Non-Elementary Compression of First-Order Proofs in Deep Inference Using Epsilon-Terms

Cameron Allett
ca821@bath.ac.uk
University of Bath
Bath, England

ABSTRACT

I introduce the falsifier calculus, a new deep-inference proof system for first-order predicate logic in the language of Hilbert's epsilon-calculus. It uses a new inference rule, the falsifier rule, to introduce epsilon-terms into a proof, distinct from the critical axioms of the traditional epsilon-calculus. The falsifier rule is a generalisation of one of the quantifier-shifts, inference rules for shifting quantifiers inside and outside of formulae. Like the epsilon-calculus and proof systems which include quantifier-shifts, the falsifier calculus admits non-elementarily shorter cut-free proofs of certain first-order theorems than the sequent calculus.

Analogous to the way in which Herbrand's Theorem decomposes a proof into a first-order and a propositional part, connected by a Herbrand disjunction as an intermediate formula, I prove a decomposition theorem for the falsifier calculus which gives rise to a new notion of intermediate formula in the epsilon-calculus, falsifier disjunctions. I then prove that certain first-order theorems admit non-elementarily smaller falsifier disjunctions than Herbrand disjunctions.

CCS CONCEPTS

• **Theory of computation → Proof theory**; *Logic and verification*; Automated reasoning.

KEYWORDS

deep inference, epsilon-calculus, Herbrand's Theorem, first-order logic

ACM Reference Format:

Cameron Allett. 2024. Non-Elementary Compression of First-Order Proofs in Deep Inference Using Epsilon-Terms. In *39th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS '24), July 8–11, 2024, Tallinn, Estonia.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3661814.3662101

1 INTRODUCTION

1.1 Non-Elementary Compression

A remarkable aspect of first-order proof theory is that some proof systems admit cut-free proofs of certain theorems which are non-elementarily shorter than in other systems [2, 7]. In this work, I introduce the *falsifier calculus*, a new deep-inference proof system in the language of Hilbert's *epsilon-calculus* [20] which admits this non-elementary compression. This system uses a new inference rule, called the *falsifier rule*, that introduces ε-terms into a proof and is distinct from the critical axioms used in the traditional epsilon-calculus. I further prove a decomposition theorem for this system, analogous to Herbrand's Theorem, which gives rise to a new notion that I call *falsifier disjunctions*. Falsifier disjunctions are analogues to Herbrand disjunctions in the language of the epsilon-calculus such that certain first-order theorems admit non-elementarily shorter falsifier disjunctions than Herbrand disjunctions, providing a new perspective on the structure of Herbrand disjunctions. The aim of this work is to better understand the properties which make a first-order proof system admit the non-elementary compression and to provide a treatment of the epsilon-calculus from a modern perspective, where the primary concerns are complexity and normalisation rather than completeness and consistency.

The falsifier calculus is defined using *deep inference* [16], a design methodology for proof systems which allows inference rules to apply at arbitrary depth inside formulae, offering a more flexible composition mechanism for composing derivations and more freedom in permuting inference rules around a proof. In recent years, there has been interest in developing proof systems which include the deep-inference rules known as *quantifier-shifts*, inference rules for logical equivalences of the form

$$\exists x A(x) \lor B \equiv \exists x (A(x) \lor B) \qquad \exists x A(x) \land B \equiv \exists x (A(x) \land B)$$
$$\forall x A(x) \lor B \equiv \forall x (A(x) \lor B) \qquad \forall x A(x) \land B \equiv \forall x (A(x) \land B)$$

where x does not occur free in B. In [2], Aguilera and Baaz demonstrate that extending Gentzen's **LK** [14] by quantifier-shifts results in a system $\mathbf{LK}_{shift}$ for which there is no elementary function bounding the length of the shortest cut-free **LK** proof of a formula in terms of the length of its shortest cut-free $\mathbf{LK}_{shift}$ proof. Since quantifier-shifts involve rewriting inside a formula, they are natural deep-inference rules and it follows that deep-inference proof systems for first-order predicate logic admit the non-elementary compression for cut-free proofs.

In a deep-inference setting, most quantifier-shifts are trivial, in that they are derivable using other inference rules with linear complexity (Proposition 3.6), with the exceptions of the rule $\mathsf{r1}{\downarrow}\,\frac{\forall x (A(x) \lor B)}{\forall x A(x) \lor B}$ and its dual $\mathsf{r1}{\uparrow}\,\frac{\exists x A(x) \land B}{\exists x (A(x) \land B)}$. To understand why this is the case, consider the following derivation including $\mathsf{r1}{\downarrow}$:

$$\mathsf{r1}{\downarrow}\,\frac{\forall x \left[A(x) \lor \left[\exists\, \frac{B(x)}{\exists y B(y)} \right] \right]}{\forall x A(x) \lor \exists y B(y)}$$

ACM ISBN 979-8-4007-0660-8/24/07.
https://doi.org/10.1145/3661814.3662101

When the existential quantifier $\exists y$ is instantiated, it is witnessed by the variable x, which is bound by a universal quantifier $\forall x$. The existential quantifier $\exists y$ in the conclusion however is not in the scope of $\forall x$ and hence is not witnessed by x – the r1↓ rule alters the witness to the existential quantifier. To assign an explicit witness to such existential quantifiers, I introduce the falsifier rule

$$\varepsilon\,\frac{\forall x(A(x) \vee B(x))}{\forall x A(x) \vee B(\varepsilon_y \overline{A}(y))}$$

in the language of Hilbert's epsilon-calculus, where for a semantics with domain $\mathbb{D}$, $\varepsilon_y \overline{A}(y)$ takes the value of e if there exists some $e \in \mathbb{D}$ such that $\overline{A}(e)$ and takes an arbitrary value otherwise. The falsifier rule is sound since if $\forall x A(x)$ is false, there must exist some element e of the domain such that $\overline{A}(e)$. Since $\forall x(A(x) \vee B(x))$ by the premise of the rule, it follows that $B(e)$. By replacing the r1↓ quantifier-shift with a falsifier rule ε, we can permute the existential quantifier down through the quantifier-shift and obtain an explicit witness for the existential quantifier $\exists y$ in the conclusion, like so

$$\text{r1}{\downarrow}\,\frac{\forall x \boxed{A(x) \vee \exists\,\frac{B(x)}{\exists y B(y)}}}{\forall x A(x) \vee \exists y B(y)} \quad \to \quad \varepsilon\,\frac{\forall x(A(x) \vee B(x))}{\boxed{\forall x A(x) \vee \exists\,\frac{B(\varepsilon_y \overline{A}(y))}{\exists y B(y)}}}$$

I shall demonstrate that in a deep-inference setting, the falsifier rule alone is enough to yield the non-elementary compression for cut-free proofs, without the presence of quantifier-shifts.

The epsilon-calculus extends the language of first-order predicate logic by *ε-terms* $\varepsilon_x A$ for all variables x and formulae A. For a given semantics $[\![-]\!]_{\mathbb{D}}$ with domain $\mathbb{D}$, each ε-term is assigned a witness in $\mathbb{D}$ by

$$[\![\varepsilon_x A(x)]\!]_{\mathbb{D}} = \begin{cases} e & \text{if there exists some } e \in \mathbb{D} \text{ such that } [\![A(e)]\!]_{\mathbb{D}} \\ a & \text{for some arbitrary } a \in \mathbb{D}, \text{ otherwise} \end{cases}$$

where e is chosen by a choice function on $\mathcal{P}(\mathbb{D})$ and a is fixed. In the traditional epsilon-calculus, ε-terms are introduced into a proof by *critical axioms* (inferences of the form $A(t) \to A(\varepsilon_x A(x))$ for all terms t) and quantifiers are encoded by ε-terms using the logical equivalences $\exists x A(x) \equiv A(\varepsilon_x A(x))$ and $\forall x A(x) \equiv A(\varepsilon_x \overline{A}(x))$. It is also known that the traditional epsilon-calculus admits non-elementarily shorter cut-free proofs than **LK** for certain theorems [7]. In this work, I propose a new approach to the epsilon-calculus using falsifiers, guided by these complexity considerations. My system and results also do not use the encodings of quantifiers by ε-terms described above, circumventing some of the cumbersomeness of notation associated with the traditional epsilon-calculus.

1.2 Case Analysis Extraction

In the proof theory of first-order predicate logic, contractions on existential formulae may be understood as case analyses on the witnesses to the existential quantifiers in the premise. A natural operation to perform on a first-order proof is thus to extract these case analyses, deriving a disjunction of terms which witness the existential quantifiers in the conclusion. This is an essential notion of *Herbrand's Theorem* [19], a fundamental theorem of classical proof theory, and the propositional disjunction collecting the term witnesses is called a *Herbrand disjunction*. In the sequent calculus, Herbrand's Theorem is traditionally proved as a corollary to cut elimination, such as in a recent exposition of a proof of the theorem due to Buss [11] (with a correction due to McKinley [25]). In a deep-inference setting, Brünnler [9] has presented a proof of the general version of Herbrand's Theorem in the form of a decomposition theorem, which does not require cuts to be eliminated from the proof. Brünnler presents a procedure which transforms a first-order proof into a factorised proof of the form

$$\begin{array}{c} \forall x_1 \ldots \forall x_n \boxed{\begin{array}{l} \Vert \text{Propositional rules} \\ A' \end{array}} \\ \Vert\;\{\exists\} \\ A'' \\ \Vert\;\{\text{r1}{\downarrow}, \text{r2}{\downarrow}, \text{r3}{\downarrow}, \text{r4}{\downarrow}\} \\ A''' \\ \Vert\;\{\text{qc}{\downarrow}\} \\ A \end{array}$$

(see Section 2 for inference rule names) where the formula $\forall x_1 \ldots \forall x_n A'$ is a Herbrand disjunction for A.

Statman [35] has shown that, in general, there is no elementary bound on the size of the smallest Herbrand disjunction for a first-order theorem in terms of the size of its smallest proof. A natural question then is whether it is possible to extract the case analyses contained in existential contraction rules from a proof without producing the non-elementary blowups incurred by Herbrand's Theorem. As will be shown, this is possible in the falsifier calculus, with the result being a falsifier disjunction which gives a disjunction of witnesses for the existential quantifiers in the conclusion, of elementary size with respect to the size of the proof.

In a deep-inference setting, to extract the case analyses contained within existential contraction rules qc↓ from a proof, the rules may be permuted down a proof by recursively permuting them down through the rule immediately beneath them. However, as shown in Figure 1, when a qc↓ rule is permuted down through an r1↑ quantifier-shift, a universal cocontraction rule qc↑, the dual of qc↓, may be introduced. Dually, when a qc↑ is permuted up through an r1↓ quantifier-shift, a qc↓ rule may be introduced. Consequently, a procedure which successively permutes qc↓ rule instances down and qc↑ rule instances up a proof is non-terminating in the standard syntax of first-order predicate logic for certain proofs.

The falsifier calculus solves the problem of non-termination due to the more expressive syntax of the epsilon-calculus. By permuting existential rules down a proof, we obtain an explicit disjunction of witnesses for each existential quantifier in the proof, with ε-terms generated as witnesses when permuting down through a falsifier rule. Figure 2 then illustrates how the falsifier rule can be used to avoid introducing qc↓ rules when permuting qc↑ rules up through r1↓ rules, to give termination. In the construction, the r1↓ rule is replaced with a falsifier rule and a regular contraction rule c↓ is introduced in place of an existential contraction rule qc↓, since the epsilon-calculus syntax can express that the witnesses to the existential quantifiers in the premise of the qc↓ rule are equal.

