

Defeasible Reasoning via Datalog[⊃]

MICHAEL J. MAHER

Reasoning Research Institute

Canberra, Australia

(e-mail: michael.maher@reasoning.org.au)

submitted 08 March 2021; revised 19 August 2021; accepted 25 August 2021

Abstract

We address the problem of compiling defeasible theories to Datalog[⊃] programs. We prove the correctness of this compilation, for the defeasible logic $DL(\partial_{||})$, but the techniques we use apply to many other defeasible logics. Structural properties of $DL(\partial_{||})$ are identified that support efficient implementation and/or approximation of the conclusions of defeasible theories in the logic, compared with other defeasible logics. We also use previously well-studied structural properties of logic programs to adapt to incomplete Datalog[⊃] implementations.

KEYWORDS: defeasible logic, program transformation, well-founded semantics, meta-programming

1 Introduction

A problem faced by defeasible logics – among other logical languages – is that changing hardware and software architectures are not reflected in implementations. Hardware architectures can range from the use of GPUs and other hardware accelerators, through multicore multithreaded architectures, to shared-nothing cloud computing. Causes for failure to exploit these architectures include lack of expertise in the architectural features, lack of manpower more generally, and difficulty in updating legacy systems. Such problems can be ameliorated by mapping a logic to logic programming as an intermediate language.

This is a common strategy in the implementation of defeasible logics. The first implementation of a defeasible logic, d-Prolog, was implemented as a Prolog meta-interpreter (Covington *et al.* 1997). Courteous Logic Programs (Grosz 1997) and its successors LPDA (Wan *et al.* 2009), Rulelog (Grosz and Kifer 2013), Flora2 (Kifer *et al.* 2018), are implemented in XSB (Swift and Warren 2012).¹ The advantages of this approach are that: (1) the target language is a high-level language with many features of the logic to be implemented; (2) the problem of optimizing the implementation to take advantage of the architecture of the underlying hardware is delegated to the implementation of logic programming; (3) flexibility and portability are consequently enhanced, since there are

¹ Not all implementations of defeasible logics use this approach. Delores (Maher *et al.* 2001) and SPINdle (Lam and Governatori 2009) are implemented in imperative languages, Phobos (Billington and Rock 2001) and Deimos (Antonioni *et al.* 2000) are implemented in Haskell, while DR-DEVICE (Bassiliades *et al.* 2006) and Situated Courteous Logic Programs (Gandhe *et al.* 2002) are built on rule systems.

multiple implementations of logic programming, based on differing architectures; and (4) the design and optimization of a specific logic can proceed at a higher level of abstraction.

In this paper, we address a defeasible logic designed for large scale reasoning (Maher *et al.* 2020) and the compilation of function-free defeasible theories into logic programming – more specifically, Datalog[⊥] under the well-founded semantics (Van Gelder *et al.* 1991). There is a multitude of implementations of variations of Datalog, coming from different motivations, not all suitable for implementing a defeasible logic, and not all supporting traditional syntax. In addition to implementations developed in the database (Bishop and Fischer 2008; Wu *et al.* 2014; Wang *et al.* 2015; Aref *et al.* 2015; Brass and Stephan 2017; Condie *et al.* 2018; Fan *et al.* 2019; Alvaro *et al.* 2010; Cognitect; Marz 2013) and logic programming (Swift and Warren 2012; Costa *et al.* 2012; Cat *et al.* 2018; Gebser *et al.* 2019; Adrian *et al.* 2018; Nguyen *et al.* 2018; Wenzel and Brass 2019; Martinez-Angeles *et al.* 2013; Tachmazidis *et al.* 2014; Tachmazidis *et al.* 2014) communities, implementations have been developed to service the programming language analysis (Jordan *et al.* 2016; Madsen *et al.* 2016; Avgustinov *et al.* 2016; Hoder *et al.* 2011; Whaley *et al.* 2005; Bembenek *et al.* 2020; Madsen and Lhoták 2020), graph processing (Seo *et al.* 2015; Wang *et al.* 2017; Aberger *et al.* 2017), and artificial intelligence (Eisner and Filardo 2010; Filardo 2017; Bellomarini *et al.* 2018; Carral *et al.* 2019; Chin *et al.* 2015) communities. These implementations address a wide range of architectures, and many provide extension beyond traditional Datalog.

However, the number of currently available implementations that provide complete support for the well-founded semantics of Datalog[⊥] is quite few. This leads us to use structural properties of the compiled program to establish when an incomplete implementation of Datalog[⊥] can be used to provide a complete implementation of a defeasible theory for the defeasible logic. Specifically, we establish how properties of the initial defeasible theory and properties of the defeasible logic are reflected in the compiled program and combine to support complete execution of the compiled program using incomplete implementations of Datalog[⊥] (with respect to the well-founded semantics). We also identify methods to obtain sound approximations to the conclusions of the defeasible theory using sound but incomplete Datalog[⊥] implementations.

The compilation builds on existing work. We represent the scalable defeasible logic as a metaprogram, in the style of Antoniou *et al.* (2000). We then use a series of unfold and fold transformations (Tamaki and Sato 1984) to convert the metaprogram, applied to the defeasible theory, to a specialized logic program. The result is encapsulated as a mapping from defeasible theories to Datalog[⊥] programs, which is established as correct as a consequence of the correctness of the individual transformations. The size of the resulting program is linear in the size of the defeasible theory.

The remainder of the paper is structured as follows. Sections 2 and 3 introduce the necessary concepts from logic programming and defeasible reasoning. Section 4 defines the defeasible logic $DL(\partial_{||})$ (Maher *et al.* 2020), while Section 5 defines the corresponding metaprogram and establishes some of its properties. Section 6 compiles the metaprogram to a simpler form using fold/unfold transformations, while Section 7 establishes the correctness of the metaprogram with respect to the original, proof-theoretic definition of $DL(\partial_{||})$. Structural properties of the compiled program are established in Section 8. Section 9 then identifies how these properties can be used to establish the correctness of implementations and approximations using incomplete Datalog[⊥] systems.

2 Logic programming

We introduce the elements of logic programming that we will need. The first part defines notation and terminology for syntactic aspects of logic programs, including dependency relations. The second part defines the semantics of logic programs we will use.

2.1 Syntax and structure of logic programs

Let Π be a set of predicate symbols, Σ be a set of function symbols, and \mathcal{V} be a set of variables. Each symbol has an associated arity greater or equal to 0. A function symbol of arity 0 is called a *constant*, while a predicate of arity 0 is called a *proposition*. The *terms* are constructed inductively in the usual way: any variable or constant is a term; if $f \in \Sigma$ has arity n and t_1, \dots, t_n are terms then $f(t_1, \dots, t_n)$ is a term; all terms can be constructed in this way. An *atom* is constructed by applying a predicate $p \in \Pi$ of arity n to n terms. A *literal* is either an atom or a negated atom *not* A , where A is an atom.

A *logic program* is a collection of *clauses* of the form

$$A :- B_1, \dots, B_m, \text{not } C_1, \dots, \text{not } C_n,$$

where $A, B_1, \dots, B_m, C_1, \dots, C_n$ are atoms ($m \geq 0, n \geq 0$). The positive literals and the negative literals are grouped separately purely for notational convenience. A is called the *head* of the clause and the remaining literals form the *body*. The set of all clauses with predicate symbol p in the head are said to be the clauses defining p . We use *ground* as a synonym for variable-free. The set of all variable-free instances of clauses in a logic program P is denoted by $\text{ground}(P)$. $\text{ground}(P)$ can be considered a propositional logic program, but it is generally infinite. For the semantics we are interested in, P and $\text{ground}(P)$ are equivalent with respect to inference of ground literals.

A logic program P is *range-restricted* if every variable in the head of a clause also appears in a positive body literal. P is *negation-safe* if, for every clause, every variable in a negative body literal also appears in a positive body literal. P is *safe* (or *allowed*) if every variable in a clause also appears in a positive body literal of that clause. Equivalently, P is safe if it is range-restricted and negation-safe. Safety is a property that ensures domain independence (Abiteboul *et al.* 1995), but also it can simplify the execution of a logic program. Range-restriction guarantees that all inferred atoms are ground; data structures and algorithms do not need to address general atoms. Safety ensures, in addition, that if negative literals are evaluated only after the positive part of the clause then, again, only ground atoms must be treated. These points apply to both top-down and bottom-up execution.

A *Datalog⁻ program* is a logic program where the only terms are constants and variables.

We present some notions of dependence among predicates that are derived purely from the syntactic structure of a logic program P . We follow the definitions and notation of Kunen (1989). p, q and r range over predicates. We define $p \sqsupseteq_{+1} q$ if p appears in the head of a rule and q is the predicate of a positive literal in the body of that rule. $p \sqsupseteq_{-1} q$ if p appears in the head of a rule and q is the predicate of a negative literal in the body of that rule.

$p \sqsupseteq q$ iff $p \sqsupseteq_{+1} q$ or $p \sqsupseteq_{-1} q$. We say p *directly depends* on q . \geq is the transitive closure of \sqsupseteq . If $p \geq q$ we say p *depends* on q . $p \approx q$ iff $p \geq q$ and $q \geq p$, expressing that p and q are

mutually recursive. $p > q$ iff $p \geq q$ and not $q \geq p$. If, for all p in P , $q \geq p$, then we say q is a \geq -largest predicate. \geq_{+1} and \geq_{-1} are defined inductively as the least relations such that $p \geq_{+1} p$, and $p \sqsupseteq_i q$ and $q \geq_j r$ implies $p \geq_{i \cdot j} r$, where $i \cdot j$ denotes multiplication of i and j . Essentially \geq_{+1} denotes a relation of dependence through an even number of negations and \geq_{-1} denotes dependence through an odd number of negations. As is usual, we will write $p \leq q$ when $q \geq p$, and similarly for the other relations. \geq_0 denotes the transitive closure of \sqsupseteq_{+1} .

A program P is *stratified* if for no predicates p and q in Π does $p \approx q$ and $p \geq_{-1} q$. Let a *predicate-consistent mapping* be a function that maps atoms to the non-negative integers such that, for every predicate, all atoms involving that predicate are mapped to the same value. Then, alternatively, P is stratified if there is a predicate-consistent mapping m that, for every clause like the one above, $m(A) \geq m(B_i)$, for $1 \leq i \leq n$ and $m(A) > m(C_j)$, for $1 \leq j \leq m$. The i^{th} *stratum* P_i is the set of clauses in P with head predicate p such that $m(p) = i$. P is *call-consistent* if, for no predicate p , does $p \geq_{-1} p$; that is no predicate depends negatively on itself. Clearly, any stratified program is call-consistent. P is *hierarchical* if for no predicates p and q , does $p \sqsupseteq q$ and $q \geq p$, that is, no predicate symbol depends on itself. Equivalently, P is hierarchical if there is a predicate-consistent mapping m that, for every clause like the one above, $m(A) > m(B_i)$ and $m(A) > m(C_j)$. Every hierarchical program is stratified.

A set $\mathcal{P} \subseteq \Pi$ of predicates in a program P is *downward-closed* if whenever $p \in \mathcal{P}$ and $q \leq p$ then $q \in \mathcal{P}$. \mathcal{P} is *downward-closed with floor \mathcal{F}* if $\mathcal{F} \subset \mathcal{P}$ and both \mathcal{P} and \mathcal{F} are downward-closed. A *signing* for \mathcal{P} and P is a function s that maps Π to $\{-1, +1\}$ such that, for $p, q \in \mathcal{P}$, $p \leq_i q$ implies $s(p) = s(q) \cdot i$. A signing is extended to atoms by defining $s(p(\vec{a})) = s(p)$. For any signing s for a set of predicates \mathcal{P} , there is an inverted signing \bar{s} defined by $\bar{s}(p) = -s(p)$. s and \bar{s} are equivalent in the sense that they divide \mathcal{P} into the same two sets. Obviously, $\bar{\bar{s}} = s$. A program P is said to be *strict* if no predicate p depends both positively and negatively on a predicate q , that is, we never have $p \geq_{+1} q$ and $p \geq_{-1} q$. Let all the predicates of P be contained in \mathcal{P} . If \mathcal{P} has a signing then P is strict; if P has a \geq -largest predicate and P is strict, then \mathcal{P} has a signing (Kunen 1989).

A similar set of dependencies over ground atoms can be defined by applying these definitions to $ground(P)$.

Given a program P , an infinite sequence of atoms $\{q_i(\vec{a}_i)\}$ is *unfounded* wrt a set of predicates \mathcal{Q} if, for every i , $q_i \sqsupseteq_{+1} q_{i+1}$ and $q_i \in \mathcal{Q}$. We say a predicate p *avoids negative unfoundedness* wrt a signing s on \mathcal{Q} if for every negatively signed predicate q on which p depends, no q -atom starts an unfounded sequence wrt \mathcal{Q} .

2.2 Semantics of logic programs

We define the semantics of interest in this paper and identify important relationships between them.

A *3-valued Herbrand interpretation* is a mapping from ground atoms to one of three truth values: **true**, **false**, and **unknown**. This mapping can be extended to all formulas using Kleene’s 3-valued logic.

Kleene’s truth tables can be summarized as follows. If ϕ is a boolean combination of the atoms **true**, **false**, and **unknown**, its truth value is **true** iff all the possible ways of putting in **true** or **false** for the various occurrences of **unknown** lead to a value **true**

being computed in ordinary 2-valued logic: ϕ gets the value **false** iff $\neg\phi$ gets the value **true**, and ϕ gets the value **unknown** otherwise. These truth values can be extended in the obvious way to predicate logic, thinking of the quantifiers as infinite disjunction or conjunction.

Equivalently, a 3-valued Herbrand interpretation I can be represented as the set of literals $\{a \mid I(a) = \mathbf{true}\} \cup \{\text{not } a \mid I(a) = \mathbf{false}\}$. This representation is used in the following definitions. The interpretations are ordered by the subset ordering on this representation.

Some semantics are defined in terms of fixedpoints of monotonic functions over a partial order. A function F is *monotonic* if $x \leq y$ implies $F(x) \leq F(y)$. A *fixedpoint* of F is a value a such that $F(a) = a$. When F is monotonic on a complete semi-lattice there is a least (under the \leq ordering) fixedpoint. We use $\text{lfp}(F)$ to denote the least fixedpoint of F . When a is an element of the partial order, $\text{lfp}(F, a)$ denotes the least fixedpoint greater than (or equal to) a .

Fitting (1985) defined a semantics for a logic program P in terms of a function Φ_P mapping 3-valued interpretations, which we define as follows.

$$\begin{aligned}\Phi_P(I) &= \Phi_P^+(I) \cup \neg \Phi_P^-(I) \\ \Phi_P^+(I) &= \{a \mid \text{there is a rule } a :- B \text{ in } \text{ground}(P) \text{ where } I(B) = \mathbf{true}\} \\ \Phi_P^-(I) &= \{a \mid \text{for every rule } a :- B \text{ in } \text{ground}(P) \text{ with head } a, I(B) = \mathbf{false}\}\end{aligned}$$

where $\neg S$ denotes the set $\{\text{not } s \mid s \in S\}$.

Fitting's semantics associates with P the least fixedpoint of Φ_P , $\text{lfp}(\Phi_P)$. This is the least 3-valued Herbrand model of the Clark completion P^* of P . Thus, the conclusions justified under this semantics are those formulas that evaluate to **true** under all 3-valued Herbrand models of P^* . **Kunen (1987)** defined a semantics that justifies as conclusions those formulas that evaluate to **true** under all 3-valued models (Herbrand or not) of P^* . He showed that these are exactly the formulas that are consequences of $\Phi_P \uparrow n$ for some finite n .² When P is a finite Datalog⁻ program the two semantics coincide. However, in general, when function symbols are permitted, Kunen's semantics is computable, while other semantics, like Fitting's semantics, the stratified semantics, and the well-founded semantics, are not.

The *stratified semantics* (or *iterated fixedpoint semantics*) (**Apt et al. 1988**) applies only when P is stratified. It is defined in stages, by building up partial models, based on the strata, until a full model is constructed. Each stratum contains essentially a definite clause program, given that all negations refer to lower strata, which have already been defined in the current partial model. Let m be a mapping describing a stratification. Initially the partial model leaves all predicates undefined, and at each stratum, in turn, it is extended to define all the predicates defined in that stratum. On each stratum i , the predicates on lower strata have been defined and, because P is stratified, there is a least partial model extending the current partial model that defines the predicates. This model is then used as the basis for the next stratum. For a more precise description,

² \uparrow is defined inductively: $\Phi_P \uparrow 0 = I_{\mathbf{unknown}}$, where $I_{\mathbf{unknown}}$ is the interpretation that assigns each atom the value **unknown**, and $\Phi_P \uparrow (k+1) = \Phi_P(\Phi_P \uparrow k)$. For limit ordinals α , $\Phi_P \uparrow \alpha = \bigcup_{\beta < \alpha} \Phi_P \uparrow \beta$. The (possibly transfinite) sequence $\Phi_P \uparrow 0, \Phi_P \uparrow 1, \dots$ is called the Kleene sequence for Φ_P .

see the papers (Apt *et al.* 1988; Apt and Bol 1994). The stratified semantics extends Fitting’s semantics, when the program is stratified.

The well-founded semantics (Van Gelder *et al.* 1991) extends Fitting’s semantics by, roughly, considering atoms to be false if they are supported only by a “loop” of atoms. This is based on the notion of unfounded sets.

Given a logic program P and a 3-valued interpretation I , a set A of ground atoms is an *unfounded set with respect to I* iff each atom $a \in A$ satisfies the following condition: For each rule r of $\text{ground}(P)$ whose head is a , (at least) one of the following holds:

1. Some literal in the body evaluates to **false** in I .
2. Some atom in the body occurs in A

The *greatest unfounded set* of P with respect to I (denoted $U_P(I)$) is the union of all the unfounded sets with respect to I . Notice that, if we ignore the second part of the definition of unfounded set wrt I , the definition of unfounded set is the same as the expression inside the definition of $\Phi_P^-(I)$. It follows that $\Phi_P^-(I) \subseteq U_P(I)$, for every I .

