Superposition for Lambda-Free Higher-Order Logic (Technical Report)

Alexander Bentkamp[1(✉)], Jasmin Christian Blanchette[1,2,3], Simon Cruanes[4,3], and Uwe Waldmann[2]

[1] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
a.bentkamp@vu.nl
[2] Max-Planck-Institut für Informatik, Saarland Informatics Campus, Saarbrücken, Germany
[3] Université de Lorraine, CNRS, Inria, LORIA, Nancy, France
[4] Aesthetic Integration, Austin, Texas

Abstract. We introduce refutationally complete superposition calculi for intentional and extensional λ-free higher-order logic, a formalism that allows partial application and applied variables. The intentional variants perfectly coincide with standard superposition on first-order clauses. The calculi are parameterized by a well-founded term order that need not be compatible with arguments, making it possible to employ the λ-free higher-order lexicographic path and Knuth–Bendix orders. We implemented the calculi in the Zipperposition prover and evaluated them on TPTP benchmarks. They appear promising as a stepping stone towards complete, efficient automatic theorem provers for full higher-order logic.

1 Introduction

Superposition is a highly successful calculus for reasoning about first-order logic with equality. We are interested in *graceful* generalizations to higher-order logic: calculi that, as much as possible, coincide with standard superposition on first-order problems and that scale up to arbitrary higher-order problems.

As a stepping stone towards full higher-order logic, in this report we focus on *λ-free higher-order logic* (Section 2), a fragment that supports partial application and application of variables. This formalism is expressive enough to permit the axiomatization of higher-order combinators such as $\mathsf{pow}_\tau : \mathit{nat} \to (\tau \to \tau) \to \tau \to \tau$ (intended to denote the iterated application $h^n\, x$):

$$\mathsf{pow}\ 0\ h \approx \mathsf{id} \qquad\qquad \mathsf{pow}\ (\mathsf{S}\ n)\ h\ x \approx h\ (\mathsf{pow}\ n\ h\ x)$$

Conventionally, functions are applied without parentheses and commas, and variables are italicized. Notice the variable number of arguments to pow and the application of h. The expressiveness of full higher-order logic can be recovered by introducing SK-style combinators to represent λ-abstractions and proxies for the logical symbols [29, 39].

A widespread technique to support partial application and application of variables in first-order logic is to make all symbols nullary and to represent application of functions of type $\tau \to \upsilon$ by a family of binary symbols $\mathsf{app}_{\tau,\upsilon}$. Following this scheme, the higher-order term $\mathsf{f}\ (h\ \mathsf{f})$ is translated to $\mathsf{app}(\mathsf{f}, \mathsf{app}(h, \mathsf{f}))$, which can be processed by first-order methods. We call this the *applicative encoding*. The existence of such a reduction explains why λ-free higher-order terms are also called "applicative first-order terms."

Although the applicative encoding is complete [29] and is employed fruitfully in tools such as Sledgehammer [8, 33], it suffers from a number of weaknesses, all related to its gracelessness. Transforming all the function symbols into constants considerably restricts what can be achieved with term orders; for example, argument tuples cannot be compared using different methods for different symbols [30, Section 2.3.1]. In a prover, the encoding also clutters the data structures, slows down the algorithms, and neutralizes the heuristics that look at the terms' root symbols. But our chief objection is the sheer clumsiness of encodings and their poor integration with interpreted symbols. And they quickly accumulate; for example, using the traditional encoding of polymorphism relying on a distinguished binary function symbol t [7, Section 3.3] in conjunction with the applicative encoding, the term $\mathsf{S}\ x$ becomes $\mathsf{t}(\mathsf{nat}, \mathsf{app}(\mathsf{t}(\mathsf{fun}(\mathsf{nat}, \mathsf{nat}), \mathsf{S}), \mathsf{t}(\mathsf{nat}, x)))$. The term's simple structure is entirely lost in translation.

Hybrid schemes have been proposed to strengthen the applicative encoding: If a given symbol always occurs with at least k arguments, these can be passed directly [33]. However, this relies on a closed-world assumption: that all terms that will ever be compared arise in the input problem. This noncompositionality conflicts with the need for complete higher-order calculi to synthesize arbitrary terms during proof search [5]. For these reasons, the applicative encoding is not an ideal basis for higher-order automated reasoning. Instead, we propose to generalize superposition to *intensional* and *extensional* λ-free higher-order logic. In the extensional version of the logic, the property $(\forall x.\ h\,x \approx k\,x) \longrightarrow h \approx k$ holds for all functions h, k of the same type. For each logic, we present two calculi (Section 3). The intentional calculi perfectly coincide with standard superposition on first-order clauses; the extensional calculi depend on an extra axiom.

Superposition is parameterized by a term order, which has a dramatic impact on search space exploration. If we assume that the term order is a simplification order enjoying totality on ground terms, the standard calculus rules and the completeness proof can be lifted verbatim. The only necessary changes concern the basic definitions of terms and substitutions. Unlike for full higher-order logic, most general unifiers exist for λ-free higher-order logic, just as they do for applicatively encoded first-order terms.

However, there is one monotonicity property that is hard to obtain unconditionally: compatibility with arguments. It states that $t \succ s$ implies $t\,u \succ s\,u$ for all terms s, t, u such that $s\,u$ and $t\,u$ are also well typed. We recently introduced graceful generalizations of the lexicographic path order (LPO) [10] and the Knuth–Bendix order (KBO) [3] with argument coefficients, but they both lack this property. For example, given a KBO with $\mathsf{g} \succ \mathsf{f}$, it may well be that $\mathsf{g}\ \mathsf{a} \prec \mathsf{f}\ \mathsf{a}$ if f has a large enough multiplier on its first argument.

Our superposition calculi are designed to be refutationally complete for such nonmonotonic orders (Section 4). To achieve this, they include an inference rule for argument congruence, which derives $C \vee s\,x \approx t\,x$ from $C \vee s \approx t$. The redundancy criterion must be defined in such a way that the larger, derived clause is not subsumed by the premise. In the completeness proof, the most difficult case is the one that normally excludes superposition at or below variables using the induction hypothesis. With nonmonotonicity, this approach no longer works, and we propose two alternatives: Perform some superposition inferences onto higher-order variables, or "purify" the clauses to circumvent the issue. We refer to the corresponding calculi as *nonpurifying* and *purifying*.

The calculi are implemented in the Zipperposition prover [18] (Section 5). We evaluate them on first- and higher-order TPTP benchmarks [47,48] and compare them with the applicative encoding (Section 6). We find that there is a substantial cost associated with the applicative encoding, that the nonmonotonicity is not particularly expensive, and that the nonpurifying calculi outperform the purifying variants.

2 Lambda-Free Higher-Order Logic

Refutational completeness of calculi for higher-order logic (also called simple type theory) [16, 24] is usually stated with respect to Henkin semantics [5, 26], in which the universes used to interpret functions need only contain the functions that are expressible as terms. Since the terms of λ-free higher-order logic exclude λ-abstractions, in "λ-free Henkin semantics" the universes interpreting functions can be even smaller.

Problematically, in a logic with applied variables but without Hilbert choice, skolemization is unsound, unless we make sure that Skolem symbols are suitably applied [34]. We achieve this using a *hybrid logic* that supports both mandatory (uncurried) and optional (curried) arguments, inspired by higher-order term rewriting [30]. Thus, if symbol sk takes two mandatory and one optional arguments, $\mathsf{sk}(x,y)$ and $\mathsf{sk}(x,y)\ z$ are valid terms, whereas sk and $\mathsf{sk}(x)$ are invalid. Nevertheless, as in our earlier work [3,10], we use the adjective "graceful" in the strong sense that we can exploit optional arguments, identifying the first-order term $\mathsf{f}(x,y)$ with the curried higher-order term $\mathsf{f}\ x\ y$.

The types of higher-order logic are defined by the grammar $\tau, \upsilon ::= o \mid \iota \mid \tau \to \upsilon$, where o is the type of Booleans, ι is an element of a fixed set of atomic types, and $\tau \to \upsilon$ is the type of functions from type τ to type υ. In our hybrid logic, a type declaration for a symbol is an expression of the form $\bar{\tau}_n \Rightarrow \tau$ (or simply τ if $n = 0$). Here and elsewhere, we write $\bar{a}_n$ or $\bar{a}$ to abbreviate the tuple $(a_1, \ldots, a_n)$ or product $a_1 \times \cdots \times a_n$, for $n \geq 0$.

We fix a set $\mathcal{V}$ of typed variables, denoted by $x : \tau$ or x. A signature consists of a nonempty set Σ of symbols with type declarations, written as $\mathsf{f} : \bar{\tau} \Rightarrow \tau$ or f. We reserve the letters s,t,u,v,w for terms and x,y,z for variables and write $: \tau$ to indicate their type. The set of λ-free higher-order terms $\mathcal{T}_\Sigma^X$ over X is defined inductively as follows. Every variable in $X \subseteq \mathcal{V}$ is a term. If $\mathsf{f} : \bar{\tau}_n \Rightarrow \tau$ and $u_i : \tau_i$ for all $i \in \{1,\ldots,n\}$, then $\mathsf{f}(\bar{u}_n) : \tau$ is a term. If $t : \tau \to \upsilon$ and $u : \tau$, then $t\ u : \upsilon$ is a term, called an *application*. Non-application terms ζ are called *heads*. Using the spine notation [15], terms can be decomposed in a unique way as a head ζ applied to zero or more arguments: $\zeta\ s_1 \ldots s_n$ or $\zeta\ \bar{s}_n$ (abusing notation). Substitution and unification are generalized in the obvious way, without the complexities associated with λ-abstractions; for example, the most general unifier of $x\ \mathsf{b}\ z$ and $\mathsf{f}\ \mathsf{a}\ y\ \mathsf{c}$ is $\{x \mapsto \mathsf{f}\ \mathsf{a}, y \mapsto \mathsf{b}, z \mapsto \mathsf{c}\}$.

Formulas φ, ψ are terms of type o, extended with quantifications $\forall x.\ \varphi$ and $\exists x.\ \varphi$, where x is a variable and φ is a formula. The familiar logical symbols $\bot$, $\top$, $\neg$, $\vee$, $\wedge$, $\longrightarrow$, and $\approx_\tau$ are interpreted. We let $s \not\approx t$ abbreviate $\neg\ s \approx t$. For superposition, we are interested in a *clausal logic* fragment based on standard clauses, which are either $\bot$ or disjunctions of literals $[\neg]\, s \approx t$, where the terms $s,t : \tau$ are built without using o. We normally view equations $s \approx t$ as unordered pairs and clauses as finite multisets of such (dis)equations.

Loosely following Fitting [23], an *interpretation* $\mathcal{J} = (\mathcal{U}, \mathcal{E}, \mathcal{J})$ consists of a type-indexed family of nonempty sets $\mathcal{U}_\tau$, called *universes*; a family of functions $\mathcal{E}_{\tau,\upsilon} : \mathcal{U}_{\tau\to\upsilon} \to (\mathcal{U}_\tau \to \mathcal{U}_\upsilon)$, one for each pair of types τ, υ; and a function $\mathcal{J}$ that maps each symbol with type declaration $\bar{\tau}_n \Rightarrow \tau$ to an element of $\overline{\mathcal{U}}_{\tau_n} \to \mathcal{U}_\tau$. We require $\mathcal{U}_o = \{0,1\}$. An interpretation is *extensional* if $\mathcal{E}_{\tau,\upsilon}$ is injective for all τ, υ. This semantics is *standard* if $\mathcal{E}_{\tau,\upsilon}$ is bijective. A *valuation* ξ is a function that maps variables $x : \tau$ to elements of $\mathcal{U}_\tau$.

For an interpretation $(\mathcal{U}, \mathcal{E}, \mathcal{J})$ and a valuation ξ, the denotation of a term is defined as follows: $[\![x]\!]^\xi_\mathcal{J} = \xi(x)$; $[\![\mathsf{f}(\bar{t})]\!]^\xi_\mathcal{J} = \mathcal{J}(\mathsf{f})([\![\bar{t}]\!]^\xi_\mathcal{J})$; $[\![s\ t]\!]^\xi_\mathcal{J} = \mathcal{E}([\![s]\!]^\xi_\mathcal{J})([\![t]\!]^\xi_\mathcal{J})$. The truth value $[\![\varphi]\!]^\xi_\mathcal{J} \in \{0,1\}$ of a formula φ is defined as in first-order logic:

$$[\![\forall(x:\tau)\ \psi]\!]^\xi_\mathcal{J} = \min_{a\in\mathcal{U}_\tau}\{[\![\psi]\!]^{\xi[x\mapsto a]}_\mathcal{J}\} \qquad [\![\exists(x:\tau)\ \psi]\!]^\xi_\mathcal{J} = \max_{a\in\mathcal{U}_\tau}\{[\![\psi]\!]^{\xi[x\mapsto a]}_\mathcal{J}\}$$

A formula φ is true in $\mathcal{J} = (\mathcal{U}, \mathcal{E}, \mathcal{J})$ under valuation ξ and we write $\mathcal{J}, \xi \models \varphi$ if $[\![\varphi]\!]^\xi_\mathcal{J} = 1$. The interpretation $\mathcal{J}$ is a model of φ, written $\mathcal{J} \models \varphi$, if $\mathcal{J}, \xi \models \varphi$ for all valuations ξ into $\{\mathcal{U}_\tau\}_\tau$.

3 The Inference Systems

We introduce four versions of the superposition calculus, varying along two axes: intentional versus extensional, and nonpurifying versus purifying. To avoid repetitions, our presentation unifies them into a single framework.

3.1 The Inference Rules

The calculi are parameterized by a partial order $\succ$ on terms that is well founded, total on ground terms, and stable under substitutions and that has the subterm property. It must also be *compatible with argument contexts*, meaning that $t' \succ t$ implies both $\mathsf{f}(\bar{s}, t', \bar{u})\ \bar{v} \succ \mathsf{f}(\bar{s}, t, \bar{u})\ \bar{v}$ and $s\ t'\ \bar{u} \succ s\ t\ \bar{u}$. On the other hand, it need not be *compatible with* (*optional*) *arguments*: $s' \succ s$ need not imply $s'\ t \succ s\ t$. Argument contexts correspond to *argument subterms*, defined as the reflexive transitive closure of the relation inductively specified by $\mathsf{f}(\bar{s})\ \bar{t} \rhd s_i$ and $\zeta\ \bar{t} \rhd t_i$ for all i. We write $s\langle u\rangle$ to indicate that u is an argument subterm in $s[u]$. For example, f and $\mathsf{f\ a}$ are subterms of $\mathsf{f\ a\ b}$, but not argument subterms. The literal and clause orders are defined from the term order as multiset extensions in the usual way.