The main result of this work, Theorem 3.3 the *Falsifier Decomposition Theorem*, is proved in this way, using a terminating procedure

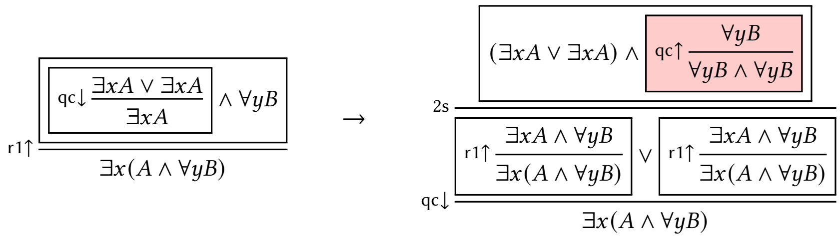


Figure 1: Reduction rule for permuting qc↓ **down through** r1↑

in the falsifier calculus which first permutes existential rules down to the bottom of a proof and then permutes qc↑ rules up the proof until they are eliminated. The resultant proof is in a normal form which I call *falsifier normal form*, that gives rise to the notion of falsifier disjunctions. The Falsifier Decomposition Theorem is analogous to Herbrand's Theorem in that case analyses are extracted from the proof but does not result in non-elementary blowups, since r1↓ quantifier-shifts in the proof are left intact in the form of falsifier rules which introduce ε-terms. The non-elementary difference in complexity between the two theorems may be seen as resulting from their difference in constructivity, since ε-terms represent elements which are drawn from the domain non-constructively.

1.3 Related Work

Since the formalisation of first-order predicate logic by Frege [13], a range of systems and techniques have been developed to investigate the shape and structure of first-order proofs, such as Gentzen's sequent calculus and natural deduction [14], Hilbert's epsilon-calculus [20] and Herbrand's Theorem [19]. More recent systems and approaches include Miller's *expansion proofs* [27], Coquand's *semantics of evidence* [12], Brünnler, Guglielmi and Ralph's work on first-order deep inference [8–10, 31–33], Heijltjes' *proof forests* [18], McKinley's *Herbrand nets* [26], Hughes' *first-order combinatorial proofs* [21, 22] and a game-semantic approach due to Alcolei, Clairambault, Hyland and Winskel [3].

The epsilon-calculus was initially introduced as part of Hilbert's program, with the goal of establishing a consistency proof for arithmetic, but has also seen a renewal of interest in recent years [4, 6, 7, 28–30].

Following Statman's original proof [35], the non-elementary compression of cut-free proofs and Herbrand disjunctions has also been studied by Aguilera, Baaz, Leitsch, Lolić and others [2, 5].

2 PRELIMINARIES

I begin by recalling some standard definitions for first-order logic and the epsilon-calculus.

Definition 2.1. Fix three disjoint countably infinite sets of symbols $\{x, y, z, \dots\}$, $\{f, g, h, \dots\}$, $\{P, Q, R, \dots\}$, whose respective elements are *variable symbols*, *function symbols* and *predicate symbols*, where every function symbol and predicate symbol has an associated non-negative integer arity and every predicate symbol P has a corresponding *dual* predicate symbol $\overline{P}$ of the same arity such that $\overline{\overline{P}} = P$ and $\overline{P} \neq P$.

I define *terms* t and *formulae* A by the following grammars:

$$t ::= x \mid f(t, \dots, t) \mid \varepsilon_x A$$
$$A ::= \mathbf{t} \mid \mathbf{f} \mid P(t, \dots, t) \mid A \vee A \mid A \wedge A \mid \exists x A \mid \forall x A$$

where x is a variable symbol, called a *variable*, f is a function symbol of arity n, each $f(t_1, \dots, t_n)$ is called a *function term*, function terms of arity 0 are called *constant terms*, $\varepsilon_x A$ is called an *ε-term*, $\mathbf{t}$ (true) and $\mathbf{f}$ (false) are called *units*, P is a predicate and each $P(t_1, \dots, t_n)$ is called an *atomic formula*.

The duals of formulae are defined using standard De Morgan duals.

Definition 2.2. The *dual* $\overline{A}$ of formulae A are defined recursively as follows. For all formulae A and B, all atomic formulae $P(t_1, \dots, t_n)$ and all variables x: $\overline{\mathbf{t}} \equiv \mathbf{f}$, $\overline{\mathbf{f}} \equiv \mathbf{t}$, $\overline{P(t_1, \dots, t_n)} \equiv \overline{P}(t_1, \dots, t_n)$, $\overline{\overline{P}(t_1, \dots, t_n)} \equiv P(t_1, \dots, t_n)$, $\overline{A \vee B} \equiv \overline{A} \wedge \overline{B}$, $\overline{A \wedge B} \equiv \overline{A} \vee \overline{B}$, $\overline{\exists x A} \equiv \forall x \overline{A}$ and $\overline{\forall x A} \equiv \exists x \overline{A}$.

In this work, I give a primarily syntactic treatment of first-order proofs and the epsilon-calculus. For the purposes of this work, either the extensional or intensional semantics for the epsilon-calculus presented in [36] may be used.

I introduce the following definitions to distinguish formulae in which quantifiers occur inside, but not outside, the scope of ε-terms.

Definition 2.3. A formula is said to be *weakly quantifier-free* if it is generated by the grammar

$$A ::= \mathbf{t} \mid \mathbf{f} \mid P(t_1, \dots, t_n) \mid A \vee A \mid A \wedge A$$

and is said to be *weakly existential-free* if it is generated by the grammar

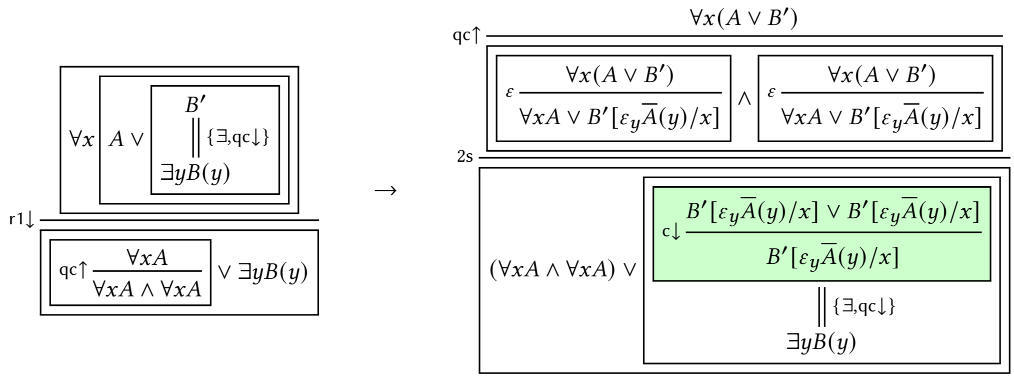
$$A ::= \mathbf{t} \mid \mathbf{f} \mid P(t_1, \dots, t_n) \mid A \vee A \mid A \wedge A \mid \forall x A$$

where $P(t_1, \dots, t_n)$ is an atomic formula and x is a variable.

The following definitions will be used when reasoning about terms inside of formulae.

Definition 2.4. A term t is said to *occur in* a term or formula as follows:

- t occurs in itself.
- If t occurs in a term s, then t occurs in all function terms of the form $f(t_1, ..., s, ..., t_n)$ and all atomic formulae of the form $P(t_1, ..., s, ..., t_n)$.
- If t occurs in a formula A, then t occurs in the formulae $A \vee B$, $B \vee A$, $A \wedge B$, $B \wedge A$, $\forall x A$ and $\exists x A$ and the term $\varepsilon_x A$ for all formulae B and all variable x.



where B' is the formula $B(t_1) \vee \cdots \vee B(t_n)$ for some terms $t_1, \ldots, t_n$.

Figure 2: Reduction rule for permuting qc↑ up through r1↓ in the presence of falsifiers

Definition 2.5. An occurrence of a variable x in a term or formula is said to be a *free occurrence* if it does not occur inside the scope of any $\exists x$, $\forall x$ or ε_x symbols.

If there is a free occurrence of a variable x in a term or formula, x is said to *occur free* in that term or formula.

If a formula is denoted $A(x_1, \ldots, x_n)$ for variables $x_1, \ldots, x_n$, I denote by $A(t_1, \ldots, t_n)$ the formula obtained by replacing every free occurrence of x_i in $A(x_1, \ldots, x_n)$ by the term t_i, for $i \in \{1, \ldots, n\}$. Likewise, if a term is denoted $t(x_1, \ldots, x_n)$, I denote by $t(s_1, \ldots, s_n)$ the term obtained by replacing every free occurrence of x_i in $t(x_1, \ldots, x_n)$ by the term s_i, for $i \in \{1, \ldots, n\}$.

Definition 2.6. A term t is said to be *free for* a variable x in a formula A if for all variables y which occur free in t, no free occurrence of x in A occurs inside the scope of a $\exists y$, $\forall y$ or ε_y symbol.

The following definition will be useful when distinguishing ε-terms which occur in a formula and contain free variables that are not bound by quantifiers or ε symbols in that formula.

Definition 2.7. If x occurs free in a formula $A(x)$ and $t(y)$ is a term such that y occurs free in $t(y)$ and $t(y)$ is free for x in $A(x)$, then $t(y)$ is said to *occur with* y *free in* $A(t(y))$.

Example 2.8. Let $A(y)$, $B(y)$ and $C(y)$ be formulae in which y occurs free. Then y occurs free in the ε-term $\varepsilon_x C(y)$ and $\varepsilon_x C(y)$ occurs in the formula $D(y)$ given by $A(y) \wedge \exists y B(\varepsilon_x C(y))$, but does not occur with y free in $D(y)$, since all occurrences of y in $\varepsilon_x C(y)$ are bound by an existential quantifier in $D(y)$. Conversely, $\varepsilon_x C(y)$ occurs with y free in the formula $D(\varepsilon_x C(y))$ given by $A(\varepsilon_x C(y)) \wedge \exists y B(\varepsilon_x C(y))$ since some occurrences of y in $A(\varepsilon_x C(y))$ occur free in $D(\varepsilon_x C(y))$.

I define the size of terms and formulae in the usual way and introduce the notion of *ε-size*.

Definition 2.9. The *size* $|t|$, *ε-size* $|t|_\varepsilon$ of terms t and *size* $|A|$ and *ε-size* $|A|_\varepsilon$ of formulae A are defined recursively as follows:

- For constant terms c, $|c| = |c|_\varepsilon = 1$.
- For variables x, $|x| = |x|_\varepsilon = 1$.
- For function terms $f(t_1, \ldots, t_n)$, $|f(t_1, \ldots, t_n)| = 1 + \sum_{i=1}^n |t_i|$ and $|f(t_1, \ldots, t_n)|_\varepsilon = 1 + \sum_{i=1}^n |t_i|_\varepsilon$.
- For ε-terms $\varepsilon_x A$, $|\varepsilon_x A| = 1$ and $|\varepsilon_x A|_\varepsilon = |A|_\varepsilon$.
- For formulae A and B, $|A \vee B| = |A \wedge B| = |A| + |B| + 1$ and $|A \vee B|_\varepsilon = |A \wedge B|_\varepsilon = |A|_\varepsilon + |B|_\varepsilon + 1$.
- For formulae A and variables x, $|\exists x A| = |\forall x A| = |A| + 1$ and $|\exists x A|_\varepsilon = |\forall x A|_\varepsilon = |A|_\varepsilon + 1$.
- For atomic formulae $P(t_1, \ldots, t_n)$, $|P(t_1, \ldots, t_n)| = 1 + \sum_{i=1}^n |t_i|$ and $|P(t_1, \ldots, t_n)|_\varepsilon = 1 + \sum_{i=1}^n |t_i|_\varepsilon$.
- $|\mathbf{t}| = |\mathbf{f}| = |\mathbf{t}|_\varepsilon = |\mathbf{f}|_\varepsilon = 1$.

Remark. In this work, complexity is of interest for the sake of proving elementary bounds for the proof and formula size of various constructions. I have thus chosen to measure the complexity of ε-terms in the maximal reasonable way, by the size of the formula bounded by the epsilon operator. Note that this differs from traditional complexity measures of ε-terms, such as *rank* and *degree* (see [30]).