The function $\mathcal{W}_P(I)$ is defined by $\mathcal{W}_P = \Phi_P^+(I) \cup \neg U_P(I)$. The *well-founded semantics* of a program P is represented by the least fixedpoint of \mathcal{W}_P . This is a 3-valued Herbrand model of P^* . Because $\Phi_P^-(I) \subseteq U_P(I)$, for every I , we have $\Phi_P(I) \subseteq \mathcal{W}_P(I)$, for every I and hence $\text{lfp}(\Phi_P) \subseteq \text{lfp}(\mathcal{W}_P)$. That is, Fitting’s semantics is weaker than the well-founded semantics. If P is stratified then the well-founded semantics is a 2-valued Herbrand model of P^* and equal to the stratified semantics.

Often, as in this paper, we are mainly interested in the positive literals of predicate(s) p that are consequences of the program, rather than the negative literals of p or the literals of other predicates. In such cases, we can avoid computing parts of the well-founded model (Maher 2021). In particular, given a signing s in which $s(p) = +1$, the positive literals of p depend on only the positive literals of predicates with a positive sign, and the negative literals of predicates with a negative sign. Let WF^{+s} denote the semantics $\{q \mid s(q) = +1, q \in \text{lfp}(\mathcal{W}_P)\} \cup \{\text{not } q \mid s(q) = -1, \text{not } q \in \text{lfp}(\mathcal{W}_P)\}$ which computes only such parts of the well-founded model. This can be computed as the least fixedpoint of a function \mathcal{UU}_P^s (Maher 2021).

For a ground literal q , we define:

- $P \models_{WF} q$ iff $q \in \text{lfp}(\mathcal{W}_P)$
- $P \models_F q$ iff $q \in \text{lfp}(\Phi_P)$
- $P \models_{WF^{+s}} q$ iff $q \in \text{lfp}(\mathcal{UU}_P^s)$

It follows from the discussion above that $P \models_F q$ implies $P \models_{WF} q$, and $P \models_{WF^{+s}} q$ implies $P \models_{WF} q$, for every program P and ground literal q .

Let I be a 3-valued Herbrand interpretation of predicates in a set \mathcal{F} and let X be a semantics based on Herbrand models. Then \models_X^I denotes consequence in the semantics X after all predicates in \mathcal{F} are interpreted according to I . Such a notion is interesting, in general, because some predicates might be defined outside the logic programming setting, or in a different module.

We now summarize two results from the paper (Maher 2021) in the following theorem. In the first part we see that only the positive or only the negative conclusions for each predicate need to be computed. The second part establishes conditions under which the

well-founded semantics and Fitting semantics agree on the truth value of some ground literals (even if they may disagree on other literals).

Theorem 1 (Maher 2021)

Let P be a logic program, $\mathcal{P} \subseteq \Pi$ be a downward-closed set of predicates with floor \mathcal{F} , let \mathcal{Q} be $\mathcal{P} \setminus \mathcal{F}$, and s be a signing for \mathcal{Q} . Let I be a fixed semantics for \mathcal{F} and $p \in \mathcal{Q}$.

1. For any ground atom $p(\vec{a})$:
 - If $s(p) = +1$ then $P \models_{WF}^I p(\vec{a})$ iff $P \models_{WF+s}^I p(\vec{a})$
 - If $s(p) = -1$ then $P \models_{WF}^I \text{not } p(\vec{a})$ iff $P \models_{WF+s}^I \text{not } p(\vec{a})$
2. Suppose, additionally, that p avoids negative unfoundedness wrt s .
 For any ground atom $p(\vec{a})$:
 - If $s(p) = +1$ then $P \models_{WF}^I p(\vec{a})$ iff $P \models_F^I p(\vec{a})$
 - If $s(p) = -1$ then $P \models_{WF}^I \text{not } p(\vec{a})$ iff $P \models_F^I \text{not } p(\vec{a})$

3 Defeasible logics

A *defeasible theory* D is a triple $(F, R, >)$ where F is a finite set of facts (literals), R a finite set of labelled rules, and $>$ a superiority relation (a binary acyclic relation) on R (expressed on the labels), specifying when one rule overrides another, given that both are applicable.

A rule r consists (a) of its antecedent (or body) $A(r)$ which is a finite set of literals, (b) an arrow, and, (c) its consequent (or head) $C(r)$ which is a literal. Rules also have distinct *labels* which are used to refer to the rule in the superiority relation. There are three types of rules: strict rules, defeasible rules and defeaters represented by a respective arrow \rightarrow , \Rightarrow and \rightsquigarrow . Strict rules are rules in the classical sense: whenever the premises are indisputable (e.g., facts) then so is the conclusion. Defeasible rules are rules that can be defeated by contrary evidence. Defeaters are rules that cannot be used to draw any conclusions; their only use is to provide contrary evidence that may prevent some conclusions. We use \leftrightarrow to range over the different kinds of arrows used in a defeasible theory. Given a set R of rules, we denote the set of all strict rules in R by R_s , and the set of strict and defeasible rules in R by R_{sd} . $R[q]$ denotes the set of rules in R with consequent q . A literal is a possibly negated (using \neg) predicate symbol applied to a sequence of variables and constants. If q is a literal, $\sim q$ denotes the complementary literal (if q is a positive literal p then $\sim q$ is $\neg p$; and if q is $\neg p$, then $\sim q$ is p).

We will focus on defeasible theories such that any variable in the head of a rule also occurs in the body, and that every fact is variable-free, a property we call *range-restricted* in analogy to the same property in logic programs. When no rule or fact contains a function symbol, except for constants, we say the defeasible theory is *function-free*. Given a fixed finite set of constants in a function-free defeasible theory, any rule is equivalent to a finite set of variable-free rules, and any defeasible theory D is equivalent to a variable-free defeasible theory $ground(D)$, for the purpose of semantical analysis. We refer to variable-free defeasible theories, etc. as *propositional*, since there is only a syntactic difference between such theories and true propositional defeasible theories. Consequently, we will formulate definitions and semantical analysis in propositional terms. However, for computational analyses and implementation we will also address defeasible theories that are not propositional.

A defeasible theory is *hierarchical* (or *acyclic* or *stratified*) if there is a predicate-consistent mapping m which maps atoms to the non-negative integers such that, for every rule, the head is mapped to a greater value than any body atom. That is, there is no recursion in the rules of the defeasible theory, not even through a literal's complement. A defeasible theory D is *locally hierarchical* if $ground(D)$ is hierarchical, where we treat each variable-free atom as a proposition/0-ary predicate symbol.

Example 2

To demonstrate defeasible theories, we consider the familiar Tweety problem and its representation as a defeasible theory. The defeasible theory D consists of the rules and facts

$$\begin{aligned}
 r_1 : \quad & \text{bird}(X) \Rightarrow \text{fly}(X) \\
 r_2 : \quad & \text{penguin}(X) \Rightarrow \neg\text{fly}(X) \\
 r_3 : \quad & \text{penguin}(X) \rightarrow \text{bird}(X) \\
 r_4 : \quad & \text{injured}(X) \rightsquigarrow \neg\text{fly}(X) \\
 f : \quad & \text{penguin}(\text{tweety}) \\
 g : \quad & \text{bird}(\text{freddie}) \\
 h : \quad & \text{injured}(\text{freddie})
 \end{aligned}$$

and a priority relation $r_2 > r_1$.

Here $r_1, r_2, r_3, r_4, f, g, h$ are labels and r_3 is (a reference to) a strict rule, while r_1 and r_2 are defeasible rules, r_4 is a defeater, and f, g, h are facts. Thus $F = \{f, g, h\}$, $R_s = \{r_3\}$, $R_{sd} = \{r_1, r_2, r_3\}$ and $R = \{r_1, r_2, r_3, r_4\}$ and $>$ consists of the single tuple (r_2, r_1) . The rules express that birds usually fly (r_1), penguins usually dont fly (r_2), that all penguins are birds (r_3), and that an injured animal may not be able to fly (r_4). In addition, the priority of r_2 over r_1 expresses that when something is both a bird and a penguin (that is, when both rules can fire) it usually cannot fly (that is, only r_2 may fire, it overrules r_1). Finally, we are given the facts that *tweety* is a penguin, and *freddie* is an injured bird.

This defeasible theory is hierarchical. One function that demonstrates this maps *injured* and *penguin* to 0, *bird* to 1, and *fly* to 2.

A *conclusion* takes the forms $+dq$ or $-dq$, where q is a literal and d is a tag indicating which inference rules were used. Given a defeasible theory D , $+dq$ expresses that q can be proved via inference rule d from D , while $-dq$ expresses that it can be established that q cannot be proved from D .

For example, in the paper by Antoniou *et al.* (2001), a defeasible logic, now called $DL(\partial)$, is defined with the following inference rules, phrased as conditions on proofs³

- | | |
|---|--|
| $+\Delta$) If $P(i + 1) = +\Delta q$ then either
(1) $q \in F$; or
(2) $\exists r \in R_s[q] \forall a \in A(r), +\Delta a \in P[1..i]$. | $-\Delta$) If $P(i + 1) = -\Delta q$ then
(1) $q \notin F$, and
(2) $\forall r \in R_s[q] \exists a \in A(r), -\Delta a \in P[1..i]$. |
|---|--|

These two inference rules concern reasoning about definitive information, involving only strict rules and facts. They define conventional monotonic inference ($+\Delta$) and provable inability to prove from strict rules and facts ($-\Delta$). For example, $+\Delta$ says that $+\Delta q$

³ Here, D is a defeasible theory $(F, R, >)$, q is a variable-free literal, P denotes a proof (a sequence of conclusions constructed by the inference rules), $P[1..i]$ denotes the first i elements of P , and $P(i)$ denotes the i^{th} element of P .

can be added to a proof P at position $i + 1$ only if $q \in F$ or there is a strict rule r with head q where each literal a in the antecedent $A(r)$ has been proved $(+\Delta a)$ earlier in the proof ($P[1..i]$).

The next inference rules refer to defeasible reasoning.

- | | |
|--|--|
| <p>$+\partial$) If $P(i + 1) = +\partial q$ then either</p> <ul style="list-style-type: none"> (1) $+\Delta q \in P[1..i]$; or (2) The following three conditions all hold. <ul style="list-style-type: none"> (2.1) $\exists r \in R_{sd}[q] \forall a \in A(r),$
$\quad +\partial a \in P[1..i]$, and (2.2) $-\Delta \sim q \in P[1..i]$, and (2.3) $\forall s \in R[\sim q]$ either <ul style="list-style-type: none"> (2.3.1) $\exists a \in A(s), -\partial a \in P[1..i]$; or (2.3.2) $\exists t \in R_{sd}[q]$ such that
$\quad \forall a \in A(t), +\partial a \in P[1..i]$, and
$\quad t > s$. | <p>$-\partial$) If $P(i + 1) = -\partial q$ then</p> <ul style="list-style-type: none"> (1) $-\Delta q \in P[1..i]$, and (2) either <ul style="list-style-type: none"> (2.1) $\forall r \in R_{sd}[q] \exists a \in A(r),$
$\quad -\partial a \in P[1..i]$; or (2.2) $+\Delta \sim q \in P[1..i]$; or (2.3) $\exists s \in R[\sim q]$ such that <ul style="list-style-type: none"> (2.3.1) $\forall a \in A(s), +\partial a \in P[1..i]$,
and (2.3.2) $\forall t \in R_{sd}[q]$ either
$\quad \exists a \in A(t), -\partial a \in P[1..i]$;
or
$\quad \text{not}(t > s)$. |
|--|--|

$+\partial q$ is a *consequence* of a defeasible theory D if there is a proof containing $+\partial q$.

In the $+\partial$ inference rule, (1) ensures that any monotonic consequence is also a defeasible consequence. (2) allows the application of a rule (2.1) with head q , provided that monotonic inference provably cannot prove $\sim q$ (2.2) and every competing rule either provably fails to apply (2.3.1) or is overridden by an applicable rule for q (2.3.2). The $-\partial$ inference rule is the strong negation (Antoniou *et al.* 2000) of the $+\partial$ inference rule. For other properties of this and other defeasible logics, the reader is referred to the paper by Billington *et al.* (2010).

Example 3

The above inference rules make several inferences from the Tweety defeasible theory in Example 2.

The $+\Delta$ inference rule infers $+\Delta penguin(tweety)$, $+\Delta bird(freddie)$, $+\Delta injured(freddie)$ from the facts, and $+\Delta bird(tweety)$ using r_3 . Such inferences are definite conclusions from the theory. The $-\Delta$ inference rule infers, among others $-\Delta penguin(freddie)$, $-\Delta injured(tweety)$, $-\Delta \neg injured(tweety)$, and $-\Delta \neg bird(tweety)$, indicating that the theory is provably unable to come to a definite conclusion about these statements, because there is no rule (and no fact) for these literals. It also infers $-\Delta fly(freddie)$, $-\Delta \neg fly(freddie)$, $-\Delta fly(tweety)$, and $-\Delta \neg fly(tweety)$ because there is no strict rule for fly or $\neg fly$ and consequently (2) of the $-\Delta$ inference rule is vacuously true.

The $+\partial$ inference rule infers $+\partial penguin(tweety)$, $+\partial bird(freddie)$, and $+\partial injured(freddie)$, and $+\partial bird(tweety)$ because these statements are known definitely. It also concludes $+\partial \neg fly(tweety)$ using rule r_2 in (2.1), the previous conclusion $-\Delta fly(tweety)$ in (2.2), and, despite the presence of r_1 as s in (2.3), using r_2 as t in (2.3.2) with the priority statement $r_2 > r_1$ to overrule r_1 . It is unable to similarly conclude $+\partial fly(freddie)$, because of the presence of r_3 and the lack of a priority statement to overrule it.

The $-\partial$ inference rule infers, among others, $-\partial penguin(freddie)$ and $-\partial injured(tweety)$ because these statements are known unprovable definitely (1), and (2.1) is satisfied

vacuously because there is no rule for these predicates. It also infers $-\partial \text{fly}(\text{freddie})$, $-\partial \neg \text{fly}(\text{freddie})$, and $-\partial \text{fly}(\text{tweety})$.

For clarity, we refer to elements of defeasible theories as rules, and elements of logic programs as clauses. Also, note the distinction between rules (syntactic elements of a defeasible theory) and inference rules (criteria for extending proofs). Logic programming predicates will be written in **teletype** font, while defeasible logic predicates will be written in *italics*. We use *not* or alternatively **not** for negation-as-failure in logic programs, and \neg for classical negation in defeasible theories. However, “predicate”, “atom” and “literal” may be used to refer to elements of either a defeasible theory or a logic program.

To avoid the confusion of existing names with names generated during transformations, we need a character that is not used in D . For readability in this paper, we choose the underscore $_$ as this character, but any other character would suffice.

4 The scalable defeasible logic $DL(\partial_{||})$

The defeasible logic $DL(\partial_{||})$ (Maher *et al.* 2020) was designed to allow defeasible inference to be scalable to very large data sets. We present the inference rules of that logic here.

$DL(\partial_{||})$ involves three tags: Δ , which expresses conventional monotonic inference; λ , an auxiliary tag; and $\partial_{||}$, which is the main notion of defeasible proof in this logic. The inference rules are presented below, phrased as conditions on proofs⁴.

- $+\Delta$) If $P(i + 1) = +\Delta q$ then either
 - (1) $q \in F$; or
 - (2) $\exists r \in R_s[q] \forall a \in A(r), +\Delta a \in P[1..i]$.

This inference rule concerns reasoning about definitive information, involving only strict rules and facts. It is identical to the rule for monotonic inference in $DL(\partial)$.

For a defeasible theory D , we define P_Δ to be the set of consequences in the largest proof satisfying the proof condition $+\Delta$, and call this the Δ closure. It contains all $+\Delta$ consequences of D .

Once P_Δ is computed, we can apply the $+\lambda$ inference rule. $+\lambda q$ is intended to mean that q is potentially defeasibly provable in D . The $+\lambda$ inference rule is as follows.

- $+\lambda$: If $P(i + 1) = +\lambda q$ then either
 - (1) $+\Delta q \in P_\Delta$ or
 - (2) (2.1) $\exists r \in R_{sd}[q] \forall \alpha \in A(r) : +\lambda \alpha \in P(1..i)$ and
 (2.2) $+\Delta \sim q \notin P_\Delta$

Using this inference rule, and given P_Δ , we can compute the λ closure P_λ , which contains all $+\lambda$ consequences of D .

⁴ As in the inference rules for $DL(\partial)$ in the previous section, D is a defeasible theory $(F, R, >)$, q is a variable-free literal, P denotes a proof, $P[1..i]$ denotes the first i elements of P , and $P(i)$ denotes the i^{th} element of P .

$+∂_{||}q$ is intended to mean that q is defeasibly provable in D . Once $P_Δ$ and $P_λ$ are computed, we can apply the $+∂_{||}$ inference rule.

$+∂_{||}$: If $P(i + 1) = +∂_{||}q$ then either

- (1) $+Δq ∈ P_Δ$ or
- (2) (2.1) $∃r ∈ R_{sd}[q] ∃α ∈ A(r) : +∂_{||}α ∈ P(1..i)$ and
 - (2.2) $+Δ∼q ∉ P_Δ$ and
 - (2.3) $∃s ∈ R[∼q]$ either
 - (2.3.1) $∃α ∈ A(s) : +λα ∉ P_λ$ or
 - (2.3.2) $∃t ∈ R_{sd}[q]$ such that
 - $∃α ∈ A(t) : +∂_{||}α ∈ P(1..i)$ and $t > s$

The $∂_{||}$ closure $P_{∂_{||}}$ contains all $∂_{||}$ consequences of D . We say a set of tagged literals Q is $∂_{||}$ -deductively closed if, given closures $P_Δ$ and $P_λ$ as defined above, for every literal q that may be appended to $Q ∪ P_Δ ∪ P_λ$ by the inference rule $+∂_{||}$ (treating $Q ∪ P_Δ ∪ P_λ$ as a proof), $q ∈ Q$. Clearly, $P_{∂_{||}}$ is the smallest $∂_{||}$ -deductively closed set.