Literal selection is supported. The selection function maps each clause C to a subclause of C consisting of negative literals. A literal L is (*strictly*) *eligible* in C if it is selected in C or there are no selected literals in C and L is (strictly) maximal in C.

We start with the **extensional nonpurifying** calculus, which consists of the five rules and the extensionality axiom given on page 5. We view positive and negative superposition as two cases of one rule called SUP. We have two rules and one axiom in addition to the standard first-order rules and their usual order conditions.

Definition 1. A term of the form $x\ \bar{s}_n$, for $n \geq 0$, *jells* with a literal $t \approx t' \in D$ if $t = \tilde{t}\ \bar{y}_n$ and $t' = \tilde{t}'\ \bar{y}_n$ for some terms $\tilde{t}$, $\tilde{t}'$ and distinct variables $\bar{y}_n$ that do not occur elsewhere in D.

Positive superposition:

$$\frac{\overbrace{D' \vee t \approx t'}^{D} \quad \overbrace{C' \vee s\langle u\rangle \approx s'}^{C}}{(D' \vee C' \vee s\langle t'\rangle \approx s')\sigma}\ \textsc{Sup}$$

- $\sigma = \mathrm{mgu}(t, u)$
- $t\sigma \not\preceq t'\sigma$ and $s\langle u\rangle\sigma \not\preceq s'\sigma$
- $(t \approx t')\sigma$ is strictly eligible in $D\sigma$
- $(s\langle u\rangle \approx s')\sigma$ is strictly eligible in $C\sigma$
- $C\sigma \not\preceq D\sigma$
- the variable conditions holds

Negative superposition:

$$\frac{\overbrace{D' \vee t \approx t'}^{D} \quad \overbrace{C' \vee s\langle u\rangle \not\approx s'}^{C}}{(D' \vee C' \vee s\langle t'\rangle \not\approx s')\sigma}\ \textsc{Sup}$$

- $\sigma = \mathrm{mgu}(t, u)$
- $t\sigma \not\preceq t'\sigma$ and $s\langle u\rangle\sigma \not\preceq s'\sigma$
- $(t \approx t')\sigma$ is strictly eligible in $D\sigma$
- $(s\langle u\rangle \not\approx s')\sigma$ is eligible in $C\sigma$
- $C\sigma \not\preceq D\sigma$
- the variable condition holds

Equality resolution:

$$\frac{C' \vee s \not\approx s'}{C'\sigma}\ \textsc{EqRes}$$

- $\sigma = \mathrm{mgu}(s, s')$
- $(s \not\approx s')\sigma$ is eligible in the premise

Equality factoring:

$$\frac{C' \vee s' \approx t' \vee s \approx t}{(C' \vee t \not\approx t' \vee s \approx t')\sigma}\ \textsc{EqFact}$$

- $\sigma = \mathrm{mgu}(s, s')$
- $s'\sigma \not\preceq t'\sigma$ and $s\sigma \not\preceq t\sigma$
- $(s \approx t)\sigma$ is eligible in the premise

Argument congruence:

$$\frac{C' \vee s \approx s'}{C' \vee s\ \bar{x} \approx s'\ \bar{x}}\ \textsc{ArgCong}$$

- $\bar{x}$ contains fresh variables
- $s \approx s'$ is strictly eligible in the premise

Positive extensionality:

$$\frac{C' \vee s\ \bar{x} \approx s'\ \bar{x}}{C' \vee s \approx s'}\ \textsc{PosExt}$$

- $\bar{x}$ is a tuple of variables that occur nowhere else in the premise
- $s\ \bar{x} \approx s'\ \bar{x}$ is strictly eligible in the premise

Extensionality axiom: For every function type $\tau \to \upsilon$, we introduce a Skolem symbol $\mathsf{diff}_{\tau,\upsilon} : (\tau \to \upsilon)^2 \Rightarrow \tau$ characterized by the axiom

$$x\,(\mathsf{diff}(x,y)) \not\approx y\,(\mathsf{diff}(x,y)) \vee x \approx y$$

We add the following *variable condition* as a side condition to the SUP rules, to further prune the search space, using the naming convention from Definition 1 for $\tilde{t}'$:

> If u has a variable head x and jells with the literal $t \approx t' \in D$, there exists a ground substitution θ with $t\sigma\theta \succ t'\sigma\theta$ and $C\sigma\theta \prec C''\sigma\theta$, where $C'' = C[x \mapsto \tilde{t}']$.

This condition generalizes the standard condition that $u \notin \mathcal{V}$. The two coincide if C is first-order. In some cases involving nonmonotonicity, the variable condition effectively mandates SUP inferences at variable positions, but never below.

The second calculus is the **intensional nonpurifying** variant. We obtain it by removing the POSEXT rule and the extensionality axiom and by replacing the variable condition with "if $u \in \mathcal{V}$, there exists a ground substitution θ with $t\sigma\theta \succ t'\sigma\theta$ and $C\sigma\theta \prec C[u \mapsto t']\sigma\theta$." For monotone term orders, this condition amounts to $u \notin \mathcal{V}$.

By contrast, the purifying calculi never perform superposition at variable positions. Instead, they rely on purification [13, 20, 40, 43] (also called abstraction) to circumvent nonmonotonicity. The idea is to rename apart problematic occurrences of a variable x in a clause to $x_1, \ldots, x_n$ and to add *purification literals* $x_1 \not\approx x$, $\ldots$, $x_n \not\approx x$ to connect the new variables. We must then ensure that all clauses are purified, by processing the initial clause set and the conclusion of every inference or simplification.

In the **extensional purifying** calculus, the purification $\mathit{pure}(C)$ of clause C is defined as the result of the following iterative procedure. Consider the literals of C excluding those of the form $y \not\approx z$. If these literals contain both $x\ \bar{u}$ and $x\ \bar{v}$ as distinct argument subterms, replace all argument subterms $x\ \bar{v}$ with $x'\ \bar{v}$, where x' is fresh, and add the purification literal $x' \not\approx x$. This calculus variant contains the POSEXT rule and the extensionality axiom. The conclusion E of each rule is changed to $\mathit{pure}(E)$, except for POSEXT, which preserves purity. Moreover, the variable condition is replaced by "either u has a non-variable head or u does not jell with the literal $t \approx t' \in D$."

In the **intensional purifying** calculus, we define $\mathit{pure}(C)$ iteratively as follows: If a variable x occurs both applied and unapplied in C, replace all unapplied occurrences of x by a fresh variable x' and add the purification literal $x' \not\approx x$. We remove the POSEXT rule and the extensionality axiom. The variable condition is replaced by "$u \notin \mathcal{V}$." The conclusion C of ARGCONG is changed to $\mathit{pure}(C)$; the other rules preserve purity.

Finally, we impose some additional restrictions on literal selection. In the nonpurifying variants, a literal may not be selected if $x\ \bar{u}$ is a maximal term of the clause and the literal contains an argument subterm $x\ \bar{v}$ with $\bar{v} \neq \bar{u}$. In the extensional purifying calculus, a literal may not be selected if it contains a variable that is applied to different arguments in the clause. In the intensional purifying calculus, a literal may not be selected if the literal contains an unapplied variable that also appears applied in the clause. These restrictions are needed for our completeness proof, but it might be possible to avoid them at the cost of a more elaborate argument.

3.2 Rationale for the Inference Rules

A key restriction of all four calculi is that they superpose only onto argument subterms, mirroring the requirement that the term order enjoy compatibility with argument contexts. The ARGCONG rule then makes it possible to simulate superposition onto non-argument subterms. However, in conjunction with the SUP rules, ARGCONG can exhibit an unpleasant behavior, which we call *argument congruence explosion*:

$$\textsc{Sup–}\dfrac{\textsc{ArgCong}\dfrac{\mathsf{g} \approx \mathsf{f}}{\mathsf{g}\,x \approx \mathsf{f}\,x} \quad h\,\mathsf{a} \not\approx \mathsf{b}}{\mathsf{f}\,\mathsf{a} \not\approx \mathsf{b}} \qquad \textsc{Sup–}\dfrac{\textsc{ArgCong}\dfrac{\mathsf{g} \approx \mathsf{f}}{\mathsf{g}\,x\,y\,z \approx \mathsf{f}\,x\,y\,z} \quad h\,\mathsf{a} \not\approx \mathsf{b}}{\mathsf{f}\,x\,y\,\mathsf{a} \not\approx \mathsf{b}}$$

In both cases, the higher-order variable h is effectively the target of a SUP inference. Such derivations essentially amount to superposition at variable positions (as shown on the left) or even superposition below variable positions (as shown on the right), both of which can be extremely prolific. In standard superposition, the explosion is averted by the condition on the SUP rule that $u \notin \mathcal{V}$. In the extensional purifying calculus, the variable condition tests that either u has a non-variable head or u does not jell with the literal $t \approx t' \in D$, which prevents derivations such as the above. In the corresponding nonpurifying variant, some such derivations may need to be performed when the term order exhibits nonmonotonicity for the terms of interest.

In the intensional calculi, the explosion can arise even for monotonic orders, and it must be tamed by heuristics. The reason is connected to the absence of the POSEXT rule (which would be unsound). The variable condition in the extensional calculi is designed to prevent derivations such as those shown above, but since it only considers the shape of the clauses, it might also block SUP inferences whose side premises do not originate from ARGCONG. Consider a left-to-right LPO [10] instance with precedence $\mathsf{h} \succ \mathsf{g} \succ \mathsf{f} \succ \mathsf{b} \succ \mathsf{a}$, and consider the following unsatisfiable clause set:

$$\mathsf{g}\,(x\,\mathsf{b})\,x \approx \mathsf{a} \qquad\qquad \mathsf{g}\,(\mathsf{f}\,\mathsf{b})\,\mathsf{h} \not\approx \mathsf{a} \qquad\qquad \mathsf{h}\,x \approx \mathsf{f}\,x$$

The only possible inference from these clauses is POSEXT, showing its necessity. It is unclear whether POSEXT is necessary for the extensional purifying variant as well, but our completeness proof suggests that it is. Our proof also suggests that to achieve refutational completeness, due to nonmonotonicity, we need either to purify the clauses or to allow some superposition at variable positions, as mandated by the respective variable conditions. However, we have yet to find an example that demonstrates the necessity of these measures.

A considerable advantage of our calculi over the use of standard superposition on applicatively encoded problems is the flexibility they offer in orienting equations. The following example gives two definitions of addition on Peano numbers:

$$\begin{array}{cc} \mathsf{add_L}\,0\,y \approx y & \mathsf{add_R}\,x\,0 \approx x \\ \mathsf{add_L}\,(\mathsf{S}\,x)\,y \approx \mathsf{add_L}\,x\,(\mathsf{S}\,y) & \mathsf{add_R}\,x\,(\mathsf{S}\,y) \approx \mathsf{add_R}\,(\mathsf{S}\,x)\,y \end{array}$$

Let $\mathsf{add_L}\,(\mathsf{S}^{100}\,0)\,\mathsf{n} \not\approx \mathsf{add_R}\,\mathsf{n}\,(\mathsf{S}^{100}\,0)$ be the negated conjecture. With LPO, we can use a left-to-right comparison for $\mathsf{add_L}$'s arguments and a right-to-left comparison for $\mathsf{add_R}$'s arguments to orient all four equations from left to right. Then the negated conjecture can be simplified to $\mathsf{S}^{100}\,\mathsf{n} \not\approx \mathsf{S}^{100}\,\mathsf{n}$ by rewriting (demodulation), and $\perp$ can be derived with a single inference. If we use the applicative encoding instead, there is no instance of LPO or KBO that can orient both recursive equations from left to right. For at least one of the two sides of the negated conjecture, the rewriting is replaced by 100 SUP inferences, which is much less efficient, especially in the presence of additional axioms. More precisely, suppose we can simplify one side of the negated conjecture by rewriting, e.g. to $\mathsf{S}^{100}\ y \not\approx \mathsf{add_R}\ y\ (\mathsf{S}^{100}\ 0)$. After that, we must perform 100 SUP inferences to derive $\mathsf{add_R}\ y\ (\mathsf{S}\ 0) \approx \mathsf{S}\ y$, $\mathsf{add_R}\ y\ (\mathsf{S}\ (\mathsf{S}\ 0)) \approx \mathsf{S}\ (\mathsf{S}\ y)$, …, $\mathsf{add_R}\ y\ (\mathsf{S}^{100}\ 0) \approx \mathsf{S}^{100}\ y$. From this last clause we can then derive $\bot$ easily.

3.3 Redundancy Criterion

For our calculi, a redundant (or composite) clause cannot simply be defined as a clause whose ground instances are entailed by smaller ($\prec$) ground instances of existing clauses, because this would make all ARGCONG inferences redundant. Our solution is to base the redundancy criterion on a weaker ground logic in which argument congruence does not hold. This logic also plays a central role in our completeness proof, to reason about the nonmonotonicity emerging from the lack of compatibility with arguments.

The weaker logic is defined via an encoding $\lfloor\ \rfloor$ of ground hybrid λ-free higher-order terms into uncurried terms, with $\lceil\ \rceil$ as its inverse. Accordingly, we refer to clausal λ-free higher-order logic as the *ceiling logic* and to its weaker relative as the *floor logic*. Essentially, the encoding indexes each symbol occurrence with its argument count. Thus, $\lfloor \mathsf{f} \rfloor = \mathsf{f}_0$ and $\lfloor \mathsf{f\ a} \rfloor = \mathsf{f}_1(\mathsf{a}_0)$. This is enough to disable argument congruence; for example, $\{\mathsf{f} \approx \mathsf{g},\ \mathsf{f\ a} \not\approx \mathsf{g\ a}\}$ is unsatisfiable, whereas its encoding $\{\mathsf{f}_0 \approx \mathsf{g}_0,\ \mathsf{f}_1(\mathsf{a}_0) \not\approx \mathsf{g}_1(\mathsf{a}_0)\}$ is satisfiable. For clauses built from fully applied ground terms, the two logics are isomorphic, as we would expect from a graceful generalization.

