 39, there is $x \in T^\varphi$ such that (for $\mathbf{v_1} = (\psi, \alpha)$) $\mathcal{T}_2^{\mathbf{v_1}}, r_2^{\mathbf{v_1}}, x \models \alpha$ and $x \notin [z]_{\pi^\varphi}$ for all z such that $\mathcal{T}_2^{\mathbf{v_1}}, r_2^{\mathbf{v_1}}, z \models \beta$ for some $(\psi, \beta) \in \mathbf{V}_{=,\neg\neq}$ (The argument is similar to the ones used in the proof of Fact 47 to make conclusions from Lemma 36). We also know by construction that $\mathcal{T}_2^{\mathbf{v_1}}, r_2^{\mathbf{v_1}} \models \psi$. In order to conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi \models \langle\epsilon \neq \downarrow[\psi]\alpha\rangle$, it only remains to observe that $[r^\varphi]_{\pi^\varphi} \neq [x]_{\pi^\varphi}$ (for a sketch of the proof see Sketch 53 in Appendix A).
- In the second case, by Rule 3 plus Lemma 26, there exists $x \in T^\varphi$ such that (for $\mathbf{v_3} = (\psi, \alpha)$) $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x \models \alpha$. We also know by construction that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3} \models \psi$. In order to conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi \models \langle\epsilon \neq \downarrow[\psi]\alpha\rangle$, it only remains to observe that $[r^\varphi]_{\pi^\varphi} \neq [x]_{\pi^\varphi}$ (the proof follows the same sketch than the previous case).

Verification of (C4) Suppose $\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$ is a conjunct of φ. By the consistency of φ plus NeqAx6, neither (ψ, α) nor (ρ, β) can be in $\mathbf{V}_{\neg=,\neg\neq}$. By the consistency of φ plus NeqAx6, it cannot be the case that one of them belongs to $\mathbf{V}_{=,\neg\neq}$ and the other one to $\mathbf{V}_{\neg=,\neq}$. Then there are five possibilities to consider (we are omitting symmetric cases):

- If $(\psi, \alpha) \in \mathbf{V}_{=,\neq}$ and $(\rho, \beta) \in \mathbf{V}_{=,\neq}$, by construction, there is $x^{\mathbf{v}_1} \in T^\varphi$ such that $\mathcal{T}_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}} \models \psi$, $\mathcal{T}_1^{\mathbf{v_1}}, r_1^{\mathbf{v_1}}, x^{\mathbf{v}_1} \models \alpha$ and $[r^\varphi]_{\pi^\varphi} = [x^{\mathbf{v}_1}]_{\pi^\varphi}$, with $\mathbf{v}_1 = (\psi, \alpha)$. Since the same happens with (ρ, β), we can conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi \models \langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$.

- If $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{=,\neg\neq}$, by construction, there is $x^{\mathbf{v}_1}\in T^\varphi$ such that $\mathcal{T}_1^{\mathbf{v}_1}, r_1^{\mathbf{v}_1}\models\psi$, $\mathcal{T}_1^{\mathbf{v}_1}, r_1^{\mathbf{v}_1}, x^{\mathbf{v}_1}\models\alpha$ and $[r^\varphi]_{\pi^\varphi}=[x^{\mathbf{v}_1}]_{\pi^\varphi}$, with $\mathbf{v}_1=(\psi,\alpha)$. By Lemma 26 plus Rule 2, there is $x\in T^{\mathbf{v}_2}$ (with $\mathbf{v}_2=(\rho,\beta)$) such that $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2}\models\rho$ and $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2}, x\models\beta$. Then, by construction, $[r^\varphi]_{\pi^\varphi}=[x]_{\pi^\varphi}$ and so $[x^{\mathbf{v}_1}]_{\pi^\varphi}=[x]_{\pi^\varphi}$. We conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$.
- If $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$, by the consistency of φ plus NeqAx6, $(\psi,\alpha,\rho,\beta)=\mathbf{u}\in\mathbf{U}$. Then, by construction, there are $x^{\mathbf{u}}\in T_1^{\mathbf{u}}$, $y^{\mathbf{u}}\in T_2^{\mathbf{u}}$ such that $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}\models\psi$, $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}\models\rho$, $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, x^{\mathbf{u}}\models\alpha$, $\mathcal{T}_2^{\mathbf{u}}, r_2^{\mathbf{u}}, y^{\mathbf{u}}\models\beta$ and $[x^{\mathbf{u}}]_{\pi^\varphi}=[y^{\mathbf{u}}]_{\pi^\varphi}$ (If $\mathbf{u}\in\mathbf{U_2}$ the assertion is straightforward and if $\mathbf{u}\in\mathbf{U_1}$ these nodes exist because of the gluing related to the set Z). Then, we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$.
- If $(\psi,\alpha)\in\mathbf{V}_{=,\neg\neq}$ and $(\rho,\beta)\in\mathbf{V}_{=,\neg\neq}$, by Rule 2 plus Lemma 26, there are $x\in T^{\mathbf{v}_2}$ (with $\mathbf{v}_2=(\psi,\alpha)$) and $y\in T^{\mathbf{v}_2'}$ (with $\mathbf{v}_2'=(\rho,\beta)$) such that $\mathcal{T}^{\mathbf{v_2}}, r^{\mathbf{v_2}}\models\psi$, $\mathcal{T}^{\mathbf{v}_2}, r^{\mathbf{v}_2}, x\models\alpha$, $\mathcal{T}^{\mathbf{v}_2'}, r^{\mathbf{v}_2'}\models\rho$ and $\mathcal{T}^{\mathbf{v}_2'}, r^{\mathbf{v}_2'}, y\models\beta$. By construction, $[x]_{\pi^\varphi}=[r^\varphi]_{\pi^\varphi}=[y]_{\pi^\varphi}$ and so, we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$.
- If $(\psi,\alpha)\in\mathbf{V}_{\neg=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$, then $(\psi,\alpha,\rho,\beta)=\mathbf{u}\in\mathbf{U}$ or $(\psi,\alpha,\rho,\beta)=\mathbf{z}\in\mathbf{Z}$. In the first case, the proof is exactly the same given for the case that $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$. In the other case, by Rule 3 plus Lemma 26, there are $x\in T^{\mathbf{v}_3}$ (with $\mathbf{v}_3=(\psi,\alpha)$), $y\in T^{\mathbf{v}_3'}$ (with $\mathbf{v}_3'=(\rho,\beta)$) such that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}\models\psi$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x\models\alpha$, $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}\models\rho$ and $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}, y\models\beta$. Observe that $[x]_{\pi^\varphi}=[y]_{\pi^\varphi}$ because of the way in which we have defined the partition π^φ. Then we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$.

