G. Moser
R. Zach

The Epsilon Calculus and Herbrand Complexity

Abstract. Hilbert's ε-calculus is based on an extension of the language of predicate logic by a term-forming operator ε_x. Two fundamental results about the ε-calculus, the first and second epsilon theorem, play a rôle similar to that which the cut-elimination theorem plays in sequent calculus. In particular, Herbrand's Theorem is a consequence of the epsilon theorems. The paper investigates the epsilon theorems and the complexity of the elimination procedure underlying their proof, as well as the length of Herbrand disjunctions of existential theorems obtained by this elimination procedure.

Keywords: Hilbert's ε-calculus, epsilon theorems, Herbrand's theorem, proof complexity

1. Introduction

Hilbert's ε-calculus [14, 16] is based on an extension of the language of predicate logic by a term-forming operator ε_x. This operator is governed by the *critical axiom*

$$A(t) \rightarrow A(\varepsilon_x A(x)) \ ,$$

where t is an arbitrary term. Within the ε-calculus, quantifiers become definable by $\exists x A(x) \Leftrightarrow A(\varepsilon_x A(x))$ and $\forall x A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$.

The ε-calculus was originally developed in the context of Hilbert's program of consistency proofs. Early work in proof theory (before Gentzen) concentrated on the ε-calculus and the ε-substitution method and was carried out by Ackermann [1, 2] (see also [19]), von Neumann [24], and Bernays [14] (see also [25, 26]). The ε-calculus is of independent and lasting interest, however, and a study from a computational and proof-theoretic point of view is particularly worthwhile.

The aim of this paper is to present the central notions of and basic results about the ε-calculus in a streamlined form and with attention to questions of proof complexity. This, we hope, will make the ε-calculus more easily accessible to a broader audience, and will make clearer the merits and disadvantages of the ε-calculus as a formalization of first-order logic.

One simple example of the merits of the ε-calculus is that by encoding quantifiers on the term-level, formalizing (informal) proofs is sometimes easier in the ε-calculus, compared to formalizing proofs in, e.g., sequent calculi. This is possible as the ε-calculus allows more condensed representation of proofs than standard sequent- or natural deduction calculus. However, it should be pointed out that the encoding of quantifiers on the term-level can

forthcoming in *Studia Logica* (2006).

come at a significant cost, as the transformation of quantified formulas may result in rather complicated term-structures.

Hilbert's ε-calculus is primarily a classical formalism, and we will restrict our attention to classical first-order logic. (But see the work of Bell [7, 8], DeVidi [11], Fitting [13], Mostowski [20] for a complementary view.) Our study is also motivated by the recent renewed interest in the ε-calculus and the ε-substitution method in, e.g., the work of Arai [3, 4], Avigad [5], and Mints et al., [17, 18]. The ε-calculus also allows the incorporation of choice construction into logic [9].

Our main focus will be the presentation and analysis of two *conservativity results* for the ε-calculus: the first and second ε-theorems. Our proofs of these results are essentially those due to Bernays [14], but we are also concerned with bounds on the length of proofs and the length of Herbrand disjunctions. Let T denote a finitely axiomatized open theory with axioms $P_1, \ldots, P_t$ containing no quantifiers or ε-terms, and let PC_ε be a usual formulation of the predicate calculus extended by the ε-operator and its characteristic axiom. Then the first ε-theorem states that any formula without quantifiers or ε-terms provable in PC_ε from T is already provable from T in the quantifier- and ε-free fragment EC of PC (the so-called elementary calculus of free variables). The second ε-theorem says that any formula without ε-terms provable in PC_ε from T is also provable from T in PC, the pure predicate calculus.

We prove the first ε-theorem in Section 5. An important extension of the first ε-theorem is the so-called extended first ε-theorem, which yields one direction of Herbrand's theorem: if A is an existential formula provable in PC_ε, then there is a disjunction of instances of the matrix of A provable in EC. Call the minimal length of such a disjunction the *Herbrand complexity* $\mathrm{HC}(A)$ of A. We obtain an upper bound on $\mathrm{HC}(A)$ by analyzing the complexity of the procedure used to eliminate critical formulas from the proof of A; this bound is hyperexponential in the number of quantifier axioms in the derivation of A. In Section 6 we show that there is also a hyperexponential lower bound on $\mathrm{HC}(A)$, and thus the procedure for generating the Herbrand disjunction based on Bernays's proof of the extended first ε-theorem is essentially optimal. In Section 7, we prove the second ε-theorem. Its proof also contains a proof of the second direction of Herbrand's theorem, i.e., that given a (valid) Herbrand disjunction of A, A is provable.

In this paper we only consider the ε-calculus without equality. The case of the ε-theorems with equality is much more involved than the proofs we consider here, and cannot be dealt with adequately in the space available. It raises important and interesting issues for the suitability of formalisms based on the epsilon calculus for automated theorem proving. The complexity of the epsilon theorems in the presence of equality will be the subject of future work.

2. The Epsilon Calculus: Syntax

The syntax of the epsilon calculus is essentially that of a standard first-order language. We will, however, frequently pass back and forth between different calculi formulated in slightly different languages. Let us first dispense with some pedantry: The language $L(\mathrm{EC})$ of the elementary calculus consists of: (1) free variables: a, b, c, …, (2) bound variables: x, y, z, …, (3) constant and function symbols: f, g, h, … with arities $\mathrm{ar}(f)$, …, (4) predicate symbols: P, Q, R, … with arities $\mathrm{ar}(P)$, …, and the propositional connectives: $\wedge$, $\vee$, $\to$, $\neg$. The language $L(\mathrm{PC})$ of the predicate calculus is $L(\mathrm{EC})$ plus the quantifiers $\forall$, $\exists$. The language $L(\mathrm{EC}_\varepsilon)$ of the pure epsilon calculus is $L(\mathrm{EC})$ plus the epsilon operator ε. The language $L(\mathrm{PC}_\varepsilon)$ of the predicate calculus with epsilon is $L(\mathrm{PC})$ together with ε.

We will distinguish between terms and semi-terms, and between formulas and semi-formulas. The definitions are:

Definition 1. *(Semi)terms*, *(semi)formulas*, and *sub-(semi)terms* are defined as follows:

1. Any free variable a is a (semi)term. Its only sub-(semi)term is a itself. It has no immediate sub-(semi)terms.

2. Any bound variable x is a semi-term. It has no sub-terms or immediate sub-(semi)terms. Its only sub-semiterm is x itself.

3. If f is a function symbol with $\mathrm{ar}(f) = 0$, then f is a (semi)term. Its only sub-(semi)term is f itself. It has no immediate sub-(semi)terms.

4. If f is a function symbol with $\mathrm{ar}(f) = n > 0$, and t_1, …, t_n are (semi)terms, then $f(t_1, \dots, t_n)$ is a (semi)term. Its immediate sub-semiterms are t_1, …, t_n, and its immediate sub-terms are those among t_1, …, t_n which are terms, plus the immediate subterms of those among $t_1, \dots, t_n$ which are not terms. Its sub-semiterms are $f(t_1, \dots, t_n)$ and the sub-semiterms of t_1, …, t_n; its subterms are those of its sub-semiterms which are terms.

5. If P is a predicate symbol with $\mathrm{ar}(P) = n > 0$, and t_1, …, t_n are (semi)terms, then $P(t_1, \dots, t_n)$ is an (atomic) (semi)formula. Its immediate sub-semiterms are t_1, …, t_n. Its immediate subterms are

those among $t_1, \ldots, t_n$ which are terms, plus the immediate subterms of those among $t_1, \ldots, t_n$ which are not terms. Its sub-(semi)terms are the sub-(semi)terms of $t_1, \ldots, t_n$.

6. If A and B are (semi)formulas, then $\neg A$, $A \land B$, $A \lor B$ and $A \to B$ are (semi)formulas. Its (immediate) sub-(semi)terms are those of A and B.

7. If $A(a)$ is a (semi)formula containing the free variable a and x is a bound variable not occurring in $A(a)$, then $\forall x\, A(x)$ and $\exists x\, A(x)$ are (semi)formulas. Its (immediate) sub-(semi)terms are those of $A(x)$.

8. If $A(a)$ is a (semi)formula containing the free variable a and x is a bound variable not occurring in $A(a)$, then $\varepsilon_x A(x)$ is a (semi)term. Its sub-(semi)terms are $\varepsilon_x A(x)$ and the sub-(semi)terms of $A(x)$. Its immediate sub-(semi)terms are those of $A(x)$.