Given a signature Σ in the ceiling logic, we define a signature $\Sigma^{\downarrow}$ in the floor logic as follows. For each higher-order type τ, we introduce an atomic type $\lfloor \tau \rfloor$ in the floor logic. For each symbol $\mathsf{f} : \bar{\tau}_k \Rightarrow \tau_{k+1} \to \cdots \to \tau_n \to \upsilon$ in Σ, where υ is atomic, we introduce symbols $\mathsf{f}_m : \lfloor \bar{\tau}_m \rfloor \Rightarrow \lfloor \tau_{m+1} \to \cdots \to \tau_n \to \upsilon \rfloor$ for $m \in \{k, \ldots, n\}$. Here and elsewhere, we write $\lfloor \bar{a} \rfloor$ for the componentwise application of $\lfloor\ \rfloor$ to the tuple $\bar{a}$. The translation of ground terms is given by $\lfloor \mathsf{f}(\bar{u}_k)\ u_{k+1} \ldots u_m \rfloor = \mathsf{f}_m(\lfloor \bar{u}_m \rfloor)$. This mapping can be extended to ground literals and ground clauses:

$$\begin{aligned} \lfloor s \approx t \rfloor &= \lfloor s \rfloor \approx \lfloor t \rfloor \\ \lfloor s \not\approx t \rfloor &= \lfloor s \rfloor \not\approx \lfloor t \rfloor \\ \lfloor L_1 \vee \cdots \vee L_n \rfloor &= \lfloor L_1 \rfloor \vee \cdots \vee \lfloor L_n \rfloor \end{aligned}$$

The $\lfloor\ \rfloor$ mapping is bijective with $\lceil\ \rceil$:

$$\begin{aligned} \lceil \mathsf{f}_{k+i}(a_1, \ldots, a_{k+i}) \rceil &= \mathsf{f}(\lceil a_1 \rceil, \ldots, \lceil a_k \rceil)\ \lceil a_{k+1} \rceil\ \ldots\ \lceil a_{k+i} \rceil \\ &\qquad \text{(where } k \text{ is determined by the type declaration of } \mathsf{f}) \\ \lceil s \approx t \rceil &= \lceil s \rceil \approx \lceil t \rceil \\ \lceil s \not\approx t \rceil &= \lceil s \rceil \not\approx \lceil t \rceil \\ \lceil L_1 \vee \cdots \vee L_n \rceil &= \lceil L_1 \rceil \vee \cdots \vee \lceil L_n \rceil \end{aligned}$$

Using $\lceil\ \rceil$, the clause order $\succ$ can be transferred to the floor logic by defining $t \succ s$ as equivalent to $\lceil t \rceil \succ \lceil s \rceil$. The property that $\succ$ on clauses is the multiset extension of $\succ$

on literals, which in turn is the multiset extension of $\succ$ on terms, is maintained because $\lceil\ \rceil$ maps the multiset representations elementwise.

Crucially, argument subterms in the ceiling logic correspond to argument subterms in the floor logic, whereas non-argument subterms in the ceiling logic are not subterms at all in the floor logic:

Lemma 2. *For all terms s and t in the floor logic,* $\lceil t[s]_p \rceil = \lceil t \rceil \langle \lceil s \rceil \rangle_p$.

Proof. By induction on p.

If $p = \varepsilon$, then $s = t[s]_p$. Hence $\lceil t[s]_p \rceil = \lceil s \rceil = \lceil t \rceil \langle \lceil s \rceil \rangle_p$.

If $p = i.p'$, then $t[s]_p = \mathsf{f}_n(u_1, \dots, u_n)$ with $u_i = u_i[s]_{p'}$. Applying $\lceil\ \rceil$, we obtain that $\lceil t[s]_p \rceil$ equals

$$\mathsf{f}(\lceil u_1 \rceil, \dots, \lceil u_{i-1} \rceil, \lceil u_i \rceil \langle \lceil s \rceil \rangle_{p'}, \lceil u_{i+1} \rceil, \dots, \lceil u_k \rceil)\ \lceil u_{k+1} \rceil\ \dots\ \lceil u_n \rceil$$

or

$$\mathsf{f}(\lceil u_1 \rceil, \dots, \lceil u_k \rceil)\ \lceil u_{k+1} \rceil\ \dots\ \lceil u_{i-1} \rceil\ \lceil u_i \rceil \langle \lceil s \rceil \rangle_{p'}\ \lceil u_{i+1} \rceil\ \dots\ \lceil u_n \rceil$$

by the induction hypothesis. It follows that $\lceil t[s]_p \rceil = \lceil t \rceil \langle \lceil s \rceil \rangle_p$. □

Lemma 3. *Well-foundedness, totality on ground terms, compatibility with* all *contexts, and the subterm property hold for* $\succ$ *in the floor logic.*

Proof. COMPATIBILITY WITH CONTEXTS: We want to show that $s \succ s'$ implies $t[s]_p \succ t[s']_p$ for floor terms t, s and s'. Assuming $s \succ s'$, we have $\lceil s \rceil \succ \lceil s' \rceil$. By compatibility with argument contexts in the ceiling logic, we have $\lceil t \rceil \langle \lceil s \rceil \rangle_p \succ \lceil t \rceil \langle \lceil s' \rceil \rangle_p$. By Lemma 2, we have $t[s]_p \succ t[s']_p$.

WELL-FOUNDEDNESS: Assume that there exists an infinite descending chain $t_1 \succ t_2 \succ \cdots$ of floor terms. By applying $\lceil\ \rceil$, we then obtain an infinite descending chain of ceiling terms $\lceil t_1 \rceil \succ \lceil t_2 \rceil \succ \cdots$, contradicting well-foundedness in the ceiling logic.

TOTALITY ON GROUND TERMS: Let s, t be ground terms of the floor logic. Then $\lceil t \rceil$ and $\lceil s \rceil$ must be comparable by totality on ground ceiling terms. Hence, t and s are comparable.

SUBTERM PROPERTY: By Lemma 2 and the subterm property in the ceiling logic, $\lceil t[s]_p \rceil = \lceil t \rceil \langle \lceil s \rceil \rangle_p \succ \lceil s \rceil$. Hence, $t[s]_p \succ s$. □

In standard superposition, redundancy relies on the entailment relation $\models$ on ground clauses. We define redundancy of ceiling clauses in the same way, but using the floor logic's entailment relation. This notion of redundancy gracefully generalizes the first-order notion.

For SUP, EQFACT, and EQRES, we can use the more precise notion of redundancy of inferences instead of redundancy of clauses, a ground inference being redundant if the conclusion follows from existing clauses that are smaller than the largest premise. For ARGCONG and POSEXT, we must use redundancy of clauses.

More precisely we define redundancy as follows: A ground ceiling clause C is *redundant with respect to a set of ceiling ground clauses N* if $\lfloor C \rfloor$ is entailed by clauses from $\lfloor N \rfloor$ that are smaller than $\lfloor C \rfloor$. A possibly nonground ceiling clause C is *redundant with respect to a set of ceiling clauses N* if all its ground instances are redundant with respect to $\mathcal{G}_\Sigma(N)$, the set of ground instances of clauses in N.

For all inference rules except for ArgCong and PosExt, a ground inference with conclusion E and right (or only) premise C is *redundant with respect to a set of ground clauses N* if one of its premises is redundant with respect to N, or if $\lfloor E \rfloor$ is entailed by clauses in $\lfloor N \rfloor$ that are smaller than $\lfloor C \rfloor$. A nonground inference is *redundant with respect to a clause set N* if all its ground instances are redundant with respect to $\mathcal{G}_\Sigma(N)$.

An ArgCong or PosExt inference is *redundant with respect to a clause set N* if its premise is redundant with respect to N or if its conclusion is contained in N or redundant with respect to N.

We call N *saturated up to redundancy* if every inference from clauses in N is redundant with respect to N.

3.4 Skolemization

A problem expressed in λ-free higher-order logic must be transformed into clausal normal form before the calculi can be applied. This process works as in the first-order case, except for skolemization. The issue is that skolemization, when performed naively, is unsound for λ-free higher-order logic with a Henkin semantics. For example, given a predicate symbol $\mathsf{p} : \tau \to \tau \to o$, the formula $(\forall x. \exists y.\ \mathsf{p}\, x\, y) \land (\forall z. \neg\, \mathsf{p}\, \mathsf{a}\, (z\, \mathsf{a}))$ has a model interpreting p as equality and such that none of the functions in the image of $\mathcal{E}_{\tau,\tau}$ map $\mathcal{I}(\mathsf{a})$ to $\mathcal{I}(\mathsf{a})$. Yet, naive skolemization would yield the clause set $\{\mathsf{p}\, x\, (\mathsf{sk}\, x), \neg\, \mathsf{p}\, \mathsf{a}\, (z\, \mathsf{a})\}$, whose unsatisfiability can be shown by taking $x := \mathsf{a}$ and $z := \mathsf{sk}$. The crux of the issue is that sk denotes a new function that can be used to instantiate z.

Inspired by Miller [34, Section 6], we adapt skolemization as follows. An existentially quantified variable $x : \tau$ in a context with universally quantified variables $\bar{x}_n$ of types $\bar{\tau}_n$ is replaced by a fresh symbol $\mathsf{sk} : \bar{\tau}_n \Rightarrow \tau$ applied to the tuple $\bar{x}_n$. For the example above, we obtain $\{\mathsf{p}\, x\, (\mathsf{sk}(x)),\ \neg\, \mathsf{p}\, \mathsf{a}\, (z\, \mathsf{a})\}$. Syntactically, z cannot be instantiated by sk, which is not even a term. Semantically, the clause set is satisfiable because we can have $\mathcal{I}(\mathsf{sk})(\mathcal{I}(\mathsf{a})) = \mathcal{I}(\mathsf{a})$ even if the image of $\mathcal{E}_{\tau,\tau}$ contains no such function.

4 Refutational Completeness

The proof of refutational completeness of the four calculi introduced in Section 3.1 follows the same general idea as for standard superposition [2, 50]. Given a clause set $N \not\ni \bot$ saturated up to redundancy, we construct a term rewriting system R based on the set of ground instances $\mathcal{G}_\Sigma(N)$. From R, we define an interpretation. We show, by induction on the clause order, that this interpretation is a model of $\mathcal{G}_\Sigma(N)$ and hence of N.

To circumvent the term order's potential nonmonotonicity, our Sup inference rule only considers the argument subterms u of a maximal term $s\langle u \rangle$. This is reflected in our proof by the reliance of the floor logic from Section 3.3. In that logic, the equation $\mathsf{g}_0 \approx \mathsf{f}_0$ cannot be used directly to rewrite the clause $\mathsf{g}_1(\mathsf{a}_0) \not\approx \mathsf{f}_1(\mathsf{a}_0)$; instead, we first need to apply ArgCong to derive $\mathsf{g}_1(x) \approx \mathsf{f}_1(x)$ and then use that equation. The floor logic is a device that enables us to reuse the traditional model construction almost verbatim, including its reliance on a first-order term rewriting system.

Following the traditional proof, we obtain a model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$. Since N is saturated up to redundancy with respect to ArgCong, the model $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ can easily be turned into a model of $\mathcal{G}_\Sigma(N)$ by conflating the interpretations of the members $\mathsf{f}_k, \ldots, \mathsf{f}_n$ of a same symbol family. For this section, we fix a set $N \not\ni \bot$ of λ-free higher-order clauses that is saturated up to redundancy. For the purifying calculi, we additionally require that all clauses in N are purified. To avoid empty Herbrand universes, we assume that the signature Σ contains, for each type τ, a symbol of type τ.

4.1 Candidate Interpretation

The construction of the candidate interpretation is as in the first-order proof, except that it is based on $\lfloor \mathcal{G}_\Sigma(N) \rfloor$. We first define sets of rewrite rules E_C and R_C for all $C \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ by induction on the clause order. Assume that E_D has already been defined for all $D \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ with $D \prec C$. Then $R_C = \bigcup_{D \prec C} E_D$. Let $E_C = \{s \to t\}$ if the following conditions are met: (a) $C = C' \vee s \approx t$; (b) $s \approx t$ is strictly maximal in C; (c) $s \succ t$; (d) C is false in R_C; (e) C' is false in $R_C \cup \{s \to t\}$; and (f) s is irreducible with respect to R_C. Then C is *productive*. Otherwise, $E_C = \emptyset$. Finally, $R_\infty = \bigcup_D E_D$.

A rewrite system R defines an interpretation $\mathcal{T}_\Sigma^\emptyset/R$ such that for every *ground* equation $s \approx t$, we have $\mathcal{T}_\Sigma^\emptyset/R \models s \approx t$ if and only if $s \leftrightarrow_R^* t$. Moreover, $\mathcal{T}_\Sigma^\emptyset/R$ is term-generated—that is, for every element a of a universe of this interpretation, there exists a ground term t such that $[\![t]\!]^\xi_{\mathcal{T}_\Sigma^\emptyset/R} = a$. To lighten notation, we will write R to refer to both the term rewriting system R and the interpretation $\mathcal{T}_\Sigma^\emptyset/R$.

The following properties of the candidate interpretations can be shown exactly as in Waldmann's first-order proof [50].

Lemma 4. *The rewrite systems R_C and R_∞ are convergent* (*i.e., terminating and confluent*)*.*

Lemma 5. *If $D \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ is true in R_D, then D is true in R_∞ and R_C for all $C \succ D$.*

Lemma 6. *If $D = D' \vee u \approx v$ is productive, then D' is false and D is true in R_∞ and R_C for all $C \succ D$.*

4.2 Lifting Lemmas

Following Waldmann's proof [50], we proceed by lifting inferences from the ground to the nonground level. We also need to lift ArgCong. A complication that arises when lifting purifying inferences is that the nonground conclusions may contain purification literals (corresponding to applied variables) not present in the ground conclusions. Given an inference I of the form $\bar{C} \vdash \mathit{pure}(E)$, we refer to the ground instances of $\bar{C} \vdash E$ as ground instances of *I up to purification*.

This auxiliary lemma is useful in the lifting lemma proofs:

Lemma 7. *Let σ be the most general unifier of s and s' (which can be assumed idempotent). Let θ be an arbitrary unifier of s and s'. Then $\sigma\theta = \theta$.*

Proof. Since σ is most general, there exists a substitution ρ, such that $\sigma\rho = \theta$. Therefore, by idempotence, $\sigma\theta = \sigma\sigma\rho = \sigma\rho = \theta$. $\square$

Lemma 8 (Lifting of non-SUP inferences). *Let C be a clause of the ceiling logic, and let θ be a substitution such that $C\theta$ is ground. Let $C\theta$ inherit C's selected literals. Then every* EQRES *or* EQFACT *inference from $C\theta$ and every ground instance of an* ARGCONG *inference from $C\theta$ is a ground instance of an inference from C up to purification.*

Proof. EQRES: We assume that there is a EQRES inference from $C\theta$. This means that $C\theta$ is of the form $C\theta = C'\theta \vee s\theta \not\approx s'\theta$ where $C = C' \vee s \not\approx s'$, and $s\theta \not\approx s'\theta$ is selected or no literal of $C\theta$ is selected and $s\theta \not\approx s'\theta$ is maximal. Then the ground inference is

$$\frac{C'\theta \vee s\theta \not\approx s'\theta}{C'\theta}\,\text{EQRES}$$

where $s\theta$ and $s'\theta$ are unifiable and ground; hence $s\theta = s'\theta$. Since $s\theta \not\approx s'\theta$ is maximal (and nothing is selected) or selected in $C\theta$, $s \not\approx s'$ is maximal (and nothing is selected) or selected in C. Hence we have the inference

$$\frac{C' \vee s \not\approx s'}{C'\sigma \vee C_P}\,\text{EQRES}$$

where $\sigma = \text{mgu}(s, s')$ and C_P are purification literals. By Lemma 7, we have $C'\sigma\theta = C'\theta$. Thus, the ground inference is the θ-ground instance of the nonground inference up to purification literals.