Example 4

We now apply these inference rules to the Tweety defeasible theory in Example 2.

As before, the $+Δ$ inference rule infers $+Δ penguin(tweety)$, $+Δ bird(freddie)$, and $+Δ injured(freddie)$ from the facts, and $+Δ bird(tweety)$ using r_3 . We have no need of the $-Δ$ inference rule.

Using the $+λ$ inference rule we infer all the literals inferred by the $+Δ$ inference rule as $+λ$ conclusions. In addition, the rule infers $+λ fly(tweety)$, $+λ fly(freddie)$, and $+λ ¬fly(tweety)$.

Using the $+∂_{||}$ inference rule, again all the $+Δ$ conclusions are inferred as $+∂_{||}$ conclusions. The only other conclusion that can be drawn with this rule is $+∂_{||} ¬fly(tweety)$. The potential inference of $fly(tweety)$ is overruled by r_2 inferring $+∂_{||} ¬fly(tweety)$. On the other hand, a potential inference of $fly(freddie)$ is not obtained because r_1 cannot overrule r_4 .

Inference rules $∂$ and $∂_{||}$ employ the notion of “team defeat”, where it doesn’t matter which rule overrides an opposing rule, as long as all opposing rules are overridden. This is expressed in (2.3.2). We can also have a version of $∂_{||}$ with “individual defeat”, where all opposing rules must be overridden by the same rule, which we denote by $∂_{||}^*$. The inference rule for $+∂_{||}^*$ replaces (2.3.2) in $∂_{||}$ by $r > s$.

$+∂_{||}^*$: If $P(i + 1) = +∂_{||}^*q$ then either

- (1) $+Δq ∈ P_Δ$ or
- (2) (2.1) $∃r ∈ R_{sd}[q] ∃α ∈ A(r) : +∂_{||}^*α ∈ P(1..i)$ and
 - (2.2) $+Δ∼q ∉ P_Δ$ and
 - (2.3) $∃s ∈ R[∼q]$ either
 - (2.3.1) $∃α ∈ A(s) : +λα ∉ P_λ$ or
 - (2.3.2) $r > s$

Example 4 does not display the distinction between $∂_{||}$ and $∂_{||}^*$: both have the same consequences. The distinction is visible when there are multiple applicable rules for both some literal q and its negation $∼q$. The following example originates from the paper by Antoniou *et al.* (2001).

Example 5

Consider some rules of thumb about animals and, particularly, mammals. An egg-laying animal is generally not a mammal. Similarly, an animal with webbed feet is generally not a mammal. On the other hand, an animal with fur is generally a mammal. Finally, the monotremes are a subclass of mammal. These rules are represented as defeasible rules below.

Furthermore, animals with fur and webbed feet are generally mammals, so r_2 should overrule r_4 . And monotremes are a class of egg-laying mammals, so r_1 should overrule r_3 .

Finally, it happens that a platypus is a furry, egg-laying, web-footed monotreme. Is it a mammal? (That is, is $mammal(platypus)$ a consequence of the defeasible theory below?)

$r_1 : monotreme(X) \Rightarrow mammal(X)$	$r_3 : laysEggs(X) \Rightarrow \neg mammal(X)$
$r_2 : hasFur(X) \Rightarrow mammal(X)$	$r_4 : webFooted(X) \Rightarrow \neg mammal(X)$
$r_1 > r_3$	$r_2 > r_4$
$monotreme(platypus)$	$laysEggs(platypus)$
$hasFur(platypus)$	$webFooted(platypus)$

It is obvious that all four rules are applicable to the question of $mammal(platypus)$. Under team defeat, each rule for $\neg mammal(platypus)$ is overcome by some rule for $mammal(platypus)$, so $mammal(platypus)$ is inferred (which is zoologically correct). However, there is no single rule for $mammal(platypus)$ that overcomes all rules for $mammal(platypus)$, so under individual defeat we cannot infer $mammal(platypus)$ (nor $\neg mammal(platypus)$).

These logics are amenable to the techniques used to establish model-theoretic (Maher 2002) and argumentation (Governatori et al. 2004) semantics for other defeasible logics. However, the proof-based definitions defined above are sufficient for this paper.

A key feature of $DL(\partial_{||})$ and $DL(\partial_{||}^*)$ inference rules is that they do not use negative inference rules, unlike $DL(\partial)$ which uses $-\Delta$ and $-\partial$. Instead they use expressions $+\Delta q \notin P_\Delta$ and $+\lambda q \notin P_\lambda$. This choice was founded on practical difficulties in scalably implementing existing non-propositional defeasible logics (Maher et al. 2020; Tachmazidis et al. 2012). However, it has implications for the structure and semantics of the corresponding metaprograms of the logics, as discussed in the next section.

5 Metaprogram for $DL(\partial_{||})$

Following earlier work (Maher and Governatori 1999; Antoniou et al. 2000), we can map $DL(\partial_{||})$ to a logic program by expressing the inference rules of $DL(\partial_{||})$ as a metaprogram. The metaprogram assumes that a defeasible theory $D = (F, R, >)$ is represented as a set of unit clauses, as follows.

1. **fact**(p). if $p \in F$
2. **strict**($r, p, [q_1, \dots, q_n]$). if $r : q_1, \dots, q_n \rightarrow p \in R$
3. **defeasible**($r, p, [q_1, \dots, q_n]$). if $r : q_1, \dots, q_n \Rightarrow p \in R$

4. **defeater**($r, p, [q_1, \dots, q_n]$). if $r : q_1, \dots, q_n \rightsquigarrow p \in R$
 5. **sup**(r, s). for each pair of rules such that $r > s$

Note that predicates in D are represented as function symbols. For convenience, we assume that negated atoms $\neg p(\vec{a})$ in the defeasible theory are represented by *not_p*(\vec{a}).

As in papers (Maher and Governatori 1999; Antoniou *et al.* 2000), the metaprogram is presented for ease of understanding, rather than formal logic programming syntax. The full details are available in the appendix.

- c1 **definitely**(X) :-
 fact(X).
- c2 **definitely**(X) :-
 strict($R, X, [Y_1, \dots, Y_n]$),
 definitely(Y_1), ..., **definitely**(Y_n).
- c3 **lambda**(X) :-
 definitely(X).
- c4 **lambda**(X) :-
 not definitely($\sim X$),
 strict_or_defeasible($R, X, [Y_1, \dots, Y_n]$),
 lambda(Y_1), ..., **lambda**(Y_n).
- c5 **defeasibly**(X) :-
 definitely(X).
- c6 **defeasibly**(X) :-
 not definitely($\sim X$),
 strict_or_defeasible($R, X, [Y_1, \dots, Y_n]$),
 defeasibly(Y_1), ..., **defeasibly**(Y_n),
 not overruled(X).
- c7 **overruled**(X) :-
 rule($S, \sim X, [U_1, \dots, U_n]$),
 lambda(U_1), ..., **lambda**(U_n),
 not defeated($S, \sim X$).
- c8 **defeated**($S, \sim X$) :-
 sup(T, S),
 strict_or_defeasible($T, X, [V_1, \dots, V_n]$),
 defeasibly(V_1), ..., **defeasibly**(V_n).

For $DL(\partial_{||}^*)$, clauses c6, c7 and c8 are replaced by the following. (Clauses c6 and c7 are modified only slightly.)

- c9 **defeasibly**(X):-
 not definitely($\sim X$),
 strict_or_defeasible($R, X, [Y_1, \dots, Y_n]$),
 defeasibly(Y_1), ..., **defeasibly**(Y_n),
 not overruled(R, X).

```

c10  overruled(R, X):-
      rule(S, ~X, [U1, ..., Un]),
      lambda(U1, ..., lambda(Un),
      not defeats(R, S).

c11  defeats(R, S):-
      sup(R, S).

```

We use $\mathcal{M}_{\partial_{||}}$ to denote the metaprogram for $DL(\partial_{||})$ (clauses [c1–c8](#)) and $\mathcal{M}_{\partial_{||}^*}$ to denote the metaprogram for $DL(\partial_{||}^*)$ (clauses [c1–c5](#) and [c9–c11](#)), or simply \mathcal{M} to refer to either metaprogram. The combination of \mathcal{M} and the representation of D is denoted by $\mathcal{M}(D)$.

We interpret $\mathcal{M}(D)$ under the well-founded semantics ([Van Gelder et al. 1991](#)). Notice that [c1–c4](#), together with the representation of D , form a stratified logic program, with `definitely` and `lambda` in different strata. The choice of well-founded semantics means that the explicit ordering of computing P_{Δ} , then P_{λ} , then $P_{\partial_{||}}$ is implemented implicitly by the stratification. For weaker semantics that are not equal to the stratified semantics on stratified programs, like Fitting’s semantics, the ordering would need to be expressed explicitly.

A careful comparison of the parts of the inference rules $+\Delta$, $+\lambda$, $+\partial_{||}$, and $+\partial_{||}^*$ with the clauses of the metaprogram strongly suggests the correctness of the metaprogram representation. Clauses [c1](#) and [c2](#) represent the $+\Delta$ inference rule, and clauses [c3](#) and [c4](#) represent the $+\lambda$ inference rule. Clauses [c5–c8](#) represent the $+\partial_{||}$ inference rule: clause [c5](#) corresponds to (1) of the inference rule; [c5](#) corresponds to (2), with the body expressing (2.1) and (2.2) and the negated call to `overruled` representing (2.3); [c7](#) corresponds to (2.3.1); and [c8](#) corresponds to (2.3.2). Negations are used to express the universal quantifier in (2.3) via the implicit logic programming quantifier and $\forall = \neg\exists\neg$. However, despite this close correspondence, we still need to establish correctness formally, and that proof will be easier using a transformed program. Consequently, the proof that $\mathcal{M}(D)$ under the well-founded semantics correctly represents the logic defined in [Section 4](#) is deferred to [Section 7](#).

There have been several mappings of defeasible logics to logic programs. An early implementation of a defeasible logic, d-Prolog ([Nute 1993](#); [Covington et al. 1997](#)), was defined as a Prolog metaprogram. Courteous Logic Programs ([Grosz 1997](#)) were originally implemented by directly modifying rules to express overriding, but later versions (such as the work by [Wan et al. \(2009\)](#)) used a metaprogramming approach. Inspired by d-Prolog, [Maher and Governatori \(1999\)](#) defined the metaprogram approach for $DL(\partial)$, mapping defeasible theories to logic programs under Kunen’s semantics. (See also the paper by [Antoniou et al. \(2006\)](#).) A key point of the approach of [Maher and Governatori \(1999\)](#) was the decomposition of defeasible logics into a conflict resolution method, expressed by the metaprogram, and a notion of failure, expressed by the logic programming semantics applied to the metaprogram. That paper also introduced well-founded defeasible logic (*WFDL*) and mapped it to logic programs under the well-founded semantics, as an example of this decomposition. [Antoniou et al. \(2000\)](#) extended that approach to other defeasible logics, and developed a principled framework for defeasible logics. [Governatori and Maher \(2017\)](#) further extended this work to a single formalism supporting multiple methods of conflict resolution.

Antoniou et al. (Antoniou and Maher 2002; Antoniou et al. 2006) investigated mappings of defeasible theories in $DL(\partial)$ to logic programs under the stable semantics, but achieved only partial results. These works used the metaprogram as a basis, but Antoniou and Maher (2002) provided a simpler mapping of propositional defeasible theories in $DL(\partial)$ to logic programs, under the assumption that D was simplified by the transformations of Antoniou et al. (2001). Brewka (2001) mapped the ambiguity propagating defeasible logic $DL(\delta)$ to prioritized logic programs under the well-founded semantics (Brewka 1996), and showed that the mapping is not sound, nor complete. He concluded that differences in the treatment of strict rules affected completeness, while differences in semantics (well-founded versus Kunen) and in treatment of rule priorities affected soundness. More recently, Nute and Maier (Maier and Nute 2006; 2010; Maier 2013) provided mappings, based on earlier mappings, of the defeasible logic ADL to logic programs under the well-founded semantics, and another mapping in the reverse direction. ADL is similar to the well-founded defeasible logic $WFDL(\delta^*)$ (Maher 2014), but with a more sophisticated treatment of inconsistency of literals. Maier (2013) also discusses a stable set semantics for these logics, and relates it to logic programs under the stable semantics. From the viewpoint of Maher and Governatori's approach Maher and Governatori (1999), these works demonstrate that a structured mapping of a defeasible logic under one semantics to logic programming under a different semantics is difficult, while mapping of a defeasible logic to logic programming under the same semantics is achievable.

These works all focussed on the correctness of the mapping. Beyond d-Prolog and Courteous Logic Programs, there is little attention to implementation, and there is no analysis of the structure of the resulting logic programs. That may be because there is little useful structure to those programs. In contrast, both $\mathcal{M}_{\partial_{||}}$ and $\mathcal{M}_{\partial_{||}^*}$ have a convenient structure.⁵

Proposition 6

For any defeasible theory D , $\mathcal{M}_{\partial_{||}}(D)$ is call-consistent and $\mathcal{M}_{\partial_{||}^*}(D)$ is stratified.

Proof

Only the predicates `defeasibly`, `overruled`, `defeated`, and `loop_defeasibly` in $\mathcal{M}_{\partial_{||}}$ are related by \approx . From the clauses we only have `defeasibly` \supseteq_{-1} `overruled`, `overruled` \supseteq_{-1} `defeated`, `defeasibly` \supseteq_{+1} `loop_defeasibly`, `defeated` \supseteq_{+1} `loop_defeasibly`, and `loop_defeasibly` \supseteq_{+1} `defeasibly`. Thus, for none of these predicates (or any others) do we have $p \supseteq_{-1} p$. Consequently $\mathcal{M}_{\partial_{||}}$ is call-consistent.

Consider the following ordering on predicates in $\mathcal{M}_{\partial_{||}^*}(D)$.

$$\begin{array}{c} \text{defeasibly, loop_defeasibly} \\ \vee \\ \text{lambda, loop_lambda, overruled} \\ \vee \\ \text{definitely, loop_definitely, defeats} \\ \vee \\ \text{neg, sup, fact, strict, defeasible, defeater, strict_or_defeasible, rule} \end{array}$$

It is straightforward to verify that this provides a stratification of $\mathcal{M}_{\partial_{||}^*}(D)$. □

⁵ Recall that $\mathcal{M}_{\partial_{||}}$ is defined in Appendix A, with a more readable version expressed above. The following propositions and discussion refer to the syntax in the appendix.

In contrast, the metaprograms for $DL(\partial)$ and $DL(\partial^*)$ (Antoniou *et al.* 2000) are not call-consistent, resulting from the use of $-\partial$ ($-\partial^*$) in (2.3.1).

The metaprogram $\mathcal{M}_{\partial_{||}}$ is not strict: both `defeasibly` and `lambda` depend on `definitely` both positively and negatively. Furthermore, `defeasibly` depends on `strict_or_defeasible` and `neg` both positively (via `defeated`) and negatively (via `lambda`). Thus $\mathcal{M}_{\partial_{||}}$ does not have a signing. However, if we take a floor consisting of `definitely`, `lambda` and all supporting predicates then the remainder of $\mathcal{M}_{\partial_{||}}$ does have a signing.

Proposition 7

Let D be a defeasible theory, and let $\mathcal{Q} = \{\text{defeasibly, overruled, defeated, loop_defeasibly}\}$.

- $\mathcal{M}_{\partial_{||}}(D)$ has a signing s for \mathcal{Q} such that `defeasibly` has sign +1 and avoids negative unfoundedness wrt s and \mathcal{Q} .
- $\mathcal{M}_{\partial_{||}^*}(D)$ has a signing s for \mathcal{Q} such that `defeasibly` has sign +1 and avoids negative unfoundedness wrt s and \mathcal{Q} .

Proof

Let \mathcal{P} be the set of predicates occurring in $\mathcal{M}_{\partial_{||}}$ and let \mathcal{F} be $\mathcal{P} \setminus \mathcal{Q}$. Then \mathcal{P} is a downward-closed set of predicates with floor \mathcal{F} . Furthermore, \mathcal{Q} has a signing s that maps `defeasibly`, `defeated` and `loop_defeasibly` to +1 and `overruled` to −1. `defeasibly` avoids negative unfoundedness wrt s and \mathcal{Q} , because the only negatively signed predicate is `overruled`, and `overruled` does not depend positively on any predicate in \mathcal{Q} .

The same reasoning applies for $\mathcal{M}_{\partial_{||}^*}$ with the same signing s . □

Notice that we can get a signing for $\mathcal{M}_{\partial_{||}}$ and $\mathcal{M}_{\partial_{||}^*}$ with smaller floor (by excluding `lambda` and `loop_lambda`), but then `defeasibly` does not necessarily avoid negative unfoundedness. For example, if D consists of $p \Rightarrow p$ then `lambda`, which would have a negative sign, has an unfounded sequence `lambda(p), loop_lambda([p]), lambda(p), ...`

As a consequence of the previous proposition, Theorem 1 part 2 applies. This means that, once the values of the predicates in \mathcal{F} (particularly `definitely` and `lambda`) are determined, the true `defeasibly` atoms can be computed under either the Fitting or well-founded semantics. This reflects the original definition of $DL(\partial_{||})$ in Section 4, where P_{Δ} and P_{λ} must be computed before computing $P_{\partial_{||}}$.