Note that terms and formulas are just semiterm and semiformula which contain no bound variables without a matching $\forall$, $\exists$, or ε. We will call a semi-term of the form $\varepsilon_x A(x)$ an *ε-expression*, or, if it is a term, an *ε-term*.

The definition above does not allow a quantifier or epsilon binding a variable x to occur in the scope of another quantifier or epsilon which binds the same variable. In order for some substitutions of terms to result in well-formed formulas, it will often be necessary to rename bound variables. Conversely, we cannot assume that all epsilon terms of the same form occurring in a proof are syntactically identical. Hence we will (a) adopt the convention that bound variables must be renamed when substituting terms so as to avoid clashes of bound variables, and (b) we tacitly identify epsilon terms which differ only by a renaming of bound variables. Thus, if $A(a)$ is a formula with free variable a, then $A(\varepsilon_x A(x))$ is a formula obtained from $A(a)$ by replacing every indicated occurrence of a by an epsilon term resulting from $\varepsilon_x A(x)$ by renaming each bound variable y occurring in it so as to ensure that y does not appear in the scope of a quantifier or epsilon in $A(a)$. For instance, if $A(a) \equiv \exists y\, P(a, y)$, then $\varepsilon_x A(x) \equiv \varepsilon_x \exists y\, P(x, y)$ contains the bound variable y, and a literal substitution for a in $A(a)$ would result in $\exists y\, P(\varepsilon_x \exists y\, P(x, y), y)$, which is not well formed. The variable y must be renamed, e.g., thus: $\exists y\, P(\varepsilon_x \exists z\, P(x, z), y)$.

An instruction to replace every occurrence of an epsilon-term e in a formula or proof by some other term is to be understood as an instruction to replace every ε-term equal to e up to renaming of bound variables; e.g., "replace $\varepsilon_x \exists y\, P(x, y)$ in the above formula by t" results in $\exists y\, P(t, y)$. It will sometimes be useful to be explicit about substitutions of terms for variables. We will write $A(a)\,\{a \leftarrow t\}$ to denote the result of replacing the indicated occurrences of a in $A(a)$ by t.

3. Axiomatization of the Epsilon Calculus

Definition 2. The set of axioms $\mathrm{AxEC}(L)$ of axioms of the *elementary calculus* for a language L consists of all propositional tautologies in the language L. To obtain the set of axioms $\mathrm{AxEC}_\varepsilon$ of the *pure epsilon calculus* we add to $\mathrm{AxEC}(L(\mathrm{EC}_\varepsilon))$ all substitution instances of

$$A(t) \to A(\varepsilon_x A(x)) \; . \tag{1}$$

An axiom of the form (1) is called a *critical formula.* We say that the critical formula *belongs* to the ε-term $\varepsilon_x A(x)$.

The set AxPC of axioms of the *predicate calculus*, and the set $\mathrm{AxPC}_\varepsilon$ of axioms of the *extended predicate calculus* consist of $\mathrm{AxEC}(L(\mathrm{PC}))$ and $\mathrm{AxEC}_\varepsilon$, respectively, together with all instances of $A(t) \to \exists x\, A(x)$ and $\forall x\, A(x) \to A(t)$ in the respective language.

Definition 3. A *proof* in EC (EC_ε) is a sequence $A_1, \dots, A_n$ of formulas such that each A_i is either in AxEC ($\mathrm{AxEC}_\varepsilon$) or it follows from formulas preceding it by modus ponens, i.e., there are j, $k < i$ so that $A_k \equiv A_j \to A_i$.

A proof in PC (PC_ε) is a sequence $A_1, \dots, A_n$ of formulas such that each A_i is either in AxPC ($\mathrm{AxPC}_\varepsilon$) or follows from formulas preceding it by modus ponens, or follows from a preceding formula by generalization, i.e., there is a $j < i$ so that either $A_j \equiv B \to C(a)$ and $A_i \equiv B \to \forall x\, C(x)$ or $A_j \equiv B(a) \to C$ and $A_i \equiv \exists x\, B(x) \to C$. In the latter case we also require that the free variable a does not occur in A_i or in any A_k with $k > i$. Such a variable a is called an *eigenvariable.* The restriction guarantees that each variable is only used as an eigenvariable in a generalization inference at most once.

A formula A is called *provable* in EC (EC_ε, PC, PC_ε) if there is a proof in EC (EC_ε, PC, PC_ε, respectively) which has A as its last formula. To indicate that A is provable in, say, EC_ε by a proof π, we write $\mathrm{EC}_\varepsilon \vdash_\pi A$.

Definition 4. The *size* $\mathrm{sz}(\pi)$ of a proof π is the number of steps in π. If π is a proof in EC_ε or PC_ε, we define the *critical count* $\mathrm{cc}(\pi)$ of π as the number of distinct critical formulas and quantifier axioms in π plus 1.

4. The Embedding Lemma

The epsilon operator allows the treatment of quantifiers in a quantifier-free system: using ε-terms, it is possible to define $\exists x$ and $\forall x$ as follows:

$$\begin{array}{rcl} \exists x\, A(x) & \Leftrightarrow & A(\varepsilon_x A(x)) \\ \forall x\, A(x) & \Leftrightarrow & A(\varepsilon_x \neg A(x)) \end{array}$$

We define a mapping ${}^{\varepsilon}$ of semiformulas and semiterms in $L(\mathrm{PC}_\varepsilon)$ to semiformulas and semiterms in $L(\mathrm{EC}_\varepsilon)$ as follows:

$$f(t_1,\ldots,t_n)^\varepsilon = f(t_1^\varepsilon,\ldots,t_n^\varepsilon) \qquad P(t_1,\ldots,t_n)^\varepsilon = P(t_1^\varepsilon,\ldots,t_n^\varepsilon)$$

$$\begin{array}{rclrclrcl} x^\varepsilon & = & x & (A \to B)^\varepsilon & = & A^\varepsilon \to B^\varepsilon & [\varepsilon_x A(x)]^\varepsilon & = & \varepsilon_x A(x)^\varepsilon \\ a^\varepsilon & = & a & (A \lor B)^\varepsilon & = & A^\varepsilon \lor B^\varepsilon & (\exists x\, A(x))^\varepsilon & = & A^\varepsilon(\varepsilon_x A(x)^\varepsilon) \\ (\neg A)^\varepsilon & = & \neg A^\varepsilon & (A \land B)^\varepsilon & = & A^\varepsilon \land B^\varepsilon & (\forall x\, A(x))^\varepsilon & = & A^\varepsilon(\varepsilon_x \neg A(x)^\varepsilon) \end{array}$$

Example 5. Consider

$$\begin{aligned} \exists x(P(x) \quad \lor \quad & \forall y\, Q(y))^\varepsilon = \\ = \quad & [P(x) \lor \forall y\, Q(y)]^\varepsilon \quad \{x \leftarrow \varepsilon_x[P(x) \lor \forall y\, Q(y)]^\varepsilon\} \\ & \quad [P(x) \lor \forall y\, Q(y)]^\varepsilon = P(x) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \\ = \quad & P(x) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \quad \{x \leftarrow \underbrace{\varepsilon_x[P(x) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}\} \\ = \quad & P(\underbrace{\varepsilon_x[P(x) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \end{aligned}$$

Example 6. Consider

$$\begin{aligned} [\exists x \quad \exists y \quad & A(x,y)]^\varepsilon = \\ = \quad & [\exists y\, A(x,y)]^\varepsilon \quad \{x \leftarrow \varepsilon_x[\exists y\, A(x,y)]^\varepsilon\} \\ & \quad [\exists y\, A(x,y)]^\varepsilon = A(x, \underbrace{\varepsilon_y A(x,y)}_{e'(x)}) \\ = \quad & A(x, \underbrace{\varepsilon_y A(x,y)}_{e'(x)})\{x \leftarrow \underbrace{\varepsilon_x A(x, \varepsilon_z A(x,z))}_{e_3}\} \\ = \quad & A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x,z))}_{e_3}, \underbrace{\varepsilon_y A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x,z))}_{e_3}, y)}_{e_4 = e'(e_3)}) \end{aligned}$$

Lemma 7 (Embedding Lemma). *If π is a* PC_ε*-proof of A then there is an* EC_ε*-proof π^ε of A^ε with* $\mathrm{sz}(\pi^\varepsilon) \leq 3 \cdot \mathrm{sz}(\pi)$ *and* $\mathrm{cc}(\pi^\varepsilon) \leq \mathrm{cc}(\pi)$*.*

Proof. We show that for all proofs π consisting of formulas $A_1, \ldots, A_n$, there is a proof π^ε containing $A_1^\varepsilon, \ldots, A_n^\varepsilon$ (plus perhaps some extra formulas) of the required size and critical count. We proceed by induction on n. The case $n = 0$ is trivial. Suppose the claim holds for the proof consisting of A_1, $\ldots$, A_n, i.e., there is a proof π^* containing A_1^ε, $\ldots$, A_n^ε, and consider the proof $\pi = A_1, \ldots, A_n, A$. If A is a propositional tautology, then A^ε is also a propositional tautology. (Note that $(\cdot)^\varepsilon$ leaves the propositional structure of A intact.) If A is a critical formula, then A^ε is also a critical formula. In both cases, we can take π^ε to be π^* extended by A_n^ε.