EQFACT: We assume that there is a EQFACT inference from $C\theta$. This means that $C\theta$ is of the form $C\theta = C'\theta \vee s'\theta \approx t'\theta \vee s\theta \approx t\theta$ where $s\theta \approx t\theta$ is maximal, no literal is selected in $C\theta$, $s\theta \not\prec t\theta$, and $C = C' \vee s' \approx t' \vee s \approx t$. Then the ground inference is

$$\frac{C'\theta \vee s'\theta \approx t'\theta \vee s\theta \approx t\theta}{C'\theta \vee t\theta \not\approx t'\theta \vee s\theta \approx t'\theta}\,\text{EQFACT}$$

where $s\theta$ and $s'\theta$ are unifiable and ground; hence $s\theta = s'\theta$. Since $s\theta \approx t\theta$ is maximal in $C\theta$, nothing is selected in $C\theta$, and $s\theta \not\prec t\theta$, $s \approx t$ is maximal in C, nothing is selected in C, and $s \not\prec t$. Hence we have the inference

$$\frac{C' \vee s' \approx t' \vee s \approx t}{(C' \vee t \not\approx t' \vee s \approx t')\sigma \vee C_P}\,\text{EQFACT}$$

where $\sigma = \text{mgu}(s, s')$ and C_P are purification literals. By Lemma 7, we have $(C' \vee t \not\approx t' \vee s \approx t')\sigma\theta = C'\theta \vee t\theta \not\approx t'\theta \vee s\theta \approx t'\theta$. Thus, the ground inference is the θ-ground instance of the nonground inference up to purification literals.

ArgCong: We assume that there is an ArgCong inference from $C\theta$. This means that $C\theta$ is of the form $C\theta = C'\theta \vee s\theta \approx s'\theta$, where $s\theta \approx s'\theta$ is strictly maximal, no literal is selected in $C\theta$, and $C = C' \vee s \approx s'$. Then the inference from $C\theta$ is

$$\frac{C'\theta \vee s\theta \approx s'\theta}{C'\theta \vee s\theta\,\bar{x} \approx s'\theta\,\bar{x}}\,\textsc{ArgCong}$$

There cannot be purification literals because the premise is ground. Every ground instance of this inference has the form

$$\frac{C'\theta \vee s\theta \approx s'\theta}{C'\theta \vee s\theta\,\bar{v} \approx s'\theta\,\bar{v}}\,\textsc{ArgCong}$$

Since $s\theta \not\approx s'\theta$ is strictly maximal in $C\theta$, $s \approx s'$ is strictly maximal in C. Since nothing is selected in $C\theta$, nothing is selected in C. Hence we have the inference

$$\frac{C' \vee s \approx s'}{C' \vee s\,\bar{x} \approx s'\,\bar{x} \vee C_P}\,\textsc{ArgCong}$$

where C_P are purification literals. Thus, the ground inference is the $\theta[x_1 \mapsto v_1, \dots x_n \mapsto v_n]$-ground instance of this inference from C up to purification literals. □

The conditions of the lifting lemma for Sup differ slightly from the first-order version. For standard superposition, the lemma applies if the superposed term is not at or under a variable. This condition is replaced by the following "liftability" criterion.

Definition 9. We call a ground Sup inference from $D\theta$ and $C\theta$ *liftable* if there exists a corresponding inference from D and C.

Lemma 10 (Lifting of Sup inferences). *Let $D = D' \vee t \approx t'$ and $C = C' \vee [\neg]\, s \approx s'$ be clauses in the ceiling logic (without common variables), and let θ be a ground substitution. Then every liftable* Sup *inference between $D\theta$ and $C\theta$ is a ground instance of a* Sup *inference from D and C up to purification.*

Proof. We assume that there is a ground Sup inference of $D\theta$ in $C\theta$. This means that $t\theta \approx t'\theta$ is strictly maximal and nothing is selected in $D\theta$. For positive superposition, $s\theta \approx s'\theta$ is strictly maximal and nothing is selected in $C\theta$. For negative superposition, either $s\theta \not\approx s'\theta$ is maximal and nothing is selected or $s\theta \not\approx s'\theta$ is selected in $C\theta$. Moreover, $D\theta \not\succeq C\theta$, $t\theta \not\prec t'\theta$, and $s\theta \not\prec s'\theta$. The ground inference is

$$\frac{D'\theta \vee t\theta \approx t'\theta \quad C'\theta \vee [\neg]\, s\theta\langle t\theta\rangle_p \approx s'\theta}{D'\theta \vee C'\theta \vee [\neg]\, s\theta\langle t'\theta\rangle_p \approx s'\theta}\,\textsc{Sup}$$

The inference conditions can be lifted: That $t\theta \approx t'\theta$ is strictly maximal in $D\theta$ implies that $t \approx t'$ is strictly maximal in D. That nothing is selected in $D\theta$ implies that nothing is selected in D. If $[\neg]\, s\theta \approx s'\theta$ is (strictly) maximal and nothing is selected in $C\theta$, then $[\neg]\, s \approx s'$ is (strictly) maximal and nothing is selected in C. If $s\theta \not\approx s'\theta$ is selected in $C\theta$, then $s \not\approx s'$ is selected in C. $D\theta \not\succeq C\theta$ implies $D \not\succeq C$. $t\theta \not\prec t'\theta$ implies $t \not\prec t'$. $s\theta \not\prec s'\theta$ implies $s \not\prec s'$.

Note that the variable condition holds and that p is a position of s, because the ground inference is liftable. The argument subterm u of s at position p is unifiable with t, because θ is a unifier. So we have the nonground inference

$$\frac{D' \vee t \approx t' \quad C' \vee [\neg]\, s\langle u\rangle_p \approx s'}{(D' \vee C' \vee [\neg]\, s\langle t'\rangle_p \approx s')\sigma \vee C_P}\ \textsc{Sup}$$

where $\sigma = \mathrm{mgu}(t,u)$ and C_P are purification literals. By Lemma 7, we have $(D' \vee C' \vee [\neg]\, s\langle t'\rangle_p \approx s')\sigma\theta = D'\theta \vee C'\theta \vee [\neg]\, s\theta\langle t'\theta\rangle_p \approx s'\theta$. Thus, the ground inference is the θ-ground instance of the nonground inference up to purification literals. □

4.3 Main Result

The model construction theorem states that the candidate interpretation R_∞ is a model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$. Like in the first-order proof, this is shown by induction on the clause order. For the induction step, we fix some clause $\lfloor C\theta \rfloor \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ and assume that all smaller clauses are true in $R_{C\theta}$. We distinguish several cases, most of which amount to showing that $C\theta$ can be used to perform a certain inference. Then we deduce that $\lfloor C\theta \rfloor$ is true in $R_{C\theta}$ to complete the induction step.

The next two lemmas are slightly adapted from Waldmann's version of the first-order proof [50]. The justification for Lemma 11, about liftable inferences, is essentially as in the first-order case. The proof of Lemma 12, about nonliftable inferences, is more problematic. The standard argument involves defining a substitution θ' such that $C\theta'$ and $C\theta$ are equivalent and $C\theta' \prec C\theta$. But due to nonmonotonicity, we might have $C\theta' \succ C\theta$, blocking the application of the induction hypothesis. This is where the variable conditions, purification, and the PosExt rule come into play.

Lemma 11. *Let $C, D \in N$, and let θ be a ground substitution. We consider a liftable* Sup *inference from $D\theta$ and $C\theta$ or an* EqRes *or* EqFact *inference from $C\theta$. Let E be the conclusion. Assume that $C\theta$ and $D\theta$ are nonredundant with respect to $\mathcal{G}_\Sigma(N)$. Then $\lfloor E \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$.*

Proof. We have a liftable Sup inference from $D\theta$ and $C\theta$ or an EqRes or EqFact inference from $C\theta$. As shown in the lifting lemmas (Lemmas 8 and 10), up to purification literals in the conclusion, this inference is an instance of an inference from C (or from D and C for Sup inferences). Let $\tilde{E}$ be its conclusion. Since N is saturated up to redundancy, this inference is redundant with respect to N and hence the θ-ground instance of this inference is redundant with respect to $\mathcal{G}_\Sigma(N)$. By definition of inference redundancy, since $C\theta$ is not redundant with respect to $\mathcal{G}_\Sigma(N)$, $\lfloor \tilde{E}\theta \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$.

By Lemma 7, we have $\tilde{E}\theta = E$ for the nonpurifying variants. In the purifying variants, we extend θ to the purification variables by copying the values of the original variable. Then the literals of $\lfloor \tilde{E}\theta \rfloor$ corresponding to purification literals are trivially true and hence $\lfloor E \rfloor$ is equivalent to $\lfloor \tilde{E}\theta \rfloor$. In all variants, it follows that $\lfloor E \rfloor$ is entailed by clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. □

Lemma 12. *Let $C, D \in N$, and let θ be a ground substitution. We consider a nonliftable* SUP *inference from $D\theta$ and $C\theta$. Assume that $C\theta$ and $D\theta$ are nonredundant with respect to $\mathcal{G}_\Sigma(N)$. Let $D'\theta$ be the clause $D\theta$ without the literal involved in the inference. Then $\lfloor C\theta \rfloor$ is entailed by $\neg\lfloor D'\theta \rfloor$ and the clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$.*

Proof. Let $C\theta = C'\theta \vee [\neg]\, s\theta \approx s'\theta$ and $D\theta = D'\theta \vee t\theta \approx t'\theta$, where $[\neg]\, s\theta \approx s'\theta$ and $t\theta \approx t'\theta$ are the literals involved in the inference, $s\theta \succ s'\theta$, $t\theta \succ t'\theta$, and C', s, s', t, t' are the respective subclauses and terms in C and D.

Let R be an interpretation such that $\lfloor D'\theta \rfloor$ is false and the clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$ are true. Since $C\theta \succ D\theta$ by the SUP order conditions, it follows that $R \models \lfloor t\theta \approx t'\theta \rfloor$. We must show that $R \models \lfloor C\theta \rfloor$.

The inference from $D\theta$ and $C\theta$ can be nonliftable either because it is a superposition below a variable or because the variable condition does not hold for the corresponding inference between D and C.

CASE 1: We assume that it is a superposition below a variable, say, x. Let $t\theta \approx t'\theta$ be the strictly maximal literal of D, where $t\theta \succ t'\theta$. Then $t\theta$ is an argument subterm of $x\theta$ and hence an argument subterm of $x\theta\, \bar{w}$ for any arguments $\bar{w}$. Let v be the term that we obtain by replacing $t\theta$ by $t'\theta$ in $x\theta$ at the relevant position. It follows from the preconditions that $R \models \lfloor t\theta \approx t'\theta \rfloor$ and by congruence, $R \models \lfloor x\theta\, \bar{w} \approx v\, \bar{w} \rfloor$ for any arguments $\bar{w}$. Hence, $R \models \lfloor C\theta \rfloor$ if and only if $R \models \lfloor C[x \mapsto v]\theta \rfloor$. By the inference conditions we have $t\theta \succ t'\theta$, which implies $\lfloor C\theta \rfloor \succ \lfloor C[x \mapsto v]\theta \rfloor$ by compatibility with argument contexts. Therefore, by the assumption about R, we have $R \models \lfloor C[x \mapsto v]\theta \rfloor$ and hence $R \models \lfloor C\theta \rfloor$.

CASE 2: The variable condition does not hold in the corresponding inference between D and C. Let u denote the superposed subterm of s.

Since the variable condition does not hold, u, t, and t' have the following form: $u = x\, \bar{v}$ for some variable x and n terms $\bar{v}$, for $n \geq 0$; $t = \tilde{t}\, \bar{x}_n$ and $t' = \tilde{t}'\, \bar{x}_n$, where $\bar{x}_n$ are variables that do not occur elsewhere in D. For the intensional variants, we additionally have $n = 0$. For the nonpurifying variants, we additionally have $C\theta \succeq C''\theta$, where $C'' = C[x \mapsto \tilde{t}']$.

CASE 2.1 (PURIFYING CALCULI): First, we assume that x occurs only with arguments $\bar{v}$ in C. For the intensional variant, this must be the case because $n = 0$ and hence x can only occur without arguments by the definition of *pure* due to the selection restriction. Define a substitution θ' by $x\theta' = \tilde{t}'\theta$ and $y\theta' = y\theta$ for other variables y. Since $t\theta \succ t'\theta$, we have $C\theta \succ C\theta'$. Moreover, $C\theta' \in \mathcal{G}_\Sigma(N)$. Then $R \models \lfloor C\theta \rfloor$ by congruence, because $R \models \lfloor C\theta' \rfloor$ and $R \models \lfloor t\theta \approx t'\theta \rfloor$.

Now we assume that x occurs with arguments other than $\bar{v}$ in C. This can only happen in the extensional variant and by the selection restrictions, L may not be selected. Therefore, $s\theta$ is the maximal term in $C\theta$. Then $s \neq x$ and hence $\bar{v} \neq \varepsilon$ because otherwise $s\theta = x\theta$ would be smaller than the applied occurrence of $x\theta$ in $C\theta$.

Define a substitution θ'' such that $x\theta'' = \tilde{t}'\theta$, $y\theta'' = \tilde{t}'\theta$ for other variables y if $y\theta = s\theta$ and C contains the literal $x \not\approx y$, and $y\theta'' = y\theta$ otherwise.