Verification of (C5) Suppose $\langle\downarrow[\psi]\alpha\neq\downarrow[\rho]\beta\rangle$ is a conjunct of φ. By the consistency of φ plus NeqAx6, neither (ψ,α) nor (ρ,β) can be in $\mathbf{V}_{\neg=,\neg\neq}$. By the consistency of φ plus NeqAx7, it cannot be the case that they both belong to $\mathbf{V}_{=,\neg\neq}$. Then there are five possibilities to consider:

- If $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{=,\neq}$, by items (C2) and (C3), there exist $x,y\in T^\varphi$, such that $\mathcal{T}^\varphi, r^\varphi, x\models\downarrow[\psi]\alpha$, $\mathcal{T}^\varphi, r^\varphi, y\models\downarrow[\rho]\beta$, $[r^\varphi]_{\pi^\varphi}=[x]_{\pi^\varphi}$ and $[r^\varphi]_{\pi^\varphi}\neq[y]_{\pi^\varphi}$. Then we conclude that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha\neq\downarrow[\rho]\beta\rangle$.
- If $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{=,\neg\neq}$, or $(\psi,\alpha)\in\mathbf{V}_{=,\neq}$ and $(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$ or $(\psi,\alpha)\in\mathbf{V}_{=,\neg\neq}$ and $(\rho,\beta)\in\mathbf{V}_{\neg=,\neq}$, the proof is analogous to the previous one.
- If $(\psi,\alpha)=\mathbf{v}_3\in\mathbf{V}_{\neg=,\neq}$ and $(\rho,\beta)=\mathbf{v_3'}\in\mathbf{V}_{\neg=,\neq}$.
  - In case $(\psi,\alpha)\neq(\rho,\beta)$: If $\langle\alpha\neq\alpha\rangle$ is a conjunct of ψ (if $\langle\beta\neq\beta\rangle$ is a conjunct of ρ, the proof is analogous), by Lemma 26, Rule 3 and Fact 49 there exist $x,y\in T^{\mathbf{v}_3}$, $z\in T^{\mathbf{v}_3'}$ such that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}\models\psi$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x\models\alpha$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, y\models\alpha$, $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}\models\rho$, $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}, z\models\beta$ and $[x]_{\pi^\varphi}\neq[y]_{\pi^\varphi}$. Then we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha\neq\downarrow[\rho]\beta\rangle$ (either x or y is not in $[z]_{\pi^\varphi}$).
    Suppose then that $\neg\langle\alpha\neq\alpha\rangle$ is a conjunct of ψ and $\neg\langle\beta\neq\beta\rangle$ is a conjunct of ρ. Then, as before, there exist $x\in T^{\mathbf{v}_3}$, $z\in T^{\mathbf{v}_3'}$ such that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}\models\psi$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x\models\alpha$, $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}\models\rho$, $\mathcal{T}^{\mathbf{v}_3'}, r^{\mathbf{v}_3'}, z\models\beta$. To conclude the proof, it only remains to observe that, in this case, $[x]_{\pi^\varphi}\neq[z]_{\pi^\varphi}$ (for a sketch of the proof see Sketch 54 in Appendix A). Then we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha\neq\downarrow[\rho]\beta\rangle$.
  - In case $(\psi,\alpha)=(\rho,\beta)$, by consistency of φ, we have that $(\psi,\alpha,\psi,\alpha)=\mathbf{u}\in\mathbf{U}$. If $\langle\alpha\neq\alpha\rangle$ is a conjunct of ψ, by Lemma 26, Rule 3 and Fact 49 there exist $x,y\in T^{\mathbf{v}_3}$ such that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}\models\psi$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x\models\alpha$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, y\models\alpha$ and $[x]_{\pi^\varphi}\neq[y]_{\pi^\varphi}$. Then we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha\neq\downarrow[\psi]\alpha\rangle$.
    Suppose then that $\neg\langle\alpha\neq\alpha\rangle$ is a conjunct of ψ. Then, as before, there exist $x\in T^{\mathbf{v}_3}$, $z\in T_1^{\mathbf{u}}$ such that $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v_3}}\models\psi$, $\mathcal{T}^{\mathbf{v}_3}, r^{\mathbf{v}_3}, x\models\alpha$, $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}\models\psi$, $\mathcal{T}_1^{\mathbf{u}}, r_1^{\mathbf{u}}, z\models\alpha$. To conclude the proof, it only remains to observe that, in this case, $[x]_{\pi^\varphi}\neq[z]_{\pi^\varphi}$ (for a sketch of the proof see Sketch 55 in Appendix A). Then we conclude from Fact 51 that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha\neq\downarrow[\rho]\beta\rangle$.