If A is an instance of a quantifier axiom, its translation A^ε either is of the form

$$[A(t) \to \exists x\, A(x)]^\varepsilon \equiv A^\varepsilon(t^\varepsilon) \to A^\varepsilon(\varepsilon_x A(x)^\varepsilon)\ ,$$

which is a critical formula, or is of the form

$$[\forall x\, A(x) \to A(t)]^\varepsilon \equiv A^\varepsilon(\varepsilon_x \neg A(x)) \to A^\varepsilon(t^\varepsilon)\ ,$$

which is the contrapositive of, and hence provable from, a critical formula. In the latter case, the size of the resulting proof increases by two additional steps.

Now suppose A follows by modus ponens from A_i and $A_j \equiv A_i \to A$. Since π^* contains A_i^ε and $A_j^\varepsilon \equiv A_i^\varepsilon \to A^\varepsilon$, adding A^ε to π^* is also a proof.

If A follows by generalization, i.e., $A \equiv B \to \forall x\, C(x)$ and $A_i \equiv B \to C(a)$ (where a satisfies the conditions on eigenvariables), then by induction hypothesis the proof π^* contains $A_i^\varepsilon \equiv B^\varepsilon \to C(a)^\varepsilon$. Replacing a everywhere in π^* by $\varepsilon_x \neg A(x)$ results in a proof containing

$$[B \to \forall x\, C(x)]^\varepsilon \equiv B^\varepsilon \to A^\varepsilon(\varepsilon_x \neg A(x)^\varepsilon)\ .$$

in place of $A_i^m eps$. Similarly, if the last inference derives $A \equiv \exists x\, B(x) \to C$ from $B(a) \to C$, by induction hypothesis π^* contains $B(a)^\varepsilon \to C^\varepsilon$, and we obtain a proof of A^ε by replacing a everywhere in π^* by $\varepsilon_x B(x)$. ■

5. The First Epsilon Theorem

We begin our discussion of the ε-theorems by a detailed proof of the first ε-theorem. This theorem states that if a formula E without quantifiers or epsilon is provable in the (extended) ε-calculus, then it is already provable in the elementary calculus. In other words, the (extended) epsilon calculus is conservative over the elementary calculus for elementary formulas. A relatively simple corollary of the first epsilon theorem is the extended first ε-theorem, which is a version of Herbrand's theorem for prenex formulas. This section is dedicated to proofs of these results. The argument we use is essentially Bernays's [14], which gives a procedure by which critical formulas in proofs of E are eliminated step-wise. We analyze the procedure and thereby obtain upper bounds on the complexity of the Herbrand disjunction obtained in the extended first ε-theorem. These bounds are given in terms of the *hyperexponential function* 2^x_y, defined by $2^x_0 = x$ and $2^x_{i+1} = 2^{2^x_i}$.

The proof will proceed by induction on the rank and degree and number of ε-terms in critical formulas in the proof of E. Rank and degree are two measures of complexity of ε-expressions: Degree applies to ε-terms only, and measures the depth of nesting of ε-terms. Rank, on the other hand, measures the complexity of cross-binding of ε-expressions.

Definition 8. The *degree* of an ε-term is inductively defined as follows:

1. If $A(x)$ contains no ε-subterms, then $\deg(\varepsilon_x A(x)) = 1$.
2. If $e_1, \ldots, e_n$ are all immediate ε-subterms of $A(x)$, then
$$\deg(\varepsilon_x A(x)) = \max\{\deg(e_1), \ldots, \deg(e_n)\} + 1 .$$

Definition 9. An ε-expression e is *subordinate* to $\varepsilon_x A(x)$ if e is a proper sub-semiterm of $A(x)$ and contains x.

Definition 10. The *rank* of an ε-expression e is defined as follows:

1. If e contains no subordinate ε-expressions, then $\mathrm{rk}(e) = 1$.
2. If $e_1, \ldots, e_n$ are all the ε-expressions subordinate to e, then
$$\mathrm{rk}(e) = \max\{\mathrm{rk}(e_1), \ldots, \mathrm{rk}(e_n)\} + 1 .$$

Example 11. First, consider the formula

$$P(\underbrace{\varepsilon_x[P(x) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}) \lor Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) .$$

Here, e_1 is the only immediate ε-subterm of e_2 and has no ε-subterms itself, so $\deg(e_1) = 1$ and $\deg(e_2) = 2$. Neither e_1 nor e_2 contains subordinate

ε-expressions, hence $\mathrm{rk}(e_1) = \mathrm{rk}(e_2) = 1$. In

$$[\exists x \exists y\, A(x, y)]^\varepsilon = A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x, z))}_{e_3}, \underbrace{\varepsilon_y A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x, z))}_{e_3}, y)}_{e_4}) \ ,$$

e_3 contains no ε-subterms, but e_4 contains e_3 as a subterm, so $\deg(e_3) = 1$ and $\deg(e_4) = 2$. On the other hand, e_3 contains the subordinate ε-expression $\varepsilon_z A(x, z)$, hence $\mathrm{rk}(e_3) = 2$. Since y does not occur in the scope of another ε, e_4 contains no subordinate ε-expressions, and $\mathrm{rk}(e_4) = 1$.

Definition 12. Suppose π is a proof in EC_ε. If e is an ε-term belonging to a critical formula $A(t) \to A(e)$ of π, then we call e a *critical epsilon term* of π, and $\mathrm{rk}(e)$ ($\deg(e)$) the rank (the degree) of that critical formula.

The *rank* $\mathrm{rk}(\pi)$ of π is the maximum rank of its critical formulas. The *degree* $\deg(\pi, r)$ of π with respect to rank r is the maximum degree of its critical ε-terms of rank r. The *order* $o(\pi, r)$ of π with respect to rank r is the number of different critical ε-terms of rank r.

Lemma 13. *Let π be a EC_ε-proof, let $r = \mathrm{rk}(\pi)$ be the maximal rank of critical formulas in π, and let e be a critical ε-term of π of maximal degree among the critical ε-terms of rank r.*

Suppose that $A(t) \to A(e)$ is a critical formula belonging to e and that $B^* \equiv B(s) \to B(\varepsilon_y B(y))$ is a critical formula in π belonging to a different ε-term $\varepsilon_y B(y)$, and suppose C is the result of replacing e by t in B^*. Then (a) if $\mathrm{rk}(B^*) = r$, then C and B^* have the the same critical ε-term $\varepsilon_y B(y)$ belonging to them, and (b) $\mathrm{rk}(C) = \mathrm{rk}(B^*)$.

Proof. We first consider and exclude a preliminary case. If the indicated occurrences of s (on the left-hand side) or the indicated occurrences of $\varepsilon_y B(y)$ (on the right-hand side) lie *inside* e, replacing e by t would result in a formula which is not of the form of a critical formula. This, however, can never be the case. For suppose it were, i.e., suppose $B(a)$ is of the form $B'(e'(a))$ and either $e \equiv e'(s)$ or $e \equiv e'(\varepsilon_y B(y))$. If $e \equiv e'(s)$, the left-hand-side $B(s)$ of the critical formula is of the form $B'(e'(s))$, and consequently the right-hand-side is $B'(e'(\varepsilon_y B'(e'(y))))$. Conversely, if $e \equiv e'(\varepsilon_y B(y))$, then the right-hand-side $B(\varepsilon_y B(y))$ would be of the form $B'(e'(\varepsilon_y B'(e'(y))))$. In either case we have an ε-term $e'(a)$ which is of the same rank as e, since $e = e'(t')$ for some term t'. (Note that by our conventions on renaming of variables, no subexpression of t' can be subordinate to e.) On the other hand, $e'(y)$ is subordinate to the critical ε-term $\varepsilon_y B(y) = \varepsilon_y B'(e'(y))$, and so $\mathrm{rk}(\varepsilon_y B(y)) > \mathrm{rk}(e'(y)) = \mathrm{rk}(e)$. But e was assumed to be a critical ε-term of maximal rank.