Similarly, by Theorem 1 part 1, if computing using the well-founded semantics, only the positive parts of `defeasibly`, `defeated` and `loop_defeasibly` need be computed, and only the negative part of `overruled` is needed. In contrast, in the corresponding metaprogram \mathcal{M}_{∂} for $DL(\partial)$ (see the definition in the papers (Maher and Governatori 1999; Antoniou *et al.* 2000)) `defeasibly` depends negatively on itself, via `overruled`. Consequently, \mathcal{M}_{∂} is not call-consistent and there is no useful signing for \mathcal{M}_{∂} ⁶, nor for \mathcal{M}_{∂^*} . Thus, we see that the structure of $DL(\partial_{||})$ (and, hence, of $\mathcal{M}_{\partial_{||}}$) allows optimizations, like those of Theorem 1, that are unavailable to \mathcal{M}_{∂} and \mathcal{M}_{∂^*} .

⁶ We could set $\mathcal{Q} = \{\text{defeasibly}\}$, to get a signing, but then $\mathcal{P} \setminus \mathcal{Q}$ is not downward-closed (because `overruled` depends on `defeasibly`), and so cannot serve as a floor. Thus, Theorem 1 cannot be applied.

Not all structural aspects of the metaprogram are convenient. $\mathcal{M}_{\partial_{||}}$ is safe, but the representation of D may not be. Furthermore, $\mathcal{M}_{\partial_{||}^*}$ is not safe. In fact, clause [c10](#) is neither range-restricted nor negation-safe, because the variable R appears in the head and negative literal, but not in a positive body literal. Furthermore, even when D is function-free, if it is not propositional then $\mathcal{M}_{\partial_{||}}(D)$ is not a Datalog[⊖] program, because the predicates of D are represented as functions in $\mathcal{M}_{\partial_{||}}(D)$. These problems will be addressed by the transformation of $\mathcal{M}_{\partial_{||}}(D)$ in the next section.

6 Transforming the metaprogram

We seek a logic program that is equivalent to $\mathcal{M}(D)$ on **definitely**, **defeasibly** and **lambda** atoms, but is in Datalog[⊖] form. We manipulate $\mathcal{M}(D)$, using transformations that preserve the semantics of the program, to achieve this end. Specifically, we use a series of fold and unfold transformations ([Tamaki and Sato 1984](#); [Maher 1988](#)). These are known to preserve the well-founded semantics ([Maher 1990](#); [Aravindan and Dung 1995](#)). We also introduce rules for new predicates, and delete rules for predicates that are no longer used. These preserve the semantics of the important predicates ([Maher 1990](#)).

6.1 Partial evaluation of the metaprogram

The following transformations are applied to the full metaprogram in [Appendix A](#) and the representation of a theory D . Unfold all occurrences of **rule**, **strict_or_defeasible**, **fact**, **strict**, **defeasible**, and **defeater**, to create specialized versions of the clauses for each applicable rule. Unfold all occurrences of **neg**, implementing $\sim X$, and unroll (i.e. repeatedly unfold) all occurrences of **loop_definitely**, **loop_defeasibly** and **loop_lambda**, implementing iteration over each body literal. Because all lists are (terminated) finite lists, unrolling will terminate.

After this process, for any strict or defeasible rule, say

$$t : p(X, Z), \neg p(Z, Y) \Rightarrow q(X, Y),$$

we have a corresponding version of clause [c8](#)

```
defeated(S, not_q(X, Y)) :-
  sup(t, S),
  defeasibly(p(X, Z)), defeasibly(not_p(Z, Y)).
```

and similar versions of clause [c7](#), etc. Unfolding all occurrences of **sup** leaves us with clauses

```
defeated(s, not_q(X, Y)) :-
  defeasibly(p(X, Z)), defeasibly(not_p(Z, Y)).
```

for every rule s with $t > s$ in D . Note that if t is not superior to any rule then no clause is generated for t .

We then delete all clauses defining the predicates we have unfolded. Such a deletion is correct because the three predicates we are interested in no longer depend on the deleted predicates.

After all these transformations, the only function symbols left are the predicates p , and their counterparts not_p , that originate in the representation of D . The resulting program P_D is a partial evaluation of $\mathcal{M}_{\partial_{||}}$ wrt the representation of D . It is a particularly transparent translation of the defeasible theory into a logic program.

Example 8

Consider a defeasible theory consisting of the rules

$$\begin{aligned} s : \quad & p(X, Y), q(Y, X) \Rightarrow \neg q(X, Y), \\ t : \quad & p(X, Z), \neg p(Z, Y) \Rightarrow q(X, Y), \end{aligned}$$

with $t > s$.

The transformed program contains

```

defeasibly(not_q(X, Y)) :-
    not definitely(q(X, Y)),
    defeasibly(p(X, Y)), defeasibly(q(Y, X)),
    not overruled(not_q(X, Y)).

overruled(q(X, Y)) :-
    lambda(p(X, Y), lambda(q(Y, X))),
    not defeated(s, not_q(X, Y)).

defeasibly(q(X, Y)) :-
    not definitely(not_q(X, Y)),
    defeasibly(p(X, Z)), defeasibly(not_p(Z, Y)),
    not overruled(q(X, Y)).

overruled(not_q(X, Y)) :-
    lambda(p(X, Z), lambda(not_p(Z, Y))),
    not defeated(t, q(X, Y)).

defeated(s, not_q(X, Y)) :-
    defeasibly(p(X, Z)), defeasibly(not_p(Z, Y)).
    
```

as well as other clauses defining `lambda`. The first two clauses are derived from rule s and the last three come from rule t . There is no `defeated` clause from s because unfolding $\text{sup}(s, S)$ eliminates the clause, there being no rule that s is superior to.

The correctness of the modified program is straightforward.

Proposition 9

Let D be a defeasible theory, and let P_D be the transformed version of $\mathcal{M}_{\partial_{||}}(D)$. Let q be a literal.

- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{definitely}(q)$ iff $P_D \models_{WF} \text{definitely}(q)$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{lambda}(q)$ iff $P_D \models_{WF} \text{lambda}(q)$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{defeasibly}(q)$ iff $P_D \models_{WF} \text{defeasibly}(q)$

Proof

$\mathcal{M}(D)$ is transformed to P_D by a series of unfolding and deletion transformations that preserve the well-founded semantics (Maher 1990; Aravindan and Dung 1995). □

6.2 More transformation

We now further transform P_D , to make it more compact and to convert it to Datalog⁻. To begin, we introduce new predicates and clauses. For any literal A (of the form $q(\vec{a})$ or $\text{not}_q(\vec{a})$), we use $\text{args}(A)$ to denote \vec{a} . For each rule $r : B_1, \dots, B_n \leftrightarrow A$ in D , we add the clauses

c12 $\text{body}_r^d(\text{args}(A)) :-$
 $\text{defeasibly}(B_1), \dots, \text{defeasibly}(B_n).$

c13 $\text{body}_r^\lambda(\text{args}(A)) :-$
 $\text{lambda}(B_1), \dots, \text{lambda}(B_n).$

If the rule is strict we also add

c14 $\text{body}_r^\Delta(\text{args}(A)) :-$
 $\text{definitely}(B_1), \dots, \text{definitely}(B_n).$

We then fold clauses derived from clauses c6 and c8 by clause c12, fold clauses derived from clauses c4 and c7 by clause c13, and fold clauses derived from clauses c2 by clause c14.

This results in clauses of the form

$\text{definitely}(A) :-$
 $\text{body}_r^\Delta(\text{args}(A)).$

$\text{lambda}(A) :-$
 $\text{not definitely}(\sim A),$
 $\text{body}_r^\lambda(\text{args}(A)).$

$\text{defeasibly}(A) :-$
 $\text{not definitely}(\sim A),$
 $\text{body}_r^d(\text{args}(A)),$
 $\text{not overruled}(A).$

$\text{overruled}(A) :-$
 $\text{body}_s^\lambda(\text{args}(A)),$
 $\text{not defeated}(s, \sim A).$

$\text{defeated}(s, \sim A) :-$
 $\text{body}_t^d(\text{args}(A)).$

For example, clauses derived from t in the previous example are now

$\text{lambda}(q(X, Y)) :-$
 $\text{not definitely}(\text{not}_q(X, Y)),$
 $\text{body}_t^\lambda(X, Y).$

$\text{defeasibly}(q(X, Y)) :-$
 $\text{not definitely}(\text{not}_q(X, Y)),$
 $\text{body}_t^d(X, Y),$
 $\text{not overruled}(q(X, Y)).$

$$\begin{aligned} \text{overruled}(\text{not_}q(X, Y)) & :- \\ & \quad \text{body}_t^\lambda(X, Y), \\ & \quad \text{not defeated}(t, q(X, Y)). \\ \text{defeated}(s, \text{not_}q(X, Y)) & :- \\ & \quad \text{body}_t^d(X, Y). \end{aligned}$$

Next, we unfold clauses **c3** and **c5** using the clauses for **definitely**. This results in additional clauses (including unit clauses) for **lambda**, and **defeasibly** corresponding to the existing clauses for **definitely**.

At this stage, the transformed program consists of the following clauses. Recall that we use \leftrightarrow to range over the different kinds of arrows used in a defeasible theory.

The introduced clauses **c13–c14** remain.

For every fact F in D , we have the unit clauses **definitely**(F), **lambda**(F), and **defeasibly**(F) in the program.

For every strict rule $r : B_1, \dots, B_n \rightarrow A$ in D , we have the clauses

c15 **definitely**(A) :-
 $\text{body}_r^\Delta(\text{args}(A)).$

c16 **lambda**(A) :-
 $\text{body}_r^\Delta(\text{args}(A)).$

c17 **defeasibly**(A) :-
 $\text{body}_r^\Delta(\text{args}(A)).$

For every strict or defeasible rule $r : B_1, \dots, B_n \leftrightarrow A$ in D , we have the clause

c18 **lambda**(A) :-
 $\text{not definitely}(\sim A),$
 $\text{body}_r^\lambda(\text{args}(A)).$

and the clause

c19 **defeasibly**(A) :-
 $\text{not definitely}(\sim A),$
 $\text{body}_r^d(\text{args}(A)),$
 $\text{not overruled}(A).$

For every rule $s : B_1, \dots, B_n \leftrightarrow \sim A$ in D , we have the clause

c20 **overruled**(A) :-
 $\text{body}_s^\lambda(\text{args}(A)),$
 $\text{not defeated}(s, \sim A).$

For every strict or defeasible rule $t : B_1, \dots, B_n \leftrightarrow A$ in D that is superior to a rule s for $\sim A$, we have the clause

c21 **defeated**($s, \sim A$) :-
 $\text{body}_t^d(\text{args}(A)).$

We now eliminate the function symbols in the clauses. For each predicate p in D , of arity n , we introduce the clauses

- c22 $\text{definitely_p}(X_1, \dots, X_n) :-$
 $\text{definitely}(p(X_1, \dots, X_n)).$
- c23 $\text{definitely_not_p}(X_1, \dots, X_n) :-$
 $\text{definitely}(\text{not_}p(X_1, \dots, X_n)).$
- c24 $\text{lambda_p}(X_1, \dots, X_n) :-$
 $\text{lambda}(p(X_1, \dots, X_n)).$
- c25 $\text{lambda_not_p}(X_1, \dots, X_n) :-$
 $\text{lambda}(\text{not_}p(X_1, \dots, X_n)).$
- c26 $\text{defeasibly_p}(X_1, \dots, X_n) :-$
 $\text{defeasibly}(p(X_1, \dots, X_n)).$
- c27 $\text{defeasibly_not_p}(X_1, \dots, X_n) :-$
 $\text{defeasibly}(\text{not_}p(X_1, \dots, X_n)).$
- c28 $\text{overruled_p}(X_1, \dots, X_n) :-$
 $\text{overruled}(p(X_1, \dots, X_n)).$
- c29 $\text{overruled_not_p}(X_1, \dots, X_n) :-$
 $\text{overruled}(\text{not_}p(X_1, \dots, X_n)).$
- c30 $\text{defeated_p}(S, X_1, \dots, X_n) :-$
 $\text{defeated}(S, p(X_1, \dots, X_n)).$
- c31 $\text{defeated_not_p}(S, X_1, \dots, X_n) :-$
 $\text{defeated}(S, \text{not_}p(X_1, \dots, X_n)).$

We now fold every body literal (except those in the above clauses) in the program containing a function symbol by the appropriate clause (c22–c31). This is essentially the replacement of atoms involving a function symbol with an atom with a predicate name incorporating the function symbol. Then we unfold the body atom of each of the above clauses (c22–c31). There are now no occurrences of the original predicates in $\mathcal{M}(D)$ (**definitely**, **lambda**, **defeasibly**, **overruled** and **defeated**) in the bodies of rules. We denote the resulting program by T_D .

The introduced predicates, such as **defeasibly_p**, together fully represent the original predicates, such as **defeasibly** in T_D .

Proposition 10

For every literal $q(\vec{a})$,

- $T_D \models_{WF} \text{definitely}(q(\vec{a}))$ iff $T_D \models_{WF} \text{definitely_q}(\vec{a})$
- $T_D \models_{WF} \text{lambda}(q(\vec{a}))$ iff $T_D \models_{WF} \text{lambda_q}(\vec{a})$
- $T_D \models_{WF} \text{defeasibly}(q(\vec{a}))$ iff $T_D \models_{WF} \text{defeasibly_q}(\vec{a})$

Proof

These statements are true once the clauses c22–c31 are introduced, by inspection of those clauses. The subsequent transformations preserve the semantics of the program. \square

6.3 The resulting program

In the following, q (possibly subscripted) may have the form p or not_p . We write $\sim q$ as part of a predicate name to represent, respectively, not_p or p . We can now outline what the final program T_D looks like.

For every fact $q(\vec{a})$ in D , there are unit clauses $\text{definitely}_q(\vec{a})$, $\text{lambda}_q(\vec{a})$, and $\text{defeasibly}_q(\vec{a})$.

For every strict rule $r : q_1(\vec{a}_1), \dots, q_n(\vec{a}_n) \rightarrow q(\vec{a})$ in D , we have the clauses

c32 $\text{definitely}_q(\vec{a}) :-$
 $\text{body}_r^\Delta(\vec{a}).$

c33 $\text{lambda}_q(\vec{a}) :-$
 $\text{body}_r^\Delta(\vec{a}).$

c34 $\text{defeasibly}_q(\vec{a}) :-$
 $\text{body}_r^\Delta(\vec{a}).$

c35 $\text{body}_r^\Delta(\vec{a}) :-$
 $\text{definitely}_{q_1}(\vec{a}_1), \dots, \text{definitely}_{q_n}(\vec{a}_n).$

For every strict or defeasible rule $r : q_1(\vec{a}_1), \dots, q_n(\vec{a}_n) \leftrightarrow q(\vec{a})$ in D , we have the clauses

c36 $\text{lambda}_q(\vec{a}) :-$
 $\text{not definitely}_{\sim q}(\vec{a}),$
 $\text{body}_r^\lambda(\vec{a}).$

c37 $\text{defeasibly}_q(\vec{a}) :-$
 $\text{not definitely}_{\sim q}(\vec{a}),$
 $\text{body}_r^d(\vec{a}),$
 $\text{not overruled}_q(\vec{a}).$

c38 $\text{body}_r^\lambda(\vec{a}) :-$
 $\text{lambda}_{q_1}(\vec{a}_1), \dots, \text{lambda}_{q_n}(\vec{a}_n).$

For every rule $s : q_1(\vec{a}_1), \dots, q_n(\vec{a}_n) \leftrightarrow q(\vec{a})$ in D , we have the clauses

c39 $\text{body}_r^d(\vec{a}) :-$
 $\text{defeasibly}_{q_1}(\vec{a}_1), \dots, \text{defeasibly}_{q_n}(\vec{a}_n).$

c40 $\text{overruled}_{\sim q}(\vec{a}) :-$
 $\text{body}_s^\lambda_q(\vec{a}),$
 $\text{not defeated}_q(s, \vec{a}).$

For every strict or defeasible rule $t : q_1(\vec{a}_1), \dots, q_n(\vec{a}_n) \leftrightarrow q(\vec{a})$ in D that is superior to a rule s for $\sim q$, we have the clause

c41 $\text{defeated}_{\sim q}(s, \vec{a}) :-$
 $\text{body}_t^d(\vec{a}).$

The transformed program T_D also contains clauses for **definitely**, **lambda**, **defeasibly**, **overruled** and **defeated** (the original predicates) which are not needed for computation of the consequences of D in $DL(\partial_{||})$ but are convenient to prove

correctness of the result. These consist of unit clauses `definitely(F)`, `lambda(F)`, and `defeasibly(F)`, for each fact F in D , and clauses `c15–c21`. These clauses will be deleted shortly.

The size of T_D is almost clearly linear in the size of D .⁷ Almost every clause is derived from a rule of D (or fact) and is linear in the size of that rule. Further, each rule gives rise to at most 9 clauses (excluding clauses of the form `c41`). The only problems are the clauses of the form `c21` and `c41`. There is one of each such clause for each superiority statement $s < t$, but the size of these clauses is not necessarily constant; the size of the arguments \vec{a} is bounded by the size of s and t , but that is not constant. We need to address such clauses more carefully.

For each predicate p in D , let D_p be the restriction of D to rules defining p and $\neg p$. Let K_p be the number of distinct superiority statements in D about rules for p and $\neg p$, M_p be the number of rules involved in those superiority statements, and A_p be the arity of p . Then $K_p + 1 \leq M_p \leq 2K_p$, so $O(K_p) = O(M_p)$. D_p must contain at least M_p rules, each of size greater than A_p . That is, $M_p A_p < size(D_p)$. The size of the clauses in T_D derived from `c41` is bounded by the product of K_p (the number of clauses) and a linear term in A_p (the size of each clause). It follows that the size of the clauses derived from `c41` is linear in the size of D_p . Thus, the total size of all clauses in T_D derived from `c41` is $O(\sum_{p \in \Pi} K_p A_p) \leq O(\sum_{p \in \Pi} size(D_p)) \leq O(size(D))$.