Verification of (C6) Suppose $\neg\langle\epsilon=\downarrow[\psi]\alpha\rangle$ is a conjunct of φ. Aiming for a contradiction, suppose that $\mathcal{T}^\varphi, r^\varphi\models\langle\epsilon=\downarrow[\psi]\alpha\rangle$. Then there is a successor z of r^φ in which ψ holds, and, by construction plus Lemma 27, z is the root of some copy of the tree $\mathcal{T}^\psi$, i.e. $z=r^\psi$ (it might be $\widetilde{\mathcal{T}^\psi}$ and $\widetilde{r^\psi}$ but, in that case, the argument is the same). Moreover, there is $x\in T^\psi$ such that $\mathcal{T}^\psi, r^\psi, x\models\alpha$, with $[x]_{\pi^\varphi}=[r^\varphi]_{\pi^\varphi}$. In addition to this, $(\psi,\alpha)\in\mathbf{V}_{\neg=,\neq}$ or $(\psi,\alpha)\in\mathbf{V}_{\neg=,\neg\neq}$. If the latter occurs, by construction of $\mathcal{T}^\varphi$ plus Lemma 34 and Lemma 27, we have that $\neg\langle\alpha=\alpha\rangle$ is a conjunct of ψ which is a contradiction. In the former, observe that $[x]_{\pi^\varphi}\neq[r^\varphi]_{\pi^\varphi}$ (for a sketch of the proof see Sketch 56 in Appendix A) which is a contradiction.

Verification of (C7) Suppose $\neg\langle\epsilon\neq\downarrow[\psi]\alpha\rangle$ is a conjunct of φ. Aiming for a contradiction, suppose that $\mathcal{T}^\varphi, r^\varphi\models\langle\epsilon\neq\downarrow[\psi]\alpha\rangle$. Then there is a successor z of r^φ in which ψ holds, and by construction and Lemma 27, z is the root of some copy of the tree $\mathcal{T}^\psi$, i.e. $z=r^\psi$ (it might be $\widetilde{\mathcal{T}^\psi}$ and $\widetilde{r^\psi}$ but, in that case, the argument is the same). Moreover, there is $x\in T^\psi$ such that $\mathcal{T}^\psi, r^\psi, x\models\alpha$, with $[x]_{\pi^\varphi}\neq[r^\varphi]_{\pi^\varphi}$. Then, by construction, $(\psi,\alpha)\notin\mathbf{V}_{=,\neg\neq}$. Since $\neg\langle\epsilon\neq\downarrow[\psi]\alpha\rangle$ is a conjunct of φ, the only remaining possibility is that $(\psi,\alpha)\in\mathbf{V}_{\neg=,\neg\neq}$ but this is a contradiction by construction plus Lemma 34 and Lemma 27.

Verification of (C8) Suppose $\neg\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$ is a conjunct of φ. By the consistency of φ plus NeqAx6, it cannot be the case that both $(\psi,\alpha),(\rho,\beta)$ are in $\mathbf{V}_{=,\neq}\cup\mathbf{V}_{=,\neg\neq}$. In case $(\psi,\alpha)\in\mathbf{V}_{\neg=,\neg\neq}$ (if $(\rho,\beta)\in\mathbf{V}_{\neg=,\neg\neq}$, the proof is analogous), suppose that $\mathcal{T}^\varphi, r^\varphi\models\langle\downarrow[\psi]\alpha=\downarrow[\rho]\beta\rangle$. In particular, there is a successor of r^φ, z and a descendant w such that $\mathcal{T}^\varphi, z, w\models$

$[\psi]\alpha$. But this is a contradiction by construction plus Lemma 34 and Lemma 27. Then $\mathcal{T}^\varphi, r^\varphi \models \neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$. We then have three remaining cases to analyze:

- If $(\psi,\alpha) \in \mathbf{V}_{=,\neg\neq}$ and $(\rho,\beta) \in \mathbf{V}_{\neg=,\neq}$, then, by items (C6) and (C7), we have the result.
- If $(\psi,\alpha) \in \mathbf{V}_{=,\neq}$ and $(\rho,\beta) \in \mathbf{V}_{\neg=,\neq}$ or $(\psi,\alpha),(\rho,\beta) \in \mathbf{V}_{\neg=,\neq}$. In order to conclude that $\mathcal{T}^\varphi, r^\varphi \models \neg\langle\downarrow[\psi]\alpha = \downarrow[\rho]\beta\rangle$, one only have to observe that, if $x, y \in T^\varphi$ are such that $\mathcal{T}^\varphi, r^\varphi, x \models \downarrow[\psi]\alpha$ and $\mathcal{T}^\varphi, r^\varphi, y \models \downarrow[\rho]\beta$, then $[x]_{\pi^\varphi} \neq [y]_{\pi^\varphi}$ (for a sketch of the proof see Sketch 57 in Appendix A).