I define the following sets of deep-inference inference rules from Figure 3 which will be used to construct derivations and proofs. For all of the symmetric relations $= \in \{=_{\mathbf{P}}, =_\exists, =_\forall\}$ in Figure 3, the corresponding inference rules are given by $= \dfrac{A}{B}$ if $A = B$.

Definition 2.10. I define the following sets of inference rules:

SKSg$_{\boldsymbol{P}}$ the set of inference rules given in the top row of Figure 3, with the restriction that the rules i↓, i↑, w↓, w↑, c↓ and c↑ may include only weakly quantifier-free formulae in the premise and conclusion

SKSg1 the set of all inference rules given in Figure 3, without any such restriction

where:

- Every inference rule $\rho \dfrac{A}{B}$ has a corresponding *dual* inference rule $\overline{\rho}$ given by $\overline{\rho} \dfrac{\overline{B}}{\overline{A}}$
- The inference rules corresponding to $A = QxA$ for $Qx \in \{\exists x, \forall x\}$ are called *vacuous* = rules
- For any instance $\forall \dfrac{\forall x A(x)}{A(t)}$ of the $\forall$ rule, the term t is said to *instantiate* the instance of the $\forall$ rule and for all terms $s(x)$ which occur with x free in $A(x)$, the term $s(t)$ is said to be *constructed by* the instance of the $\forall$ rule

- For any instance $\exists \frac{A(t)}{\exists x A(x)}$ of the $\exists$ rule, the term t is said to *witness* the instance of the $\exists$ rule

Note that $\mathsf{SKSg}_{\boldsymbol{P}}$ and SKSg1 are closed under dual rules.

For ease of expression, I will often denote instances of the $=_{\mathbf{P}}$, $=_\exists$ and $=_\forall$ rules simply by $=$ and sometimes omit instances of $=$ rules when displaying derivations.

A notable and useful characteristic of deep-inference proof systems is the ability to decompose inference rules into derivations of smaller rules. I introduce a set SKS1 of decomposed inference rules, shown later to be equivalent to SKSg1 (Propositions 3.5 and 3.6).

Definition 2.11. I define atomic variants ai↓, ai↑, ac↓, ac↑, aw↓, aw↑ of the rules i↓, i↑, c↓, c↑, w↓, w↑, which are identical to the standard variants except that the formulae A and $\overline{A}$ in the premise and conclusion of each rule as presented in Figure 3 must be atomic formulae.

I further define the *quantifier contraction* rules qc↓ and qc↑ by

$$\mathsf{qc}\!\downarrow \frac{\exists x A \vee \exists x A}{\exists x A} \qquad \mathsf{qc}\!\uparrow \frac{\forall x A}{\forall x A \wedge \forall x A}$$

I define the set of inference rules SKS1 = {ai↓, ai↑, ac↓, ac↑, aw↓, aw↑, s, m, $\exists$, $\forall$, $=_{\mathbf{P}}$, $=_\exists$, $=_\forall$, qc↓, qc↑, r1↓, r1↑} with the restriction that SKS1 does not contain any = rules of the form $= \frac{\exists x \exists y A}{\exists y \exists x A}$, $= \frac{\forall x \forall y A}{\forall y \forall x A}$, $= \frac{A}{\exists x A}$ or $= \frac{\forall x A}{A}$.

I define derivations in the *open deduction* formalism [17] as follows.

Definition 2.12. I define *derivations* ϕ with formula premises A and conclusions B, denoted $\phi \left\| \begin{smallmatrix} A \\ \\ B \end{smallmatrix}\right.$, their *size* $|\phi|$, *ε-size* $|\phi|_\varepsilon$, their *duals* $\overline{\phi} \left\| \begin{smallmatrix} \overline{A} \\ \\ \overline{B} \end{smallmatrix}\right.$, their *subderivations* and the terms which *occur (free)* in them inductively as follows:

- Every derivation is a subderivation of itself.
- Every formula A is a derivation, with premise A, conclusion A, size $|A|$, ε-size $|A|_\varepsilon$ and dual $\overline{A}$.

For all derivations $\psi \left\| \begin{smallmatrix} A \\ \\ A' \end{smallmatrix}\right.$ and $\chi \left\| \begin{smallmatrix} B \\ \\ B' \end{smallmatrix}\right.$, we have the following:

- Composition by inference: if $\rho \frac{A'}{B}$ is an instance of an inference rule ρ,

$$\psi;_\rho\chi \left\| \begin{matrix} A \\ \\ B' \end{matrix}\right. \equiv \rho \frac{\boxed{\psi \left\| \begin{matrix} A \\ A' \end{matrix}\right.}}{\boxed{\chi \left\| \begin{matrix} B \\ B' \end{matrix}\right.}}$$

is a derivation with $|\psi;_\rho \chi| = |\psi| + |\chi|$, $|\psi;_\rho \chi|_\varepsilon = |\psi|_\varepsilon + |\chi|_\varepsilon$ and $\overline{\psi;_\rho \chi} \equiv \overline{\chi};_{\overline{\rho}} \overline{\psi}$. Every subderivation of ψ and every subderivation of χ is a subderivation of $\psi;_\rho \chi$. Every term which occurs in ψ or χ also occurs in $\psi;_\rho \chi$ and every free occurrence of a term in ψ or χ is also a free occurrence in $\psi;_\rho \chi$.
- Composition by connective: for $\star \in \{\vee, \wedge\}$,

$$\psi\star\chi \left\| \begin{matrix} A \star B \\ \\ A' \star B' \end{matrix}\right. \equiv \left(\boxed{\psi \left\| \begin{matrix} A \\ A' \end{matrix}\right.} \star \boxed{\chi \left\| \begin{matrix} B \\ B' \end{matrix}\right.} \right)$$

is a derivation with $|\psi\star\chi| = |\psi|+|\chi|+1$ $|\psi\star\chi|_\varepsilon = |\psi|_\varepsilon+|\chi|_\varepsilon+1$ and $\overline{\psi \star \chi} \equiv \overline{\psi} \mathbin{\overline{\star}} \overline{\chi}$, where $\overline{\vee} = \wedge$ and $\overline{\wedge} = \vee$. Every subderivation of ψ and every subderivation of χ is a subderivation of $\psi \star \chi$. Every term which occurs in ψ or χ also occurs in $\psi \star \chi$ and every free occurrence of a term in ψ or χ is also a free occurrence in $\psi \star \chi$.
- Composition by quantifier: for $Qx \in \{\forall x, \exists x\}$, where x is any variable,

$$Qx\psi \left\| \begin{matrix} QxA \\ \\ QxA' \end{matrix}\right. \equiv Qx \left(\psi \left\| \begin{matrix} A \\ \\ A' \end{matrix}\right. \right)$$

is a derivation with $|Qx\psi| = |\psi| + 1$, $|Qx\psi|_\varepsilon = |\psi|_\varepsilon + 1$ and $\overline{Qx\psi} \equiv \overline{Qx}\,\overline{\psi}$, where $\overline{\exists x} = \forall x$ and $\overline{\forall x} = \exists x$. Every subderivation of ψ is a subderivation of $Qx\psi$. Every term which occurs in ψ also occurs in $Qx\psi$ and every free occurrence of a term in ψ is also a free occurrence in $Qx\psi$, unless the term is x, which has no free occurrences in $Qx\psi$.

Composition by inference and composition by connective are defined to be associative: for $\star \in \{\vee, \wedge\}$, all inference rules ρ_1 and ρ_2 and all derivations ψ, ϕ, χ, $(\psi \star \phi) \star \chi \equiv \psi \star (\phi \star \chi) \equiv \psi \star \phi \star \chi$ and $\psi;_{\rho_1} (\phi;_{\rho_2} \chi) \equiv (\psi;_{\rho_1} \phi);_{\rho_2} \chi \equiv \psi;_{\rho_1} \phi;_{\rho_2} \chi$.

$$\phi \left\| \begin{matrix} A \\ \\ B \end{matrix}\right. \mathcal{S}$$

denotes that every inference rule in the derivation ϕ is an element of the set $\mathcal{S}$.

If a derivation ϕ with conclusion A has premise $\mathbf{t}$, it is called a *proof*, denoted $\phi \left\| \begin{smallmatrix} \\ A \end{smallmatrix}\right.$.

Remark. Observe that the definitions of size and ε-size are such that for any formula A, $|A|_\varepsilon \leq |\varepsilon_x B|_\varepsilon |A|$, where $\varepsilon_x B$ is the largest ε-term which occurs in A and, similarly, for any derivation ϕ, $|\phi|_\varepsilon \leq |\varepsilon_x B|_\varepsilon |\phi|$, where $\varepsilon_x B$ is the largest ε-term which occurs in ϕ.

Definition 2.13. A derivation ϕ is said to be *epsilon-free* if no ε-terms occur in ϕ.

Soundness and completeness of the open deduction system with rules SKSg1 follows by translating into the system SKSgq presented in [8], since SKSg1 locally simulates every inference rule in SKSgq.

Theorem 2.14. *Every valid epsilon-free formula has a proof in* SKSg1.

I further introduce the following, mostly standard, definitions.

Definition 2.15. A formula A which is a subderivation of a formula B is said to be a *subformula* of B.

The inference rules of $\mathsf{SKSg}_{\boldsymbol{p}}$:

$$\mathsf{i}{\downarrow}\frac{\mathbf{t}}{A \vee \bar{A}} \quad \mathsf{w}{\downarrow}\frac{\mathbf{f}}{A} \quad \mathsf{c}{\downarrow}\frac{A \vee A}{A} \quad \mathsf{s}\frac{A \wedge (B \vee C)}{(A \wedge B) \vee C}$$

$$\mathsf{i}{\uparrow}\frac{A \wedge \bar{A}}{\mathbf{f}} \quad \mathsf{w}{\uparrow}\frac{A}{\mathbf{t}} \quad \mathsf{c}{\uparrow}\frac{A}{A \wedge A} \quad \mathsf{m}\frac{(A \wedge B) \vee (C \wedge D)}{(A \vee C) \wedge (B \vee D)}$$

$$A \vee \mathbf{f} =_{\mathbf{P}} A \qquad \mathbf{f} \wedge \mathbf{f} =_{\mathbf{P}} \mathbf{f} \qquad \mathbf{t} \vee \mathbf{t} =_{\mathbf{P}} \mathbf{t} \qquad A \wedge \mathbf{t} =_{\mathbf{P}} A$$

$$A \vee B =_{\mathbf{P}} B \vee A \qquad (A \vee B) \vee C =_{\mathbf{P}} A \vee (B \vee C)$$

$$A \wedge B =_{\mathbf{P}} B \wedge A \qquad (A \wedge B) \wedge C =_{\mathbf{P}} A \wedge (B \wedge C)$$

The remaining inference rules of SKSg1:

$$\exists\frac{A(t)}{\exists x A(x)} \quad \mathsf{r1}{\downarrow}\frac{\forall x(A(x) \vee B)}{\forall x A(x) \vee B} \quad \mathsf{r2}{\downarrow}\frac{\forall x(A(x) \wedge B)}{\forall x A(x) \wedge B} \quad \mathsf{r3}{\downarrow}\frac{\exists x(A(x) \vee B)}{\exists x A(x) \vee B} \quad \mathsf{r4}{\downarrow}\frac{\exists x(A(x) \wedge B)}{\exists x A(x) \wedge B}$$

$$\forall\frac{\forall x A(x)}{A(t)} \quad \mathsf{r1}{\uparrow}\frac{\exists x A(x) \wedge B}{\exists x(A(x) \wedge B)} \quad \mathsf{r2}{\uparrow}\frac{\exists x A(x) \vee B}{\exists x(A(x) \vee B)} \quad \mathsf{r3}{\uparrow}\frac{\forall x A(x) \wedge B}{\forall x(A(x) \wedge B)} \quad \mathsf{r4}{\uparrow}\frac{\forall x A(x) \vee B}{\forall x(A(x) \vee B)}$$

where t is free for x in $A(x)$ in the $\exists$ and $\forall$ rules and x does not occur free in B in the remaining rules.

$$A =_{\exists} \exists x A \qquad \exists x \exists y B =_{\exists} \exists y \exists x B \qquad \exists x C(x) =_{\exists} \exists y C(y)$$

$$A =_{\forall} \forall x A \qquad \forall x \forall y B =_{\forall} \forall y \forall x B \qquad \forall x C(x) =_{\forall} \forall y C(y)$$

where x does not occur free in A and y is free for x in $C(x)$.