There are then only two ways in which e can occur in a critical formula: either (i) e occurs only in the indicated occurrences of s but not in $B(y)$ at all, or (ii) e occurs in $B(y)$ (and perhaps also in s).

Case (i): e occurs only in s, i.e., $s \equiv s'(e)$. Replacing e by t results in the critical formula $C \equiv B(s'(t)) \to B(\varepsilon_y B(y))$. The new critical formula C belongs to the same ε-term as the original formula, hence we obtain (a) $\mathrm{rk}(C) = \mathrm{rk}(B^*) = \mathrm{rk}(\varepsilon_y B(y))$, and (b) holds trivially.

Case (ii): e occurs in $B(y)$ (and perhaps also in s). In this case, $B(y) \equiv B'(y, e)$ and the critical formula has the form

$$B'(s, e) \to B'(\varepsilon_y B'(y, e), e) \ .$$

Then the ε-term belonging to this critical formula, $e' \equiv \varepsilon_y B'(y, e)$, contains e as a proper subterm and hence is of higher degree than e. Since e is a critical ε-term of maximal degree among the critical ε-terms of maximal rank in π, this implies that $\mathrm{rk}(e') < \mathrm{rk}(e)$. Replacing e by t yields the critical formula

$$C \equiv B'(s', t) \to B'(\varepsilon_y B'(y, t), t) \ ,$$

belonging to the ε-term $\varepsilon_y B'(y, t)$. This term has the same rank as e' and hence a lower rank than e itself (although it might have a degree higher than $\deg(e)$). Hence, $\mathrm{rk}(C) < r$. ■

Lemma 14. *Suppose* $\mathrm{EC}_\varepsilon \vdash_\pi E$. *Let* $r = \mathrm{rk}(\pi)$ *be the maximal rank of critical formulas in* π, *and let* e *be a critical* ε*-term of* π *of maximal degree among the critical* ε*-terms of rank* r. *Then there is a proof* π_e *so that* $\mathrm{EC}_\varepsilon \vdash_{\pi_e} E$ *with* $\mathrm{rk}(\pi_e) \leq r$, $\deg(\pi_e, r) \leq \deg(\pi, r)$ *and* $o(\pi_e, r) = o(\pi, r) - 1$.

Proof. The ε-expression e is of the form $\varepsilon_x A(x)$, and suppose that $A(t_1) \to A(e)$, …, $A(t_n) \to A(e)$ are all the critical formulas in π belonging to e. For each $i = 1, \ldots, n$, we obtain a proof π_i of $A(t_i) \to E$ as follows:

1. Replace e everywhere in π by t_i. Every critical formula $A(t_j) \to A(e)$ belonging to e thus turns into a formula of the form $A(t'_j) \to A(t_i)$. To see this, note that e cannot occur in $A(x)$, for otherwise $e \equiv \varepsilon_x A(x)$ would be a proper subterm of itself, which is impossible.

2. Add $A(t_i)$ to the axioms. Now every one of the new formulas $A(t'_j) \to A(t_i)$ is derivable using modus ponens from the tautology

$$A(t_i) \to (A(t'_j) \to A(t_i)) \ ,$$

3. Apply the deduction theorem to obtain π_i.

We verify that π_i is indeed an EC_ε-proof with the required properties. In the construction of π_i, we replaced e by t_i throughout the proof. Such a substitution obviously preserves tautologies. By Lemma 13, it also turns critical formulas into critical formulas (belonging, perhaps, to different critical ε-terms). Lemma 13 also guarantees that replacing e by t_i everywhere does not change the rank of critical formulas, and that it does not change the critical ε-terms of maximal rank r at all (in particular, it does not increase the degree of critical ε-terms of rank r).

We started with critical formulas $A(t_i) \to A(e)$, and obtained a proof π_i which does not contain any critical formulas belonging to e. Hence e is no longer a *critical* ε-term in π_i. The ranks of all other critical formulas (and the corresponding critical ε-terms) remain unchanged. Thus $o(\pi_i, r) = o(\pi, r) - 1$.

Secondly, we construct a proof π' of $\bigwedge_{i=1}^n \neg A(t_i) \to E$ as follows:

1. Add $\bigwedge \neg A(t_i)$ to the axioms. Now every critical formula $A(t_i) \to A(e)$ belonging to e is provable using the propositional tautology $\neg A(t_i) \to (A(t_i) \to A(e))$.

2. Apply the deduction theorem for the propositional calculus to obtain a proof π', which contains exactly the same critical formulas as π except those belonging to e.

Now combine the proofs π_i of $A(t_i) \to E$ and π' of $\bigwedge_i \neg A(t_i) \to E$ to get the proof π_e of E (using case distinction).

None of the proofs π_i, π' contain critical formulas belonging to e, and no critical ε-terms of rank r other than those in π. Thus $\mathrm{rk}(\pi_e) \leq r$, $\deg(\pi_e, r) \leq \deg(\pi, r)$, and $o(\pi_e, r) = o(\pi, r) - 1$ hold. ■

Theorem 15 (First Epsilon Theorem). *If E is a formula without bound variables (no quantifiers, no epsilons) and $\mathrm{PC}_\varepsilon \vdash E$ then $\mathrm{EC} \vdash E$.*

Proof. First, use the embedding lemma to obtain π so that $\mathrm{EC}_\varepsilon \vdash_\pi A$. The theorem then follows from the preceding lemma by induction on $r = \mathrm{rk}(\pi)$ and $d = o(\pi, r)$. If $r = 0$, there are no critical formulas, so there is nothing to prove. If $r > 0$, then d-fold application of the lemma results in a proof π' of rank $< r$. ■

Theorem 16 (Extended First Epsilon Theorem). *Suppose $E(e_1, \dots, e_m)$ is a quantifier-free formula containing only the ε-terms $s_1, \dots, s_m$, and*

$$\mathrm{EC}_\varepsilon \vdash_\pi E(s_1, \dots, s_m) \ ,$$

then there are ε-free terms t^i_j $(1 \le i \le n,\ 1 \le j \le m)$ such that

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \ldots, t^i_m)$$

where $n \le 2^{3\cdot\mathrm{cc}(\pi)}_{2\cdot\mathrm{cc}(\pi)}$.

Corollary 17 (Herbrand's Theorem). *If $\exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m)$ is a purely existential formula containing only the bound variables x_1, …, x_m, and*

$$\mathrm{PC}_\varepsilon \vdash_\pi \exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m)\ ,$$

then there are ε-free terms t^i_j $(1 \le i \le n,\ 1 \le j \le m)$ such that

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \ldots, t^i_m)$$

where $n \le 2^{3\cdot\mathrm{cc}(\pi)}_{2\cdot\mathrm{cc}(\pi)}$.

Proof. Immediate from Theorem 16 using the Embedding Lemma. ■

The rest of this section is devoted to the proof of of Theorem 16. First, some additional notation.

Definition 18. Suppose π is a proof in EC_ε. The *width* $\mathrm{wd}_\pi(e)$ of π with respect to e is the number of different critical formulas in π belonging to the ε-term e. The *width* $\mathrm{wd}(\pi, r)$ of π with respect to rank r is given by

$$\mathrm{wd}(\pi, r) = \max\{\mathrm{wd}_\pi(e) \mid e \text{ of rank } r \text{ occurs in } \pi\} + 1\ .$$

Definition 19. Let $E(a_1, \ldots, a_m)$ be a formula in $L(\mathrm{EC})$ without bound variables, and let s_1, …, s_m be terms in $L(\mathrm{EC}_\varepsilon)$. An $\lor$-*expansion* (of $E \equiv E(s_1, \ldots, s_m)$) is a finite disjunction

$$E' \equiv E_1 \lor \cdots \lor E_l\ ,$$

where each $E_i \equiv E(s^i_1, \ldots, s^i_m)$ for terms s^i_j $(1 \le i \le l,\ 1 \le j \le m)$. We call l the *length* $\mathrm{len}(E', E)$ of the expansion.