Verification of (C9) Suppose $\neg\langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$ is a conjunct of φ. By the consistency of φ plus NeqAx6, it cannot be the case that one of $(\psi,\alpha),(\rho,\beta)$ is from $\mathbf{V}_{=,\neq}$ and the other from $\mathbf{V}_{=,\neg\neq}$, neither can one be from $\mathbf{V}_{=,\neq}$ and the other from $\mathbf{V}_{\neg=,\neq}$, or one from $\mathbf{V}_{=,\neg\neq}$ and the other from $\mathbf{V}_{\neg=,\neq}$, or both from $\mathbf{V}_{=,\neq}$. In case $(\psi,\alpha) \in \mathbf{V}_{\neg=,\neg\neq}$ (if $(\rho,\beta) \in \mathbf{V}_{\neg=,\neg\neq}$, the proof is analogous), suppose that $\mathcal{T}^\varphi, r^\varphi \models \langle\downarrow[\psi]\alpha \neq \downarrow[\rho]\beta\rangle$. In particular, there is a successor z of r^φ and a descendant w such that $\mathcal{T}^\varphi, z, w \models [\psi]\alpha$. But this is a contradiction by construction plus Lemma 34 and Lemma 27. We then have two remaining cases to analyze:

- If $(\psi,\alpha),(\rho,\beta) \in \mathbf{V}_{=,\neg\neq}$, by item (C7), we have the result.
- If $(\psi,\alpha),(\rho,\beta) \in \mathbf{V}_{\neg=,\neq}$, by the consistency of φ plus NeqAx6, $(\psi,\alpha,\rho,\beta) \in \mathbf{Z}$ and the result follows immediately from the construction of the model.

5. Conclusions

The addition of an equivalence relation on top of a tree-like Kripke model, and the ability of the modal language to compare if two nodes at the end of path expressions are in the *same* or in *different* equivalence classes has proved to change remarkably the canonical model construction of the basic modal logic. When the language has only comparisons by 'equality', the situation is somewhat simpler, based on the fact that 'equality' is a transitive relation. Also notice that while

"all pairs of paths with certain properties end in *different* equivalence classes" (2)

is expressible when tests by equality are present,

"all pairs of paths with certain properties end in the *same* equivalence classes" (3)

is only expressible when tests by inequality are also present. Both properties are universal. However, in the construction of the canonical model, (2) is compatible with adding many disjoint copies of subtrees with disjoint partitions, while (3) is not. The axiomatization for the fragment containing both the operators of 'equality' and 'inequality' proved to be much more involved than the one containing only 'equality', as witnessed by the large amount of axioms reflecting the intricate relationships between both binary operators.

In this research we have considered XPath$_=(\downarrow)$ over arbitrary data trees. Furthermore, XPath$_=(\downarrow)$ is also suitable for reasoning about (finite or infinite) data *graphs*, as it is done in [20,1]. In either of the alternatives (finite vs. infinite data trees vs. data graphs) it can be shown that XPath$_=(\downarrow)$ is also axiomatizable by the system given in this paper —notice there are no specific axioms of an underlying tree topology. Since our construction of canonical models gives us a recursively bounded finite data tree, we conclude:

Corollary 52 *(Bounded tree model property). There is a primitive recursive function f such that any satisfiable node or path expression φ of XPath$_=(\downarrow)$ of size n over the class of finite/arbitrary data trees/data graphs is satisfiable in a data tree of size at most $f(n)$.*

Hence, although in the database community one may restrict to the finite case, the above corollary shows that allowing or disallowing infinite models does not make any difference.

This already shows that the satisfiability problem of XPath$_=(\downarrow)$ is decidable over any of the classes of models stated above. Of course, this result —at least for XPath$_=(\downarrow)$ over finite data trees— is not new, as mentioned in the introduction [11]. However, the canonical model construction may give us new insights into obtaining sequent calculus axiomatizations, as done in [5], which might be useful for obtaining alternative proofs of complexity for the satisfiability problem of fragments or extensions of XPath$_=(\downarrow)$.

On the application side, the axioms may help to define effective rewrite rules for query optimization in XPath$_=(\downarrow)$.

The study of XPath with 'descendant' instead of 'child' axis seems to be much harder. This question, or the addition of other axes such as 'parent' or 'sibling' (in the case of ordered trees), constitute future lines of research.

Acknowledgments

This work was partially supported by grant ANPCyT-PICT-2013-2011, UBACyT 20020150100002BA, and 16-STIC-04 FoG.

Appendix A. Missing proofs

Lemma 29. *Let* $* \in \{=, \neq\}$, $\gamma \in P_n$, $\psi_i \in N_{n-i}$ *for* $i = 1, \dots, i_0$, $\alpha, \beta \in P_{n-i_0}$ such that

$$\langle \gamma * \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle \wedge \neg \langle \gamma * \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle$$

is consistent and $\neg\langle \alpha \neq \alpha \rangle$ *is a conjunct of* ψ_{i_0}. *Then* $\neg\langle \alpha = \beta \rangle$ *is a conjunct of* ψ_{i_0}.

Proof. Let us start with the case of $*$ being $\neq$. Aiming for a contradiction, suppose that $\langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle$ is consistent and that both $\neg\langle \alpha \neq \alpha \rangle$ and $\langle \alpha = \beta \rangle$ are conjuncts of ψ_{i_0}.