Figure 3: The inference rules of SKSg1

Definition 2.16. A derivation is said to be *cut-free* if it contains no instances of the rules i↑ or ai↑.

Definition 2.17. For every derivation ϕ, terms s, t and variable x, I denote by $\phi[t/x]$ the derivation obtained by replacing every free occurrence of x in ϕ with t and $s[t/x]$ the term obtained by replacing every free occurrence of x in s with t.

Definition 2.18. A *formula context* $K\{\}$ is a function from derivations to derivations which is a formula with exactly one occurrence of the *hole* $\{-\}$ in the position of an atomic formula. For all derivations ϕ, $K\{\phi\}$ is given by replacing the hole in $K\{\}$ with ϕ.

For convenience in normalisation and describing the structure of derivations, I introduce the notion of *sequential composition.*

Definition 2.19. Let $\psi \begin{array}{c} A \\ \| \\ A' \end{array}$ and $\phi \begin{array}{c} A' \\ \| \\ B \end{array}$ be derivations. The *sequential composition* $\psi;\phi$ of ψ and ϕ is defined recursively as follows:

- If ψ is a formula, then $\psi;\phi \equiv \phi$. Likewise, if ϕ is a formula, then $\psi;\phi \equiv \psi$.
- If $\psi \equiv \chi;_\rho \omega$, where χ and ω are derivations and ρ is an inference rule, then $\psi;\phi \equiv \chi;_\rho (\omega;\phi)$. Likewise, if $\phi \equiv \chi;_\rho \omega$, then $\psi;\phi \equiv (\psi;\chi);_\rho \omega$.
- If $\psi \equiv \chi \star \omega$ and $\phi \equiv \chi' \star \omega'$, where $\star \in \{\vee, \wedge\}$ and χ, χ', ω and ω' are derivations such that the conclusion of χ is the premise of χ' and the conclusion of ω is the premise of ω', then $\psi;\phi \equiv (\chi;\chi') \star (\omega;\omega')$.
- If $\psi \equiv Qx\chi$ and $\phi \equiv Qx\chi'$, where $Qx \in \{\forall x, \exists x\}$ for some variable x and χ and χ' are derivations, then $\psi;\phi \equiv Qx(\chi;\chi')$.

I will write $\begin{array}{c} A \\ \psi \| \\ A' \\ \phi \| \\ B \end{array}$ to mean $\psi;\phi$.

I define the notion of an inference rule *occurring above* or *occurring below* another instance of an inference rule in a derivation as follows.

Definition 2.20. Let $\rho_1 \frac{A_1}{B_1}$ and $\rho_2 \frac{A_2}{B_2}$ be instances of inference rules ρ_1 and ρ_2 in a derivation ϕ. If ϕ may be expressed in the form $\psi; K\{A_1;_{\rho_1} B_1\}; \psi'; J\{A_2;_{\rho_2} B_2\}; \psi''$ for some derivations ψ, ψ', ψ'' and formula contexts $K\{\}$ and $J\{\}$ but not in the form $\chi; K'\{A_2;_{\rho_2} B_2\}; \chi'; J'\{A_1;_{\rho_1} B_1\}; \chi''$ for any derivations χ, χ', χ'' and formula contexts $K'\{\}$ and $J'\{\}$, then the instance of ρ_1 is said to *occur above* the instance of ρ_2 in ϕ and the instance of ρ_2 is said to *occur below* the instance of ρ_1 in ϕ.

An instance of an inference rule ρ in a derivation ϕ is said to be a *lowermost rule instance (of ρ)* in ϕ if it does not occur above any other rule instances (of ρ) in ϕ. Likewise, it is said to be an *uppermost rule instance (of ρ)* in ϕ if it does not occur below any other rule instances (of ρ) in ϕ.

3 THE FALSIFIER CALCULUS

I now introduce the falsifier calculus $\mathsf{SKSg}\varepsilon$ as the system comprised of propositional rules $\mathsf{SKSg}_{\mathbf{P}}$, the universal instantiation rule $\forall$, universal equality rules $=_\forall$ and the falsifier rule ε, given as follows.

Definition 3.1. The *falsifier rule* ε is given by

$$\varepsilon\,\frac{\forall x(A(x)\vee B(x))}{\forall xA(x)\vee B(\varepsilon_y\overline{A}(y))}$$

for all formulae $A(x)$, $B(x)$ and all variables y such that $\varepsilon_y\overline{A}(y)$ is free for x in $B(x)$.

For any instance of the ε rule as above, the ε-term $\varepsilon_y\overline{A}(y)$ is called the *critical term* of the instance of the ε rule. For all terms $t(x)$ which occur with x free in $B(x)$, the term $t(\varepsilon_y\overline{A}(y))$ is said to be *constructed by* the instance of ε.

Definition 3.2. The *falsifier calculus* $\mathsf{SKSg}\varepsilon$ is given by $\mathsf{SKSg}\varepsilon = \mathsf{SKSg}_{\mathbf{P}}\cup\{\varepsilon,\forall,=_\forall\}$.

I can now state the main result of this work, the Falsifier Decomposition Theorem, which decomposes any first-order proof into an upper segment in the falsifier calculus $\mathsf{SKSg}\varepsilon$ and a lower segment in $\{\exists,\mathsf{qc}\!\downarrow\}$, with a falsifier disjunction as an intermediate formula connecting the two segments.

Theorem 3.3 (The Falsifier Decomposition Theorem). *For every epsilon-free proof ϕ with conclusion A in $\mathsf{SKSg1}$, there exists a proof of the form*

$$\begin{array}{c} \phi'\,\Big\Vert\,\mathsf{SKSg}\varepsilon \\ A' \\ \Big\Vert\,\{\exists,\mathsf{qc}\!\downarrow\} \\ A \end{array} \tag{1}$$

such that the following elementary bounds hold

$$|\phi'| = \exp^{10}(O(|\phi|^2\ln|\phi|))$$
$$|A'| = \exp^{7}(O(|\phi|^2\ln|\phi|))$$
$$|\phi'|_\varepsilon = \exp^{12}(O(|\phi|^2\ln|\phi|))$$
$$|A'|_\varepsilon = \exp^{12}(O(|\phi|^2\ln|\phi|))$$

Furthermore, if ϕ is cut-free, then ϕ' may be chosen to be cut-free.

It is expected that smaller bounds exist for the sizes and ε-sizes of ϕ' and A' than those given above, but the present bounds have been chosen for the sake of exposition of the complexity assessment.

Definition 3.4. The normal form for proofs given by (1) is called *falsifier normal form* and the formula A' is called a *falsifier disjunction for A.*

I defer the proof of Theorem 3.3 and the statements of some of its consequences to Section 3.2.

3.1 Rule Admissibility and Permutations

In order to prove Theorem 3.3 the Falsifier Decomposition Theorem, I first establish some lemmas and propositions.

I note the following standard property of first-order deep-inference proof systems, that inference rules may be decomposed into derivations in $\mathsf{SKS1}$.

Proposition 3.5 (Atomicity). *For every instance $\rho\frac{A}{B}$ of an inference rule $\rho\in\{\mathsf{i}\!\downarrow,\mathsf{i}\!\uparrow,\mathsf{c}\!\downarrow,\mathsf{c}\!\uparrow,\mathsf{w}\!\downarrow,\mathsf{w}\!\uparrow,=_\exists,=_\forall\}$ such that if ρ is an $=$ rule it is of the form $=\frac{\exists x\exists yA}{\exists y\exists xA}$, $=\frac{\forall x\forall yA}{\forall y\forall xA}$, $=\frac{A}{\exists xA}$ or $=\frac{\forall xA}{A}$, there exists a derivation*

$$\begin{array}{c} A \\ \rho'\,\Big\Vert\,\mathcal{S} \\ B \end{array}$$

of size $O((|A|+|B|)^2)$ if $\rho\in\{\mathsf{i}\!\downarrow,\mathsf{i}\!\uparrow,\mathsf{c}\!\downarrow,\mathsf{c}\!\uparrow\}$ or $O(|A|+|B|)$ if $\rho\in\{\mathsf{w}\!\downarrow,\mathsf{w}\!\uparrow,=_\exists,=_\forall\}$, where $\mathcal{S}\subseteq\mathsf{SKS1}$ is given by

- $\{\mathsf{ai}\!\downarrow,\exists,\mathsf{r1}\!\downarrow,\mathsf{s},=_{\mathbf{P}},=_\exists,=_\forall\}$ *if ρ is* $\mathsf{i}\!\downarrow$
- $\{\mathsf{ai}\!\uparrow,\forall,\mathsf{r1}\!\uparrow,\mathsf{s},=_{\mathbf{P}},=_\exists,=_\forall\}$ *if ρ is* $\mathsf{i}\!\uparrow$
- $\{\mathsf{ac}\!\downarrow,\mathsf{m},\mathsf{qc}\!\downarrow,\forall,=_{\mathbf{P}},=_\forall\}$ *(or* $\{\mathsf{ac}\!\downarrow,\mathsf{m},\forall,=_{\mathbf{P}},=_\forall\}$ *if A is weakly existential-free) if ρ is* $\mathsf{c}\!\downarrow$
- $\{\mathsf{ac}\!\uparrow,\mathsf{m},\mathsf{qc}\!\uparrow,\exists,=_{\mathbf{P}},=_\exists\}$ *(or* $\{\mathsf{ac}\!\uparrow,\mathsf{m},\mathsf{qc}\!\uparrow,=_{\mathbf{P}}\}$ *if A is weakly existential-free) if ρ is* $\mathsf{c}\!\uparrow$
- $\{\mathsf{aw}\!\downarrow,\exists,=_{\mathbf{P}},=_\forall\}$ *if ρ is* $\mathsf{w}\!\downarrow$
- $\{\mathsf{aw}\!\uparrow,\forall,=_{\mathbf{P}},=_\exists\}$ *(or* $\{\mathsf{aw}\!\uparrow,\forall,=_{\mathbf{P}}\}$ *if A is weakly existential-free) if ρ is* $\mathsf{w}\!\uparrow$
- $\{\exists,=_\exists\}$ *if ρ is* $=_\exists$
- $\{\forall,=_\forall\}$ *if ρ is* $=_\forall$

Proof. Omitted. See [32], Lemmas 3.17–19 for similar decompositions. ◻

Remarkably, apart from $\mathsf{r1}\!\downarrow$ and $\mathsf{r1}\!\uparrow$, the quantifier-shifts may also be decomposed into derivations in $\mathsf{SKS1}$, as follows.

Proposition 3.6 (Decomposition of quantifier-shifts). *For every instance $\rho\frac{A}{B}$ of an inference rule $\rho\in\{\mathsf{r2}\!\downarrow,\mathsf{r2}\!\uparrow,\mathsf{r3}\!\downarrow,\mathsf{r3}\!\uparrow,\mathsf{r4}\!\downarrow,\mathsf{r4}\!\uparrow\}$, there exists a derivation*

$$\begin{array}{c} A \\ \Big\Vert\,\{\mathsf{aw}\!\downarrow,\mathsf{aw}\!\uparrow,\mathsf{qc}\!\downarrow,\mathsf{qc}\!\uparrow,\exists,\forall,=_{\mathbf{P}},=_\exists,=_\forall\} \\ B \end{array}$$

of size $O(|A|)$, which does not contain any $=$ rules of the form $=\frac{\exists x\exists yA}{\exists y\exists xA}$, $=\frac{\forall x\forall yA}{\forall y\forall xA}$, $=\frac{A}{\exists xA}$ or $=\frac{\forall xA}{A}$.