Thus the modifications can lead only to a linear blow-up from D to T_D .

Proposition 11

The size of the resulting program T_D is linear in the size of D .

Let us consider now the earlier example of rules in D .

Example 12

Consider a defeasible theory consisting of the rules

$$\begin{aligned} s : \quad & p(X, Y), q(Y, X) \Rightarrow \neg q(X, Y), \\ t : \quad & p(X, Z), \neg p(Z, Y) \Rightarrow q(X, Y), \end{aligned}$$

with $t > s$, and some facts for p .

The transformed program contains

```

lambda_q(X, Y) :-
    not definitely_not_q(X, Y),
    body_t^lambda(X, Y).
defeasibly_q(X, Y) :-
    not definitely_not_q(X, Y),
    body_t^d(X, Y),
    not overruled_q(X, Y).
overruled_not_q(X, Y) :-
    body_t^lambda(X, Y),
    not defeated_q(t, X, Y).
defeated_not_q(s, X, Y) :-
    body_t^d(X, Y).
    
```

⁷ We measure size by the number of symbols in a defeasible theory or logic program.

```

bodyλt(X, Y) :-
    lambda_p(X, Z), lambda_not_q(Z, Y).
bodydt(X, Y) :-
    defeasibly_p(X, Z), defeasibly_not_q(Z, Y).
lambda_not_q(X, Y) :-
    not_definitely_q(X, Y),
    bodyλs(X, Y).
defeasibly_not_q(X, Y) :-
    not_definitely_q(X, Y),
    bodyds(X, Y),
    not_overruled_q(X, Y).
overruled_q(X, Y) :-
    bodyλs(X, Y),
    not_defeated_not_q(s, X, Y).
bodyλs(X, Y) :-
    lambda_p(X, Y), lambda_q(Y, X).
bodyds(X, Y) :-
    defeasibly_p(X, Y), defeasibly_q(Y, X).
    
```

The first six clauses are derived from *t*, and the last five from *s*.

Notice that $DL(\partial_{||})$ as defined in Section 4 only refers to positive consequences. As a result, it is only the positive consequences of $\mathcal{M}_{\partial_{||}}(D)$ for the predicates **defeasibly**, **definitely** and, to a lesser extent, **lambda** that we must be concerned with. This extends to the transformed program.

$\mathcal{M}_{\partial_{||}}(D)$ and the transformed program T_D are equivalent under the well-founded semantics.

Proposition 13

Let D be a defeasible theory, and let T_D be the transformed version of $\mathcal{M}_{\partial_{||}}(D)$. Let q be a literal.

- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{definitely}(q)$ iff $T_D \models_{WF} \text{definitely}(q)$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{lambda}(q)$ iff $T_D \models_{WF} \text{lambda}(q)$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{defeasibly}(q)$ iff $T_D \models_{WF} \text{defeasibly}(q)$

Proof

$\mathcal{M}(D)$ is transformed to T_D by a series of unfolding and folding transformations and additions of clauses defining new predicates, that preserve the well-founded semantics (Maher 1990; Aravindan and Dung 1995). □

Similar results apply for $\mathcal{M}_{\partial_{||}^*}(D)$. The structure of the two metaprograms is largely the same, with minor variations in clauses c9 and c10, and a substantial simplification in c11. Thus the same transformations apply, except to clauses derived from c11. The resulting program contains clauses as described in c32–c36, c38–c39, as well as clauses defining **defeasibly_q**, **overruled_~q**, and **defeated_q**.

Reflecting the minor difference between c9 and c6, the transformed $\mathcal{M}_{\partial_{||}^*}(D)$ contains clauses like c42 rather than c37. Similarly, c43 is only a minor variation of c40, while c44

is a set of unit clauses `defeated`(r, s) corresponding to the superiority statements $r > s$ in D .

c42 `defeasibly_q`(\vec{a}) :-
 `not definitely_~q`(\vec{a}),
 `body` ^{d} _{r} (\vec{a}),
 `not overruled_q`(r, \vec{a}).

c43 `overruled_~q`(r, \vec{a}) :-
 `body` ^{λ} _{s} - q (\vec{a}),
 `not defeats_q`(r, s).

c44 `defeats_q`(r, s).

The size of the transformed program is clearly linear in the size of D since the only problematic clauses c21 and c41 in T_D for $\mathcal{M}_{\partial_{||}}(D)$ are unit clauses in the transformed program for $\mathcal{M}_{\partial_{||}^*}(D)$.

With these variations, we have similar results to Propositions 13 and 11.

Proposition 14

Let D be a defeasible theory, and let T_D^* be the transformed version of $\mathcal{M}_{\partial_{||}^*}(D)$. Let q be a literal.

- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{definitely}(q)$ iff $T_D^* \models_{WF} \text{definitely}(q)$
- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{lambda}(q)$ iff $T_D^* \models_{WF} \text{lambda}(q)$
- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{defeasibly}(q)$ iff $T_D^* \models_{WF} \text{defeasibly}(q)$
- The size of T_D^* is linear in the size of D .

Combining the previous results, we summarize the relationship between the applied metaprogram and the compiled version as follows. Let S_D be the transformed program T_D after deleting the clauses for `definitely`, `lambda`, `defeasibly`, `overruled` and `defeated`, and let S_D^* be T_D^* after deleting similar clauses.

Theorem 15

Let D be a defeasible theory and S_D (S_D^*) be the transformed program. Let $q(\vec{a})$ be a ground literal. q has the form p or $\neg p$ in D , but `p` or `not_p` in the transformed metaprogram.

- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{definitely}(q(\vec{a}))$ iff $S_D \models_{WF} \text{definitely_q}(\vec{a})$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{lambda}(q(\vec{a}))$ iff $S_D \models_{WF} \text{lambda_q}(\vec{a})$
- $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \text{defeasibly}(q(\vec{a}))$ iff $S_D \models_{WF} \text{defeasibly_q}(\vec{a})$
- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{definitely}(q(\vec{a}))$ iff $S_D^* \models_{WF} \text{definitely_q}(\vec{a})$
- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{lambda}(q(\vec{a}))$ iff $S_D^* \models_{WF} \text{lambda_q}(\vec{a})$
- $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \text{defeasibly}(q(\vec{a}))$ iff $S_D^* \models_{WF} \text{defeasibly_q}(\vec{a})$

Proof

By Propositions 9 and 13, $\mathcal{M}_{\partial_{||}}(D)$ is equivalent to P_D and T_D on the predicates `definitely`, `lambda`, and `defeasibly`. Similarly, $\mathcal{M}_{\partial_{||}^*}(D)$ is equivalent to P_D^* and T_D^* , by Proposition 14. By Proposition 10, literals such as `defeasibly_q`(\vec{a}) are inferred iff

defeasibly($q(\vec{a})$) is inferred in T_D (and similarly for T_D^*). Finally, the deletion of predicates from T_D to get S_D (and from T_D^* to get S_D^*) preserves the equivalence because the predicates of interest in S_D do not depend on the deleted predicates. \square

Additionally, following Proposition 11, S_D and S_D^* are linear in the size of D .

The results in this section depend only on the correctness of the transformations used. Consequently, they extend beyond the well-founded semantics to many other logic programming semantics. Indeed, Aravindan and Dung (1995) showed that these transformations preserve the regular models (You and Yuan 1994), stable theory semantics (Kakas and Mancarella 1991), and stable semantics (Gelfond and Lifschitz 1988), as well as the well-founded semantics. Maher (2017) extended this approach to partial stable models (Przymusiński 1990) and L-stable models (Eiter et al. 1997). Earlier work showed that these transformations preserved the (two-valued) Clark-completion semantics (Maher 1988) and Fitting’s and Kunen’s semantics (Bossi et al. 1992). However, these latter semantics do not express the ordering of computation (that is, P_Δ then P_λ then $P_{\partial_{||}}$).

The results are also independent of whether D is function-free or not. Moreover, if constraints (in the sense of constraint logic programming (Jaffar and Maher 1994)) are permitted in the defeasible theory, they can be expressed in the corresponding CLP language, and the results still apply.

Finally, a similar sequence of transformations would apply to the metaprogram of almost any (sensible) defeasible logic. Thus this approach provides a provably correct compilation of defeasible logics to Datalog[−] using only the metaprogram representation and a simple fold/unfold transformation system.

This section has established the correctness of the mapping from the metaprogram to Datalog[−]. We now turn to establishing the correctness of the metaprogram.

7 Correctness of the metaprogram

To establish the correctness of the compilation of $DL(\partial_{||})$ (and $DL(\partial_{||}^*)$) to Datalog[−], we need to verify that the metaprogram presented in Section 5 is correct with respect to the proof theory defined in Section 4.

In general, D is not propositional, but the inference rules in Section 4 are formulated for propositional defeasible theories, so we consider D to be a schema defining the sets of ground instances of rules in D . As a result, the atoms are essentially propositional and the inference rules can be applied. Under the well-founded semantics, a logic program and the ground instances of all its clauses are equivalent. As a result, in the following proofs we consider only ground rules and ground clauses.

The proofs for Δ and λ are straightforward inductions.

Theorem 16

Let D be a defeasible theory and q be a ground literal. q has the form $p(\vec{a})$ or $\neg p(\vec{a})$ in D , but $\mathbf{p}(\vec{a})$ or $\mathbf{not_p}(\vec{a})$ in $\mathcal{M}_{\partial_{||}}(D)$.

- $q \in P_\Delta$ iff $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \mathbf{definitely}(q)$
- $q \in P_\lambda$ iff $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \mathbf{lambda}(q)$

Proof

It is convenient to prove this from the program P_D , rather than $\mathcal{M}_{\partial_{||}}(D)$. As shown in Proposition 13, these two programs are equivalent for the predicates of interest. For simplicity of notation, we write P instead of $ground(P_D)$.

Part 1 \Rightarrow

The proof is by induction on the length of a proof for q , with induction hypothesis: if $+\Delta q$ has proof of length $\leq n$ then $\mathbf{definitely}(q) \in lfp(\mathcal{W}_P)$. For $n = 1$, if $+\Delta q$ has a proof of length 1 then either q is a fact or q is the head of a strict rule with an empty body. In either case, $\mathbf{definitely}(q)$ is a unit clause in P , so $\mathbf{definitely}(q) \in lfp(\mathcal{W}_P)$. If $+\Delta q$ has a proof of length $n + 1$, then there is an instance $q_1, \dots, q_k \rightarrow q$ of a strict rule in D such that each q_i has a proof of length $\leq n$. By the induction hypothesis, $\mathbf{definitely}(q_i) \in lfp(\mathcal{W}_P)$, for each i . P must contain a clause

$$\mathbf{definitely}(q) \text{ :- } \mathbf{definitely}(q_1), \dots, \mathbf{definitely}(q_k),$$

from the definition of P . Hence $\mathbf{definitely}(q) \in lfp(\mathcal{W}_P)$. Thus, by induction, if $+\Delta q$ has a proof then $\mathbf{definitely}(q) \in lfp(\mathcal{W}_P)$.

Part 1 \Leftarrow

The proof is by induction on the length of the Kleene sequence, with induction hypothesis: if $\mathbf{definitely}(q) \in \mathcal{W}_P \uparrow n$ then $q \in P_\Delta$. If $\mathbf{definitely}(q) \in \mathcal{W}_P \uparrow 1$ then $\mathbf{definitely}(q)$ is a unit clause in P . It follows that q is either a fact or the head of a strict rule with empty body in D . Consequently, $q \in P_\Delta$. If $\mathbf{definitely}(q) \in \mathcal{W}_P \uparrow (n + 1)$ then P must have a clause with head $\mathbf{definitely}(q)$, say

$$\mathbf{definitely}(q) \text{ :- } \mathbf{definitely}(q_1), \dots, \mathbf{definitely}(q_k),$$

where $\mathbf{definitely}(q_i) \in \mathcal{W}_P \uparrow n$, for $i = 1, \dots, k$. By the induction hypothesis, $+\Delta q_i \in P_\Delta$, for $i = 1, \dots, k$. Hence $+\Delta q \in P_\Delta$.

Part 2 \Rightarrow

The proof is by induction on the length of a proof for q , with induction hypothesis: if $+\lambda q$ has proof of length $\leq n$ then $\mathbf{lambda}(q) \in lfp(\mathcal{W}_P)$. For $n = 1$, if $+\lambda q$ has a proof of length 1 then either q is a fact or q is the head of a strict or defeasible rule with an empty body. In either case, $\mathbf{lambda}(q)$ is a unit clause in P , so $\mathbf{lambda}(q) \in lfp(\mathcal{W}_P)$. If $+\lambda q$ has a proof of length $n + 1$, then there is an instance $q_1, \dots, q_k \rightarrow q$ of a strict or defeasible rule in D such that each q_i has a proof of length $\leq n$ and $+\Delta \sim q \notin P_\Delta$. By the induction hypothesis, $\mathbf{lambda}(q_i) \in lfp(\mathcal{W}_P)$, for each i . By the first part of this theorem, $\mathbf{definitely}(\sim q) \notin lfp(\mathcal{W}_P)$, and hence *not* $\mathbf{definitely}(\sim q) \in lfp(\mathcal{W}_P)$ (because the well-founded semantics is total on stratified downward-closed subprograms like the clauses defining $\mathbf{definitely}$). P must contain a clause

$$\mathbf{lambda}(q) \text{ :- } \mathbf{not\ definitely}(\sim q), \mathbf{lambda}(q_1), \dots, \mathbf{lambda}(q_k),$$

from the definition of P . Hence $\mathbf{lambda}(q) \in lfp(\mathcal{W}_P)$. Thus, by induction, if $+\lambda q$ has a proof then $\mathbf{lambda}(q) \in lfp(\mathcal{W}_P)$.

Part 2 \Leftarrow

The proof is by induction on the length of the Kleene sequence, with induction hypothesis: if $\mathbf{lambda}(q) \in \mathcal{W}_P \uparrow n$ then $q \in P_\lambda$. If $\mathbf{lambda}(q) \in \mathcal{W}_P \uparrow 1$ then $\mathbf{lambda}(q)$ is a unit clause in P . It follows that q is either a fact or the head of a strict or defeasible rule with

empty body in D . Consequently, $q \in P_\lambda$. If $\text{lambda}(q) \in \mathcal{W}_P \uparrow (n + 1)$ with $n > 0$, then P must have a clause with head $\text{lambda}(q)$, say

$$\text{lambda}(q) \text{ :- not definitely}(\sim q), \text{lambda}(q_1), \dots, \text{lambda}(q_k),$$

where $\text{lambda}(q_i) \in \mathcal{W}_P \uparrow n$, for $i = 1, \dots, k$ and $\text{not definitely}(\sim q) \in \text{lfp}(\mathcal{W}_P)$. By the induction hypothesis, $+\lambda q_i \in P_\lambda$, for $i = 1, \dots, k$ and, by the first part of this theorem, $+\Delta q \notin P_\Delta$. D must have a rule $q_1, \dots, q_k \hookrightarrow q$, where \hookrightarrow is \rightarrow or \Rightarrow , from which the clause in P mentioned above arises. Thus part (2) of the inference rule for λ applies, and $+\lambda q$ can be proved. Hence $+\lambda q \in P_\lambda$. Thus, by induction, if $\text{lambda}(q) \in \text{lfp}(\mathcal{W}_P)$ then $+\lambda q \in P_\lambda$. \square

As a consequence of this theorem and Theorem 15, the compilation is correct with respect to the inference rules $+\Delta$ and $+\lambda$.

Theorem 17

Let D be a defeasible theory and $S_D (S_D^*)$ be the transformed program. Let $q(\vec{a})$ be a ground literal. q has the form p or $\neg p$ in D , but p or $\text{not } p$ in the transformed metaprogram.

- $q \in P_\Delta$ iff $S_D \models_{WF} \text{definitely-}q(\vec{a})$ iff $S_D^* \models_{WF} \text{definitely-}q(\vec{a})$
- $q \in P_\lambda$ iff $S_D \models_{WF} \text{lambda-}q(\vec{a})$ iff $S_D^* \models_{WF} \text{lambda-}q(\vec{a})$

Demonstrating the correctness of the compilation for the main tags is more difficult, largely because of the greater complexity of the inference rules. We will use the following lemma to structure the proof.

Lemma 18

Let (L_1, \leq_1) and (L_2, \leq_2) be partial orders. Let $\Psi : L_1 \rightarrow L_2$ and $\Gamma : L_2 \rightarrow L_1$ be monotonic functions. Let $X_1 \in L_1$ and $X_2 \in L_2$.

If the following conditions hold

1. $X_2 \leq_2 \Psi(X_1)$
2. $X_1 \leq_1 \Gamma(X_2)$
3. $\Psi(\Gamma(X_2)) = X_2$

then $X_2 = \Psi(X_1)$.

Proof

$\Psi(X_1) \leq_2 \Psi(\Gamma(X_2)) = X_2 \leq_2 \Psi(X_1)$, using monotonicity of Ψ and the conditions of the theorem. Hence $X_2 = \Psi(X_1)$. \square

In particular, let $f_1 : L_1 \rightarrow L_1$ and $f_2 : L_2 \rightarrow L_2$ be monotonic functions, and let (L_1, \leq_1) and (L_2, \leq_2) be complete partial orders, so that $X_1 = \text{lfp}(f_1)$ and $X_2 = \text{lfp}(f_2)$ exist. Then, under the conditions of this lemma, $\Psi(\text{lfp}(f_1)) = \text{lfp}(f_2)$

In the application of the lemma, (L_1, \leq_1) is the set of Herbrand interpretations of P_D under the containment ordering and (L_2, \leq_2) is the set of sets of tagged literals from D , again under the containment ordering. f_1 is \mathcal{W}_{P_D} , so X_1 is the well-founded model of P_D and f_2 is the function that applies the inference rules of $DL(\partial_{||})$ in every way possible, so X_2 is $P_{\partial_{||}} \cup P_\lambda \cup P_\Delta$, the least deductively closed set under the $DL(\partial_{||})$ inference rules.