First, let us prove some facts that will be useful in the rest of the proof:

1. The following derivation:

$$\begin{aligned}
\langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle &\leq \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle && (\text{NeqAx4}) \\
&\leq \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}] \rangle && (\textbf{Der12} \text{ Fact 5}) \\
&\leq \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha = \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{EqAx5 \& } \textbf{Der21} \text{ (Fact 5)}) \\
&\leq \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{EqAx4})
\end{aligned}$$

   In particular, by **Der13** (Fact 5), we have that $\langle \downarrow[\psi_{i_0}]\beta \rangle$ is consistent and so $\langle \beta = \beta \rangle$ is a conjunct of ψ_{i_0} (by Lemma 26).
2. From the second line of Item 1, we have that $\langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}] \rangle$ is consistent, and then, by **Der13** (Fact 5), ψ_{i_0} is consistent.
3. Aiming for a contradiction, let us suppose that $\langle \beta \neq \beta \rangle$ is a conjunct of ψ_{i_0}. Then

$$\begin{aligned}
&\langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle \\
&\leq \langle \gamma \rangle \wedge \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}] \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{NeqAx4 \& Item 1}) \\
&\leq \langle \gamma \rangle \wedge \langle \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{NeqAx5 \& } \textbf{Der21} \text{ (Fact 5)}) \\
&\equiv \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{NeqAx7}) \\
&\equiv \text{FALSE} && (\text{Boolean})
\end{aligned}$$

   which is a contradiction. Then $\neg\langle \beta \neq \beta \rangle$ is a conjunct of ψ_{i_0}.
4. Because ψ_{i_0} is consistent (Item 2), by the previous Item plus NeqAx7 , $\neg\langle \alpha \neq \beta \rangle$ is a conjunct of ψ_{i_0}.

Then we have

$$\begin{aligned}
&\langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\alpha \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle \\
&\leq \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle \wedge \neg \langle \gamma \neq \downarrow[\psi_1] \dots \downarrow[\psi_{i_0}]\beta \rangle && (\text{Items 1 and 4 \& NeqAx9 \& } \textbf{Der21} \text{ (Fact 5)}) \\
&\equiv \text{FALSE} && (\text{Boolean})
\end{aligned}$$

which is contradiction, from the assumption that $\langle \alpha = \beta \rangle$ was a conjunct of ψ_{i_0}. Therefore, $\neg\langle \alpha = \beta \rangle$ is a conjunct of ψ_{i_0}.

For the case of $*$ being $=$, use NeqAx10 plus **Der21** of Fact 5. $\square$

Lemma 30. *Let* $\psi \in N_n$, $\alpha, \beta \in P_n$ *such that* $\langle \downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha \rangle \wedge \neg \langle \downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma \rangle$ *is consistent and* $\neg\langle \alpha \neq \alpha \rangle$ *is a conjunct of* ψ. *Then* $\neg\langle \alpha = \gamma \rangle$ *is a conjunct of* ψ.

Proof. Aiming for a contradiction, suppose that $\langle \downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha \rangle \wedge \neg \langle \downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma \rangle$ is consistent and both $\neg\langle \alpha \neq \alpha \rangle$ and $\langle \alpha = \gamma \rangle$ are conjuncts of ψ.

Let us prove some facts that will be useful in the rest of the proof:

1. The following derivation:

$$\begin{aligned}
\langle \downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha \rangle &\leq \langle \downarrow[\psi]\alpha \rangle && (\text{NeqAx4}) \\
&\leq \langle \downarrow[\psi] \rangle && (\textbf{Der12} \text{ (Fact 5)}) \\
&\leq \langle \downarrow[\psi]\alpha = \downarrow[\psi]\gamma \rangle && (\text{EqAx5 \& } \textbf{Der21} \text{(Fact 5)}) \\
&\leq \langle \downarrow[\psi]\gamma \rangle && (\text{EqAx4})
\end{aligned}$$

   In particular, we have that $\langle \gamma = \gamma \rangle$ is a conjunct of ψ (by Lemma 26).
2. From the second line of Item 1, we have that $\langle \downarrow[\psi] \rangle$ is consistent, and by **Der13** (Fact 5), ψ is consistent.

3. Aiming for a contradiction, let us suppose that $\langle\gamma \neq \gamma\rangle$ is a conjunct of ψ. Then

$$\begin{aligned}
&\langle\downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle \\
&\quad \leq \langle\downarrow[\psi]\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle && \text{(Item 1)} \\
&\quad \leq \langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle && \text{(NeqAx5 \& \textbf{Der21} (Fact 5))} \\
&\quad \equiv \text{FALSE} && \text{(Boolean)}
\end{aligned}$$

which is a contradiction. Then $\neg\langle\gamma \neq \gamma\rangle$ is a conjunct of ψ.

4. Because ψ is consistent (Item 2), by the previous item plus NeqAx7, $\neg\langle\alpha \neq \gamma\rangle$ is a conjunct of ψ.

Then we have

$$\begin{aligned}
&\langle\downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle \\
&\quad \leq \langle\downarrow[\psi]\gamma\rangle \wedge \langle\downarrow[\psi]\alpha \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle && \text{(Item 1)} \\
&\quad \leq \langle\downarrow[\psi]\alpha \neq \downarrow[\psi]\gamma\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle && \text{(NeqAx7)} \\
&\quad \leq \langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle \wedge \neg\langle\downarrow[\psi]\gamma \neq \downarrow[\psi]\gamma\rangle && \text{(Items 1 and 4, \& NeqAx9 \& \textbf{Der21} (Fact 5))} \\
&\quad \equiv \text{FALSE} && \text{(Boolean)}
\end{aligned}$$

which is contradiction, from the assumption that $\langle\alpha = \gamma\rangle$ was a conjunct of ψ. Therefore, $\neg\langle\alpha = \gamma\rangle$ is a conjunct of ψ. □

Lemma 34. *Let $\psi \in N_n$, $\alpha \in P_n$, $\gamma \in P_{n+1}$. If $\neg\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma\rangle \wedge \langle\downarrow[\psi]\rangle$ is consistent, then $\neg\langle\alpha = \alpha\rangle$ is a conjunct of ψ.*