Proof. I present derivations for $\mathsf{r2}\!\downarrow$, $\mathsf{r3}\!\downarrow$ and $\mathsf{r4}\!\downarrow$. The remaining rules may be derived dually.

$$\mathsf{r2}\!\downarrow\frac{\forall x(A(x)\wedge B)}{\forall xA(x)\wedge B}$$

$$\downarrow$$

$$\mathsf{qc}\!\uparrow\frac{\forall x(A(x)\wedge B)}{\boxed{=\frac{\forall x\boxed{A(x)\wedge\boxed{\mathsf{w}\!\uparrow\frac{B}{\mathbf{t}}}}\wedge\boxed{\forall\frac{\forall x\boxed{\boxed{\mathsf{w}\!\uparrow\frac{A(x)}{\mathbf{t}}}\wedge B}}{\mathbf{t}\wedge B}}}{\forall xA(x)\wedge B}}}$$

where the instances of $\mathsf{w}\!\uparrow$ in the derivation above are replaced with derivations in $\{\mathsf{aw}\!\uparrow,\forall,=_{\mathbf{P}},=_\exists\}$ using Proposition 3.5 and sequential composition.

$$\text{r3}{\downarrow}\,\frac{\exists x(A(x)\vee B)}{\exists x A(x)\vee B} \quad\longrightarrow\quad =\frac{\exists x\,\boxed{\boxed{\exists\,\frac{A(x)}{\exists y A(y)}}\vee B}}{\exists x A(x)\vee B}$$

where y is some variable that is free for x in $A(x)$.

$$\text{r4}{\downarrow}\,\frac{\exists x(A(x)\wedge B)}{\exists x A(x)\wedge B} \quad\longrightarrow\quad =\frac{\exists x\,\boxed{\boxed{\exists\,\frac{A(x)}{\exists y A(y)}}\wedge B}}{\exists x A(x)\wedge B}$$

where y is some variable that is free for x in $A(x)$.

Observe that the size of each construction is linear with respect to the size of the premise of each inference rule. The result follows. □

The following lemma will be used to establish bounds for the ε-size of formulae and derivations in terms of their size. In particular, it provides an elementary bound for the ε-size of a derivation in the falsifier calculus terms of its size.

Lemma 3.7. *Let ϕ be a derivation in* SKSgε *such that every ε-term which occurs in ϕ is constructed by some instance of the ε rule or the $\forall$ rule in ϕ and for every instance $\rho\,\frac{A}{B}$ of an inference rule in ϕ:*

1. *every ε-term which occurs in A is constructed by some instance of the ε rule or the $\forall$ rule which occurs above the instance of ρ in ϕ*
2. *if ρ is $\forall$, every ε-term which occurs in the term t that instantiates the instance of ρ is constructed by some instance of the ε rule or the $\forall$ rule which occurs above the instance of ρ in ϕ*
3. *if ρ is an inference rule other than ε or $\forall$, every ε-term which occurs in B is constructed by some instance of the ε rule or the $\forall$ rule which occurs above the instance of ρ in ϕ*

Then $|\phi|_\varepsilon = O(\exp(\exp|\phi|\ln|\phi|))$.

Proof. By induction on the number of instances of ε and $\forall$. □

The proof of Theorem 3.3 will proceed in three phases. In the first phase, existential contraction rules qc↓ are permuted down to the bottom of the proof. In the second phase, existential instantiation rules $\exists$ are permuted down the proof to separate the proof into an upper segment of weakly existential-free formulae and a lower segment in $\{\exists, \text{qc}{\downarrow}\}$. In the third phase, universal cocontraction rules qc↑ are permuted up the proof until they are eliminated. The procedure will use the following lemmas to locally rewrite subderivations of the proof when performing the permutations.

The following lemma provides reduction rules for permuting qc↓ rules down through other inference rules during the first phase of the procedure.

Lemma 3.8. *For every inference rule $\rho \in (\text{SKS1}\setminus\{\text{qc}{\downarrow}, \text{qc}{\uparrow}\})\cup\{\text{c}{\uparrow}\}$ and every derivation ϕ of the form*

$$\rho\,\frac{\boxed{K\left\{\boxed{\text{qc}{\downarrow}\,\frac{\exists xA\vee\exists xA}{\exists xA}}\right\}}}{B}$$

where $K\{\}$ is a formula context, there exists a derivation of the form

$$\begin{array}{c} K\{\exists xA\vee\exists xA\} \\ \rho'\ \Big\Vert\ (\text{SKS1}\setminus\{\text{qc}{\downarrow},\text{qc}{\uparrow},\text{ai}{\uparrow}\})\cup\{\text{c}{\uparrow}\} \\ B' \\ \Big\Vert\ \{\text{qc}{\downarrow}\} \\ B \end{array}$$

such that $|\rho'| \le k|\phi|^2$ for some constant k.

Proof. The cases for most inference rules ρ are omitted. I present constructions for when ρ is a vacuous $=_\exists$ rule, which results in the largest derivations ρ', and for the rules r1↑ and c↑, which are responsible for most of the complexity during the first phase of the procedure:

$$\text{qc}{\downarrow}\,\frac{\exists xA\vee\exists xA}{=\,\frac{\exists xA}{A}} \quad\longrightarrow\quad \text{c}{\downarrow}\,\frac{\boxed{\boxed{=\,\frac{\exists xA}{A}}\vee\boxed{=\,\frac{\exists xA}{A}}}}{A}$$

where the instance of c↓ in the derivation above is replaced with a derivation in $\{\text{ac}{\downarrow}, \text{m}, \text{qc}{\downarrow}, \forall, =_{\mathbf{P}}, =_\forall\}$ using Proposition 3.5 and sequential composition. By Proposition 3.5, the resultant derivation is of size $O(|A|^2)$.

$$\text{r1}{\uparrow}\,\frac{\boxed{\boxed{\text{qc}{\downarrow}\,\frac{\exists xA(x)\vee\exists xA(x)}{\exists xA(x)}}\wedge C}}{\exists x(A(x)\wedge C)}$$

$$\downarrow$$

$$\text{qc}{\downarrow}\,\frac{\text{s}\,\frac{\text{s}\,\frac{\boxed{(\exists xA(x)\vee\exists xA(x))\wedge\boxed{\text{c}{\uparrow}\,\frac{C}{C\wedge C}}}}{(\exists xA(x)\vee(\exists xA(x)\wedge C))\wedge C}}{\boxed{\boxed{\text{r1}{\uparrow}\,\frac{\exists xA(x)\wedge C}{\exists x(A(x)\wedge C)}}\vee\boxed{\text{r1}{\uparrow}\,\frac{\exists xA(x)\wedge C}{\exists x(A(x)\wedge C)}}}}}{\exists x(A(x)\wedge C)}$$

$$\text{c}{\uparrow}\,\frac{\boxed{K\left\{\boxed{\text{qc}{\downarrow}\,\frac{\exists xA\vee\exists xA}{\exists xA}}\right\}}}{K\{\exists xA\}\wedge K\{\exists xA\}}$$

$$\downarrow$$

$$\text{c}{\uparrow}\,\frac{K\{\exists xA\vee\exists xA\}}{\boxed{K\left\{\boxed{\text{qc}{\downarrow}\,\frac{\exists xA\vee\exists xA}{\exists xA}}\right\}\wedge K\left\{\boxed{\text{qc}{\downarrow}\,\frac{\exists xA\vee\exists xA}{\exists xA}}\right\}}}$$

□

The following lemma provides reduction rules for permuting $\exists$ rules down through other inference rules during the second phase of the procedure. Prior to this phase, instances of r1↓ in the proof are replaced with equivalent instances of the falsifier rule ε to ensure that the permutations are possible. When an instance of $\exists$ is permuted down through an instance of ε, an ε-term may be introduced into the proof.

Lemma 3.9. *For every inference rule* $\rho \in (\mathsf{SKS1} \setminus \{\mathsf{qc}\downarrow, \mathsf{qc}\uparrow, \mathsf{r1}\downarrow, \exists\}) \cup \{\mathsf{c}\uparrow, \varepsilon\}$ *and every derivation* ϕ *of the form*

$$\rho\,\frac{\boxed{K\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\}}}{B}$$

where $K\{\}$ *is a formula context, there exists a derivation of the form*

$$\begin{array}{c} K\{A(t)\} \\ \rho' \,\Big\|\, \{\rho\} \\ B' \\ \Big\|\, \{\exists\} \\ B_\varepsilon \end{array}$$

such that $|\rho'| \leq |t||\phi|$, ρ' *contains at most one instance of* ρ *and* B_ε *is a formula obtained by replacing some* ε*-terms of the form* $\varepsilon_z(\overline{J\{\exists x A(x)\}[z/y]})$ *which occur in* B *with* $\varepsilon_z(\overline{J\{A(t)\}[z/y]})$ *for some formula context* $J\{\}$ *and variables* y, z.

Proof. The cases for most inference rules ρ are omitted. I present constructions for when ρ is $\forall$, which is responsible for most of the complexity during the second phase of the procedure, and for when ρ is ε, which introduces ε-terms into the proof:

$$\forall\,\frac{\boxed{\forall y J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\}}}{J\{\exists x A(x)\}[s/y]} \quad\longrightarrow\quad \forall\,\frac{\forall y J\{A(t)\}}{\boxed{J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\}[s/y]}}$$

$$\varepsilon\,\frac{\boxed{\forall y\left(C(y) \vee J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\}\right)}}{\forall y C(y) \vee J\{\exists x A\}[\varepsilon_z \overline{C}(z)/y]}$$
$$\downarrow$$
$$\varepsilon\,\frac{\forall y(C(y) \vee J\{A(t)\})}{\boxed{\forall y C(y) \vee J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\}[\varepsilon_z \overline{C}(z)/y]}}$$

$$\varepsilon\,\frac{\boxed{\forall y\left(J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\} \vee C(y)\right)}}{\forall y J\{\exists x A(x)\} \vee C(\varepsilon_z(\overline{J\{\exists x A(x)\}[z/y]}))}$$
$$\downarrow$$
$$\varepsilon\,\frac{\forall y(J\{A(t)\} \vee C(y))}{\boxed{\forall y J\left\{\boxed{\exists\,\dfrac{A(t)}{\exists x A(x)}}\right\} \vee C(\varepsilon_z(\overline{J\{A(t)\}[z/y]}))}}$$

□

The following lemma provides reduction rules for permuting universal cocontraction rules $\mathsf{qc}\uparrow$ up through most other inference rules during the third phase of the procedure.

Lemma 3.10. *For every inference rule* $\rho \in \mathsf{SKS1} \setminus \{\mathsf{qc}\downarrow, \mathsf{qc}\uparrow, \mathsf{r1}\downarrow, \mathsf{r1}\uparrow, \exists, =_\exists\}$ *and every derivation* ϕ *of the form*

$$\rho\,\frac{B}{\boxed{K\left\{\boxed{\mathsf{qc}\uparrow\,\dfrac{\forall x A}{\forall x A \wedge \forall x A}}\right\}}}$$

where $K\{\}$ *is a formula context and* B *is weakly existential-free, there exists a derivation of the form*

$$\begin{array}{c} B \\ \Big\|\, \{\mathsf{qc}\uparrow\} \\ B' \\ \rho' \,\Big\|\, \mathsf{SKS1} \setminus \{\mathsf{qc}\downarrow, \mathsf{qc}\uparrow, \mathsf{ai}\uparrow, \mathsf{r1}\downarrow, \mathsf{r1}\uparrow, \exists, =_\exists\} \\ K\{\forall x A \wedge \forall x A\} \end{array}$$

such that $|\rho'| \leq k|\phi|^2$ *for some constant* k.