Proposition 20. *Suppose A' is an $\lor$-expansion of A and A'' is an $\lor$-expansion of A'. Then A'' is also an $\lor$-expansion of A, and* $\mathrm{len}(A'', A) \le \mathrm{len}(A'', A') \cdot \mathrm{len}(A', A)$.

Proof. Obvious. ■

Lemma 21. *Suppose $E(a_1, \dots, a_m)$ is a formula in $L(\mathrm{EC})$ without bound variables and π is an EC_ε-proof of $E(s_1, \dots, s_m)$ where $s_1, \dots, s_m$ are terms in $L(\mathrm{EC}_\varepsilon)$. Let $r = \mathrm{rk}(\pi)$ and let e be a critical ε-term of π of maximal degree among the critical ε-terms of rank r, and let $n = \mathrm{wd}_\pi(e)$ be the number of critical formulas belonging to e. Then there are terms s_j^i $(0 \le i \le n+1$, $1 \le j \le m)$ and a proof π_e with end formula*

$$E(s_1^1, \dots, s_m^1) \lor \dots \lor E(s_1^{n+1}, \dots, s_m^{n+1}) \; ,$$

so that $\mathrm{rk}(\pi_e) \le r$, $\deg(\pi_e, r) \le \deg(\pi, r)$, and $o(\pi_e, r) = o(\pi, r) - 1$. Furthermore $\mathrm{cc}(\pi_e) \le \mathrm{cc}(\pi) \cdot (n+1)$ and $\mathrm{wd}(\pi_e, r) \le \mathrm{wd}(\pi, r) \cdot (n+1) \le \mathrm{wd}(\pi, r)^2$.

Proof. We adapt the construction of π_i in Lemma 14. The only difference to the previous construction is that when replacing e by t_i throughout π, the end-formula $E(s_1, \dots, s_m)$ may change. However, e can only occur in $s_1, \dots, s_m$ since $E(a_1, \dots, a_m)$ contains no bound variables and hence no ε-terms. For each critical formula $A(t_i) \to A(e)$ we obtain a proof π_i of $A(t_i) \to E(s_1^i, \dots, s_m^i)$. The construction of π' as before yields a proof of

$$\bigwedge_{i=1}^{n} \neg A(t_i) \to E(s_1^{n+1}, \dots, s_m^{n+1}) \; ,$$

if we take $s_j^{n+1} = s_j$. Then, since obviously for each i

$$E(s_1^i, \dots, s_m^i) \to \bigvee_{i=1}^{n+1} E(s_1^i, \dots, s_m^i) \; ,$$

is provable, we obtain a proof π_e of $\bigvee_{i=1}^{n+1} E(s_1^i, \dots, s_m^i)$ with the desired properties. Observe that the length of the $\lor$-expansion $\bigvee E(s_1^i, \dots, s_m^i)$ is $n + 1 = \mathrm{wd}_\pi(e) + 1 \le \mathrm{wd}(\pi, r)$.

It remains to verify the bounds on $\mathrm{cc}(\pi_e)$ and $\mathrm{wd}(\pi_e, r)$. By Lemma 13, replacing e by t_i to obtain π_i does not introduce new critical ε-terms of rank r. (Critical formulas belonging to ε-terms of $\mathrm{rk}(e)$ may be altered, but the corresponding critical ε-terms remain the same.) New critical ε-terms can only appear at a rank $< \mathrm{rk}(e)$, and if they do, their rank is equal to the rank of a some critical ε-term already in π. The total number of critical formulas in π_i is at most that of π less the number n of critical formulas belonging to e, i.e., $\mathrm{cc}(\pi_i) \le \mathrm{cc}(\pi) - n$. Moreover, obviously $\mathrm{cc}(\pi') \le \mathrm{cc}(\pi) - n$ holds.

When we combine the $n+1$ proofs π_i and π' to obtain π_e, the worst case is that every critical formula in π_i has been changed. Thus $\mathrm{cc}(\pi_e) \leq (\mathrm{cc}(\pi) - n)(n+1) \leq \mathrm{cc}(\pi)(n+1)$. Now looking more closely at the critical formulas of rank r in π_i, we see that whenever case (i) in the proof of Lemma 13 applies, a critical formula belonging to some ε-term e' of rank r in π turns into a potentially new critical formula in π_i also belonging to e'. However, these are the only new critical formulas belonging to e'. Hence, there may be up to $\mathrm{wd}_\pi(e') \cdot (n+1)$ different critical formulas belonging to e' in π_e. Thus $\mathrm{wd}(\pi_e, r) \leq \mathrm{wd}(\pi, r) \cdot (n+1)$, which is $\leq \mathrm{wd}(\pi, r)^2$ since $n+1 = \mathrm{wd}_\pi(e) + 1 \leq \mathrm{wd}(\pi, r)$. ■

We now iterate the elimination of ε-terms of highest rank and estimate the critical count of the proof resulting from the elimination of all ε-terms of rank $\mathrm{rk}(e)$.

Lemma 22. *Suppose $E(a_1, \ldots, a_m)$ is a formula in $L(\mathrm{EC})$ without bound variables and π is an EC_ε-proof of $E(s_1, \ldots, s_m)$, where $s_1, \ldots, s_m$ are terms in $L(\mathrm{EC}_\varepsilon)$. Then there is a proof σ of an $\vee$-expansion E' of $E(s_1, \ldots, s_m)$, so that $\mathrm{rk}(\sigma) < \mathrm{rk}(\pi)$. Furthermore,*

$$\mathrm{cc}(\sigma) \leq 2^{2^{2 \cdot \mathrm{cc}(\pi)}} \qquad \textit{and} \qquad \mathrm{len}(E', E(s_1, \ldots, s_m)) \leq 2^{2^{2 \cdot \mathrm{cc}(\pi)}} .$$

Proof. Let $d = o(\pi, r)$ and let $e_1, \ldots, e_d$ be all critical ε-terms of rank r in π. We assume the sequence $e_1, \ldots, e_d$ is ordered so that the degree never increases. Let $k = \mathrm{cc}(\pi)$ and $r = \mathrm{rk}(\pi)$. As observed in the preceding proof, an application of Lemma 21 cannot increase the number of critical ε-terms of maximal rank. Thus let $\sigma^0 = \pi$ and $\sigma^j = \sigma^{j-1}_{e_j}$ for $j > 0$. We thus obtain $\sigma = \sigma^d$ by d-fold iteration of Lemma 21. Let $E^0 \equiv E(s_1, \ldots, s_m)$. By construction, the critical ε-terms of rank r in σ^j are $e_j, \ldots, e_d$ and the end-formula E^j of σ^j is an $\vee$-expansion of $E \equiv E(s_1, \ldots, s_m)$; we set $E' \equiv E^d$.

By induction on j we prove:

$$\begin{aligned} \mathrm{wd}(\sigma^j, r) &\leq k^{2^j} , \\ \mathrm{cc}(\sigma^j) &\leq k \cdot k^{\sum_{l=0}^{j-1} 2^l} , \\ \mathrm{len}(E^j, E) &\leq k^{\sum_{l=0}^{j-1} 2^l} . \end{aligned}$$

For $j = 0$, we have $\mathrm{wd}(\sigma^0, r) = \mathrm{wd}(\pi, r) \leq k$, $\mathrm{cc}(\sigma^0) = \mathrm{cc}(\pi) = k$ and $\mathrm{len}(E^0, E(s_1, \ldots, s_m)) = 1$, by definition.