Proof. Aiming for a contradiction, suppose that $\langle\alpha = \alpha\rangle$ is a conjunct of ψ. Then

$$\begin{aligned}
&\neg\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma\rangle \wedge \langle\downarrow[\psi]\rangle \\
&\quad \leq \neg\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma\rangle \wedge \langle\downarrow[\psi]\alpha = \downarrow[\psi]\alpha\rangle && \text{(EqAx5 \& \textbf{Der21} (Fact 5))} \\
&\quad \equiv \neg\langle\gamma\rangle \wedge \langle\gamma\rangle && \text{(NeqAx6)} \\
&\quad \equiv \text{FALSE} && \text{(Boolean)}
\end{aligned}$$

and this concludes the proof. □

Lemma 36. *Let $* \in \{=, \neq\}$, $\psi \in N_n$, $\alpha, \beta \in P_n$, $\gamma \in P_{n+1}$. If $\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma * \downarrow[\psi]\beta\rangle$ is consistent, then $\neg\langle\alpha * \beta\rangle$ is a conjunct of ψ.*

Proof. Let us first prove the case for $*$ being $\neq$. Suppose that $\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle$ is consistent. Aiming for a contradiction, suppose that $\langle\alpha \neq \beta\rangle$ is a conjunct of ψ. Then

$$\begin{aligned}
&\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle \\
&\quad \equiv \langle\gamma = \downarrow[\psi \wedge \langle\alpha \neq \beta\rangle]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle && \text{(Hypothesis)} \\
&\quad \leq \langle\gamma\rangle \wedge \langle\downarrow[\psi \wedge \langle\alpha \neq \beta\rangle]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle && \text{(EqAx1 \& EqAx4)} \\
&\quad \leq \langle\gamma\rangle \wedge \langle\downarrow[\psi \wedge \langle\alpha \neq \beta\rangle]\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle && \text{(\textbf{Der12} (Fact 5))} \\
&\quad \leq \langle\gamma\rangle \wedge \langle\downarrow[\psi]\alpha \neq \downarrow[\psi]\beta\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle && \text{(NeqAx5 \& \textbf{Der21} (Fact 5))} \\
&\quad \leq \langle\gamma \neq \downarrow[\psi]\beta\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\beta\rangle && \text{(NeqAx7)} \\
&\quad \equiv \text{FALSE} && \text{(Boolean)}
\end{aligned}$$

which is a contradiction. Then $\langle\alpha \neq \beta\rangle$ is a conjunct of ψ. For the case in which $*$ is $=$, the proof is similar but instead of NeqAx5 we use EqAx5 and instead of NeqAx7 we use NeqAx6. □

Lemma 39. *Let $* \in \{=, \neq\}$, $\psi \in N_n$, $\alpha, \beta \in P_n$, $\gamma \in P_{n+1}$. If $\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma * \downarrow[\psi]\beta\rangle$ is consistent, then $\langle\alpha * \beta\rangle$ is a conjunct of ψ.*

Proof. Let us first prove the case for $*$ being $=$. Suppose that $\langle\gamma = \downarrow[\psi]\alpha\rangle \wedge \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma = \downarrow[\psi]\beta\rangle$ is consistent. Aiming for a contradiction, suppose that $\neg\langle\alpha = \beta\rangle$ is a conjunct of ψ. Also, since $\langle\downarrow[\psi]\alpha\rangle$ is consistent (by EqAx4), then by Lemma 26 $\langle\alpha = \alpha\rangle$ is a conjunct of ψ. Then

$$
\begin{aligned}
&\langle\gamma = \downarrow[\psi]\alpha\rangle\wedge\neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle\wedge\langle\gamma = \downarrow[\psi]\beta\rangle \\
&\quad\leq \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma = \downarrow[\psi\wedge\neg\langle\alpha = \beta\rangle\wedge\langle\alpha\rangle]\beta\rangle && (\text{EqAx6}) \\
&\quad\leq \neg\langle\gamma \neq \downarrow[\psi]\alpha\rangle \wedge \langle\gamma \neq \downarrow[\psi]\alpha\rangle && (\text{NeqAx8 \& }\mathbf{Der21}\text{ (Fact 5)}) \\
&\quad\equiv \text{FALSE} && (\text{Boolean})
\end{aligned}
$$

which is a contradiction. Then $\langle\alpha = \beta\rangle$ is a conjunct of ψ. For the case in which $*$ is $\neq$, the proof is similar but using NeqAx9 instead of NeqAx8. □

Sketch 53. Thinking in terms of sequences as in the proofs of Facts 47 and 49, one only has to observe that:

- $[x]_{\pi_2^{\mathbf{v_1}}} \neq [z]_{\pi_2^{\mathbf{v_1}}}$ for all $z \in T_2^{\mathbf{v_1}}$ in a class that was glued to the class of the root via a root$_{=,\neg\neq}$-kind gluing.
- root$_{=,\neq}$-kind gluings are made in different subtrees.
- By the same arguments given in the proofs of Facts 47 and 49, we can't have a sequence containing any of the following:

  – root$_{=,\neq}$-Z, – Z-root$_{=,\neq}$, – root$_{=,\neg\neq}$-Z,
  – Z-root$_{=,\neg\neq}$, – root$_{=,\neq}$-U_2, – U_2-root$_{=,\neq}$,
  – root$_{=,\neg\neq}$-U_2, – U_2-root$_{=,\neg\neq}$.