Proof. The cases for most inference rules ρ are omitted. I present a construction for when ρ is a vacuous $=_\forall$ rule, which results in the largest derivations ρ':

$$\mathsf{qc}\uparrow\,\frac{=\,\dfrac{A}{\forall x A}}{\forall x A \wedge \forall x A} \quad\longrightarrow\quad \mathsf{c}\uparrow\,\frac{A}{\boxed{\boxed{=\,\dfrac{A}{\forall x A}} \wedge \boxed{=\,\dfrac{A}{\forall x A}}}}$$

where the instance of $\mathsf{c}\uparrow$ in the derivation above is replaced with a derivation in $\{\mathsf{ac}\uparrow, \mathsf{m}, \mathsf{qc}\uparrow, =_\forall\}$ using Proposition 3.5 and sequential composition. By Proposition 3.5, the resultant derivation is of size $O(|A|^2)$. □

During the third phase of the procedure, when universal cocontraction rules $\mathsf{qc}\uparrow$ are permuted up the proof, the greatest source of complexity and most troublesome case is when $\mathsf{qc}\uparrow$ rules are permuted up through falsifier rules ε. In this case, we use the construction given in the following lemma, which is invariant under the permutation.

Lemma 3.11. *For all variables* x *and* y *and all weakly existential-free formulae* $A(x)$ *and* $B(x)$, *let* $D(A(x), B(x), x, y, n)$ *denote the derivation*

$$\begin{array}{c} \boxed{\boxed{\varepsilon\,\dfrac{\forall x(A(x) \vee B(x))}{\forall x A(x) \vee B(\varepsilon_y \overline{A}(y))}} \wedge \cdots \wedge \boxed{\varepsilon\,\dfrac{\forall x(A(x) \vee B(x))}{\forall x A(x) \vee B(\varepsilon_y \overline{A}(y))}}} \\ \Big\|\, \{\mathsf{s}\} \\ \boxed{(\forall x A(x) \wedge \cdots \wedge \forall x A(x)) \vee \boxed{\begin{array}{c} B(\varepsilon_y \overline{A}(y)) \vee \cdots \vee B(\varepsilon_y \overline{A}(y)) \\ \Big\|\, \{\mathsf{c}\downarrow\} \\ B(\varepsilon_y \overline{A}(y)) \end{array}}} \end{array}$$

where the premise of the derivation is a conjunction of n *copies of the formula* $\forall x(A(x) \vee B(x))$.

For all derivations of the form

$$\begin{array}{c} \forall x(A(x) \vee B(x)) \wedge \cdots \wedge \forall x(A(x) \vee B(x)) \\ {\scriptstyle D(A(x),B(x),x,y,n)} \Big\| \\ \boxed{K\left\{\boxed{\mathsf{qc}\uparrow \dfrac{\forall zC}{\forall zC \wedge \forall zC}}\right\}} \end{array}$$

where $K\{\}$ *is a formula context, there exists a derivation of the form*

$$\begin{array}{c} \forall x(A(x) \vee B(x)) \wedge \cdots \wedge \forall x(A(x) \vee B(x)) \\ \Big\| {\scriptstyle \{\mathsf{qc}\uparrow\}} \\ \forall x(A'(x) \vee B'(x)) \wedge \cdots \wedge \forall x(A'(x) \vee B'(x)) \\ {\scriptstyle D(A'(x),B'(x),x,y,n')} \Big\| \\ (\forall x A'(x) \wedge \cdots \wedge \forall x A'(x)) \vee B'(\varepsilon_y \overline{A'}(y)) \\ \Big\| {\scriptstyle \{\mathsf{aw}\uparrow, \forall, =_P\}} \\ (\forall x A(x) \wedge \cdots \wedge \forall x A'(x) \wedge \cdots \wedge \forall x A(x)) \vee B'(\varepsilon_y \overline{A'}(y)) \end{array}$$

of size $O(n^2(|A(x)| + |B(x)|))$*, where* $A'(x)$*,* $B'(x)$ *and* n' *are given by one of the following*

(1) $A'(x)$ *is obtained by replacing a subformula* $\forall zC$ *of* $A(x)$ *with* $\forall zC \wedge \forall zC$*,* $B'(x)$ *is the formula* $B(x)$ *and* $n' = n$
(2) $A'(x)$ *is the formula* $A(x)$*,* $B'(x)$ *is obtained by replacing a subformula* $\forall zC$ *of* $B(x)$ *with* $\forall zC \wedge \forall zC$ *and* $n' = n$
(3) $A'(x)$ *is the formula* $A(x)$*,* $B'(x)$ *is the formula* $B(x)$ *and* $n' = n + 1$

Proof. Omitted. □

3.2 Consequences and Proof of Main Result

Before proving Theorem 3.3 the Falsifier Decomposition Theorem, I note some of its corollaries and provide an example.

It follows from the Falsifier Decomposition Theorem that the falsifier calculus SKSgε admits non-elementarily smaller cut-free proofs than **LK** for certain formulae and that there exist first-order theorems with non-elementarily smaller falsifier disjunctions than Herbrand disjunctions, as follows.

Corollary 3.12. *Every valid epsilon-free formula has a proof in* (SKSg$\varepsilon \setminus \{\mathsf{i}\uparrow\}) \cup \{\exists, \mathsf{qc}\downarrow\}$ *and there is no elementary function bounding the size of the smallest cut-free **LK** proof of a formula in terms of the (ε-)size of its smallest* (SKSg$\varepsilon \setminus \{\mathsf{i}\uparrow\}) \cup \{\exists, \mathsf{qc}\downarrow\}$ *proof.*

Proof. By Theorem 3.3 of [2], there is no elementary function bounding the size of the smallest cut-free **LK** proof of a formula in terms of the size of its smallest cut-free $\mathbf{LK}_{shift}$ proof, where $\mathbf{LK}_{shift}$ is the system presented in [2]. It is a standard exercise to show that SKSg1 $\setminus \{\mathsf{i}\uparrow\}$ polynomially simulates cut-free $\mathbf{LK}_{shift}$ so that there is no elementary function bounding the size of the smallest cut-free **LK** proof of a formula in terms of the size of its smallest SKSg1 $\setminus \{\mathsf{i}\uparrow\}$ proof. The result follows by Theorem 3.3. □

Corollary 3.13. *There is no elementary function bounding the size of the smallest Herbrand disjunction of a valid epsilon-free formula in terms of the (ε-)size of its smallest falsifier disjunction.*

Proof. For any Herbrand disjunction A' for a valid formula A, a cut-free **LK** proof of A exists of size $O(\exp |A'|)$, since exponentially-sized cut-free **LK** proofs exist for all propositional tautologies. Hence if such an elementary function did exist, it would contradict Corollary 3.12. □

Remark. The fact that there is no elementary bound on the size of the smallest Herbrand disjunction for a formula in terms of the ε-size of its smallest falsifier disjunction demonstrates that the ε-terms in falsifier disjunctions compress the complexity of Herbrand disjunctions, rather than simply rearranging their complexity.

Example 3.14. The following is a proof of the drinker's paradox in falsifier normal form.

$$= \dfrac{\mathbf{t}}{\exists \dfrac{\varepsilon \dfrac{\forall x \boxed{\mathsf{i}\downarrow \dfrac{\mathbf{t}}{P(x) \vee \overline{P}(x)}}}{\forall x P(x) \vee \overline{P}(\varepsilon_y \overline{P}(y))}}{\exists y(\forall x P(x) \vee \overline{P}(y))}}$$

In this example, the falsifier disjunction for the formula $\exists y(\forall x P(x) \vee \overline{P}(y))$ is $\forall x P(x) \vee \overline{P}(\varepsilon_y \overline{P}(y))$. The smallest Herbrand disjunction for the formula is $\forall x_1 \forall x_2 (P(x_1) \vee \overline{P}(c) \vee P(x_2) \vee \overline{P}(x_1))$, reflective of the compression seen in falsifier disjunctions over Herbrand disjunctions.

I now prove the Falsifier Decomposition Theorem.

Proof of Theorem 3.3. I present a procedure for transforming ϕ into the desired form. The procedure is separated into three phases. In Phase 1, we permute instances of the existential contraction rule qc↓ down to the bottom of the proof. In Phase 2, we permute instances of the existential instantiation rule ∃ down to the bottom of the proof, separating the proof into an upper segment of weakly existential-free formulae and a lower segment in $\{\exists, \mathsf{qc}\downarrow\}$. In Phase 3, we permute instances of the universal cocontraction rule qc↑ up the proof until they are eliminated.

To begin, we replace all instances of the rules i↓, i↑, w↓, w↑, c↓, r2↓, r2↑, r3↓, r3↑, r4↓, r4↑ and = rules of the forms $= \dfrac{\exists x \exists y A}{\exists y \exists x A}$, $= \dfrac{\forall x \forall y A}{\forall y \forall x A}$, $= \dfrac{A}{\exists x A}$ and $= \dfrac{\forall x A}{A}$ in the proof with derivations in SKS1 using Propositions 3.5 and 3.6 and sequential composition. Note that in order to avoid introducing unnecessary vacuous $=_\exists$ rule instances into the proof (since they duplicate instances of qc↓ during Phase 1), we do not decompose instances of c↑ in this way, but will do so at a later stage (see Phase 3 below). For ease of expression, we replace every instance of qc↑ introduced by this decomposition with an instance of c↑. By Propositions 3.5 and 3.6, the resultant proof ψ_0 is of size $O(|\phi|^2)$ and if ϕ is cut-free then ψ_0 is cut-free.

To ensure that instances of inference rules may be permuted around the proof without creating variable binding conflicts, we rename (α-convert) variables and quantifiers in the proof such that for all variables x, no $\exists x$ or $\forall x$ in the proof occurs in the scope of another $\exists x$ or $\forall x$ symbol. We then extend the proof with a derivation in $\{=_\exists, =_\forall\}$ using sequential composition to ensure that this renaming does not alter the conclusion of the proof.

Phase 1

We permute all instances of qc↓ down the proof using the rewriting system defined as follows. At each inductive step, we select a lowermost instance of qc↓ in the proof which occurs above some instance of a rule other than qc↓. We then permute this instance of qc↓ down through a rule instance ρ immediately beneath it in the proof in the following manner:

If ρ occurs inside the context of qc↓, we apply the following rewrite, replacing the subderivation in the proof using sequential composition:

$$\mathsf{qc}{\downarrow}\,\frac{\exists x K\{A\} \vee \exists x K\{A\}}{\boxed{\exists x K\left\{\boxed{\rho\,\frac{A}{B}}\right\}}}$$

$$\downarrow$$

$$\mathsf{qc}{\downarrow}\,\frac{\boxed{\boxed{\exists x K\left\{\boxed{\rho\,\frac{A}{B}}\right\}} \vee \boxed{\exists x K\left\{\boxed{\rho\,\frac{A}{B}}\right\}}}}{\exists x K\{B\}}$$

Otherwise, if ρ occurs outside the context of qc↓, in a subderivation of the form

$$\rho\,\frac{\boxed{K\left\{\boxed{\mathsf{qc}{\downarrow}\,\frac{\exists x A \vee \exists x A}{\exists x A}}\right\}}}{B}$$

for some formula context $K\{\}$, we replace the above subderivation in the proof with the derivation given by Lemma 3.8 using sequential composition.

The procedure terminates once every instance of qc↓ in the proof is above only other instances of qc↓. Termination is guaranteed since the height of the selected instance of qc↓ is reduced after each inductive step.

The resultant proof is of the form

$$\begin{array}{c} \psi_1 \Big\| (\mathsf{SKS1} \setminus \{\mathsf{qc}{\downarrow}, \mathsf{qc}{\uparrow}\}) \cup \{\mathsf{c}{\uparrow}\} \\ A_1' \\ \Big\| \{\mathsf{qc}{\downarrow}\} \\ A \end{array}$$

By Lemma 3.8, if ψ_0 is cut-free then the rewrites presented do not introduce any further instances of ai↑ into the proof. Therefore, if ϕ is cut-free then ψ_1 is cut-free.

Phase 2

In this phase, we permute all instances of the $\exists$ rule down the proof.