Theorem 19

Let D be a defeasible theory and q be a ground literal. q has the form $p(\vec{a})$ or $\neg p(\vec{a})$ in D , but $\mathbf{p}(\vec{a})$ or $\mathbf{not_p}(\vec{a})$ in $\mathcal{M}_{\partial_{||}}(D)$.

- $D \vdash +\partial_{||}q$ iff $\mathcal{M}_{\partial_{||}}(D) \models_{WF} \mathbf{defeasibly}(q)$
- $D \vdash +\partial_{||}^*q$ iff $\mathcal{M}_{\partial_{||}^*}(D) \models_{WF} \mathbf{defeasibly}(q)$

Proof

It is convenient to prove this from the program P_D , rather than $\mathcal{M}_{\partial_{||}}(D)$. As shown in Proposition 13, these two programs are equivalent for the predicates of interest. For simplicity of notation, we write P instead of $ground(P_D)$.

Let W denote the well-founded model of P as derived in Section 6. Let $X = \Psi(W)$, where $\Psi(W)$ is defined as $\{+\Delta q \mid \mathbf{definitely}(q) \in W\} \cup \{+\lambda q \mid \mathbf{lambda}(q) \in W\} \cup \{+\partial_{||}q \mid \mathbf{defeasibly}(q) \in W\}$. Clearly Ψ is monotonic. Note that, by Proposition 9, W is also the well-founded model of $\mathcal{M}_{\partial_{||}}(D)$ as derived in Section 6.

We claim that X is $\partial_{||}$ -deductively closed from D . Consider a literal p , and suppose that $+\partial_{||}p$ can be inferred from X . Then either (1) $+\Delta p \in X$ or (2) there is a strict or defeasible rule r where all body literals p_i are tagged by $\partial_{||}$ in X ; $+\Delta \sim p \notin X$; and for every rule s for $\sim p$ either some body literal q_i has $+\lambda q_i \notin X$, or there is a strict or defeasible rule t where $t > s$ and all body literals p'_i are tagged by $\partial_{||}$ in X . If (1) then we must have $\mathbf{definitely}(p) \in W$. But then, since W is a model of P , $\mathbf{defeasibly}(p) \in W$ and hence $+\partial_{||}p$ is in X (by the definition of X).

If (2) then (2.1) $\mathbf{defeasibly}(p_i) \in W$ for each body literal p_i of r ; (2.2) $\mathbf{definitely}(\sim p) \notin W$; and (2.3) for every s either $\mathbf{lambda}(q_i) \notin W$ or there is a t with $t > s$ and $\mathbf{defeasibly}(p'_i) \in W$ for every body literal p'_i of t . By Corollary 23, $\mathbf{not_definitely}(\sim p) \in W$ and $\mathbf{not_lambda}(q_i) \in W$. Hence, using this fact and (2.1), P contains a version of **c6** instantiated by r with the entire body satisfied in W , except perhaps for $\mathbf{not_overruled}(p)$. Furthermore, for every instantiated version of **c7** by a rule s for $\sim p$ in P either $\mathbf{lambda}(q_i)$ is not satisfied for some body literal q_i of s , or there is an instantiated version of **c8** by t such that the body of this version is satisfied in W . In the former case, $\mathbf{lambda}(q_i)$ evaluates to **false** in W . In the later case, since W is a model of P , $\mathbf{defeated}(s, \sim p) \in W$. Thus, in either case, the body of the version of **c7** instantiated by s evaluates to **false** in W . It follows that $\mathbf{not_overruled}(p)$ evaluates to **true** in W . Hence the body of the version of **c6** is satisfied in W and consequently $\mathbf{defeasibly}(p)$ must be in W . Hence $+\partial_{||}p$ is in X (by the definition of X).

Since this argument applies for any literal p , X is $\partial_{||}$ -deductively closed. By Theorem 16, X is deductively closed. Hence $X \supseteq P_{\partial_{||}} \cup P_{\lambda} \cup P_{\Delta}$, which is the smallest deductively closed set. This establishes condition 1 of Lemma 18.

We define a function Γ from sets of tagged literals to 3-valued interpretations. For any set Z of tagged literals, let $Y_Z = \{\mathbf{definitely}(q) \mid +\Delta q \in Z\} \cup \{\mathbf{lambda}(q) \mid +\lambda q \in Z\} \cup \{\mathbf{defeasibly}(q) \mid +\partial_{||}q \in Z\}$ be the set of corresponding logic programming atoms. Let $\mathcal{W}_P^1(I) = \mathcal{W}_P(I) \cup I$. We define $\Gamma(Z) = lfp(\mathcal{W}_P^1, Y_Z)$, the least fixedpoint of \mathcal{W}_P^1 containing Y_Z . Notice that this is well-defined, since $\mathcal{W}_P^1(f, Y)$ is a monotonic function on the sub-complete lattice of supersets of Y when f is a monotonic function. Furthermore, Γ is monotonic, since $Z \subseteq Z'$ implies $Y_Z \subseteq Y_{Z'}$ and $lfp(\mathcal{W}_P^1, Y)$ is monotonic in Y . In addition, $\mathcal{W}_P(\Gamma(Z)) \subseteq \Gamma(Z)$ since $\Gamma(Z)$ is a fixedpoint of \mathcal{W}_P^1 .

Now, let $Z = P_{\partial_{||}} \cup P_{\lambda} \cup P_{\Delta}$ be the union of the three closures and $U = \Gamma(Z)$. Then $\mathcal{W}_P(U) \subseteq U$, that is, U is a prefixedpoint of \mathcal{W}_P . Hence $W \subseteq U = \Gamma(Z)$, that is, $X_1 \subseteq \Gamma(X_2)$. This establishes condition 2 of Lemma 18.

To apply Lemma 18 we must establish that $\Psi(\Gamma(X_2)) = X_2$, where $X_2 = P_{\partial_{||}} \cup P_{\lambda} \cup P_{\Delta}$, but we work at a greater level of generality. By definition of Γ , $\Gamma(Z) \supseteq Y_Z$, for any Z . Hence $\Psi(\Gamma(Z)) \supseteq \Psi(Y_Z) = Z$. For the other direction, we know that equality holds for Δ and λ conclusions, by Theorem 16. Thus we focus on $\partial_{||}$ conclusions. We define $(\mathcal{W}_P^1)^n(Y_Z) = Y_Z$ if $n \leq 0$.

Suppose, for some deductively closed set Z of tagged literals and some literal p , $+\partial_{||}p \in \Psi(\Gamma(Z)) \setminus Z$. Then $\text{defeasibly}(p) \in \Gamma(Z) \setminus Y_Z$. Let q be one of the first such literals generated by \mathcal{W}_P^1 . That is, $\text{defeasibly}(q) \in (\mathcal{W}_P^1)^{n+1}(Y_Z)$ and $\Psi((\mathcal{W}_P^1)^n(Y_Z)) = Z$. Hence there is a strict or defeasible rule $q_1, \dots, q_k \hookrightarrow q$ of D such that *not definitely*($\sim q$) $\in (\mathcal{W}_P^1)^n(Y_Z)$, $\{\text{defeasibly}(q_1), \dots, \text{defeasibly}(q_n)\} \subseteq (\mathcal{W}_P^1)^n(Y_Z)$, and *not overruled*(q) $\in (\mathcal{W}_P^1)^n(Y_Z)$, since the rule in P must be derived from c6. By Theorem 16 and Corollary 23 we must have $+\Delta \sim q \notin P_{\Delta}$. Because we chose q to be (one of) the first literals to be derived, $\text{defeasibly}(q_i) \in Y_Z$ and hence $+\partial_{||}q_i \in Z$, for $i = 1, \dots, k$. Because all rules for *overruled* are derived from c7, for every rule s for $\sim q$ in D , say $p_1, \dots, p_m \hookrightarrow \sim q$, either (1) $\text{lambda}(p_i) \notin (\mathcal{W}_P^1)^{n-1}(Y_Z)$, for some i , or (2) *defeated*(s, q) $\in (\mathcal{W}_P^1)^{n-1}(Y_Z)$. If (1) then, by Theorem 16 and Corollary 23, $+\lambda p_i \notin P_{\lambda}$, for some i . If (2) then for some strict or defeasible rule t in D of the form $q'_1, \dots, q'_h \hookrightarrow q$, $t > s$ and $\text{defeasibly}(q'_i) \in (\mathcal{W}_P^1)^{n-2}(Y_Z)$, for $i = 1, \dots, h$ and, hence, $+\partial_{||}q'_i \in Z$, for $i = 1, \dots, h$.

In summary, there is a strict or defeasible rule $q_1, \dots, q_k \hookrightarrow q$ of D where: $+\Delta \sim q \notin P_{\Delta}$, $+\partial_{||}q_i \in Z$, for $i = 1, \dots, k$, and for every rule s for $\sim q$ in D , say $p_1, \dots, p_m \hookrightarrow \sim q$, either $+\lambda p_i \notin P_{\lambda}$, for some i , or there is a strict or defeasible rule t in D of the form $q'_1, \dots, q'_h \hookrightarrow q$ with $t > s$ and $+\partial_{||}q'_i \in Z$, for $i = 1, \dots, h$.

Thus, by the $\partial_{||}$ inference rule, and because Z is deductively closed for D , $+\partial_{||}q \in Z$. This contradicts our initial supposition that $+\partial_{||}q \in \Psi(\Gamma(Z)) \setminus Z$. Hence there is no such q , and we must have $\Psi(\Gamma(Z)) = Z$. In particular, $X_2 = P_{\partial_{||}} \cup P_{\lambda} \cup P_{\Delta}$ is deductively closed, so $\Psi(\Gamma(X_2)) = X_2$. This establishes condition 3 of Lemma 18.

Hence, by Lemma 18, $P_{\partial_{||}} \cup P_{\lambda} \cup P_{\Delta} = \Psi(W)$, where W is the well-founded model of P_D . In particular, $D \vdash +\partial_{||}q$ iff $P_D \models_{WF} \text{defeasibly}(q)$. The first part then follows by Proposition 9.

The second part is established in a similar manner, using $\partial_{||}^*$ and P_D^* instead of $\partial_{||}$ and P_D . The argument is slightly simpler, in line with the slightly simpler inference rule of $\partial_{||}^*$ and the slightly simpler P_D^* . □

We can now show the correctness of computing with predicates such as `defeasibly.q`.

Theorem 20

Let D be a defeasible theory and S_D be the transformed program. Let $q(\vec{a})$ be a ground literal. q has the form p or $\neg p$ in D , but `p` or `not.p` in the transformed metaprogram S_D .

- $D \vdash +\Delta q(\vec{a})$ iff $S_D \models_{WF} \text{definitely.q}(\vec{a})$
- $D \vdash +\lambda q(\vec{a})$ iff $S_D \models_{WF} \text{lambda.q}(\vec{a})$
- $D \vdash +\partial_{||}q(\vec{a})$ iff $S_D \models_{WF} \text{defeasibly.q}(\vec{a})$
- $D \vdash +\partial_{||}^*q(\vec{a})$ iff $S_D^* \models_{WF} \text{defeasibly.q}(\vec{a})$

Proof

By Theorems 16 and 19, provability from D is represented by inference from $\mathcal{M}_{\partial_{||}}(D)$ (or $\mathcal{M}_{\partial_{||}^*}(D)$) under the well-founded semantics. By Theorem 15, such inferences are equivalent to the inferences from S_D stated in the statement of this theorem. \square

These results establish that the metaprograms $\mathcal{M}_{\partial_{||}}$ and $\mathcal{M}_{\partial_{||}^*}$ correctly reflect the proof-theoretic definitions of Section 4. However, the metaprograms are able to accommodate non-propositional defeasible theories, and easily extend to handle constraints, which are problematic for the proof theory when the constraint domain is infinite. There is a strong case that metaprogram formulations should be considered the canonical definitions for defeasible logics.

8 Properties of the compiled program

Some syntactic properties of $\mathcal{M}_{\partial_{||}}$ and $\mathcal{M}_{\partial_{||}^*}$ were already established in Section 5. However, S_D has these and further properties that are important to its implementation. We now establish these properties.

Theorem 21

The transformed program S_D for a defeasible theory D over $DL(\partial_{||})$ is:

1. a Datalog⁻ program iff D is function-free
2. variable-free iff D is variable-free
3. range-restricted iff D is range-restricted
4. safe iff D is range-restricted
5. call-consistent
6. stratified if D is hierarchical
7. locally stratified if D is locally hierarchical

Proof

1. If D is function-free then the only functions in $\mathcal{M}(D)$ are predicates from D . After the merging of tags and predicate names, and the deletion of clauses, there are no functions remaining. Conversely, if S_D is a Datalog⁻ program then no term involves a function. But all terms of D appear in S_D . Hence D is function-free.

2. By inspection of the final transformed program (clauses c32–c41), the only arguments have the form \vec{a} (possibly with subscript) which come from rules in D or are the argument s in clauses c40 and c41, which is variable-free. Thus S_D is variable-free if D is variable-free. Every term in D appears in S_D . Thus if S_D is variable-free then D is variable-free.

3. Suppose D is range-restricted. By inspection of the final program, all heads of clauses have arguments \vec{a} , all negative literals also have arguments \vec{a} , and all clauses have arguments \vec{a} in a positive body literal, except for the body clauses. Hence these clauses are all range-restricted. Body clauses have arguments \vec{a} from the head of a rule in the head of the clause and all the arguments \vec{a}_i from the body of that rule in positive literals in the clause body. Hence, since D is range-restricted, these body clauses are range-restricted.

Conversely, every rule in D is reflected in a clause of the form c38. Since S_D is range-restricted, c38 is range-restricted, and hence all rules in D are range-restricted. Further-

more, all facts in D are represented in unit clauses in S_D , and hence are variable-free. Thus, D is range-restricted.

4. By inspection of S_D , it is negation-safe: only clauses **c40** could cause a problem, but the argument s is a constant. By the previous part, S_D is safe iff D is range-restricted.

5. This follows from Proposition 6, and Theorem 6.7 from the paper by Maher (1993) (or Theorem 1 from the paper by Maher (1990)).

6. Let $n : \Pi \rightarrow \mathbb{N}$ a mapping demonstrating the hierarchicality of D and define the function $m : \Pi \rightarrow \mathbb{N}$ by:

Predicates of the form **definitely**_q and **body**_r^Δ are mapped to 0.

Predicates of the form **lambda**_q and **body**_r^λ are mapped to 1.

Predicates of the form **defeasibly**_q are mapped to $3 * n(q) + 5$.

Predicates of the form **overruled**_q are mapped to $3 * n(q) + 4$.

Predicates of the form **defeated**_q and **body**_r^d are mapped to $3 * n(q) + 3$.

It is straightforward to verify that this defines a stratification of S_D . For example, to verify clauses of the form **c39** satisfy the stratification condition we have $m(\text{defeasibly}_{q_i}) = 3 * n(q_i) + 5 \leq 3 * n(q) + 2 < 3 * n(q) + 3 = m(\text{body}_r^d)$ for each strict or defeasible rule $r : q_1(\vec{a}_1), \dots, q_n(\vec{a}_n) \hookrightarrow q(\vec{a})$ for q , where we use $n(q_i) < n(q)$ from the hierarchicality of D .

7. Let $n : HB \rightarrow \mathbb{N}$ a mapping demonstrating the local hierarchicality of D and define the function $m : \Pi \rightarrow \mathbb{N}$ by:

Ground atoms of the form **definitely**_q(\vec{a}) and **body**_r^Δ(\vec{a}) are mapped to 0.

Ground atoms of the form **lambda**_q(\vec{a}) and **body**_r^λ(\vec{a}) are mapped to 1.

Ground atoms of the form **defeasibly**_q(\vec{a}) are mapped to $3 * n(q(\vec{a})) + 5$.

Ground atoms of the form **overruled**_q(\vec{a}) are mapped to $3 * n(q(\vec{a})) + 4$.

Ground atoms of the form **defeated**_q(\vec{a}) and **body**_r^d(\vec{a}) are mapped to $3 * n(q(\vec{a})) + 3$.

The proof is essentially the same as part 6. □

Of these properties, most are a reflection of properties of D . However, the call-consistency of S_D is a reflection of the defeasible logic. In $DL(\partial_{||})$, in the $+\partial_{||}$ inference rule, $+\partial_{||}$ appears only positively and $-\partial_{||}$ does not appear; in $DL(\partial)$ in the $+\partial$ inference rule, $-\partial$ appears. It is this difference that leads to the call-consistency of S_D , independent of D .

For the logic $DL(\partial_{||}^*)$ we have similar properties. An important difference with the previous theorem is that the transformed program is stratified, whether or not the defeasible theory is hierarchical.

Theorem 22

The transformed program S_D^* for a defeasible theory D over $DL(\partial_{||}^*)$ is:

1. a Datalog[⊃] program iff D is function-free
2. variable-free iff D is variable-free
3. range-restricted iff D is range-restricted
4. safe iff D is range-restricted
5. stratified

Proof

The arguments are essentially the same as for Theorem 21. In particular, the proof of part 5 is the same as the proof of part 5 of Theorem 21 (notwithstanding the different properties addressed). In part 4, clauses c42, c43, and c44 are negation-safe because arguments r and s are constants. □

As a corollary, we have that the subset of clauses in S_D^* defining predicates of the form `definitely_q` and `lambda_q` (and `body_r^Δ` and `body_r^λ`) is stratified. This is of interest because the same set of clauses define `definitely_q` and `lambda_q` in S_D .

Corollary 23

Let G be the subset of S_D and S_D^* defining predicates of the form `definitely_q`, `lambda_q`, `body_r^Δ`, and `body_r^λ`. Then G is stratified.