Sketch 54. Thinking in terms of sequences as in the proofs of Facts 47 and 49, one only has to observe that:

- $[x]_{\pi^{\mathbf{v_3}}} \neq [y]_{\pi^{\mathbf{v_3}}}$ for all $y \in T^{\mathbf{v_3}}$ in a class that was glued to the class of the root via a root$_{=,\neg\neq}$-kind gluing (Use Lemma 36).
- $[z]_{\pi^{\mathbf{v}'_3}} \neq [y]_{\pi^{\mathbf{v}'_3}}$ for all $y \in T^{\mathbf{v}'_3}$ in a class that was glued to the class of the root via a root$_{=,\neg\neq}$-kind gluing (Use Lemma 36).
- root$_{=,\neq}$-kind gluings are made in different subtrees.
- $[x]_{\pi^{\mathbf{v_3}}}$ and $[z]_{\pi^{\mathbf{v}'_3}}$ can not be glued together by a sequence of all Z-kind gluings because of the consistency of φ plus Lemmas 45 and 42.
- $[x]_{\pi^{\mathbf{v_3}}}$ and $[z]_{\pi^{\mathbf{v}'_3}}$ can not be glued together by a sequence that begins or ends with a U_2-kind gluing because we use new subtrees for that kind of gluings.
- By the same arguments given in the proofs of Facts 47 and 49, we can't have a sequence containing any of the following:

  – root$_{=,\neq}$-Z, – Z-root$_{=,\neq}$, – root$_{=,\neg\neq}$-Z,
  – Z-root$_{=,\neg\neq}$, – root$_{=,\neq}$-U_2, – U_2-root$_{=,\neq}$,
  – root$_{=,\neg\neq}$-U_2, – U_2-root$_{=,\neg\neq}$.

- By the same arguments given in the proof of Fact 49, we can reduce sequences with two consecutive Z-kind gluings to sequences that not have two consecutive Z-kind gluings.
- By the same arguments given in the proof of Fact 49, we can't have sequences containing U_2-U_2.
- One can prove that $[x]_{\pi^{\mathbf{v_3}}}$ and $[z]_{\pi^{\mathbf{v}'_3}}$ are not glued together by a sequence that alternates Z-kind gluings and U_2-kind gluings (starting and ending with Z) by induction with arguments similar to the ones used in Lemma 50.

Sketch 55. Thinking in terms of sequences as in the proofs of Facts 47 and 49, one only has to observe that:

- $[x]_{\pi^{\mathbf{v_3}}} \neq [y]_{\pi^{\mathbf{v_3}}}$ for all $y \in T^{\mathbf{v_3}}$ in a class that was glued to the class of the root via a root$_{=,\neg\neq}$-kind gluing (Use Lemma 36).
- $[z]_{\pi_1^{\mathbf{u}}} \neq [y]_{\pi_1^{\mathbf{u}}}$ for all $y \in T_1^{\mathbf{u}}$ in a class that was glued to the class of the root via a root$_{=,\neg\neq}$-kind gluing (Use Lemma 36).
- root$_{=,\neq}$-kind gluings are made in different subtrees.
- $[x]_{\pi^{\mathbf{v_3}}}$ and $[z]_{\pi_1^{\mathbf{u}}}$ can not be glued together by a sequence that begins with a U_2-kind gluing because we use new subtrees for that kind of gluings.
- $[x]_{\pi^{\mathbf{v_3}}}$ and $[z]_{\pi_1^{\mathbf{u}}}$ can not be glued together by a sequence that begins with a Z-kind gluing because of the consistency of φ plus NeqAx7 and Lemma 30.

Sketch 56. Thinking in terms of sequences as in the proofs of Facts 47 and 49, one only has to observe that:

- $[x]_{\pi^\psi} \neq [y]_{\pi^\psi}$ for all $y \in T^\psi$ in a class that was glued to the class of the root via a $\text{root}_{=,\neg\neq}$-kind gluing (Use Lemma 36).
- $[x]_{\pi^\psi} \neq [y]_{\pi^\psi}$ for all $y \in T^\psi$ in a class that was glued to the class of the root via a $\text{root}_{=,\neq}$-kind gluing (Rule 1).
- By the same arguments given in the proofs of Facts 47 and 49, we can't have a sequence containing any of the following:

– $\text{root}_{=,\neq}$-Z,	– Z-$\text{root}_{=,\neq}$,	– $\text{root}_{=,\neg\neq}$-Z,
– Z-$\text{root}_{=,\neg\neq}$,	– $\text{root}_{=,\neq}$-U_2,	– U_2-$\text{root}_{=,\neq}$,
– $\text{root}_{=,\neg\neq}$-U_2,	– U_2-$\text{root}_{=,\neg\neq}$.	

Sketch 57. Thinking in terms of sequences as in the proofs of Facts 47 and 49, one only has to observe that:

- In case $\psi = \rho$, by consistency of φ plus EqAx5 and **Der21** of Fact 5, $\neg\langle\alpha = \beta\rangle$ is a conjunct of ψ.
- $[x]_{\pi^\psi}$ and $[y]_{\pi^\rho}$ can not be glued together by a sequence of all Z-kind gluings because of the consistency of φ plus Lemmas 45 and 42.
- By Lemma 36 plus construction of $\mathcal{T}^\varphi$, $[y]_{\pi^\rho} \neq [z]_{\pi^\rho}$ for all $z \in T^\rho$ in a class that was glued to the root via a $\text{root}_{=,\neg\neq}$-kind gluing.
- By Rule 1, $[y]_{\pi^\rho} \neq [z]_{\pi^\rho}$ for all $z \in T^\rho$ in a class that was glued to the root via a $\text{root}_{=,\neq}$-kind gluing.
- By the same arguments given in the proofs of Facts 47 and 49, we can't have a sequence containing any of the following:

– $\text{root}_{=,\neq}$-Z,	– Z-$\text{root}_{=,\neq}$,	– $\text{root}_{=,\neg\neq}$-Z,
– Z-$\text{root}_{=,\neg\neq}$,	– $\text{root}_{=,\neq}$-U_2,	– U_2-$\text{root}_{=,\neq}$,
– $\text{root}_{=,\neg\neq}$-U_2,	– U_2-$\text{root}_{=,\neg\neq}$.	