We show that $C\theta \succ C\theta''$ by proving that no literal of $C\theta''$ is larger than the maximal literal $[\neg]\, s\theta \approx s'\theta$ of $C\theta$ and that $[\neg]\, s\theta \approx s'\theta$ appears more often in $C\theta$ than in $C\theta''$:

For a literal of the form $x \not\approx y$, we have $x\theta'' \prec s\theta$ and $y\theta'' \prec s\theta$. For literals that are not of this form, by the definition of *pure* in the extensional variant, x occurs always with arguments $\bar{v}$. Hence these literals are equal or smaller in $C\theta''$ than in $C\theta$, because $x\theta''\,\bar{v} \prec x\theta\,\bar{v}$ and $y\theta'' \preceq y\theta$. Therefore, no literal of $C\theta''$ is larger than the maximal literal $[\neg]\, s\theta \approx s'\theta$ of $C\theta$. Moreover, these inequalities show that every occurrence of $[\neg]\, s\theta \approx s'\theta$ in $C\theta''$ corresponds to an occurrence of $[\neg]\, s\theta \approx s'\theta$ in $C\theta$ that corresponds to a literal in C without the variable x. Since at least one occurrence of $[\neg]\, s\theta \approx s'\theta$ in $C\theta$ corresponds to a literal in C containing x, $[\neg]\, s\theta \approx s'\theta$ appears more often in $C\theta$ than in $C\theta''$. This concludes the argument that $C\theta \succ C\theta''$.

A POSEXT inference from D to $D' \vee \tilde{t} \approx \tilde{t}'$ is possible. Therefore, $D' \vee \tilde{t} \approx \tilde{t}'$ is in N or redundant with respect to to N because N is saturated up to redundancy. In either case, $R \models \lfloor (D' \vee \tilde{t} \approx \tilde{t}')\theta \rfloor$ because this clause is smaller than $C\theta$. Since $D'\theta$ is false in R, we have $R \models \lfloor \tilde{t}\theta \approx \tilde{t}'\theta \rfloor$.

For every literal of the form $x \not\approx y$ where $y = s\theta$, the variable y can only occur without arguments in C because of the maximality of $s\theta$. Since $C\theta \succ C\theta''$, we have $R \models \lfloor C\theta'' \rfloor$. If for every literal of the form $x \not\approx y$ where $y = s\theta$ we have $R \models \lfloor y\theta'' \approx y\theta \rfloor$, then $R \models \lfloor C\theta \rfloor$ by congruence. If for some literal of the form $x \not\approx y$ where $y = s\theta$ we have $R \models \lfloor y\theta'' \not\approx y\theta \rfloor$, then $R \models \lfloor y\theta \not\approx y\theta'' = \tilde{t}'\theta \approx \tilde{t}\theta = x\theta \rfloor$ which means that a literal of $C\theta$ is true in R and therefore $C\theta$ is true in R.

CASE 2.2 (NONPURIFYING CALCULI): Since the variable condition does not hold, we have $C\theta \succeq C''\theta$. Equality is impossible, because $x\theta = \tilde{t}\theta \neq \tilde{t}'\theta$ and x occurs in C. Hence, we have $C\theta \succ C''\theta$.

By the definition of R, $C\theta \succ C''\theta$ implies $R \models \lfloor C''\theta \rfloor$. We will use equalities that are true in R to rewrite $\lfloor C\theta \rfloor$ into $\lfloor C''\theta \rfloor$, which implies $R \models \lfloor C\theta \rfloor$ by congruence.

By saturation up to redundancy, using a POSEXT with premise D (if $n < \text{length}(\bar{z})$) or ARGCONG inference with premise D (if $n > \text{length}(\bar{z})$) or using D itself (if $n = \text{length}(\bar{z})$), we can show that up to variable renaming $D' \vee \tilde{t}\,\bar{z} \approx \tilde{t}'\,\bar{z}$ is in $\mathcal{G}_\Sigma(N \cup \text{Red}(N))$ for any type-correct tuple of fresh variables $\bar{z}$. Hence, $D'\theta \vee \tilde{t}\theta\,\bar{u} \approx \tilde{t}'\theta\,\bar{u}$ is in $\mathcal{G}_\Sigma(N \cup \text{Red}(N))$ for any type-correct ground arguments $\bar{u}$.

First, we observe that whenever $\tilde{t}\theta\,\bar{u}$ and $\tilde{t}'\theta\,\bar{u}$ are smaller than the maximal term of $C\theta$ for some arguments $\bar{u}$, we have

$$R \models \lfloor \tilde{t}\theta\,\bar{u} \rfloor \approx \lfloor \tilde{t}'\theta\,\bar{u} \rfloor \qquad (\dagger)$$

To show this, we assume that $\tilde{t}\theta\,\bar{u}$ and $\tilde{t}'\theta\,\bar{u}$ are smaller than the maximal term of $C\theta$ and we distinguish two cases: If $t\theta$ is smaller than the maximal term of $C\theta$, all terms in $D'\theta$ are smaller than the maximal term of $C\theta$ and hence $D'\theta \vee \tilde{t}\theta\,\bar{u} \approx \tilde{t}'\theta\,\bar{u} \prec C\theta$. If, on the other hand, $t\theta$ is equal to the maximal term of $C\theta$, $\tilde{t}\theta\,\bar{u}$ and $\tilde{t}'\theta\,\bar{u}$ are smaller than $t\theta$. Hence $\tilde{t}\theta\,\bar{u} \approx \tilde{t}'\theta\,\bar{u} \prec t\theta \approx t'\theta$ and $D'\theta \vee \tilde{t}\theta\,\bar{u} \approx \tilde{t}'\theta\,\bar{u} \prec D\theta \prec C\theta$. In both cases, since $D'\theta$ is false in R, by the definition of R, $R \models \lfloor \tilde{t}\theta\,\bar{u} \rfloor \approx \lfloor \tilde{t}'\theta\,\bar{u} \rfloor$.

We proceed by a case distinction on whether $s\theta$ appears in a selected or in a maximal literal of $C\theta$.

CASE 2.2.1: $s\theta$ is the maximal side of a selected literal of $C\theta$. Then, by the selection restrictions, x cannot be the head of a maximal literal of C.

At every position where $x\,\bar{u}$ occurs in C with some (or no) arguments $\bar{u}$, we rewrite $(\tilde{t}\,\bar{u})\theta$ to $(\tilde{t}'\,\bar{u})\theta$ in $C\theta$ if the latter term is smaller. We start with the innermost occurrences of x, such that the order of the two terms at one step does not change by later rewriting.

Analogously, at every position where $x\,\bar{u}$ occurs in C with some (or no) arguments $\bar{u}$, we rewrite $(\tilde{t}'\,\bar{u})\theta$ to $(\tilde{t}\,\bar{u})\theta$ in $C''\theta$ if the latter term is smaller, again starting with the innermost occurrences.

Note that we never rewrite at the top level of the maximal term of $C\theta$ or $C''\theta$ because x cannot be the head of a maximal literal of C. The two resulting clauses are identical because $C\theta$ and $C''\theta$ only differ at positions where x occurs in C. The rewritten terms are all smaller than the maximal term of $C\theta$. With (†), this implies that $R \models C\theta$ by congruence.

CASE 2.2.2: $s\theta$ is the maximal side of a maximal literal of $C\theta$. Then, since $C\theta \succ C''\theta$, every term in $C\theta$ and in $C''\theta$ is smaller or equal to $s\theta$. Let C_0 and $\tilde{C}_0$ be the clauses resulting from rewriting $\lfloor t\theta \rfloor \to \lfloor t'\theta \rfloor$ wherever possible in $\lfloor C\theta \rfloor$ and $\lfloor C''\theta \rfloor$, respectively. Since $\lfloor t\theta \rfloor$ is a subterm of $\lfloor s\theta \rfloor$, now every term in C_0 and $\tilde{C}_0$ is strictly smaller than $\lfloor s\theta \rfloor$.

We define inductively $C_1, C_2, \ldots$ as follows: Given C_i, choose a subterm of the form $\lfloor \tilde{t}\theta\,\bar{u} \rfloor$ where $\tilde{t}\theta\,\bar{u} \succ \tilde{t}'\theta\,\bar{u}$ or of the form $\lfloor \tilde{t}'\theta\,\bar{u} \rfloor$ where $\tilde{t}'\theta\,\bar{u} \succ \tilde{t}\theta\,\bar{u}$. Let C_{i+1} be the clause resulting from rewriting that subterm $\lfloor \tilde{t}\theta\,\bar{u} \rfloor$ to $\lfloor \tilde{t}'\theta\,\bar{u} \rfloor$ or that subterm $\lfloor \tilde{t}'\theta\,\bar{u} \rfloor$ to $\lfloor \tilde{t}\theta\,\bar{u} \rfloor$ in C_i, depending on which term was chosen.

Analogously, we define $\tilde{C}_1, \tilde{C}_2, \ldots$ by applying the same algorithm to $\tilde{C}_0$. In both cases, the process terminates because $\succ$ is of compatible with argument contexts and well founded. Let C_* and $\tilde{C}_*$ be the respective final clauses.

Note that the algorithm preserves the invariant that every term in C_i and $\tilde{C}_i$ is strictly smaller than $s\theta$. By congruence via (†), applied at every step of the algorithm, we know that C_* and $\lfloor C\theta \rfloor$ are equivalent in R and that $\tilde{C}_*$ and $\lfloor C''\theta \rfloor$ are equivalent in R as well.

We will show that $C_* = \tilde{C}_*$. Assume that $C_* \neq \tilde{C}_*$. Note that the algorithm preserves a second invariant, namely that $\lceil C_i \rceil$ and $\lceil \tilde{C}_j \rceil$ are equal except for positions where one contains $\tilde{t}\theta$ and the other one contains $\tilde{t}'\theta$. Consider the deepest position where $\lceil C_* \rceil$ and $\lceil \tilde{C}_* \rceil$ are different. The respective position in C_* and $\tilde{C}_*$ then contains $\lfloor \tilde{t}\theta\,\bar{u} \rfloor$ and $\lfloor \tilde{t}'\theta\,\bar{u} \rfloor$ or vice versa. The arguments $\bar{u}$ must be equal because we consider the deepest position possible. But then $\tilde{t}\theta\,\bar{u} \succ \tilde{t}'\theta\,\bar{u}$ or $\tilde{t}\theta\,\bar{u} \prec \tilde{t}'\theta\,\bar{u}$, which contradicts the fact that the algorithm terminated in C_* and $\tilde{C}_*$.

This shows that $C_* = \tilde{C}_*$. Hence $\lfloor C\theta \rfloor$ and $\lfloor C''\theta \rfloor$ are equivalent, which proves $R \models \lfloor C\theta \rfloor$. □

Using these two lemmas, the induction argument works as in the first-order case.

Lemma 13 (Model construction). *Let $\lfloor C\theta \rfloor \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$. We have*

(i) *$E_{\lfloor C\theta \rfloor} = \emptyset$ if and only if $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$;*
(ii) *if $C\theta$ is redundant with respect to $\mathcal{G}_\Sigma(N)$, then $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$;*
(iii) *$\lfloor C\theta \rfloor$ is true in R_∞ and in R_D for every $D \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ with $D \succ \lfloor C\theta \rfloor$; and*
(iv) *if $C\theta$ has selected literals, then $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$.*

Proof. We use induction of the clause order $\succ$ on floor logic ground clauses and assume that (i)–(iv) are already satisfied for all clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$.

Note that the 'if' part of (i) is obvious from the construction and that condition (iii) follows from (i) by Lemmas 5 and 6. So it remains to show (ii), (iv), and the 'only if' part of (i), i.e., we show the following: If $E_{\lfloor C\theta \rfloor} = \emptyset$ or $C\theta$ is redundant with respect to $\mathcal{G}_\Sigma(N)$ or $C\theta$ has selected literals, then $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$.

CASE 1: $C\theta$ is redundant with respect to $\mathcal{G}_\Sigma(N)$. Then $\lfloor C\theta \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, these clauses are true in $R_{\lfloor C\theta \rfloor}$. Hence $\lfloor C\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$.

CASE 2: $C\theta$ is not redundant with respect to $\mathcal{G}_\Sigma(N)$ and $C\theta$ contains an eligible negative literal. Let $s\theta \not\approx s'\theta$ with $s\theta \succeq s'\theta$ be one of these literals and $C'\theta$ the rest of the clause.

CASE 2.1: $s\theta = s'\theta$. Then there is an EQRES inference:

$$\frac{C'\theta \vee s\theta \not\approx s'\theta}{C'\theta}\ \textsc{EqRes}$$

By Lemma 11, $\lfloor C'\theta \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, these clauses are true in $R_{\lfloor C\theta \rfloor}$, which implies that $\lfloor C'\theta \rfloor$ and hence $\lfloor C\theta \rfloor$ are true in $R_{\lfloor C\theta \rfloor}$.

CASE 2.2: $s\theta \succ s'\theta$. If $R \models \lfloor s\theta \not\approx s'\theta \rfloor$, then it follows directly that $R \models \lfloor C\theta \rfloor$. So we assume that $\lfloor s\theta \rfloor \downarrow_{R_{\lfloor C\theta \rfloor}} \lfloor s'\theta \rfloor$ (i.e., $\lfloor s\theta \rfloor$ and $\lfloor s'\theta \rfloor$ have the same normal form), which means that $R \models \lfloor s\theta \approx s'\theta \rfloor$. Since $s\theta \succ s'\theta$, $\lfloor s\theta \rfloor$ must be reducible by some rule in some $E_{\lfloor D\theta \rfloor} \subseteq R_{\lfloor C\theta \rfloor}$. Without loss of generality, we assume that C and D are variable disjoint; so we can use the same substitution θ. Let $D\theta = D'\theta \vee t\theta \approx t'\theta$ with $E_{\lfloor D\theta \rfloor} = \{\lfloor t\theta \rfloor \to \lfloor t'\theta \rfloor\}$.

There is a SUP inference

$$\frac{D'\theta \vee t\theta \approx t'\theta \quad C'\theta \vee s\theta\langle t\theta \rangle \not\approx s'\theta}{D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \not\approx s'\theta}\ \textsc{Sup}$$

If this inference is not liftable, by Lemma 12, $\neg\lfloor D'\theta \rfloor$ and the clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$ imply $\lfloor C\theta \rfloor$. Since $\lfloor D\theta \rfloor$ is productive, $\lfloor D'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$ by Lemma 6. By part (iii) of the induction hypothesis, all clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$ are true in $R_{\lfloor C\theta \rfloor}$. Therefore, $\lfloor C\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$.

If this inference is liftable, by Lemma 11, $\lfloor D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \not\approx s'\theta \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, $\lfloor D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \not\approx s'\theta \rfloor$ is then true in $R_{\lfloor C\theta \rfloor}$. Since $t\theta \to t\theta' \in R_{\lfloor C\theta \rfloor}$ and $\lfloor D'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$, it follows that $R_{\lfloor C\theta \rfloor} \models C\theta$.

CASE 3: $C\theta$ is not redundant and contains no eligible negative literal. Then nothing is selected in $C\theta$ and $C\theta$ can be written as $C'\theta \vee s\theta \approx s'\theta$, where $s\theta \approx s'\theta$ is a maximal literal. If $E_{\lfloor C\theta \rfloor} = \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$ or $R_{\lfloor C\theta \rfloor} \models \lfloor C'\theta \rfloor$ or $s\theta = s'\theta$, there is nothing to show, so assume that $E_{\lfloor C\theta \rfloor} = \emptyset$ and that $\lfloor C'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$. Without loss of generality, $s\theta \succ s'\theta$.