Let W be the well-founded model of S_D (or S_D^*). Then, for every predicate q in D and every \vec{a} , `definitely_q`(\vec{a}) $\notin W$ iff *not* `definitely_q`(\vec{a}) $\in W$, and `lambda_q`(\vec{a}) $\notin W$ iff *not* `lambda_q`(\vec{a}) $\in W$.

Proof

G is stratified because S_D^* stratified and G is a subset of S_D^* . G is downward-closed, and the well-founded model is total on stratified programs (Van Gelder *et al.* 1991). The second part then follows. □

We saw earlier that computing the `defeasibly` predicate in $\mathcal{M}_{\partial_1}(D)$ can be achieved by first computing `definitely` and `lambda`, and then applying Fitting’s semantics to `defeasibly` and related predicates. The same basic idea applies to S_D .

Theorem 24

Let D be a defeasible theory and S_D be the program transformed from $\mathcal{M}_{\partial_1}(D)$. Let \mathcal{P} denote the predicates of S_D , and \mathcal{F} be the set of predicates in S_D of the form `definitely_q`, `body_r^Δ`, `lambda_q`, and `body_r^λ`. Let $\mathcal{Q} = \mathcal{P} \setminus \mathcal{F}$. Then

- \mathcal{P} is downward-closed with floor \mathcal{F}
- \mathcal{Q} has a signing s where all predicates `defeasibly_q` are assigned +1.
- For any predicate q in D , the predicate `defeasibly_q` avoids negative unfoundedness wrt s .

Proof

It is clear, by inspection of S_D , that \mathcal{P} is downward-closed with floor \mathcal{F} . Let s assign all predicates `overruled_q` the value -1 , and all other predicates in \mathcal{Q} the value $+1$. It is straightforward to verify that this is a signing for \mathcal{Q} . Furthermore, any predicate `overruled_q` does not depend positively on any predicate in \mathcal{Q} (it only depends positively on predicates `body_r^λ`, which are in \mathcal{F}). Thus, trivially, every predicate in \mathcal{Q} avoids negative unfoundedness wrt s and, in particular, the predicates `defeasibly_q`. □

Thus, the predicates `defeasibly_q` can be computed by first computing `definitely_q`, then `lambda_q` under the stratified approach, and then applying Fitting’s semantics.

This same result holds for any S_D^* derived from $\mathcal{M}_{\partial_1^*}(D)$, by essentially the same proof. However, it is less useful because we have already established that S_D^* is stratified.

Corollary 25

Let D be a defeasible theory and S_D^* be the program transformed from $\mathcal{M}_{\partial_{||}}(D)$. Let \mathcal{P} denote the predicates of S_D^* , and \mathcal{F} be the set of predicates in S_D^* of the form `definitely_q`, `bodyrΔ`, `lambda_q`, and `bodyrλ`. Let $\mathcal{Q} = \mathcal{P} \setminus \mathcal{F}$. Then

- \mathcal{P} is downward-closed with floor \mathcal{F}
- \mathcal{Q} has a signing s where all predicates `defeasibly_q` are assigned +1.
- For any predicate q in D , the predicate `defeasibly_q` avoids negative unfoundedness wrt s .

The use of the well-founded semantics to define the meaning of $\mathcal{M}_{\partial_{||}}(D)$ might at first appear questionable, especially for propositional defeasible theories. Consequences of propositional theories can be computed in time linear in the size of the theory, while the well-founded semantics has only a quadratic upper bound. However, the above results show that the consequences of a propositional D in $DL(\partial_{||})$ (or $DL(\partial_{||}^*)$), computed by the well-founded semantics from S_D (or S_D^*), can be derived in linear time. This is because the subset of S_D (or S_D^*) defining predicates `definitely_q` and `lambda_q` is stratified, and the stratified semantics can be computed in linear time. Furthermore, the size of S_D is linear in the size of D (from Proposition 11). And then, by Theorems 24 and 1 part 2, it suffices to compute the remaining predicates under Fitting’s semantics, which can be done in linear time for essentially propositional programs.

More generally, the properties identified in this section provide a partial basis for choosing the implementation of Datalog⁻ in which to execute the transformed metaprogram, and to adapt to available implementations. This will be discussed in detail in the next section.

9 Executing compiled theories

There are three main implementation techniques for Datalog: top-down execution with tabling, like XSB; grounding with propositional inference, as used in ASP systems; and bottom-up execution, using database techniques. In addition, there are novel implementation techniques: *bddbdb* (Whaley *et al.* 2005) is based on BDDs, and there is promise in implementations based on linear algebra (Sato 2017), which may be able to exploit the considerable research on software and hardware accelerators for linear algebra. Of the three main techniques, all can be used both for query-answering and for generating all conclusions, although their relative efficiency for each scenario is not completely clear.

The compilation of defeasible logics to Datalog⁻ is motivated by the variety of Datalog systems, but the effectiveness of this approach is limited by the number of implementations of the full well-founded semantics. For example, many systems require safe programs. Fortunately, this is not a serious barrier in implementing defeasible logics because defeasible theories are usually range-restricted and so, as shown in the previous section, the compiled program is safe.

However, many systems have other shortcomings. While some implementations aim to compute the well-founded model, others only apply to stratified programs, and others are in between. Nevertheless, they may be sufficient for executing some defeasible theories and/or providing an approximation of the conclusions for theories. In this section we explore these possibilities. We will briefly discuss the various systems and how the

properties presented in the previous section can be used to adapt to limitations of the systems. The systems are organized by the semantics that they can compute.

9.1 Well-founded semantics

XSB (Swift and Warren 2012) is based on top-down execution with tabling, SLG-resolution. Although SLG-resolution is complete for function-free programs (Chen and Warren 1996), the current version of XSB does not implement answer completion (Swift *et al.* (2017), page 109), one of the operations employed by SLG-resolution. Although it appears that this is only needed in pathological examples, the result is that, in general, XSB is sound but not complete (Swift *et al.* 2017).⁸

Early versions of Smodels (Niemelä and Simons 1997) computed the well-founded model, but the system is no longer at the state of the art. DLV (Adrian *et al.* 2018) is primarily aimed at computing answer sets for disjunctive logic programs, but the switch `-wf` allows the computation of the well-founded model. On the other hand, *clingo* (Gebser *et al.* 2019) does not provide direct access to the well-founded model, but it does provide an indirect way to compute the well-founded model of S_D . It has the switch `-e cautious` which generates all literals true in all stable models, and thus gives an upper bound of the well-founded model. Furthermore, by Theorem 10 in the appendix of the paper by Maher (2021) (adapted from Theorem 5.11 of the paper by Dung (1995)), for S_D this gives exactly the well-founded model. However, this is likely to be a quite inefficient way to compute the consequences of D .

IRIS (Bishop and Fischer 2008) was designed as a platform for implementing languages such as RDFS and description logics. It computes the well-founded semantics, with a choice of techniques, but development seems to have stalled in 2010.

The system of Tachmazidis *et al.* (2014) provides a implementation of the well-founded semantics, based on the MapReduce framework (Dean and Ghemawat 2004). An earlier system (Tachmazidis and Antoniou 2013) implemented the semantics only for stratified programs. These are more proof-of-concept implementations than production-level systems.

All these systems can compute the consequences of a defeasible theory D in $DL(\partial_{\parallel})$. Furthermore, from Theorem 1, for most predicates only the positive or only the negative part needs to be computed, although it appears that this distinction is only useful for bottom-up implementations, such as IRIS and the system of Tachmazidis *et al.* (2014).

9.2 Stratified semantics

The stratified semantics (that is, the well-founded semantics computed only when the program is stratified) is possibly the easiest form of negation to implement, since it requires only a simple syntactic analysis and the layering of negation-free subprograms. There are numerous systems that compute this semantics, including LogicBlox (Aref *et al.* 2015), Soufflé (Jordan *et al.* 2016), QL (Avgustinov *et al.* 2016), RecStep (Fan

⁸ A complicating issue is that rules such as $p(X) \Rightarrow p(X)$, which can lead to such pathological examples, can be used in defeasible theories to ensure that some p -conclusions are undefined, rather than disproved. Thus the incompleteness of XSB might not affect this usage.

et al. 2019), VLog (Carral *et al.* 2019), and Formulog (Bembenek *et al.* 2020). Furthermore, the grounders *gringo* (Gebser *et al.* 2011) and I-DLV (Calimeri *et al.* 2017), for *clingo* and DLV respectively, will compute the well-founded model for stratified programs. Finally, any implementation of Datalog (without negation) can be used as the basis for an implementation of the stratified semantics, using a scripting language, for example. However, that would come with a significant drag on performance, compared with an integrated implementation of the stratified semantics.

Recall that we are only interested in conclusions of the form $+\Delta q(\vec{a})$ and $+\partial_{||} q(\vec{a})$, for literals $q(\vec{a})$, and hence only interested in the computation of atoms of the form `defeasibly_q(\vec{a})` or `delta_q(\vec{a})`. The latter can be computed exactly by a system supporting the stratified semantics, while the program for the former is not, in general, stratified. However, if D is hierarchical then S_D is stratified (Theorem 21.6) and such systems can compute exactly the consequences of $\mathcal{M}_{\partial_{||}}(D)$.

Even when D is not hierarchical, these systems can provide a sound approximation to $\mathcal{M}_{\partial_{||}}(D)$. As we have seen (Theorem 22), S_D^* is stratified. Consequently, these systems can compute the set of `defeasibly_q(\vec{a})` atoms computed from S_D^* . As established by Maher *et al.* (2020) (Theorem 11), $\partial_{||}^* \subset \partial_{||}$ so this set of atoms is a sound approximation of the consequences of $\mathcal{M}_{\partial_{||}}(D)$. Furthermore, stratified fragments of S_D can be used to potentially improve the sound approximation. In some cases, it can be necessary to alternate the use of S_D^* and stratified fragments of S_D .

9.3 Intermediate and ad hoc semantics

Other semantics for Datalog⁻ are less frequently targeted. Locally hierarchical defeasible theories need the implementation of well-founded semantics only for locally stratified programs (Theorem 21), which falls in between the stratified semantics and the full well-founded semantics. Few, if any, implementations specifically address this class.

Nevertheless, the output of the grounder *gringo* (Gebser *et al.* 2011) can be used to infer an underestimate of the well-founded model W of the input program. From its output we can determine $a \in W$ if the fact a is output, and $\text{not } a \in W$ if no rule for a is output; this provides us with an underestimate of W . This underestimate is, in fact, exact for stratified programs (Kaminski 2020). For locally stratified programs, repeated application of *gringo* on its output can lead to exactly the well-founded model (Kaminski 2020). Possibly I-DLV (Calimeri *et al.* 2017) can achieve the same outcome.

Theorems 1 part 2 and 24 show that the `defeasibly_q` predicates can be computed using a combination of the stratified and Fitting semantics. Although only at the research stage, the linear algebraic approach is capable of this combination, since it handles definite clauses (Sato 2017) and Fitting's semantics (Sato *et al.* 2020).

Finally, in theory, implementations need only compute one part of each predicate not in the floor (from Theorem 1). However, as mentioned earlier, it depends very much on the implementation technique whether this property can be exploited to improve performance.

10 Conclusions

In this paper we have formulated a metaprogram representation of the defeasible logic $DL(\partial_{||})$ and proved it correct. We used established transformations to derive a correct

compilation of defeasible theories to Datalog[⊥] programs. And, using properties of the consequent programs, we outlined how they can be used to adapt to limitations of an underlying Datalog system.

Although we focussed on the logic $DL(\partial_{||})$, the metaprogramming and transformation approach applies to any defeasible logic. However the design of $DL(\partial_{||})$, motivated by scalability, induced structure on the resulting Datalog[⊥] that can simplify computation and/or made it more efficient than the result of compilation for other defeasible logics (such as $DL(\partial)$). This structure also supported adaptation and approximation in response to the limited availability of implementations of the full well-founded semantics for Datalog[⊥]. It suggests that the designers of new defeasible logics should take the possibility of similar structures into account during the design of these logics.

We also provided more evidence of the usefulness of using a metaprogram to define a defeasible logic, rather than inference rules such as those in Sections 3 and 4. Those inference rules do not generalize easily to non-propositional defeasible theories over infinite domains, whereas that is a non-issue for metaprograms. Furthermore, such inference rules are difficult to reason about. Metaprograms provide access to all the tools of (constraint) logic programming for reasoning about and implementing the defeasible logics.

The original motivation for this work was to provide alternatives to the bespoke implementation of $DL(\partial_{||})$ by Maher *et al.* (2020). So, it was disappointing to realize the limited range of Datalog implementations available that support the full well-founded semantics. There remains much scope for implementations of the well-founded semantics on novel architectures.

Acknowledgements: This paper is dedicated to the memory of Christian Schulte.

The author has an adjunct position at Griffith University and an honorary position at UNSW. He thanks the reviewers for their thorough reviews and comments, which helped improve the paper.

Competing interests: The author declares none.

References

- ABERGER, C. R., LAMB, A., TU, S., NÖTZLI, A., OLUKOTUN, K. AND RÉ, C. 2017. Emptyheaded: A relational engine for graph processing. *ACM Transactions on Database Systems* 42, 4, 20:1–20:44.
- ABITEBOUL, S., HULL, R. AND VIANU, V. 1995. *Foundations of Databases*. Addison-Wesley.
- ADRIAN, W. T., ALVIANO, M., CALIMERI, F., CUTERI, B., DODARO, C., FABER, W., FUSCÀ, D., LEONE, N., MANNA, M., PERRI, S., RICCA, F., VELTRI, P. AND ZANGARI, J. 2018. The ASP system DLV: Advancements and applications. *KI* 32, 2-3, 177–179.
- ALVARO, P., MARCZAK, W. R., CONWAY, N., HELLERSTEIN, J. M., MAIER, D. AND SEARS, R. 2010. Dedalus: Datalog in time and space. In *Datalog Reloaded - First International Workshop, Datalog 2010*, O. de Moor, G. Gottlob, T. Furche, and A. J. Sellers, Eds. Lecture Notes in Computer Science, vol. 6702. Springer, 262–281.
- ANTONIOU, G., BILLINGTON, D., GOVERNATORI, G. AND MAHER, M. J. 2000. A flexible framework for defeasible logics. In *AAAI/IAAI*. AAAI Press/The MIT Press, 405–410.
- ANTONIOU, G., BILLINGTON, D., GOVERNATORI, G. AND MAHER, M. J. 2001. Representation results for defeasible logic. *ACM Transactions on Computational Logic* 2, 2, 255–287.
- ANTONIOU, G., BILLINGTON, D., GOVERNATORI, G. AND MAHER, M. J. 2006. Embedding defeasible logic into logic programming. *Theory and Practice of Logic Programming* 6, 6, 703–735.

- ANTONIOU, G., BILLINGTON, D., GOVERNATORI, G., MAHER, M. J. AND ROCK, A. 2000. A family of defeasible reasoning logics and its implementation. In *ECAI*, W. Horn, Ed. IOS Press, 459–463.
- ANTONIOU, G. AND MAHER, M. J. 2002. Embedding defeasible logic into logic programs. In *Logic Programming, 18th International Conference*. 393–404.
- APT, K. R., BLAIR, H. A. AND WALKER, A. 1988. Towards a theory of declarative knowledge. In *Foundations of Deductive Databases and Logic Programming*. Morgan Kaufmann, 89–148.
- APT, K. R. AND BOL, R. N. 1994. Logic programming and negation: A survey. *Journal of Logical Programming* 19/20, 9–71.
- ARAVINDAN, C. AND DUNG, P. M. 1995. On the correctness of unfold/fold transformation of normal and extended logic programs. *Journal of Logical Programming* 24, 3, 201–217.
- AREF, M., TEN CATE, B., GREEN, T. J., KIMELFELD, B., OLTEANU, D., PASALIC, E., VELDHUIZEN, T. L. AND WASHBURN, G. 2015. Design and implementation of the LogicBlox system. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, T. K. Sellis, S. B. Davidson, and Z. G. Ives, Eds. ACM, 1371–1382.
- AVGUSTINOV, P., DE MOOR, O., JONES, M. P. AND SCHÄFER, M. 2016. QL: object-oriented queries on relational data. In *30th European Conference on Object-Oriented Programming, ECOOP 2016*, S. Krishnamurthi and B. S. Lerner, Eds. LIPIcs, vol. 56. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2:1–2:25.
- BASSILIADIS, N., ANTONIOU, G. AND VLAHAVAS, I. P. 2006. A defeasible logic reasoner for the semantic web. *Journal on Semantic Web and Information Systems* 2, 1, 1–41.
- BELLOMARINI, L., SALLINGER, E. AND GOTTLÖB, G. 2018. The Vadalog system: Datalog-based reasoning for knowledge graphs. *Proceedings of the VLDB Endowment* 11, 9, 975–987.
- BEMBENEK, A., GREENBERG, M. AND CHONG, S. 2020. Formulog: Datalog for smt-based static analysis. *Proceedings of the ACM on Programming Languages* 4, OOPSLA, 141:1–141:31.
- BILLINGTON, D., ANTONIOU, G., GOVERNATORI, G. AND MAHER, M. J. 2010. An inclusion theorem for defeasible logics. *ACM Transactions on Computational Logic* 12, 1, 6.
- BILLINGTON, D. AND ROCK, A. 2001. Propositional plausible logic: Introduction and implementation. *Studia Logica* 67, 2, 243–269.
- BISHOP, B. AND FISCHER, F. 2008. IRIS - integrated rule inference system. In *Proc. Workshop on Advancing Reasoning on the Web: Scalability and Commonsense*. Vol. 350. CEUR Workshop Proceedings.
- BOSSI, A., COCCO, N. AND ETALLE, S. 1992. Transforming normal programs by replacement. In *Proc. Meta-Programming in Logic, 3rd International Workshop*. Lecture Notes in Computer Science, vol. 649. Springer, 265–279.
- BRASS, S. AND STEPHAN, H. 2017. Pipelined bottom-up evaluation of Datalog programs: The push method. In *Perspectives of System Informatics - 11th International Andrei P. Ershov Informatics Conference, PSI 2017*, A. K. Petrenko and A. Voronkov, Eds. Lecture Notes in Computer Science, vol. 10742. Springer, 43–58.
- BREWKA, G. 1996. Well-founded semantics for extended logic programs with dynamic preferences. *Journal of Artificial Intelligence Research* 4, 19–36.
- BREWKA, G. 2001. On the relationship between defeasible logic and well-founded semantics. In *LPNMR*. 121–132.
- CALIMERI, F., FUSCÀ, D., PERRI, S., AND ZANGARI, J. 2017. I-DLV: the new intelligent grounder of DLV. *Intelligenza Artificiale* 11, 1, 5–20.
- CARRAL, D., DRAGOSTE, I., GONZÁLEZ, L., JACOBS, C. J. H., KRÖTZSCH, M. AND URBANI, J. 2019. Vlog: A rule engine for knowledge graphs. In *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Part II*, C. Ghidini, O. Hartig, M. Maleshkova, V. Svátek, I. F. Cruz, A. Hogan, J. Song, M. Lefrançois, and F. Gandon, Eds. Lecture Notes in Computer Science, vol. 11779. Springer, 19–35.