- By the same arguments given in the proof of Fact 49, we can reduce sequences with two consecutive Z-kind gluings to sequences that not have two consecutive Z-kind gluings.
- By the same arguments given in the proof of Fact 49, we can't have sequences containing U_2-U_2.
- One can prove by induction that $[x]_{\pi^\psi}$ and $[y]_{\pi^\rho}$ are not glued together by a sequence that alternates Z-kind gluings and U_2-kind gluings (neither starting with Z or with U_2) with arguments similar to the ones used in Lemma 50.

(Notation: For $\psi, \rho \in N_n$, we use the notation $\mathcal{T}^\psi = (T^\psi, \pi^\psi), \mathcal{T}^\rho = (T^\rho, \pi^\rho)$ with roots r^ψ, r^ρ respectively to denote *any* tree in which ψ, ρ, respectively, are satisfiable.)

References

[1] S. Abriola, P. Barceló, D. Figueira, S. Figueira, Bisimulations on data graphs, in: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR, 2016, pp. 309–318.
[2] S. Abriola, M.E. Descotte, S. Figueira, Model theory of XPath on data trees. Part II: binary bisimulation and definability, Inf. Comput. (2017), http://dx.doi.org/10.1016/j.ic.2017.01.002.
[3] S. Abriola, M.E. Descotte, S. Figueira, Definability for downward and vertical XPath on data trees, in: 21th Workshop on Logic, Language, Information and Computation, in: Lecture Notes in Computer Science, vol. 6642, 2014, pp. 20–34.
[4] S.N. Artëmov, V. Krupski, Data storage interpretation of labeled modal logic, Ann. Pure Appl. Log. 78 (1–3) (1996) 57–71.
[5] D. Baelde, S. Lunel, S. Schmitz, A sequent calculus for a modal logic on finite data trees, in: 25th EACSL Annual Conference on Computer Science Logic, CSL 2016, August 29–September 1, 2016, Marseille, France, 2016, pp. 32:1–32:16.
[6] C. Baier, J. Katoen, Principles of Model Checking, MIT Press, 2008.
[7] M. Benedikt, W. Fan, G.M. Kuper, Structural properties of XPath fragments, Theor. Comput. Sci. 336 (1) (2005) 3–31.
[8] M. Bojańczyk, A. Muscholl, T. Schwentick, L. Segoufin, Two-variable logic on data trees and XML reasoning, J. ACM 56 (3) (2009) 1–48.
[9] L. Cardelli, G. Ghelli, TQL: a query language for semistructured data based on the ambient logic, Math. Struct. Comput. Sci. 14 (3) (2004) 285–327.
[10] J. Clark, S. DeRose, XML path language (XPath), Website, W3C Recommendation, http://www.w3.org/TR/xpath, 1999.
[11] D. Figueira, Decidability of downward XPath, ACM Trans. Comput. Log. 13 (4) (2012) 34.
[12] D. Figueira, S. Figueira, C. Areces, Basic model theory of XPath on data trees, in: International Conference on Database Theory, 2014, pp. 50–60.
[13] D. Figueira, S. Figueira, C. Areces, Model theory of XPath on data trees. Part I: bisimulation and characterization, J. Artif. Intell. Res. 53 (2015) 271–314.
[14] K. Fine, Normal forms in modal logic, Notre Dame J. Form. Log. 16 (2) (1975) 229–237.
[15] F. Geerts, W. Fan, Satisfiability of XPath queries with sibling axes, in: Database Programming Languages, 10th International Symposium, DBPL 2005, Trondheim, Norway, August 28–29, 2005, 2005, pp. 122–137, Revised selected papers.
[16] G. Gottlob, C. Koch, R. Pichler, Efficient algorithms for processing XPath queries, ACM Trans. Database Syst. 30 (2) (2005) 444–491.
[17] D. Harel, D. Kozen, J. Tiuryn, Dynamic Logic, MIT Press, Cambridge, MA, 2000.
[18] E.V. Huntington, Boolean algebra. A correction to: "New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's Principia mathematica" [Trans. Amer. Math. Soc. 35 (1933), no. 1, 274–304; 1501684], Trans. Am. Math. Soc. 35 (2) (1933) 557–558.
[19] E.V. Huntington, New sets of independent postulates for the algebra of logic, with special reference to Whitehead and Russell's Principia Mathematica, Trans. Am. Math. Soc. 35 (1) (1933) 274–304.
[20] L. Libkin, D. Vrgoč, Regular path queries on graphs with data, in: International Conference on Database Theory, 2012, pp. 74–85.
[21] B. ten Cate, G. Fontaine, T. Litak, Some modal aspects of XPath, J. Appl. Non-Class. Log. 20 (3) (2010) 139–171.
[22] B. ten Cate, T. Litak, M. Marx, Complete axiomatizations for XPath fragments, J. Appl. Log. 8 (2) (2010) 153–172.
[23] B. ten Cate, M. Marx, Axiomatizing the logical core of XPath 2.0, Theory Comput. Syst. 44 (4) (2009) 561–589.