CASE 3.1: $\lfloor s\theta \approx s'\theta \rfloor$ is maximal in $\lfloor C\theta \rfloor$, but not strictly maximal. Then $C\theta$ can be written as $C''\theta \vee t\theta \approx t'\theta \vee s\theta \approx s'\theta$, where $t\theta = s\theta$ and $t'\theta = s'\theta$. In this case, there is a EQFACT inference

$$\frac{C\theta \;=\; C''\theta \vee t\theta \approx t'\theta \vee s\theta \approx s'\theta}{C''\theta \vee t'\theta \not\approx s'\theta \vee t\theta \approx t'\theta}\ \textsc{EqFact}$$

By Lemma 11, its conclusion is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, these clauses are true in $R_{\lfloor C\theta \rfloor}$, which implies that $\lfloor C''\theta \vee t'\theta \not\approx s'\theta \vee t\theta \approx t'\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$. Since $t'\theta = s'\theta$ and hence $\lfloor t'\theta \not\approx s'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$, this implies that $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$.

CASE 3.2: $s\theta \approx s'\theta$ is strictly maximal in $C\theta$ and $\lfloor s\theta \rfloor$ is reducible by $R_{\lfloor C\theta \rfloor}$. Let $\lfloor t\theta \rfloor \to \lfloor t'\theta \rfloor \in R_{\lfloor C\theta \rfloor}$ be a rule that reduces $\lfloor s\theta \rfloor$. This rule stems from a productive clause $\lfloor D\theta \rfloor = \lfloor D'\theta \vee t\theta \approx t'\theta \rfloor$. Without loss of generality, we assume that C and D are variable disjoint; so we can use the same substitution θ.

We can now proceed in essentially the same way as in Case 2.2: There is a SUP inference

$$\frac{D'\theta \vee t\theta \approx t'\theta \quad C'\theta \vee s\theta\langle t\theta \rangle \approx s'\theta}{D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \approx s'\theta}\ \textsc{Sup}$$

If this inference is not liftable, by Lemma 12, $\neg\lfloor D'\theta \rfloor$ and the clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$ imply $\lfloor C\theta \rfloor$. Since $\lfloor D\theta \rfloor$ is productive, $\lfloor D'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$ by Lemma 6. By part (iii) of the induction hypothesis, all clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$ are true in $R_{\lfloor C\theta \rfloor}$. Therefore, $\lfloor C\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$.

If this inference is liftable, by Lemma 11, $\lfloor D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \approx s'\theta \rfloor$ is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, $\lfloor D'\theta \vee C'\theta \vee s\theta\langle t'\theta \rangle \approx s'\theta \rfloor$ is then true in $R_{\lfloor C\theta \rfloor}$. Since $t\theta \to t\theta' \in R_{\lfloor C\theta \rfloor}$ and $\lfloor D'\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$, it follows that $R_{\lfloor C\theta \rfloor} \models C\theta$.

CASE 3.3: $s\theta \approx s'\theta$ is strictly maximal in $C\theta$ and $\lfloor s\theta \rfloor$ is irreducible with respect to $R_{\lfloor C\theta \rfloor}$. Then there are three possibilities: $\lfloor C\theta \rfloor$ can be true in $R_{\lfloor C\theta \rfloor}$, or $\lfloor C'\theta \rfloor$ can be true in $R_{\lfloor C\theta \rfloor} \cup \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$, or $E_{\lfloor C\theta \rfloor} = \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$. In the first and the third case, there is nothing to show. Let us therefore assume that $\lfloor C\theta \rfloor$ is false in $R_{\lfloor C\theta \rfloor}$ and $\lfloor C'\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor} \cup \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$. Then $C'\theta = C''\theta \vee t\theta \approx t'\theta$, where the literal $\lfloor t\theta \approx t'\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor} \cup \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$ and false in $R_{\lfloor C\theta \rfloor}$. In other words, $\lfloor t\theta \rfloor \downarrow_{R_{\lfloor C\theta \rfloor} \cup \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}} \lfloor t'\theta \rfloor$, but not $\lfloor t\theta \rfloor \downarrow_{R_{\lfloor C\theta \rfloor}} \lfloor t'\theta \rfloor$. Consequently, there is a rewrite proof of $\lfloor t\theta \rfloor \to^* \lfloor u \rfloor \leftarrow^* \lfloor t'\theta \rfloor$ by $R_{\lfloor C\theta \rfloor} \cup \{\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor\}$ in which the rule $\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor$ is used at least once. Without loss of generality, we assume that $t\theta \succeq t'\theta$. Since $s\theta \approx s'\theta \succ t\theta \approx t'\theta$ and $s\theta \succ s'\theta$ we can conclude that $s\theta \succeq t\theta \succ t'\theta$. But then there is only one possibility how the rule $\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor$ can be used in the rewrite proof: We must have $s\theta = t\theta$ and the rewrite proof must have the form $\lfloor t\theta \rfloor \to \lfloor s'\theta \rfloor \to^* \lfloor u \rfloor \leftarrow^* \lfloor t'\theta \rfloor$, where the first step uses $\lfloor s\theta \rfloor \to \lfloor s'\theta \rfloor$ and all other steps use rules from $R_{\lfloor C\theta \rfloor}$. Consequently, $\lfloor s'\theta \approx t'\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$. Now observe that there is an EQFACT inference

$$\frac{C''\theta \vee t\theta \approx t'\theta \vee s\theta \approx s'\theta}{C''\theta \vee t'\theta \not\approx s'\theta \vee t\theta \approx t'\theta}\ \textsc{EqFact}$$

By Lemma 11, its conclusion is entailed by clauses from $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ that are smaller than $\lfloor C\theta \rfloor$. By part (iii) of the induction hypothesis, these clauses are true in $R_{\lfloor C\theta \rfloor}$, which implies that $\lfloor C''\theta \vee t'\theta \not\approx s'\theta \vee t\theta \approx t'\theta \rfloor$ is true in $R_{\lfloor C\theta \rfloor}$. Since the literal $\lfloor t'\theta \not\approx s'\theta \rfloor$ must be false in $R_{\lfloor C\theta \rfloor}$, this implies that $R_{\lfloor C\theta \rfloor} \models \lfloor C\theta \rfloor$, contradicting our assumption. □

Given a model R_∞ of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$, we construct a model $R_\infty^{\uparrow}$ of $\mathcal{G}_\Sigma(N)$. The key properties are that R_∞ is term-generated and that the interpretations of the members $\mathsf{f}_k, \ldots, \mathsf{f}_n$ of a same symbol family behave in the same way.

Lemma 14 (Argument congruence). *For all terms $\mathsf{f}_m(\bar{s})$ and $\mathsf{g}_n(\bar{t})$ if $[\![\mathsf{f}_m(\bar{s})]\!]^{\xi}_{R_\infty} = [\![\mathsf{g}_n(\bar{t})]\!]^{\xi}_{R_\infty}$, then $[\![\mathsf{f}_{m+1}(\bar{s}, u)]\!]^{\xi}_{R_\infty} = [\![\mathsf{g}_{n+1}(\bar{t}, u)]\!]^{\xi}_{R_\infty}$ for all u.*

Proof. What we want to show is equivalent to

$$R_\infty \models \mathsf{f}_m(\bar{s}) \approx \mathsf{g}_n(\bar{t}) \text{ implies } R_\infty \models \mathsf{f}_{m+1}(\bar{s}, u) \approx \mathsf{g}_{n+1}(\bar{t}, u)$$

which is equivalent to

$$\mathsf{f}_m(\bar{s}) \downarrow_{R_\infty} \mathsf{g}_n(\bar{t}) \text{ implies } \mathsf{f}_{m+1}(\bar{s}, u) \leftrightarrow^*_{R_\infty} \mathsf{g}_{n+1}(\bar{t}, u)$$

For every rewrite step rewriting a subterm, there is obviously an analogous rewrite step if the term u is appended at the top level. Therefore, it suffices to prove that

$$\mathsf{h}_k(\bar{v}) \to \mathsf{h}'_{k'}(\bar{v}') \in R_\infty \text{ implies } \mathsf{h}_{k+1}(\bar{v}, u) \leftrightarrow^*_{R_\infty} \mathsf{h}'_{k'+1}(\bar{v}', u)$$

for all function symbols h, h' and all k, k'.

Since $\mathsf{h}_k(\bar{v}) \to \mathsf{h}'_{k'}(\bar{v}') \in R_\infty$, it must come from a productive clause of the form $\lfloor C\theta \rfloor = \lfloor C'\theta \rfloor \vee \mathsf{h}_k(\bar{v}) \approx \mathsf{h}'_{k'}(\bar{v}')$. We have an ArgCong inference from $C\theta$ with the following ground instance:

$$\frac{C'\theta \vee \lceil \mathsf{h}_k(\bar{v}) \approx \mathsf{h}'_{k'}(\bar{v}') \rceil}{C'\theta \vee \lceil \mathsf{h}_k(\bar{v}, u) \approx \mathsf{h}'_{k'}(\bar{v}', u) \rceil \vee u_1 \not\approx u_1 \vee \cdots \vee u_l \not\approx u_l}\ \textsc{ArgCong}$$

(The additional literals $u_1 \not\approx u_1 \vee \cdots \vee u_l \not\approx u_l$ are due to purification. For the non-purifying variants, $l = 0$.) By the lifting lemma (Lemma 8), this is a ground instance of an inference from C. By part (ii) of Lemma 13, a productive clause is never redundant; hence $C\theta$ is not redundant and therefore C is not redundant. Hence, the conclusion E of the inference from C is in $N \cup \mathrm{Red}(N)$. Therefore, the ground instance $\lfloor C'\theta \rfloor \vee \mathsf{h}_k(\bar{v}, u) \approx \mathsf{h}'_{k'}(\bar{v}', u)$ of $\lfloor E \rfloor$ is either contained in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ or it is entailed by clauses in $\lfloor \mathcal{G}_\Sigma(N) \rfloor$. Thus, it is true in R_∞, because R_∞ is a model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$. By Lemma 6, $\lfloor C'\theta \rfloor$ is false in R_∞. The literals $\lfloor u_1 \not\approx u_1 \vee \cdots \vee u_l \not\approx u_l \rfloor$ are obviously false. So, $\mathsf{h}_k(\bar{v}, u) \approx \mathsf{h}'_{k'}(\bar{v}', u)$ must be true in R_∞ and $\mathsf{h}_{k+1}(\bar{v}, u) \leftrightarrow^*_{R_\infty} \mathsf{h}'_{k'+1}(\bar{v}', u)$. □

Definition 15. Define an interpretation $R_\infty^{\uparrow} = (\mathcal{U}^{\uparrow}, \mathcal{E}^{\uparrow}, \mathcal{J}^{\uparrow})$ in the ceiling logic as follows. Let $(\mathcal{U}, \mathcal{E}, \mathcal{J}) = R_\infty$. Let $\mathcal{U}^{\uparrow}_\tau = \mathcal{U}_{\lfloor \tau \rfloor}$ and $\mathcal{J}^{\uparrow}(\mathsf{f}) = \mathcal{J}(\mathsf{f}_k)$, where k is the number of mandatory arguments of f. Since R_∞ is term-generated, for every $a \in \mathcal{U}_{\lfloor \tau \to \upsilon \rfloor}$, there exists a ground term $s : \tau \to \upsilon$ such that $[\![\lfloor s \rfloor]\!]^{\xi}_{R_\infty} = a$. Without loss of generality, we write $s = \mathsf{f}(\bar{s}_k)\ s_{k+1} \ldots s_m$. Then define $\mathcal{E}^{\uparrow}$ as follows:

$$\begin{aligned}
\mathcal{E}^{\uparrow}_{\tau,\upsilon}(a) &= \mathcal{E}^{\uparrow}_{\tau,\upsilon}([\![\mathsf{f}_m(\lfloor \bar{s}_m \rfloor)]\!]^{\xi}_{R_\infty}) \\
&= \big(b \mapsto \mathcal{J}(\mathsf{f}_{m+1})([\![\lfloor \bar{s}_m \rfloor]\!]^{\xi}_{R_\infty}, b)\big) \\
&= \big([\![u]\!]^{\xi}_{R_\infty} \mapsto [\![\mathsf{f}_{m+1}(\lfloor \bar{s}_m \rfloor, u)]\!]^{\xi}_{R_\infty}\big)
\end{aligned}$$

This interpretation is well defined if the definition of $\mathcal{E}^{\uparrow}$ does not depend on the choice of the ground term s. To show this, we assume that there exists another ground term $t = \mathsf{g}(\bar{t}_l)\ t_{l+1} \ldots t_n$ such that $[\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty} = a$. By Lemma 14, it follows from $[\![\lfloor s \rfloor]\!]^{\xi}_{R_\infty} = [\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty}$ that

$$[\![\mathsf{f}_{m+1}(\lfloor \bar{s}_m \rfloor, u)]\!]^{\xi}_{R_\infty} = [\![\mathsf{g}_{n+1}(\lfloor \bar{t}_n \rfloor, u)]\!]^{\xi}_{R_\infty}$$

indicating that the definition of $\mathcal{E}^{\uparrow}$ is independent of the choice of s.

Since R_∞ is a term-generated model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$, we can show that $R^{\uparrow}_\infty$ is also term-generated. And using the same argument as in the first-order proof, we can lift this result to nonground clauses. For the extensional variants, we also need to show that $R^{\uparrow}_\infty$ is an extensional interpretation.

Lemma 16 (Model transfer to ceiling logic). *$R^{\uparrow}_\infty$ is a term-generated model of $\mathcal{G}_\Sigma(N)$.*

Proof. By Lemma 13, R_∞ is a model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$, i.e., for all clauses $\lfloor C \rfloor \in \lfloor \mathcal{G}_\Sigma(N) \rfloor$ we have $[\![\lfloor C \rfloor]\!]^{\xi}_{R_\infty} = 1$.

We prove by induction on ground terms t and ground formulas φ of the ceiling logic that $[\![t]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty}$ and $[\![\varphi]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor \varphi \rfloor]\!]^{\xi}_{R_\infty}$. It follows for all $C \in \mathcal{G}_\Sigma(N)$ that $[\![C]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor C \rfloor]\!]^{\xi}_{R_\infty} = 1$, and hence $(\mathcal{U}^{\uparrow}, \mathcal{E}^{\uparrow}, \mathcal{J}^{\uparrow})$ is a model of $\mathcal{G}_\Sigma(N)$.