Now assume that σ^j obeys the stated bounds, we prove that σ^{j+1} does as well. Apply Lemma 21 to eliminate the ε-term e_{j+1}. This yields a proof σ^{j+1} of E^{j+1}. By Lemma 21 and the induction hypothesis we have:

$$\begin{array}{rcl} \mathrm{wd}(\sigma^{j+1}, r) & \leq & \mathrm{wd}(\sigma^j, r)^2 \leq (k^{2^j})^2 = k^{2^{j+1}} \ , \\ \mathrm{cc}(\sigma^{j+1}) & \leq & (k \cdot k^{\sum_{l=0}^{j-1} 2^l}) \cdot k^{2^j} = k \cdot k^{\sum_{l=0}^{j} 2^l}, \text{and} \\ \mathrm{len}(E^{j+1}, E) & \leq & (k^{\sum_{l=0}^{j-1} 2^l}) \cdot k^{2^j} = k^{\sum_{l=0}^{j} 2^l} \ . \end{array}$$

since $\mathrm{wd}_{\sigma^j}(e_{j+1}) + 1 \leq \mathrm{wd}(\sigma^j, r) \leq k^{2^j}$. Thus the claim follows and we obtain

$$\begin{array}{rcl} \mathrm{cc}(\sigma) & \leq & k \cdot k^{\sum_{l=0}^{d-1} 2^l} = k \cdot k^{2^d - 1} = k^{2^d} \quad \text{and} \\ \mathrm{len}(E', E) & \leq & k^{\sum_{l=0}^{d-1} 2^l} = k^{2^d - 1} \ . \end{array}$$

The order $d = o(\pi, r)$ is the number of critical ε-terms of rank r, and hence $\leq k$. Using the inequality $k \leq 2^k$ we obtain the (rough) upper bounds $\mathrm{cc}(\sigma) \leq k^{2^d} \leq 2^{2^{2k}}$ and $\mathrm{len}(E', E) \leq k^{2^{d-1}} \leq 2^{2^{2k}}$. ■

Proof of Extended First Epsilon Theorem 16. Consider a proof π of $E(s_1, \ldots, s_m)$, where $s_1, \ldots, s_m$ are terms containing ε's and $E(a_1, \ldots, a_m)$ contains no bound variables. Let $k = \mathrm{cc}(\pi)$ and p be the number of different ranks of critical ε-terms in π, i.e., $p = |\{r : \mathrm{wd}(\pi, r) > 1\}|$. The number p is the number of times we have to apply Lemma 22 to eliminate all critical formulas from π. (Note that by the proof of Lemma 21 each elimination step can only decrease the number of different ranks.) Although obviously $p \leq r = \mathrm{rk}(\pi)$ we also have $p \leq k = \mathrm{cc}(\pi)$, so the number of times Lemma 22 must be applied is actually independent of $\mathrm{rk}(\pi)$.

Now let $\pi^0 = \pi$ and π^{j+1} be the proof σ constructed in Lemma 22 starting with π^j. Note that the end-formula of each π^j is an $\vee$-expansion of E. If we write E^j for the end-formula of π^j, then E^p is the required Herbrand disjunction

$$\bigvee_{i=1}^{n} E(t_1^i, \ldots, t_m^i) \ . \tag{2}$$

To establish a bound on the length n of (2), we apply Lemma 22 $(p - 1)$ times. This yields the following bounds:

$$\mathrm{cc}(\pi^{p-1}) \leq 2_{2(p-1)}^{2k+(p-1)} \quad \text{and} \quad \mathrm{len}(E_{p-1}, E) \leq 2_{2(p-1)}^{2k+(p-1)} \ .$$

Another application of Lemma 22 yields that

$$\begin{aligned} n &\leq \operatorname{len}(E_p, E_{p-1}) \cdot \operatorname{len}(E_{p-1}, E) \\ &\leq 2^{2^{2\cdot \mathrm{cc}(\pi^{p-1})}} \cdot \operatorname{len}(E_{p-1}, E) \\ &\leq 2^{2^{2^{\left(2^{2k+(r-1)}_{2(r-1)}\right)}}} \cdot 2^{2k+(p-1)}_{2(p-1)} \\ &\leq 2^{2^{2^{2k+p}_{2(p-1)}}} = 2^{2k+p}_{2p} \leq 2^{3k}_{2k} \,. \end{aligned}$$

As a last step, we remove the remaining (non-critical) ε-terms from the proof by replacing outermost ε-terms by free variables. Clearly, this cannot increase the length n of (2). ■

6. Lower Bounds on Herbrand Disjunctions

As noted in the proof of Theorem 16, the bound on the length of the Herbrand disjunction depends only on the critical count of the initial proof. This is in contrast to the bound we would obtain by the more standard approach of cut-elimination and the mid-sequent theorem which depends on the length and cut complexity of the original proof (see, e.g., [10, 23]). In the case of the ε-calculus, the result concerns the relation between the critical count of a proof of $\exists x\, E(x)$ in PC_ε and the length of a Herbrand disjunction $\bigvee E(t_i)$. In the case of the sequent calculus and cut-elimination the result concerns the relation between the length and cut complexity of a proof with cut, and the length of a cut-free proof, which in turn determines the length of a Herbrand disjunction obtained via the mid-sequent theorem. In both cases, the relationship is hyperexponential. Statman [22] and Orevkov [21] showed that this bound is not just an artefact of the particular cut-elimination procedure considered, but that proofs with cut essentially have hyper-exponential speedup over cut-free proofs. The question may then be raised whether the same holds true of the ε-calculus, i.e., whether the bound on the length of Herbrand disjunctions obtained in the first ε-theorem is tight. Although we do not have a result quite as optimal as Orevkov's in this regard, it can be shown that every ε-elimination procedure that yields Herbrand disjunctions must by hyperexponential.

We sketch the proof of such a lower bound theorem for the length of Herbrand disjunctions. Recall that the *Herbrand complexity* $\mathrm{HC}(E)$ of a purely existential formula $E \equiv \exists x_1 \ldots \exists x_n E'(x_1, \ldots, x_n)$ is the length of the shortest $\vee$-expansion of $E'(x_1, \ldots, x_n)$.

THEOREM 23. *There is a sequence of formulas E_k so that*

1. *for each k, there is a* PC_ε*-proof* π_k *of* E_k *with* $\mathrm{cc}(\pi_k) \leq c \cdot k$ *(for some constant c), but*

2. $\mathrm{HC}(E_k) \geq 2^1_k$.

We follow the presentation of Orevkov's Theorem in [23, §6.11]. (Statman's result requires equality, but Orevkov's does not.) Consider a language including a unary constant 0, a unary function symbol S and a ternary relation R, whose meaning is fixed by the following axioms:

$$\begin{aligned} &\mathrm{Hyp}_1 \equiv \forall x\, R(x, 0, S(x)) \ , \\ &\mathrm{Hyp}_2 \equiv \forall y \forall x \forall z \forall z_1 (R(y, x, z) \land R(z, x, z_1) \to R(y, S(x), z_1)) \ . \end{aligned}$$

Further, we define

$$C_k \equiv \exists z_k \ldots \exists z_0 (R(0, 0, z_k) \land R(0, z_k, z_{k-1}) \land \cdots \land R(0, z_1, z_0)) \ .$$

$R(n, m, k)$ expresses that $n + 2^m = k$, and C_k expresses that 2^1_k is defined. E_k is the (purely existential) prefix form of $\mathrm{Hyp}_1 \land \mathrm{Hyp}_2 \to C_k$.

Lemma 24. *For every k,* $\mathrm{PC}_\varepsilon \vdash_{\pi_k} E_k$*, where* $\mathrm{cc}(\pi_k) = c \cdot k$ *(for some constant c).*

Proof. E_k is provable in the sequent calculus (alternatively, in natural deduction) using proofs (with cut) of length linear in k, see [23]. Proofs in the sequent calculus and in natural deduction can be translated into proofs in PC_ε with linear increase in proof length. Moreover, as in the embedding lemma, only weak quantifier inferences (∃I, ∀E) increase the critical count of the PC_ε-proof. (We omit the details, which are routine.) ■

This establishes part (1) of the theorem. Orevkov's result concerns proof lengths; we have to adapt the proof to Herbrand complexity. We give a direct proof of a lower bound on $\mathrm{HC}(E_k)$; the result can also be obtained using techniques from proof complexity as in [6]. In order to simplify the presentation, we will consider Herbrand *sequents* of $\mathrm{Hyp}_1, \mathrm{Hyp}_2 \Rightarrow C_k$ instead of Herbrand disjunctions, i.e., valid sequents of the form $\Gamma_1, \Gamma_2 \Rightarrow \Delta$ where each formula in Γ_1 is a substitution instance of $R(x, 0, S(x))$, each formula in Γ_2 is a substitution instance of $R(y, x, z) \land R(z, x, z_1) \to R(y, S(x), z_1)$, and each formula in Δ is one of $R(0, 0, z_k) \land R(0, z_k, z_{k-1}) \land \cdots \land R(0, z_1, z_0)$. Then obviously $\max\{|\Gamma_1|, |\Gamma_2|, |\Delta|\} \leq \mathrm{HC}(E_k)$. In the following, $\bar{n}$ abbreviates $S^n(0)$, and $\Psi = \{\mathrm{Hyp}_1, \mathrm{Hyp}_2\}$.