- CAT, B. D., BOGAERTS, B., BRUYNOOGHE, M., JANSSENS, G. AND DENECKER, M. 2018. Predicate logic as a modeling language: the IDP system. In *Declarative Logic Programming: Theory, Systems, and Applications*, 279–323.
- CHEN, W. AND WARREN, D. S. 1996. Tabled evaluation with delaying for general logic programs. *Journal of the ACM* 43, 1, 20–74.
- CHIN, B., VON DINCKLAGE, D., ERCEGOVAC, V., HAWKINS, P., MILLER, M. S., OCH, F. J., OLSTON, C. AND PEREIRA, F. 2015. Yedalog: Exploring knowledge at scale. In *1st Summit on Advances in Programming Languages, SNAPL 2015, May 3-6, 2015, Asilomar, California, USA*, T. Ball, R. Bodík, S. Krishnamurthi, B. S. Lerner and G. Morrisett, Eds. LIPIcs, vol. 32. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 63–78.
- COGNITECT. What is datomic cloud? <https://docs.datomic.com/cloud/index.html>. Accessed 8/2/2020.
- CONDIE, T., DAS, A., INTERLANDI, M., SHKAPSKY, A., YANG, M. AND ZANIOLO, C. 2018. Scaling-up reasoning and advanced analytics on bigdata. *Theory and Practice of Logic Programming* 18, 5–6, 806–845.
- COSTA, V. S., ROCHA, R. AND DAMAS, L. 2012. The YAP Prolog system. *Theory and Practice of Logic Programming* 12, 1–2, 5–34.
- COVINGTON, M., NUTE, D. AND VELLINO, A. 1997. *Prolog Programming in Depth*. Prentice-Hall.
- DEAN, J. AND GHEMAWAT, S. 2004. MapReduce: Simplified data processing on large clusters. In *6th Symposium on Operating System Design and Implementation (OSDI)*. USENIX Association, Berkeley, CA, USA, 10–10.
- DUNG, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77, 2, 321–358.
- EISNER, J. AND FILARDO, N. W. 2010. Dyna: Extending Datalog for modern AI. In *Datalog Reloaded*. 181–220.
- EITER, T., LEONE, N. AND SACCÀ, D. 1997. On the partial semantics for disjunctive deductive databases. *Annals of Mathematics and Artificial Intelligence* 19, 1–2, 59–96.
- FAN, Z., ZHU, J., ZHANG, Z., ALBARGHOUTHI, A., KOUTRIS, P. AND PATEL, J. M. 2019. Scaling-up in-memory Datalog processing: Observations and techniques. *PVLDB* 12, 6, 695–708.
- FILARDO, N. W. 2017. Dyna 2: Towards a general weighted logic language. Ph.D. thesis, Johns Hopkins University.
- FITTING, M. 1985. A Kripke-Kleene semantics for logic programs. *Journal of Logical Programming* 2, 4, 295–312.
- GANDHE, M., FININ, T. AND GROSOFF, B. 2002. SweetJess: Translating DamlRuleML to Jess. In *International Workshop on Rule Markup Languages for Business Rules on the Semantic Web in conjunction with ISWC2002*. Sardinia, Italy.
- GEBSER, M., KAMINSKI, R., KAUFMANN, B. AND SCHAUB, T. 2019. Multi-shot ASP solving with clingo. *Theory and Practice of Logic Programming* 19, 1, 27–82.
- GEBSER, M., KAMINSKI, R., KÖNIG, A. AND SCHAUB, T. 2011. Advances in *gringo* series 3. In *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR*, 345–351.
- GELFOND, M. AND LIFSCHITZ, V. 1988. The stable model semantics for logic programming. In *Proc. JICSLP*, 1070–1080.
- GOVERNATORI, G. AND MAHER, M. J. 2017. Annotated defeasible logic. *Theory and Practice of Logic Programming* 17, 5–6, 819–836.
- GOVERNATORI, G., MAHER, M. J., ANTONIOU, G. AND BILLINGTON, D. 2004. Argumentation semantics for defeasible logic. *Journal of Logical Computing* 14, 5, 675–702.

- GROSOFF, B. AND KIFER, M. 2013. Rulelog: Syntax and semantics. <http://ruleml.org/rif/rulelog/spec/Rulelog.html>. Accessed: April 2015.
- GROSOFF, B. N. 1997. Prioritized conflict handling for logic programs. In *ILPS*. 197–211.
- HODER, K., BJØRNER, N. AND DE MOURA, L. M. 2011. μZ - an efficient engine for fixed points with constraints. In *Computer Aided Verification - 23rd International Conference, CAV*. 457–462.
- JAFFAR, J. AND MAHER, M. J. 1994. Constraint logic programming: A survey. *The Journal of Logic Programming* 19 & 20, 503–582.
- JORDAN, H., SCHOLZ, B. AND SUBOTIC, P. 2016. Soufflé: On synthesis of program analyzers. In *Computer Aided Verification*. 422–430.
- KAKAS, A. C. AND MANCARELLA, P. 1991. Negation as stable hypotheses. In *Logic Programming and Non-monotonic Reasoning, Proceedings of the First International Workshop*, 275–288.
- KAMINSKI, R. 2020. personal communication.
- KIFER, M., YANG, G., WAN, H. AND ZHAO, C. 2018. Flora-2 documentation. <http://flora.sourceforge.net/documentation.html>. Accessed 14/4/2020.
- KUNEN, K. 1987. Negation in logic programming. *Journal of Logical Programming* 4, 4, 289–308.
- KUNEN, K. 1989. Signed data dependencies in logic programs. *Journal of Logical Programming* 7, 3, 231–245.
- LAM, H. AND GOVERNATORI, G. 2009. The making of SPINdle. In *Rule Interchange and Applications, International Symposium, RuleML, Proceedings*. 315–322.
- MADSEN, M. AND LHOTÁK, O. 2020. Fixpoints for the masses: programming with first-class datalog constraints. *Proceedings of the ACM on Programming Languages* 4, OOPSLA, 125:1–125:28.
- MADSEN, M., YEE, M. AND LHOTÁK, O. 2016. From Datalog to Flix: a declarative language for fixed points on lattices. In *Proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2016*, C. Krintz and E. Berger, Eds. ACM, 194–208.
- MAHER, M. J. 1988. Correctness of a logic program transformation system. Tech. rep., IBM T.J. Watson Research Center.
- MAHER, M. J. 1990. Reasoning about stable models (and other unstable semantics). Tech. rep.
- MAHER, M. J. 1993. A transformation system for deductive databases modules with perfect model semantics. *Theoretical Computer Science* 110, 2, 377–403.
- MAHER, M. J. 2002. A model-theoretic semantics for defeasible logic. In *Paraconsistent Computational Logic*, H. Decker, J. Villadsen, and T. Waragai, Eds. Datalogiske Skrifter, vol. 95. Roskilde University, Roskilde, Denmark, 67–80.
- MAHER, M. J. 2014. Comparing defeasible logics. In *21st European Conference on Artificial Intelligence*, 585–590.
- MAHER, M. J. 2017. Relating concrete defeasible reasoning formalisms and abstract argumentation. *Fundamenta Informaticae* 155, 3, 233–260.
- MAHER, M. J. 2021. On signings and the well-founded semantics. *Theory and Practice of Logic Programming* accepted for publication.
- MAHER, M. J. AND GOVERNATORI, G. 1999. A semantic decomposition of defeasible logics. In *AAAI/IAAI*. AAAI Press, 299–305.
- MAHER, M. J., ROCK, A., ANTONIOU, G., BILLINGTON, D. AND MILLER, T. 2001. Efficient defeasible reasoning systems. *International Journal on Artificial Intelligence Tools* 10, 4, 483–501.
- MAHER, M. J., TACHMAZIDIS, I., ANTONIOU, G., WADE, S. AND CHENG, L. 2020. Rethinking defeasible reasoning: A scalable approach. *Theory and Practice of Logic Programming*. 20, 4, 552–586.
- MAIER, F. 2013. Interdefinability of defeasible logic and logic programming under the well-founded semantics. *Theory and Practice of Logic Programming* 13, 1, 107–142.

- MAIER, F. AND NUTE, D. 2006. Ambiguity propagating defeasible logic and the well-founded semantics. In *JELIA*, M. Fisher, W. van der Hoek, B. Konev, and A. Lisitsa, Eds. Lecture Notes in Computer Science, vol. 4160. Springer, 306–318.
- MAIER, F. AND NUTE, D. 2010. Well-founded semantics for defeasible logic. *Synthese* 176, 2, 243–274.
- MARTINEZ-ANGELES, C. A., DE CASTRO DUTRA, I., COSTA, V. S., AND BUENABAD-CHÁVEZ, J. 2013. A Datalog engine for GPUs. In *Declarative Programming and Knowledge Management*. 152–168.
- MARZ, N. 2013. Cascalog. <https://www.cascading.org/projects/cascalog/>. Accessed: April 2020.
- NGUYEN, H. D., SAKAMA, C., SATO, T., AND INOUE, K. 2018. Computing logic programming semantics in linear algebra. In *Multi-disciplinary Trends in Artificial Intelligence - 12th International Conference, MIWAI*, 32–48.
- NIEMELÄ, I. AND SIMONS, P. 1997. Smodels - An implementation of the stable model and well-founded semantics for normal LP. In *Logic Programming and Nonmonotonic Reasoning, 4th International Conference, LPNMR'97*, J. Dix, U. Furbach, and A. Nerode, Eds. Lecture Notes in Computer Science, vol. 1265. Springer, 421–430.
- NUTE, D. 1993. Defeasible Prolog. In *Proc. AAAI Fall Symposium on Automated Deduction in Nonstandard Logics*, 105–112.
- PRZYMUSINSKI, T. C. 1990. The well-founded semantics coincides with the three-valued stable semantics. *Fundamenta Informaticae* 13, 4, 445–463.
- SATO, T. 2017. A linear algebraic approach to Datalog evaluation. *Theory and Practice of Logic Programming* 17, 3, 244–265.
- SATO, T., SAKAMA, C. AND INOUE, K. 2020. From 3-valued semantics to supported model computation for logic programs in vector spaces. In *Proceedings of the 12th International Conference on Agents and Artificial Intelligence, ICAART 2020*. 758–765.
- SEO, J., GUO, S. AND LAM, M. S. 2015. Socialite: An efficient graph query language based on Datalog. *IEEE Transactions on Knowledge and Data Engineering* 27, 7, 1824–1837.
- SWIFT, T. AND WARREN, D. S. 2012. XSB: Extending Prolog with tabled logic programming. *Theory and Practice of Logic Programming* 12, 1-2, 157187.
- SWIFT, T., WARREN, D. S., *et al.* 2017. The XSB System, Version 3.8.x, Volume 1: Programmers Manual. Tech. rep.
- TACHMAZIDIS, I. AND ANTONIOU, G. 2013. Computing the stratified semantics of logic programs over big data through mass parallelization. In *RuleML*. Lecture Notes in Computer Science, vol. 8035. Springer, 188–202.
- TACHMAZIDIS, I., ANTONIOU, G. AND FABER, W. 2014. Efficient computation of the well-founded semantics over big data. *Theory and Practice of Logic Programming* 14, 4-5, 445–459.
- TACHMAZIDIS, I., ANTONIOU, G., FLOURIS, G., KOTOULAS, S., AND MCCLUSKEY, L. 2012. Large-scale parallel stratified defeasible reasoning. In *ECAI 2012 - 20th European Conference on Artificial Intelligence*, 738–743.
- TACHMAZIDIS, I., CHENG, L., KOTOULAS, S., ANTONIOU, G. AND WARD, T. E. 2014. Massively parallel reasoning under the well-founded semantics using X10. In *26th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2014*. IEEE Computer Society, 162–169.
- TAMAKI, H. AND SATO, T. 1984. Unfold/fold transformation of logic programs. In *Proc. ICLP*, 127–138.
- VAN GELDER, A., ROSS, K. A. AND SCHLIPF, J. S. 1991. The well-founded semantics for general logic programs. *J. ACM* 38, 3, 620–650.
- WAN, H., GROSOFF, B. N., KIFER, M., FODOR, P. AND LIANG, S. 2009. Logic programming with defaults and argumentation theories. In *ICLP*, P. M. Hill and D. S. Warren, Eds. Lecture Notes in Computer Science, vol. 5649. Springer, 432–448.

- WANG, J., BALAZINSKA, M., AND HALPERIN, D. 2015. Asynchronous and fault-tolerant recursive Datalog evaluation in shared-nothing engines. *Proceedings of the VLDB Endowment* 8, 12, 1542–1553.
- WANG, K., HUSSAIN, A., ZUO, Z., XU, G. H. AND SANI, A. A. 2017. Graspan: A single-machine disk-based graph system for interprocedural static analyses of large-scale systems code. In *Proceedings of the Twenty-Second International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS*, 389–404.
- WENZEL, M. AND BRASS, S. 2019. Declarative programming for microcontrollers - Datalog on Arduino. In *Declarative Programming and Knowledge Management - Conference on Declarative Programming, DECLARE 2019, Unifying INAP, WLP, and WFLP*, P. Hofstedt, S. Abreu, U. John, H. Kuchen and D. Seipel, Eds. Lecture Notes in Computer Science, vol. 12057. Springer, 119–138.
- WHALEY, J., AVOTS, D., CARBIN, M. AND LAM, M. S. 2005. Using Datalog with binary decision diagrams for program analysis. In *Programming Languages and Systems, Third Asian Symposium, APLAS*, 97–118.
- WU, H., DIAMOS, G. F., SHEARD, T., AREF, M., BAXTER, S., GARLAND, M. AND YALAMAN-CHILI, S. 2014. Red fox: An execution environment for relational query processing on gpus. In *12th Annual IEEE/ACM International Symposium on Code Generation and Optimization, CGO*, D. R. Kaeli and T. Moseley, Eds. ACM, 44.
- YOU, J. AND YUAN, L. 1994. A three-valued semantics for deductive databases and logic programs. *Journal of Computer and System Sciences* 49, 2, 334–361.

Appendix A The metaprogram

The presentation of $\mathcal{M}_{\partial_{||}}$ in the body of the paper sacrifices correct syntax for readability. In this appendix we present $\mathcal{M}_{\partial_{||}}$ in correct logic programming syntax. This involves a number of auxiliary predicates.

The main clauses of $\mathcal{M}_{\partial_{||}}$ are as follows:

- ```

c45 definitely(X):-
 fact(X).

c46 definitely(X):-
 strict(R, X, Y),
 loop_definitely(Y).

c47 lambda(X):-
 definitely(X).

c48 lambda(X):-
 neg(X, X'),
 not definitely(X'),
 strict_or_defeasible(R, X, Y),
 loop_lambda(Y).

c49 defeasibly(X):-
 definitely(X).

c50 defeasibly(X):-
 neg(X, X'),
 not definitely(X'),

```

```

strict_or_defeasible(R, X, Y),
loop_defeasibly(Y),
not overruled(R, X).

```

```

c51 overruled(R, X):-
 neg(X, X'),
 rule(S, X', U),
 loop_lambda(U),
 not defeated(S, X').

```

```

c52 defeated(S, X'):-
 neg(X, X'),
 sup(T, S),
 strict_or_defeasible(T, X, V),
 loop_defeasibly(V).

```

We still need to define the predicates not defined above. There are additional clauses to represent sets of rules.

```

c53 rule(R, H, B):-
 strict_or_defeasible(R, H, B).

```

```

c54 rule(R, H, B):-
 defeater(R, H, B).

```

```

c55 strict_or_defeasible(R, H, B):-
 strict(R, H, B).

```

```

c56 strict_or_defeasible(R, H, B):-
 defeasible(R, H, B).

```

To express the complement of a literal  $\sim q$  we define, for each predicate  $p$  in  $D$ ,

```

c57 neg(p(...), not_p(...)).

```

```

c58 neg(not_p(...), p(...)).

```

The auxiliary predicate `loop_defeasibly` (`loop_lambda`, `loop_definitely`) maps the representation of a rule body to a corresponding sequence of calls to `defeasibly` (respectively, `lambda`, `definitely`).

```

c59 loop_defeasibly([]).

```

```

c60 loop_defeasibly([H|T]) :- defeasibly(H), loop_defeasibly(T).

```

```

c61 loop_lambda([]).

```

```

c62 loop_lambda([H|T]) :- lambda(H), loop_lambda(T).

```

```

c63 loop_definitely([]).

```

```

c64 loop_definitely([H|T]) :- definitely(H), loop_definitely(T).

```