Let t be a ground higher-order term, and we assume that $[\![t]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty}$ for all subterms of t. If t is of the form $\mathsf{f}(\bar{t}_k)$, then

$$\begin{aligned}
[\![t]\!]^{\xi}_{R^{\uparrow}_\infty} &= \mathcal{J}^{\uparrow}(f)([\![\bar{t}_k]\!]^{\xi}_{R^{\uparrow}_\infty}) \\
&= \mathcal{J}(\mathsf{f}_k)([\![\bar{t}_k]\!]^{\xi}_{R^{\uparrow}_\infty}) \\
&\overset{\text{IH}}{=} \mathcal{J}(\mathsf{f}_k)([\![\lfloor \bar{t}_k \rfloor]\!]^{\xi}_{R_\infty}) \\
&= [\![\mathsf{f}_k(\lfloor \bar{t}_k \rfloor)]\!]^{\xi}_{R_\infty} \\
&= [\![\lfloor \mathsf{f}(\bar{t}_k) \rfloor]\!]^{\xi}_{R_\infty} = [\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty}
\end{aligned}$$

If t is an application $t = t_1\ t_2$, where t_1 is of type $\tau \to \upsilon$, then writing t_1 as $t_1 = \mathsf{f}(\bar{s}_k)\ s_{k+1} \ldots s_m$ lets us derive

$$\begin{aligned}
[\![t_1\ t_2]\!]^{\xi}_{R^{\uparrow}_\infty} &= \mathcal{E}^{\uparrow}_{\tau,\upsilon}([\![t_1]\!]^{\xi}_{R^{\uparrow}_\infty})([\![t_2]\!]^{\xi}_{R^{\uparrow}_\infty}) \\
&\overset{\text{IH}}{=} \mathcal{E}^{\uparrow}_{\tau,\upsilon}([\![\lfloor t_1 \rfloor]\!]^{\xi}_{R_\infty})([\![\lfloor t_2 \rfloor]\!]^{\xi}_{R_\infty}) \\
&= \mathcal{E}^{\uparrow}_{\tau,\upsilon}([\![\mathsf{f}_m(\lfloor \bar{s}_m \rfloor]\!]^{\xi}_{R_\infty})([\![\lfloor t_2 \rfloor]\!]^{\xi}_{R_\infty}) \\
&\overset{\text{Def }\mathcal{E}^{\uparrow}}{=} [\![\mathsf{f}_{m+1}(\lfloor \bar{s}_m \rfloor, \lfloor t_2 \rfloor)]\!]^{\xi}_{R_\infty} \\
&= [\![\lfloor t_1\ t_2 \rfloor]\!]^{\xi}_{R_\infty}
\end{aligned}$$

So we have shown that $[\![t]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor t \rfloor]\!]^{\xi}_{R_\infty}$ for all terms t. Given that, the induction on formulas φ to show that $[\![\lfloor \varphi \rfloor]\!]^{\xi}_{R^{\uparrow}_\infty} = [\![\lfloor \varphi \rfloor]\!]^{\xi}_{R_\infty}$ is trivial.

It remains to show that $R_\infty^{\uparrow}$ is term-generated. Let $a \in \mathcal{U}_\tau^{\uparrow}$. Since $\mathcal{U}_\tau^{\uparrow} = \mathcal{U}_{\lfloor \tau \rfloor}$ and R_∞ is term-generated, we have a ground term t of the floor logic with $[\![t]\!]_{R_\infty}^{\xi} = a$. Using what we showed above, we have $[\![\lfloor t \rfloor]\!]_{R_\infty^{\uparrow}}^{\xi} = [\![t]\!]_{R_\infty}^{\xi} = a$. Hence, $R_\infty^{\uparrow}$ is term-generated. $\square$

Lemma 17 (Model transfer to nonground clauses). *$R_\infty^{\uparrow}$ is a model of N.*

Proof. Let $(\forall x. C) \in N$. Then $R_\infty^{\uparrow} \models \forall x. C$ iff $[\![C]\!]_{R_\infty^{\uparrow}}^{\xi[x_i \mapsto a_i]} = 1$ for all ξ and a_i. Choose ground terms t_i such that $[\![t_i]\!]_{R_\infty^{\uparrow}}^{\xi} = a_i$; define θ such that $x_i\theta = t_i$, then $[\![C]\!]_{R_\infty^{\uparrow}}^{\xi[x_i \mapsto a_i]} = [\![C]\!]_{R_\infty^{\uparrow}}^{\xi \circ \theta} = [\![C\theta]\!]_{R_\infty^{\uparrow}}^{\xi} = 1$ since $C\theta \in \mathcal{G}_\Sigma(N)$ and $R_\infty^{\uparrow} \models \mathcal{G}_\Sigma(N)$ by Lemma 16. $\square$

Lemma 18 (Completeness of the extensionality axioms). *If N contains the extensionality axioms, $R_\infty^{\uparrow}$ is extensional.*

Proof. Assume that the clause set N contains the extensionality axioms. By Lemma 17, the extensionality axioms are hence true in $R_\infty^{\uparrow}$.

Assume that $(\mathcal{U}^{\uparrow}, \mathcal{E}^{\uparrow})$ is not extensional. Then $\mathcal{E}^{\uparrow}$ is not injective, i.e., there are $a \neq b \in \mathcal{U}_{\tau \to \upsilon}^{\uparrow}$, such that $\mathcal{E}^{\uparrow}(a) = \mathcal{E}^{\uparrow}(b)$. Let $\xi = \{x \mapsto a, y \mapsto b\}$. Then

$$[\![\forall x.\, \forall y.\; x\,(\mathsf{diff}(x,y)) \not\approx y\,(\mathsf{diff}(x,y)) \vee x \approx y]\!]_{R_\infty^{\uparrow}}^{\xi} = 0$$

because

$$[\![x\,(\mathsf{diff}(x,y))]\!]_{R_\infty^{\uparrow}}^{\xi} = \mathcal{E}^{\uparrow}(a)([\![\mathsf{diff}(x,y)]\!]_{R_\infty^{\uparrow}}^{\xi}) = \mathcal{E}^{\uparrow}(b)([\![\mathsf{diff}(x,y)]\!]_{R_\infty^{\uparrow}}^{\xi}) = [\![y\,(\mathsf{diff}(x,y))]\!]_{R_\infty^{\uparrow}}^{\xi}$$

$\square$

We summarize the results of this section in the following theorem.

Theorem 19 (Refutational completeness). *Let N be a clause set that is saturated by any of the four calculi, up to redundancy. For the purifying calculi, we additionally assume that all clauses in N are purified. Then N has a model if and only if $\bot \notin N$. Such a model is extensional if N contains the extensionality axioms.*

Proof. If $\bot \in N$, then obviously N does not have a model. If $\bot \notin N$, then the interpretation R_∞ (that is, $\mathcal{T}_\Sigma^{\emptyset}/R_\infty$) is a model of $\lfloor \mathcal{G}_\Sigma(N) \rfloor$ according to part (iii) of Lemma 13. By Lemma 16, $R_\infty^{\uparrow}$ is a term-generated model of $\mathcal{G}_\Sigma(N)$. By Lemma 17, it is a model of N. If N contains the extensionality axioms, then $R_\infty^{\uparrow}$ is even an extensional model by Lemma 18. $\square$

5 Implementation in Zipperposition

Zipperposition [17,18] is an open source superposition-based theorem prover written in OCaml.[1] It was initially designed for polymorphic first-order logic with equality, as embodied by TPTP TFF [9]. Recently, we extended it with a pragmatic higher-order mode with support for λ-abstractions and extensionality, without any completeness guarantees. Using this mode, Zipperposition entered the 2017 edition of the CADE ATP System Competition [46]. We have now also implemented a complete λ-free mode based on the four calculi described in this paper, extended with polymorphism.

[1] `https://github.com/c-cube/zipperposition`

The pragmatic higher-order mode provided a convenient basis to implement our calculi. It includes higher-order term and type representations and orders. Its ad hoc calculus extensions are similar to our calculi. Notably, they include an ARGCONG rule and a POSEXT-like rule, and SUP inferences are performed only at argument subterms. In the term indexes, which are imperfect (overapproximating), terms whose head is an applied variable and λ-abstractions are treated as fresh variables. This could be further optimized to reduce the number of unification candidates. One of the bugs we found during our implementation work occurred because argument positions shift when applying substitutions to applied variables. We resolved this by numbering argument positions in terms from right to left.

To implement the λ-free mode, we restricted the unification algorithm to non-λ-terms, and we added support for mandatory arguments to make skolemization sound, by associating the number of mandatory arguments to each symbol and incorporating this number in the unification algorithm. To satisfy the requirements on selection, we avoid selecting literals that contain higher-order variables. Finally, we disabled rewriting of non-argument subterms to comply with our redundancy notion.

For the purifying calculi, we implemented purification as a simplification rule. This ensures that it is applied aggressively on all clauses, whether initial clauses from the problem or clauses produced during saturation, before any inferences are performed.

For the nonpurifying calculi, we added the possibility to perform SUP inferences at variable positions. This means that variables must be indexed as well. In addition, we modified the variable condition. However, it is in general impossible to decide whether there exists a ground substitution θ with $t\sigma\theta \succ t'\sigma\theta$ and $C\sigma\theta \prec C''\sigma\theta$. We overapproximate the condition as follows: (1) check whether x appears with different arguments in the clause C; (2) use an order-specific algorithm (for LPO and KBO) to determine whether there might exist a ground substitution θ and terms $\bar{u}$ such that $t\sigma\theta \succ t'\sigma\theta$ and $t\sigma\theta\,\bar{u} \prec t'\sigma\theta\,\bar{u}$; and (3) check whether $C\sigma \not\succeq C''\sigma$. If these three conditions apply, we conclude that there might exist a ground substitution θ witnessing nonmonotonicity.

For the extensional calculi, we added a single extensionality axiom based on a polymorphic symbol $\mathsf{diff} : \forall\alpha\beta.\ (\alpha \to \beta)^2 \Rightarrow \alpha$. To curb the explosion associated with extensionality, this axiom and all clauses derived from it are penalized by the clause selection heuristic. Moreover, we added a negative extensionality rule that resembles Vampire's extensionality resolution rule [25].

Using Zipperposition, we can quantify the disadvantage of the applicative encoding on the problem given at the end of Section 3.2. Well-chosen LPO and KBO instances allow Zipperposition to derive $\perp$ in 4 iterations and 0.04 s. KBO or LPO with default settings needs 203 iterations and 0.5 s, whereas KBO or LPO on the applicative encoding of the problem needs 203 iterations and almost 2 s.

6 Evaluation

We evaluated Zipperposition's implementation of our four calculi on TPTP benchmarks. We compare them with Zipperposition's first-order mode on the applicative encoding with and without the extensionality axiom. The encoding is implemented as a preprocessor, which makes all function symbols nullary and replaces all applications with a binary app symbol. For simplicity, the encoder uses a single polymorphic app symbol instead of a symbol family. Our experimental data is available online.[2] We used the developer version of Zipperposition, commit number 7fe2ebeb.[3]

We instantiated all variants with LPO [10] (which is nonmonotonic) and KBO [3] without argument coefficients (which is monotonic). This gives us a rough indication of the cost of nonmonotonicity. However, when using a monotonic order, it may be more efficient (and also refutationally complete) to superpose at non-argument subterms directly instead of relying on the ArgCong rule.

We collected 671 first-order problems in TPTP TFF format and 1114 higher-order problems in TPTP THF format, both groups containing monomorphic and polymorphic problems. We excluded all problems containing λ-expressions, the quantifier constants !! ($\forall$) and ?? ($\exists$), arithmetic types, or the $distinct predicate, as well as problems that nest Boolean expressions inside terms.

Figures 1 and 2 summarize, for various configurations, the number of solved satisfiable and unsatisfiable problems within 300 s (excluding the applicative encoder). The average time and number of iterations are computed over the problems that all configurations for the respective logic and term order found to be unsatisfiable within the timeout. The evaluation was carried out on StarExec [45] using Intel Xeon E5-2609 0 CPUs clocked at 2.40 GHz.

The experimental results on first-order problems confirm our hypothesis that the applicative encoding is inefficient. For LPO, the success rate drops by 16%–18%; for both orders, the average time to show unsatisfiability roughly quadruples. In contrast, our calculi handle first-order problems gracefully. Even the extensional calculi, which include graceless extensionality axioms, is almost as effective as the first-order mode. We expect that our calculi will scale up to large, mildly higher-order problems—a practically relevant class of problems that is underrepresented in the TPTP library.

The higher-order problems considered in this evaluation have a very different flavor. They tend to be small, and many of them are satisfiable for λ-free higher-order logic, even though they may be unsatisfiable for full higher-order logic and labeled as such in the TPTP. On these problems, the nonpurifying calculi outperform their purifying relatives. The applicative encoding and the nonpurifying calculi are comparable on unsatisfiable problems, which is probably indicative of the small size of the problems. The nonpurifying calculi saturate less often than the encoding, probably because of the tight selection restriction in our implementation, but the encoding is much slower, probably due to the additional symbols in the encoding. This difference in speed is smaller for the intensional calculi, a possible consequence of the argument congruence explosion.

The nonpurifying calculi perform better with KBO than with LPO. This confirms our expectations, given that KBO is generally considered the more robust default option for superposition and that the nonmonotonic LPO triggers Sup inferences at variable positions—the price to pay for the order's nonmonotonicity.

[2] http://matryoshka.gforge.inria.fr/pubs/lfhosup_data/

[3] https://github.com/c-cube/zipperposition/tree/7fe2ebeb

		# sat		# unsat		∅ time (s)		∅ iterations	
		LPO	KBO	LPO	KBO	LPO	KBO	LPO	KBO
TFF	first-order mode	0	0	**181**	**220**	**4.0**	**4.4**	**1497**	**1473**
	applicative encoding	0	0	150	203	19.0	16.0	1698	1916
	nonpurifying calculus	0	0	**181**	219	4.2	4.6	**1497**	**1473**
	purifying calculus	0	0	**181**	218	4.3	4.8	**1497**	**1473**
THF	applicative encoding	**444**	**438**	**676**	671	0.8	**0.2**	**72**	81
	nonpurifying calculus	353	360	675	**676**	**0.6**	0.3	83	**63**
	purifying calculus	338	343	664	666	0.8	1.0	116	231

Fig. 1: Evaluation of the intensional calculi

		# sat		# unsat		∅ time (s)		∅ iterations	
		LPO	KBO	LPO	KBO	LPO	KBO	LPO	KBO
TFF	first-order mode	0	0	**181**	**220**	**2.8**	**4.3**	**1219**	**1420**
	applicative encoding	0	0	151	201	19.0	17.6	1837	1792
	nonpurifying calculus	0	0	179	215	6.2	6.8	1610	1524
	purifying calculus	0	0	180	215	5.0	7.4	1291	1464
THF	applicative encoding	**426**	**421**	**677**	671	0.7	0.8	**78**	89
	nonpurifying calculus	310	327	669	**675**	**0.6**	**0.4**	83	**66**
	purifying calculus	227	261	647	650	1.0	1.0	114	108

Fig. 2: Evaluation of the extensional calculi

7 Discussion and Related Work

Our calculi join a long list of extensions and refinements of superposition. Among the most closely related is Peltier's [37] Isabelle formalization of the refutational completeness of a superposition calculus that operates on λ-free higher-order terms and that is parameterized by a monotonic term order. Extensions with polymorphism and induction, developed by Cruanes [17, 18] and Wand [51], contribute to increasing the power of automatic provers. Detection of inconsistencies in axioms, as suggested by Schulz et al. [42], is important for large axiomatizations.