Lemma 25. *Suppose $\Psi \Rightarrow R(\bar{n}, \bar{m}, \bar{l})$ is valid. Then $l = n + 2^m$ and for each Herbrand sequent $T \equiv (\Gamma_1, \Gamma_2 \Rightarrow R(\bar{n}, \bar{m}, \bar{l}))$ of $\Psi \Rightarrow R(\bar{n}, \bar{m}, \bar{l})$, we have*

$$\{R(\bar{i}, 0, S(\bar{i})) : n \leq i < n + 2^m\} \subseteq \Gamma_1 .$$

In particular, $|\Gamma_1| \geq 2^m$.

Proof. If $m = 0$, then clearly the only possibility is $l = n + 1$. Then any Herbrand sequent of $\Psi \Rightarrow R(\bar{n}, 0, S(\bar{n}))$ can be written as $R(\bar{n}, 0, S(\bar{n})), \Gamma' \Rightarrow R(\bar{n}, 0, S(\bar{n}))$ and satisfies the conditions.

Suppose the result is established for m, and consider the case for $m + 1$. Let $\Gamma_1, \Gamma_2 \Rightarrow R(\bar{n}, S(\bar{m}), \bar{l})$ be any Herbrand sequent of $\Psi \Rightarrow R(\bar{n}, S(\bar{m}), \bar{l})$ with Γ_1 the instances corresponding to Hyp_1 and Γ_2 those corresponding to Hyp_2. As $R(\bar{n}, S(\bar{m}), \bar{l})$ cannot follow from instances of $R(x, 0, S(x))$ alone, Γ_2 is nonempty and must contain a formula of the form

$$R(\bar{n}, \bar{m}, \bar{k}) \wedge R(\bar{k}, \bar{m}, \bar{l}) \rightarrow R(\bar{n}, S(\bar{m}), \bar{l}) ,$$

such that $T_1 \equiv (\Gamma_1, \Gamma_2 \Rightarrow R(\bar{n}, \bar{m}, \bar{k}))$ and $T_2 \equiv (\Gamma_1, \Gamma_2 \Rightarrow R(\bar{k}, \bar{m}, \bar{l}))$ are both valid. That means that T_1 is a Herbrand sequent of $\Psi \Rightarrow R(\bar{n}, \bar{m}, \bar{k})$ and T_2 one of $\Psi \Rightarrow R(\bar{k}, \bar{m}, \bar{l})$. The induction hypothesis applies, and it follows that $k = n + 2^m$ and $l = k + 2^m$, thus $l = n + 2^{m+1}$. Further, Γ_1 must contain

$$\{R(\bar{i}, 0, S(\bar{i})) : n \leq i < n + 2^m\} \cup \{R(\bar{i}, 0, S(\bar{i})) : n + 2^m \leq i < n + 2^{m+1}\} ,$$

and the lemma follows. ■

Lemma 26. *Let S be a sequent of the form $\Psi \Rightarrow \exists \bar{z}\, A(\bar{z})$. Then every minimal Herbrand sequent of S is of the form $\Gamma \Rightarrow \Delta$ with $|\Delta| = 1$.*

Proof. We use of the terminology and results of Chapter XI of [12]. Suppose that $T \equiv \Gamma \Rightarrow \Delta$ is a Herbrand sequent of S with $\Delta = A(\bar{t}_1), \ldots, A(\bar{t}_n)$. Each formula in Γ is Horn. Thus the term model $\mathfrak{I}^\Gamma$ is a free model of Γ (Corollary 2.5 of [12]). Since $\Gamma \Rightarrow \Delta$ is valid and $\mathfrak{I}^\Gamma \models \Gamma$, $\mathfrak{I}^\Gamma \models A(\bar{t}_1) \vee \ldots \vee A(\bar{t}_n)$. Then there is an i so that $\mathfrak{I}^\Gamma \models A(\bar{t}_i)$. But $\mathfrak{I}^\Gamma$ is free. Hence, every model of Γ is also a model of $A(t_i)$ and $\Gamma \Rightarrow A(t_i)$ is a Herbrand sequent of S. ■

Lemma 27. *If $T \equiv (\Gamma_1, \Gamma_2 \Rightarrow \Delta)$ is a minimal Herbrand sequent of $\Psi \Rightarrow C_k$, then $|\Gamma_1| \geq 2^1_k$.*

Proof. By Lemma 26, T is of the form

$$\Gamma,\Gamma_2 \Rightarrow R(0,0,\bar{n}_k) \land R(0,\bar{n}_k,\bar{n}_{k-1}) \land \cdots \land R(0,\bar{n}_1,\bar{n}_0)\ . \tag{3}$$

(Note that by substituting 0 for free variables, terms in a Herbrand sequent may always be brought into that form). As (3) is valid, each of the sequents $\Gamma_1, \Gamma_2 \Rightarrow R(0,0,\bar{n}_k)$, $\dots$, $\Gamma_1, \Gamma_2 \Rightarrow R(0,\bar{n}_1,\bar{n}_0)$ is valid as well. Applying Lemma 25 $(k-1)$-times, we see that $n_1 = 2^1_{k-1}$. Since $\Gamma_1, \Gamma_2 \Rightarrow R(0,\bar{n}_1,\bar{n}_0)$ is a Herbrand sequent of $\Psi \Rightarrow R(0,\bar{n}_1,\bar{n}_0)$, Γ_1 contains the instances of Hyp_1 given in Lemma 25, and $n_0 = 2^{n_1} = 2^1_k$. Hence, $|\Gamma| \geq 2^1_k$. ■

7. The Second Epsilon Theorem

The Second Epsilon Theorem is a generalization of the first. It states that a formula without ε-terms provable in the extended predicate calculus is already provable in the predicate calculus (without ε-terms). Its proof utilizes no additional methods specific to the ε-calculus beyond those of the first ε-theorem.

Definition 28. Suppose $A = \mathsf{Q}_1 x_1 \dots \mathsf{Q}_n x_n\, B(x_1, \dots, x_n)$ is a prenex formula. Let $z_1, \dots, z_l$ be all the $\forall$-quantified variables among $x_1, \dots, x_n$, let $y_1, \dots, y_m$ be all the $\exists$-quantified ones, and let $f_1, \dots, f_l$ be new function symbols. The Herbrand normal form A^H of A is

$$\exists y_1 \dots \exists y_m\, C(y_1, \dots, y_m, t_1, \dots, t_l)\ ,$$

where $t_j = f_j(y_1, \dots, y_m)$.

Lemma 29. *Suppose* $\mathrm{PC}_\varepsilon \vdash A$*. Then* $\mathrm{PC}_\varepsilon \vdash A^H$*.*

Proof. Standard. ■

Theorem 30 (Second Epsilon Theorem). *If A is a formula of $L(\mathrm{PC})$ and $\mathrm{PC}_\varepsilon \vdash A$, then $\mathrm{PC} \vdash A$.*

Proof. For simplicity, we give the proof for prenex formula A with a simple quantifier structure. The general results follows similarly. Assume A has the form $\exists x \forall y \exists z\, B(x,y,z)$ with $B(x,y,z)$ quantifier-free and only the indicated variables occur in A. We apply Lemma 29 to obtain a proof of $\exists x \exists z\, B(x, f(x), z)$. The extended first ε-theorem now yields that there are ε-free terms r_i, s_i so that

$$\mathrm{EC} \vdash \bigvee_i B(r_i, f(r_i), s_i)\ . \tag{4}$$

The idea is now to replace the $f(r_i)$ by new free variables a_i and obtain from (4), that

$$\bigvee_i B(r'_i, a_i, s'_i)$$

is deducible in EC. Then the original prenex formula A can be obtained if we employ suitable quantifier-shifting rules (deducible in PC).

Let $f(r_1), \ldots, f(r_p)$ denote terms occurring in the disjunction (4). Let p_i be the number of occurrences of f in $f(r_i)$. We may assume that the disjunction is arranged so that that the sequence $f(r_1), \ldots, f(r_p)$ is ordered such that $p_i \leq p_{i+1}$. Now let $a_1, \ldots, a_p$ be new free variables. Replace each occurrence of $f(r_i)$ which does not occur as a subterm of another $f(r_j)$ by a_i. Then (4) becomes

$$\bigvee_i B(r'_i, a_i, s'_i) \ , \tag{5}$$

Observe that r'_i does not contain a_j for $j \geq i$. For if r'_i did contain a_j, then r_i must contain $f(r_j)$, and $p_j < p_i$. But we assumed that the disjunctions were ordered so that the sequence of p_i was non-decreasing.