To begin, to ensure that $\exists$ rules can be permuted down the proof, we replace every instance of r1↓ in the proof with an instance of the falsifier rule ε, using the following transformation:

$$\mathsf{r1}{\downarrow}\,\frac{\forall x(A(x) \vee B)}{\forall x A(x) \vee B} \quad \to \quad \varepsilon\,\frac{\forall x(A(x) \vee B)}{\forall x A(x) \vee B}$$

For the convenience of further normalisation, we assume that the critical terms of all instances of the ε rule in the proof use distinct variables which are not used anywhere else in the proof.

We now permute instances of the $\exists$ rule down the proof using the rewriting system defined as follows. At each inductive step, we select a lowermost instance of $\exists$ in the proof which occurs above some instance of a rule other than $\exists$ or qc↓. We then permute this instance of $\exists$ down through a rule instance ρ immediately beneath it in the proof in the following manner:

If ρ occurs inside the context of $\exists$, we apply the following rewrite, replacing the subderivation in the proof using sequential composition:

$$\exists\,\frac{K\{A\}[t/x]}{\boxed{\exists x K\left\{\boxed{\rho\,\frac{A}{B}}\right\}}} \quad \to \quad \exists\,\frac{\boxed{K\left\{\boxed{\rho\,\frac{A}{B}}\right\}[t/x]}}{\exists x K\{B\}} \qquad (2)$$

Otherwise, if ρ occurs outside the context of $\exists$, in a subderivation of the form

$$\rho\,\frac{\boxed{K\left\{\boxed{\exists\,\frac{A(t)}{\exists x A(x)}}\right\}}}{B}$$

for some formula context $K\{\}$, we replace the above subderivation in the proof with the derivation given by Lemma 3.9 using sequential composition. To maintain correctness of the proof, if an ε-term $\varepsilon_z(\overline{J\{\exists x A(x)\}[z/y]})$ is locally renamed to $\varepsilon_z(\overline{J\{A(t)\}[z/y]})$ by this reduction, we replace every occurrence of $\varepsilon_z(\overline{J\{\exists x A(x)\}[z/y]})$ in the proof with $\varepsilon_z(\overline{J\{A(t)\}[z/y]})$. When performing this replacement, in the case of nested ε-terms, we replace innermost occurrences of the term $\varepsilon_z(\overline{J\{\exists x A(x)\}[z/y]})$ before outermost occurrences.

The procedure terminates once every instance of $\exists$ in the proof is above only other instances of $\exists$ and qc↓. Termination is guaranteed since the height of the selected instance of $\exists$ is reduced after each inductive step.

The resultant proof is of the form

$$\begin{array}{c} \psi_2 \Big\| (\mathsf{SKS1} \setminus \{\mathsf{qc}{\downarrow}, \mathsf{qc}{\uparrow}, \mathsf{ai}{\uparrow}, \mathsf{r1}{\downarrow}, \mathsf{r1}{\uparrow}, \exists, =_\exists\}) \cup \{\mathsf{c}{\uparrow}, \varepsilon\} \\ A_2' \\ \Big\| \{\exists, \mathsf{qc}{\downarrow}\} \\ A \end{array}$$

where every formula in ψ_2 is weakly existential-free.

By Lemma 3.9, if ψ_1 is cut-free then the rewrites presented do not introduce any further instances of ai↑ into the proof. Therefore, if ϕ is cut-free then ψ_2 is cut-free.

Phase 3

In this phase, we permute all instances of qc↑ up the proof until they are eliminated.

To begin, we replace all instances of c↑ in the proof with derivations in $\{\mathsf{ac}{\uparrow}, \mathsf{m}, \mathsf{qc}{\uparrow}, =_{\mathbf{P}}\}$ using Proposition 3.5 and sequential composition.

We now permute instances of qc↑ up the proof using the rewriting system defined as follows. At each inductive step, we permute an uppermost instance of qc↑ up through a rule instance ρ immediately above it in the proof in the following manner:

If ρ occurs inside the context of qc↑, we apply the following rewrite, replacing the subderivation in the proof using sequential composition:

$$\mathsf{qc}\!\uparrow\dfrac{\boxed{\forall x K\left\{\boxed{\rho\,\dfrac{B}{A}}\right\}}}{\forall x K\{A\}\wedge\forall x K\{A\}}$$

$$\downarrow$$

$$\mathsf{qc}\!\uparrow\dfrac{\forall x K\{B\}}{\boxed{\boxed{\forall x K\left\{\boxed{\rho\,\dfrac{B}{A}}\right\}}\wedge\boxed{\forall x K\left\{\boxed{\rho\,\dfrac{B}{A}}\right\}}}}$$

Otherwise, if the instance of qc↑ being permuted is immediately below a derivation of the form $D(A(x), B(x), x, y, n)$ as described in Lemma 3.11, we replace the whole subderivation with the appropriate of the three derivations described in Lemma 3.11, using sequential composition. To maintain correctness of the proof, if an ε-term is locally renamed by this reduction (case (1) of Lemma 3.11), we replace every occurrence of the ε-term in the proof, in the same manner as was described in Phase 2.

Otherwise, if ρ occurs outside the context of qc↑, in a subderivation of the form

$$\rho\,\dfrac{B}{\boxed{K\left\{\boxed{\mathsf{qc}\!\uparrow\dfrac{\forall x A}{\forall x A\wedge\forall x A}}\right\}}}$$

for some formula context $K\{\}$, we replace the above subderivation in the proof with the derivation given by Lemma 3.10 using sequential composition. Note that if ρ is ε, it is of the form $D(C(x), E(x), x, y, 1)$ as described in Lemma 3.11 and hence is handled by the reduction described in the previous paragraph.

The procedure terminates once the proof contains no instances of qc↑. Termination is guaranteed since the height of an uppermost instance of qc↑ is increased after each inductive step. Every universal quantifier in the proof must be introduced by a vacuous $=_\forall$ rule and when an instance of qc↑ is permuted up through such a rule it is eliminated, introducing one further instance of qc↑ for each universal quantifier in the premise of the $=_\forall$ instance.

After termination, we replace every instance of c↓ in the proof (resulting from the constructions of Lemma 3.11) with a derivation in $\{\mathsf{ac}\!\downarrow, \mathsf{m}, \forall, =_{\mathbf{P}}, =_\forall\}$ using Proposition 3.5 and sequential composition. We then replace all atomic rules in the proof with their standard variants in $\mathsf{SKSg}_{\boldsymbol{P}}$ (ac↓ instances are replaced with c↓ instances, etc.) to obtain a proof of the form

$$\begin{array}{c} \phi' \Big\Vert \mathsf{SKSg}\varepsilon \\ A' \\ \Big\Vert \{\exists, \mathsf{qc}\!\downarrow\} \\ A \end{array}$$

as required.

By Lemmas 3.10 and 3.11, if ψ_2 is cut-free then the rewrites presented do not introduce any further instances of ai↑ into the proof. Therefore, if ϕ is cut-free then ϕ' is cut-free, as required.

Complexity (Sketch)

I now assess the size and ε-size of the relevant formulae and derivations. The bounds computed are not intended to be optimal, but demonstrate elementary complexity with respect to $|\phi|$. I proceed by computing bounds for the size of the proof at the end of each of the three phases. As shown above, the proof ψ_0 prior to Phase 1 is of size $O(|\phi|^2)$.

Phase 1

Each rewrite for permuting an instance of qc↓ down through a rule instance immediately beneath it replaces a subderivation χ of the proof with a derivation of size at most $k|\chi|^2$ for some constant k, by the rewrites described in Phase 1 and Lemma 3.8. Therefore

$$|\psi_1| \leq k^{(2^N-1)}|\psi_0|^{(2^N)} \tag{3}$$

where N is the number of qc↓ instances permuted down the proof. As such, I compute an upper bound for the number of instances of qc↓ in the proof during Phase 1.

There are three sources which introduce further instances of qc↓ into the proof during Phase 1: (A) permuting instances of qc↓ down through instances of c↑ which occur in the proof prior to Phase 1, (B) permuting instances of qc↓ down through instances of vacuous $=_\exists$ rules, and (C) permuting instances of qc↓ down through the instances of c↑ introduced when permuting instances of qc↓ down through instances of r1↑. For a given instance of an inference rule in the proof, the potentially-greatest source of further instances of qc↓ is that of type (C), when the inference rule is r1↑, since an instance of c↑ may be introduced each time an instance of qc↓ is permuted down through an instance of r1↑. It therefore suffices to compute an upper bound for the number of qc↓ instances introduced by a single instance of r1↑ and then account for every instance of an inference rule in the proof introducing that many qc↓ instances.

For a given instance of r1↑ in the proof which has N instances of qc↓ permuted down through it during Phase 1, at most N instances of c↑ are introduced into the proof. When an instance of qc↓ is permuted down through an instance of c↑, a further instance of qc↓ is introduced. Hence if N instances of qc↓ occur above a given instance of r1↑ in the proof, at most N instances of c↑ are introduced when permuting the qc↓ instances down through it, resulting in at most N^2 instances of qc↓ after the duplication from permuting down through the introduced c↑ instances.

A proof which contains N instances of qc↓ and M instances of r1↑ prior to Phase 1 will therefore contain at most N^{2M} instances of qc↓ during Phase 1 due to this duplication. Since the proof prior to Phase 1 is of size $O(|\phi|^2)$, it contains $O(|\phi|^2)$ instances of qc↓ and $O(|\phi|^2)$ total inference rule instances, yielding an upper bound of $O((|\phi|^2)^{O(|\phi|^2)}) = O(\exp(O(|\phi|^2 \ln|\phi|)))$ instances of qc↓ in the proof during Phase 1. By (3) above, since $|\psi_0| = O(|\phi|^2)$, this yields the bound

$$|\psi_1| = \exp^3(O(|\phi|^2 \ln|\phi|)) \tag{4}$$

Phase 2

Each rewrite for permuting an instance of the ∃ rule which is witnessed by a term t down through a rule instance immediately beneath it replaces a subderivation χ of the proof with a derivation of size at most $|t||\chi|$, by the rewrites described in Phase 2 and Lemma 3.9, since the existential quantifier in the conclusion is replaced by t. Therefore, since there are at most $|\psi_1|$ existential quantifiers in ψ_1,

$$|\psi_2| \leq |t|^{|\psi_1|}|\psi_1| \quad (5)$$

where t is the largest term which witnesses an instance of ∃ during Phase 2. I therefore compute an upper bound for $|t|$.

The substitutions introduced by ∀ rules can increase the size of terms which witness ∃ rules. Whenever an instance of ∃ which is witnessed by a term s is permuted down through an instance of ∀ which is instantiated by a term r, an instance of ∃ which is witnessed by $s[r/x]$ for some variable x is introduced (see the first construction in the proof of Lemma 3.9). Furthermore, the terms which instantiate instances of ∀ rules may be altered in the same way when permuting an instance of ∃ down by the rewrite (2) presented in Phase 2, when ρ is ∀. Consider an instance of ∀ in the proof which is instantiated by r and has M instances of ∃ above it during Phase 2, witnessed by $s_1, \ldots, s_M$. After the instances of ∃ have been permuted down, it will be instantiated by a term of size at most $|r[s_1/x_1]\ldots[s_M/x_M]| \leq |r||s|^M$, where s is the largest of the terms s_i. Therefore each of the M instances of ∃ will be witnessed by terms of size at most $|r||s|^{M+1}$ after being permuted down through the ∀ rule. If M instances of ∃ are permuted down through L instances of ∀, the resultant ∃ instances are thus witnessed by terms of size at most $|r|^{O(M^{L-1})}|s|^{O(M^L)}$, where r is the largest term which instantiates one of the ∀ instances.