Also of interest is Bofill and Rubio's [12] integration of nonmonotonic orders in ordered paramodulation, a precursor of superposition. Their work is a veritable tour de force, but it is also highly complicated and restricted to ordered paramodulation. Lack of compatibility with arguments being a mild form of nonmonotonicity, it seems preferable to start with superposition, enrich it with an ArgCong rule, and tune the side conditions until we obtain a complete calculus.

Most complications can be avoided by using a monotonic order such as KBO without argument coefficients, but we expect that the coefficients will play an important role to support λ-abstractions. For example, the term $\lambda x.\ x + x$ could be treated as a constant with a coefficient of 2 on its argument and a heavy weight to ensure $(\lambda x.\ x+x)\ y \succ y+y$.

LPO can also be used to good effect. This technique could allow provers to perform aggressive β-reduction in the vast majority of cases, without compromising completeness.

Many researchers have proposed or used encodings of higher-order logic constructs into first-order logic, including Robinson [39], Kerber [29], Dougherty [21], Dowek et al. [22], Hurd [28], Meng and Paulson [33], Obermeyer [36], and Czajka [19]. Encodings of types, such as those by Bobot and Paskevich [11] and Blanchette et al. [7], are also crucial to obtain a sound encoding of higher-order logic. These ideas are implemented in proof assistant tools such as HOLyHammer and Sledgehammer [8].

Another line of research has focused on the development of automated proof procedures for higher-order logic. Robinson's [38] and Huet's [27] pioneering work stands out. Andrews [1] and Benzmüller and Miller [5] provide excellent surveys. The competitive higher-order automatic theorem provers include LEO-II [6] (based on unordered paramodulation), Satallax [14] (based on a tableau calculus and a SAT solver), Agsy-HOL [32] (based on a focused sequent calculus and a generic narrowing engine), and Leo-III [44] (based on a pragmatic extension of superposition with no completeness guarantees). The Isabelle proof assistant [35] (which includes a tableau reasoner and a rewriting engine) and its Sledgehammer subsystem also participate in the higher-order division of the CADE ATP System Competition [46].

Zipperposition is a convenient vehicle for experimenting and prototyping because it is easier to understand and modify than highly-optimized C or C++ provers. Our middle-term goal is to design higher-order superposition calculi, implement them in state-of-the-art provers such as E [41], SPASS [52], and Vampire [31], and integrate these in proof assistants to provide a high level of automation. With its stratified architecture, Otter-λ [4] is perhaps the closest to what we are aiming at, but it is limited to second-order logic and offers no completeness guarantees. In preliminary work supervised by Blanchette and Schulz, Vukmirović [49] has generalized E's data structures and algorithms to λ-free higher-order logic, assuming a monotonic KBO [3].

8 Conclusion

We presented four superposition calculi for intensional and extensional λ-free higher-order logic and proved them refutationally complete. The calculi nicely generalize standard superposition and are compatible with our λ-free higher-order LPO and KBO. Our experiments confirm what one would naturally expect: that native support for partial application and applied variables outperforms the applicative encoding.

The new calculi reduce the gap between proof assistants based on higher-order logic and superposition provers. We can use them to reason about arbitrary higher-order problems by axiomatizing suitable combinators. But perhaps more importantly, they appear promising as a stepping stone towards complete, highly efficient automatic theorem provers for full higher-order logic.

Acknowledgment. We are grateful to the maintainers of StarExec for letting us use their service. We want to thank Sander Dahmen, Johannes Hölzl, Stephan Schulz, Geoff Sutcliffe, Petar Vukmirović, and all the participants in the 2017 Dagstuhl Seminar on Deduction beyond First-Order Logic for stimulating discussions. We also want to thank Mark Summerfield and Sophie Tourret for suggesting several textual improvements. Bentkamp and Blanchette's research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 713999, Matryoshka).

References

[1] Andrews, P.B.: Classical type theory. In: Robinson, J.A., Voronkov, A. (eds.) Handbook of Automated Reasoning, vol. II, pp. 965–1007. Elsevier and MIT Press (2001)

[2] Bachmair, L., Ganzinger, H.: Rewrite-based equational theorem proving with selection and simplification. J. Log. Comput. 4(3), 217–247 (1994)

[3] Becker, H., Blanchette, J.C., Waldmann, U., Wand, D.: A transfinite Knuth–Bendix order for lambda-free higher-order terms. In: de Moura, L. (ed.) CADE-26. LNCS, vol. 10395, pp. 432–453. Springer (2017)

[4] Beeson, M.: Lambda logic. In: Basin, D.A., Rusinowitch, M. (eds.) IJCAR 2004. LNCS, vol. 3097, pp. 460–474. Springer (2004)

[5] Benzmüller, C., Miller, D.: Automation of higher-order logic. In: Siekmann, J.H. (ed.) Computational Logic, Handbook of the History of Logic, vol. 9, pp. 215–254. Elsevier (2014)

[6] Benzmüller, C., Paulson, L.C., Theiss, F., Fietzke, A.: LEO-II—A cooperative automatic theorem prover for higher-order logic. In: Armando, A., Baumgartner, P., Dowek, G. (eds.) IJCAR 2008. LNCS, vol. 5195, pp. 162–170. Springer (2008)

[7] Blanchette, J.C., Böhme, S., Popescu, A., Smallbone, N.: Encoding monomorphic and polymorphic types. Log. Meth. Comput. Sci. 12(4) (2016)

[8] Blanchette, J.C., Kaliszyk, C., Paulson, L.C., Urban, J.: Hammering towards QED. J. Formaliz. Reas. 9(1), 101–148 (2016)

[9] Blanchette, J.C., Paskevich, A.: TFF1: The TPTP typed first-order form with rank-1 polymorphism. In: Bonacina, M.P. (ed.) CADE-24. LNCS, vol. 7898, pp. 414–420. Springer (2013)

[10] Blanchette, J.C., Waldmann, U., Wand, D.: A lambda-free higher-order recursive path order. In: Esparza, J., Murawski, A.S. (eds.) FoSSaCS 2017. LNCS, vol. 10203, pp. 461–479. Springer (2017)

[11] Bobot, F., Paskevich, A.: Expressing polymorphic types in a many-sorted language. In: Tinelli, C., Sofronie-Stokkermans, V. (eds.) FroCoS 2011. LNCS, vol. 6989, pp. 87–102. Springer (2011)

[12] Bofill, M., Rubio, A.: Paramodulation with non-monotonic orderings and simplification. J. Autom. Reason. 50(1), 51–98 (2013)

[13] Brand, D.: Proving theorems with the modification method. SIAM J. Comput. 4, 412–430 (1975)

[14] Brown, C.E.: Satallax: An automatic higher-order prover. In: Gramlich, B., Miller, D., Sattler, U. (eds.) IJCAR 2012. LNCS, vol. 7364, pp. 111–117. Springer (2012)

[15] Cervesato, I., Pfenning, F.: A linear spine calculus. J. Log. Comput. 13(5), 639–688 (2003)

[16] Church, A.: A formulation of the simple theory of types. J. Symb. Log. 5(2), 56–68 (1940)

[17] Cruanes, S.: Extending Superposition with Integer Arithmetic, Structural Induction, and Beyond. Ph.D. thesis, École polytechnique (2015)

[18] Cruanes, S.: Superposition with structural induction. In: Dixon, C., Finger, M. (eds.) FroCoS 2017. LNCS, vol. 10483, pp. 172–188. Springer (2017)

[19] Czajka, Ł.: Improving automation in interactive theorem provers by efficient encoding of lambda-abstractions. In: Avigad, J., Chlipala, A. (eds.) CPP 2016. pp. 49–57. ACM (2016)

[20] Digricoli, V.J., Harrison, M.C.: Equality-based binary resolution. J. ACM 33(2), 253–289 (1986)

[21] Dougherty, D.J.: Higher-order unification via combinators. Theor. Comput. Sci. 114(2), 273–298 (1993)
[22] Dowek, G., Hardin, T., Kirchner, C.: Higher-order unification via explicit substitutions (extended abstract). In: LICS '95. pp. 366–374. IEEE (1995)
[23] Fitting, M.: Types, Tableaus, and Gödel's God. Kluwer (2002)
[24] Gordon, M.J.C., Melham, T.F. (eds.): Introduction to HOL: A Theorem Proving Environment for Higher Order Logic. Cambridge University Press (1993)
[25] Gupta, A., Kovács, L., Kragl, B., Voronkov, A.: Extensional crisis and proving identity. In: Cassez, F., Raskin, J. (eds.) ATVA 2014. LNCS, vol. 8837, pp. 185–200. Springer (2014)
[26] Henkin, L.: Completeness in the theory of types. J. Symb. Log. 15(2), 81–91 (1950)
[27] Huet, G.P.: A mechanization of type theory. In: Nilsson, N.J. (ed.) IJCAI-73. pp. 139–146. William Kaufmann (1973)
[28] Hurd, J.: First-order proof tactics in higher-order logic theorem provers. In: Archer, M., Di Vito, B., Muñoz, C. (eds.) Design and Application of Strategies/Tactics in Higher Order Logics. pp. 56–68. NASA Technical Reports (2003)
[29] Kerber, M.: How to prove higher order theorems in first order logic. In: Mylopoulos, J., Reiter, R. (eds.) IJCAI-91. pp. 137–142. Morgan Kaufmann (1991)
[30] Kop, C.: Higher Order Termination: Automatable Techniques for Proving Termination of Higher-Order Term Rewriting Systems. Ph.D. thesis, Vrije Universiteit Amsterdam (2012)
[31] Kovács, L., Voronkov, A.: First-order theorem proving and Vampire. In: Sharygina, N., Veith, H. (eds.) CAV 2013. LNCS, vol. 8044, pp. 1–35. Springer (2013)
[32] Lindblad, F.: A focused sequent calculus for higher-order logic. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) IJCAR 2014. LNCS, vol. 8562, pp. 61–75. Springer (2014)
[33] Meng, J., Paulson, L.C.: Translating higher-order clauses to first-order clauses. J. Autom. Reason. 40(1), 35–60 (2008)
[34] Miller, D.A.: A compact representation of proofs. Studia Logica 46(4), 347–370 (1987)
[35] Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL: A Proof Assistant for Higher-Order Logic, LNCS, vol. 2283. Springer (2002)
[36] Obermeyer, F.H.: Automated Equational Reasoning in Nondeterministic λ-Calculi Modulo Theories $\mathscr{H}^*$. Ph.D. thesis, Carnegie Mellon University (2009)
[37] Peltier, N.: A variant of the superposition calculus. Archive of Formal Proofs (2016), `https://www.isa-afp.org/entries/SuperCalc.shtml`
[38] Robinson, J.: Mechanizing higher order logic. In: Meltzer, B., Michie, D. (eds.) Machine Intelligence, vol. 4, pp. 151–170. Edinburgh University Press (1969)
[39] Robinson, J.: A note on mechanizing higher order logic. In: Meltzer, B., Michie, D. (eds.) Machine Intelligence, vol. 5, pp. 121–135. Edinburgh University Press (1970)
[40] Schmidt-Schauß, M.: Unification in a combination of arbitrary disjoint equational theories. J. Symb. Comput. 8, 51–99 (1989)
[41] Schulz, S.: System description: E 1.8. In: McMillan, K.L., Middeldorp, A., Voronkov, A. (eds.) LPAR-19. LNCS, vol. 8312, pp. 735–743. Springer (2013)
[42] Schulz, S., Sutcliffe, G., Urban, J., Pease, A.: Detecting inconsistencies in large first-order knowledge bases. LNCS, vol. 10395, pp. 310–325. Springer (2017)
[43] Snyder, W., Lynch, C.: Goal directed strategies for paramodulation. In: Book, R.V. (ed.) RTA-91. LNCS, vol. 488, pp. 150–161. Springer (1991)
[44] Steen, A., Benzmüller, C.: The higher-order prover Leo-III, submitted
[45] Stump, A., Sutcliffe, G., Tinelli, C.: StarExec: A cross-community infrastructure for logic solving. In: Demri, S., Kapur, D., Weidenbach, C. (eds.) IJCAR 2014. LNCS, vol. 8562, pp. 367–373. Springer (2014)
[46] Sutcliffe, G.: The CADE-26 automated theorem proving system competition—CASC-26. AI Commun. 30(6), 419–432 (2017)

[47] Sutcliffe, G., Benzmüller, C., Brown, C.E., Theiss, F.: Progress in the development of automated theorem proving for higher-order logic. In: Schmidt, R.A. (ed.) CADE-22. LNCS, vol. 5663, pp. 116–130. Springer (2009)
[48] Sutcliffe, G., Schulz, S., Claessen, K., Baumgartner, P.: The TPTP typed first-order form with arithmetic. In: Bjørner, N., Voronkov, A. (eds.) LPAR-18. LNCS, vol. 7180, pp. 406–419. Springer (2012)
[49] Vukmirović, P.: Implementation of Lambda-Free Higher-Order Superposition. M.Sc. thesis, Vrije Universiteit Amsterdam (2018)
[50] Waldmann, U.: Automated reasoning II. Lecture notes, Max-Planck-Institut für Informatik (2016), `http://resources.mpi-inf.mpg.de/departments/rg1/teaching/autrea2-ss16/script-current.pdf`
[51] Wand, D.: Superposition: Types and Polymorphism. Ph.D. thesis, Universität des Saarlandes (2017)
[52] Weidenbach, C., Dimova, D., Fietzke, A., Kumar, R., Suda, M., Wischnewski, P.: SPASS version 3.5. In: Schmidt, R.A. (ed.) CADE-22. LNCS, vol. 5663, pp. 140–145. Springer (2009)