It is easy to see that (5) is also a tautology, since pairs of equal atomic formulas remain pairs of equal atomic formulas. Thus from (5) we obtain that

$$\bigvee_i \exists z B(r'_i, a_i, z) \ , \tag{6}$$

by existentially generalizing on the terms s'_i. Now consider the last disjunct in (6). By the preceding observation, a_p does not occur in any other disjunct, or in r'_p. Hence, in PC, we may deduce from

$$\bigvee_{i=1}^{p-1} \exists z\, B(r'_i, a_i, z) \lor \exists z\, B(r'_p, a_p, z) \ ,$$

the formula

$$\bigvee_{i=1}^{p-1} \exists z\, B(r'_i, a_i, z) \lor \forall y \exists z\, B(r'_p, y, z) \ .$$

Iterating these steps we eventually obtain a proof of A in PC. ■

8. Conclusion and Further Work

The above proofs of the first and second ε-theorem were formulated for theorems of PC_ε. However, as indicated in the introduction, the theorems remain valid in the presence of open (quantifier- and ε-free) theories.

Corollary 31. *Let* Ax *be a set of formulas without bound variables.*

1. *Let E be a formula without bound variables (no quantifiers, no epsilons). If* $\mathrm{Ax} \vdash E$ *in* PC_ε*, then* $\mathrm{Ax} \vdash E$ *in* EC*,*

2. *Let $\exists \overline{x} E(\overline{x})$ be a purely existential formula. If* $\mathrm{Ax} \vdash \exists \overline{x} E(\overline{x})$ *in* PC_ε*, then* $\mathrm{Ax} \vdash \bigvee_i E(t_{i1}, \ldots, t_{in})$ *for some ε-free terms t_{ij} in* EC

3. *If E is an ε-free formula and* $\mathrm{Ax} \vdash E$ *in* PC_ε*, then* $\mathrm{Ax} \vdash E$ *in* PC*.*

The discussion of the ε-calculus given here is only a first step toward a more comprehensive investigation of Hilbert's ε-calculus. The gap in the upper and lower bound for the Herbrand complexity of theorems of EC_ε suggests that a stricter analysis or a refinement of the elimination procedure for the first ε-theorem is possible. A more interesting and pressing question, however, is the analysis of ε-elimination procedures for the ε-calculus with equality. The addition of equality to the ε-calculus is not as straightforward as it is in the predicate calculus, and the first ε-theorem is significantly more complicated if equality is present than when it is not. It remains a topic for future work.

In the introduction we claimed that encoding of quantifiers on the term level using the ε-operator may allow for a more condensed representation of proofs. Let us briefly sketch the reason for this. Since modus ponens is the only inference rule in EC_ε, a formula A^ε is provable in EC_ε iff there is a tautology of the form

$$\bigwedge_{i,j} (B_i(t_j) \to B_i(\varepsilon_x B_i(x))) \to A^\varepsilon \ , \tag{7}$$

Thus it suffices to find the critical formulas $B_i(t_j) \to B_i(\varepsilon_x B_i(x))$, i.e., the substitutions involved in the proof of A, such that (7) is a tautology. This suggests that the formalization of proofs is simpler in the ε-calculus or at least that proofs in the ε-calculus can be represented more succinctly.

As pointed out above, the bound on the length of the Herbrand disjunctions obtained using the first ε-theorem depends only on the critical count of the initial proof. (If equality is present, however, the maximal rank of critical formulas will also play a role.) In other systems, such as the sequent calculus, the number of critical formulas corresponds to the number of weak quantifier inferences. Standard methods for obtaining bounds on Herbrand disjunctions ordinarily do not yield a bound only in the number of weak quantifier inferences. Consequently, the result obtained above is of independent interest. The change of input parameters is especially significant when considered in conjunction with the above remarks on formalizability. Standard methods usually only yield specific information about Herbrand disjunctions—such as their length—if a complete formal proof is available. Within the ε-calculus, we may weaken this assumption, as provability witnessed by a tautology of the form of (7) suffices. This fact was successfully employed by Kreisel in his "unwinding" of the proof of Littlewood's theorem [15].

References

[1] Ackermann, W., 'Begründung des Tertium non datur mittels der Hilbertschen Theorie der Widerspruchsfreiheit', *Mathematische Annalen* 93:1–36, 1925.

[2] Ackermann, W., 'Zur Widerspruchsfreiheit der Zahlentheorie', *Mathematische Annalen* 117:162–194, 1940.

[3] Arai, T., 'Epsilon substitution method for $ID_1(\Pi^0_1 \vee \Sigma^0_1)$', *Annals of Pure and Applied Logic* 121:163–208, 2003.

[4] Arai, T., 'Ideas in the epsilon substitution method for Π^0_1-FIX', *Annals of Pure and Applied Logic* 136:3–21, 2005.

[5] Avigad, J., 'Update procedures and the 1-consistency of arithmetic', *Mathematical Logic Quarterly* 48:3–13, 2002.

[6] Baaz, M. and A. Leitsch, 'On Skolemization and proof complexity', *Fundamenta Informaticae* 20:353–379, 1994.

[7] Bell, J. L., 'Hilbert's epsilon-operator and classical logic', *Journal of Philosophical Logic* 22:1–18, 1993.

[8] Bell, J. L., 'Hilbert's epsilon operator in intuitionistic type theories', *Mathematical Logic Quarterly* 39:323–337, 1993.

[9] Blass, A. and Y. Gurevich, 'The logic of choice', *Journal of Symbolic Logic* 65:1264–1310, 2000.

[10] Buss, S. R., 'Introduction to proof theory', in: S. R. Buss, ed., *Handbook of Proof Theory*, pp. 1–79, Elsevier, 1998.

[11] DeVidi, D., 'Intuitionistic epsilon- and tau-calculi', *Mathematical Logic Quarterly* 41:523–546, 1995.

[12] Ebbinghaus, H.-D., J. Flum, and W. Thomas, *Mathematical Logic*, Springer, 1994.

[13] Fitting, M., 'A modal logic epsilon-calculus', *Notre Dame Journal of Formal Logic* 16:1–16, 1975.

[14] Hilbert, D. and P. Bernays, *Grundlagen der Mathematik*, vol. 2, Springer, Berlin, 1939.

[15] Kreisel, G., 'Interpretation of non-finitist proofs II', *Journal of Symbolic Logic* 17:43–58, 1952.

[16] Leisenring, A., *Mathematical Logic and Hilbert's ε-symbol*, MacDonald Technical and Scientific, London, 1969.

[17] Mints, G., 'A termination proof for epsilon substitution using partial derivations', *Theoretical Computer Science* 303:187–213, 2003.

[18] Mints, G. and S. Tupailo, 'Epsilon-substitution method for the ramified language and Δ^1_1-comprehension rule', in: A. Cantini et al., eds., *Logic and Foundations of Mathematics*, pp. 107–130, Kluwer, Dordrecht, 1999.

[19] Moser, G., 'Ackermann's substitution method (remixed)', *Annals of Pure and Applied Logic* 2006, to appear.

[20] Mostowski, A., 'The Hilbert epsilon function in many-valued logics', *Acta Philos. Fenn.* 16:169–188, 1963.

[21] Orevkov, V. P., 'Lower bounds for increasing complexity of derivations after cut elimination', *Journal of Soviet Mathematics* 20:2337–2350, 1982.

[22] Statman, R., 'Lower bounds on Herbrand's theorem', *Proceedings of the American Mathematical Society* 75:104–107, 1979.

[23] Troelstra, A. and H. Schwichtenberg, *Basic Proof Theory*, Cambridge University Press, 2nd ed., 2000.

[24] von Neumann, J., 'Zur Hilbertschen Beweistheorie', *Mathematische Zeitschrift* 26:1–46, 1927.

[25] Zach, R., 'The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's program', *Synthese* 137:211–259, 2003.

[26] Zach, R., 'Hilbert's "Verunglückter Beweis", the first epsilon theorem, and consisteny proofs', *History and Philosophy of Logic* 25:79–94, 2004.

Georg Moser
Computational Logic Group
Institute of Computer Science
University of Innsbruck
A–6020 Innsbruck, Austria
georg.moser@uibk.ac.at

Richard Zach
Department of Philosophy
University of Calgary
Calgary, AB T2N 1N4, Canada
rzach@ucalgary.ca