Since all terms which witness ∃ instances and instantiate ∀ instances in ψ_1 must occur in ϕ, we have $|s|, |r| \leq |\phi|$. Since the proof contains at most $|\psi_1|$ instances of ∀, we have $L \leq |\psi_1|$ and since it contains at most $|\psi_1|$ existential quantifiers, we have $M \leq |\psi_1|$. Therefore the largest term which witnesses an instance of ∃ during Phase 2 is of size at most $|\phi|^{O(|\psi_1|^{|\psi_1|})}$. By (4) and (5), this yields the bound

$$|\psi_2| = \exp^7(O(|\phi|^2 \ln |\phi|)) \quad (6)$$

Phase 3

The complexity increase from Phase 3 is analogous to that of Phase 1, since each rewrite for permuting an instance of qc↑ up through a rule instance immediately above it replaces a subderivation χ of the proof with a derivation of size at most $k|\chi|^2$ for some constant k and the constructions of Lemma 3.11 duplicate qc↑ instances in the same manner that r1↑ instances duplicate qc↓ instances during Phase 1. Therefore

$$|\phi'| = \exp^{10}(O(|\phi|^2 \ln |\phi|))$$

Finally, since A' is obtained by renaming ε-terms in A'_2 during Phase 3, by (6), $|A'| = |A'_2| \leq |\psi_2| = \exp^7(O(|\phi|^2 \ln |\phi|))$. The resultant proof also meets the conditions of Lemma 3.7, by the rewrites of Phases 2 and 3 which alter ε-terms, and so the required bounds for $|\phi'|_\varepsilon$ and $|A'|_\varepsilon$ follow. □

4 CONCLUSION

I introduced the falsifier calculus, a new proof system for first-order predicate logic in the language of Hilbert's epsilon-calculus which admits non-elementarily shorter cut-free proofs than traditional sequent-calculus systems. I further proved the Falsifier Decomposition Theorem, giving rise to the notion of falsifier disjunctions, analogues to Herbrand disjunctions which are non-elementarily shorter than Herbrand disjunctions for certain first-order theorems.

Unlike Herbrand's Theorem or Gentzen's sharpened Hauptsatz (midsequent theorem) [14], the Falsifier Decomposition Theorem does not fully separate the propositional and first-order parts of a proof, but is a new decomposition that provides a novel insight into the normalisation theory of first-order proofs. It is also expected that the Falsifier Decomposition Theorem will be useful in establishing further normalisation results for first-order proofs within the deep-inference methodology, including extending the *experiments method* [32], a deep-inference cut-elimination procedure for propositional classical logic, to first-order predicate logic and formalising the connection between falsifier disjunctions and Herbrand disjunctions to give an independent proof of Herbrand's Theorem. I note that the ε-terms contained in falsifier disjunctions resemble the case distinctions used to derive Herbrand disjunctions in Shoenfield's variant of Gödel's functional interpretation [15, 34] (see also [1]) and Kreisel's no-counterexample interpretation [23, 24].

Central to the proof theory of the traditional critical-axiom-based epsilon-calculus are the *first epsilon theorem* and *second epsilon theorem* (see [30]), which establish conservativity of the epsilon-calculus over propositional classical logic and first-order predicate logic, respectively. The *extended first epsilon theorem* [6, 30] further establishes that for any quantifier-free, epsilon-free formula $A(x_1, \ldots, x_n)$ and ε-terms $\varepsilon_{x_1}B_1, \ldots, \varepsilon_{x_n}B_n$, if $A(\varepsilon_{x_1}B_1, \ldots, \varepsilon_{x_n}B_n)$ is provable in the epsilon-calculus, then there exist epsilon-free terms t^i_j such that $\bigvee_{i=1}^{i=m} A(t^i_1, \ldots, t^i_n)$ is a propositional tautology. This may be used to prove Herbrand's Theorem for existential theorems, by encoding existential quantifiers with ε-terms. It is not yet known whether an analogous result holds for the falsifier calculus, but this provides an interesting avenue for further research, especially since the Falsifier Decomposition Theorem is proved for general first-order theorems.

ACKNOWLEDGMENTS

This work would not have been possible without Alessio Guglielmi, who was watching over me while this research was taking place and whose loss has been felt deeply across the MathFound group in Bath. I would also like to thank my supervisor Willem Heijltjes for his invaluable guidance and support throughout the writing of this paper.

This work was funded by the United Kingdom Research & Innovation (UKRI) under Engineering and Physical Sciences Research Council (EPSRC) grant EP/T518013/1, studentship 2601979. For the purpose of Open Access the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] Bahareh Afshari, Stefan Hetzl, and Graham E. Leigh. 2020. Herbrand's theorem as higher order recursion. *Annals of Pure and Applied Logic* 171, 6 (2020), 102792. https://doi.org/10.1016/j.apal.2020.102792

[2] Juan P. Aguilera and Matthias Baaz. 2019. Unsound Inferences Make Proofs Shorter. *The Journal of Symbolic Logic* 84, 1 (2019), 102–122. https://doi.org/10.1017/jsl.2018.51

[3] Aurore Alcolei, Pierre Clairambault, Martin Hyland, and Glynn Winskel. 2018. The true concurrency of Herbrand's theorem. In *27th EACSL Annual Conference on Computer Science Logic (CSL 2018)*, Dan R. Ghica and Achim Jung (Eds.). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, GBR, 5:1–5:22. https://doi.org/10.4230/LIPIcs.CSL.2018.5

[4] Toshiyasu Arai. 2003. Epsilon Substitution Method for Id1. *Annals of Pure and Applied Logic* 121, 2-3 (2003), 163–208. https://doi.org/10.1016/s0168-0072(02)00112-4

[5] Matthias Baaz and Alexander Leitsch. 1994. On Skolemization And Proof Complexity. *Fundam. Inf.* 20, 4 (dec 1994), 353–379.

[6] Matthias Baaz, Alexander Leitsch, and Anela Lolic. 2018. A Sequent-Calculus Based Formulation of the Extended First Epsilon Theorem. In *Logical Foundations of Computer Science*, Sergei Artemov and Anil Nerode (Eds.). Springer International Publishing, Cham, 55–71.

[7] Matthias Baaz and Anela Lolic. 2024. Epsilon Calculus Provides Shorter Cut-Free Proofs. arXiv:2401.09183 [math.LO]

[8] Kai Brünnler. 2004. *Deep Inference and Symmetry in Classical Proofs.* Logos Verlag, Berlin. https://people.bath.ac.uk/ag248/kai/phd.pdf

[9] Kai Brünnler. 2006. Cut Elimination inside a Deep Inference System for Classical Predicate Logic. *Studia Logica* 82, 1 (2006), 51–71. https://doi.org/10.1007/s11225-006-6605-4

[10] Kai Brünnler and Alessio Guglielmi. 2004. A First Order System with Finite Choice of Premises. In *First-Order Logic Revisited*, Vincent Hendricks, Fabian Neuhaus, Stig Andur Pedersen, Uwe Scheffler, and Heinrich Wansing (Eds.). Logos Verlag, 59–74.

[11] Samuel R. Buss. 2000. An Introduction to Proof Theory. *Bulletin of Symbolic Logic* 6, 4 (2000), 464–465. https://doi.org/10.2307/420968

[12] Thierry Coquand. 1995. A semantics of evidence for classical arithmetic. *Journal of Symbolic Logic* 60, 1 (1995), 325–337. https://doi.org/10.2307/2275524

[13] Gottlob Frege. 1967. Begriffsschrift. In *From Frege to Gödel*, Jean Van Heijenoort (Ed.). Harvard University Press, 1–83.

[14] Gerhard Gentzen. 1964. Investigations into Logical Deduction. *American Philosophical Quarterly* 1, 4 (1964), 288–306. http://www.jstor.org/stable/20009142

[15] Philipp Gerhardy and Ulrich Kohlenbach. 2005. Extracting Herbrand disjunctions by functional interpretation. *Arch. Math. Log.* 44, 5 (2005), 633–644. http://dblp.uni-trier.de/db/journals/aml/aml44.html#GerhardyK05

[16] Alessio Guglielmi. 2007. A System of Interaction and Structure. *ACM Transactions on Computational Logic* 8, 1 (2007), 1:1–64. https://doi.org/10.1145/1182613.1182614

[17] Alessio Guglielmi, Tom Gundersen, and Michel Parigot. 2010. A Proof Calculus Which Reduces Syntactic Bureaucracy. In *21st International Conference on Rewriting Techniques and Applications (RTA) (Leibniz International Proceedings in Informatics (LIPIcs), Vol. 6)*, Christopher Lynch (Ed.). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 135–150. https://doi.org/10.4230/LIPIcs.RTA.2010.135

[18] Willem Heijltjes. 2010. Classical proof forestry. *Annals of Pure and Applied Logic* 161, 11 (2010), 1346–1366. https://doi.org/10.1016/j.apal.2010.04.006 Special Issue: Classical Logic and Computation (2008).

[19] Jacques Herbrand. 1930. *Recherches sur la théorie de la démonstration.* Ph.D. Dissertation. University of Paris.

[20] David Hilbert. 1927. The Foundations of Mathematics. (1927).

[21] Dominic J. D. Hughes. 2019. First-order proofs without syntax. arXiv:1906.11236 [math.LO]

[22] Dominic J. D. Hughes, Lutz Straßburger, and Jui-Hsuan Wu. 2021. Combinatorial Proofs and Decomposition Theorems for First-order Logic. *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)* (2021), 1–13.

[23] Ulrich Kohlenbach. 1999. On the No-Counterexample Interpretation. *The Journal of Symbolic Logic* 64, 4 (1999), 1491–1511. http://www.jstor.org/stable/2586791

[24] G. Kreisel. 1951. On the Interpretation of Non-Finitist Proofs–Part I. *The Journal of Symbolic Logic* 16, 4 (1951), 241–267. http://www.jstor.org/stable/2267908

[25] Richard McKinley. 2010. A sequent calculus demonstration of Herbrand's theorem. arXiv:1007.3414 [math.LO]

[26] Richard McKinley. 2013. Proof Nets for Herbrand's Theorem. *ACM Trans. Comput. Logic* 14, 1, Article 5 (feb 2013). https://doi.org/10.1145/2422085.2422090

[27] Dale A. Miller. 1987. A Compact Representation of Proofs. *Studia Logica: An International Journal for Symbolic Logic* 46, 4 (1987), 347–370. http://www.jstor.org/stable/20015330

[28] Grigori Mints. 2013. Epsilon substitution for first- and second-order predicate logic. *Annals of Pure and Applied Logic* 164, 6 (2013), 733–739. https://doi.org/10.1016/j.apal.2012.05.004 Classical Logic and Computation 2010(CLAC 2010).

[29] Kenji Miyamoto and Georg Moser. 2023. Herbrand complexity and the epsilon calculus with equality. *Archive for Mathematical Logic* (2023). https://doi.org/10.1007/s00153-023-00877-3

[30] Georg Moser and Richard Zach. 2006. The Epsilon Calculus and Herbrand Complexity. *Studia Logica* 82, 1 (2006), 133–155. https://doi.org/10.1007/s11225-006-6610-7

[31] B. Ralph. 2016. Decomposing First-Order Proofs using Deep Inference. *Proof, Computation, Complexity 2016* (2016).

[32] Benjamin Ralph. 2019. *Modular Normalisation of Classical Proofs.* Ph.D. Dissertation. University of Bath. https://people.bath.ac.uk/ag248/br/phd.pdf

[33] Benjamin Ralph. 2020. Herbrand Proofs and Expansion Proofs as Decomposed Proofs. *Journal of Logic and Computation* 30, 8 (10 2020), 1711–1742. https://doi.org/10.1093/logcom/exaa052 arXiv:https://academic.oup.com/logcom/article-pdf/30/8/1711/34673322/exaa052.pdf

[34] Joseph R. Shoenfield. 1967. *Mathematical Logic.* A K Peters/CRC Press.

[35] R. Statman. 1979. Lower Bounds on Herbrand's Theorem. *Proc. Amer. Math. Soc.* 75, 1 (1979), 104–107. http://www.jstor.org/stable/2042682

[36] Richard Zach. 2017. Semantics and Proof Theory of the Epsilon Calculus. In *Logic and Its Applications*, Sujata Ghosh and Sanjiva Prasad (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 27–47.