On Herbrand's Theorem

Samuel R. Buss

University of California, San Diego
La Jolla, California 92093-0112, U.S.A.

Abstract. We firstly survey several forms of Herbrand's theorem. What is commonly called "Herbrand's theorem" in many textbooks is actually a very simple form of Herbrand's theorem which applies only to $\forall\exists$-formulas; but the original statement of Herbrand's theorem applied to arbitrary first-order formulas. We give a direct proof, based on cut-elimination, of what is essentially Herbrand's original theorem. The "no-counterexample theorems" recently used in bounded and Peano arithmetic are immediate corollaries of this form of Herbrand's theorem. Secondly, we discuss the results proved in Herbrand's 1930 dissertation.

1 Introduction

This paper discusses the famous theorem of Herbrand, which is one of the central theorems of proof-theory. The theorem called "Herbrand's theorem" in modern-day logic courses is typically only a very weak version of the theorem originally stated by Herbrand in his 1930 dissertation [8]. His 1930 dissertation contains in addition a number of other fundamental results, including, the unification algorithm, the fact that equality axioms do not help in proving equality-free sentences, a main result that is very similar to the cut-elimination theorem, and even a partial statement of the completeness theorem. The dissertation also contains a serious flaw in the proof of the main theorem, which was discovered and corrected by Dreben et al. in the 1960's, as well as earlier by Gödel in unpublished work.

This author first studied Herbrand's thesis while preparing an introductory article [2]; there we restate Herbrand's theorem in an essentially equivalent form and give a direct proof based on the cut-elimination theorem (this restatement is the same as Theorem 3 of the present paper). Since Herbrand's work contains a number of interesting constructions that are not widely known, we felt it worthwhile to prepare this paper as a survey of Herbrand's main results in chapter 5 of his dissertation.

The outline of this paper is follows: first we discuss the commonly used, weak form of Herbrand's theorem that applies only to $\forall\exists$-formulas. Then we discuss two ways of extending the theorem to general formulas: firstly, using Herbrand/Skolem functions to reexpress an arbitrary formula as a $\forall\exists$ formula, and, secondly, using a method based on "strong $\vee$-expansions" to prove a theorem

⋆ Supported in part by NSF grant DMS-9205181

which is very similar to the fundamental theorem of Herbrand. We give proofs of these results based on the cut-elimination theorem for the sequent calculus. After that, we discuss the fundamental theorem as stated by Herbrand. We also discuss the relationship of Herbrand's work to the completeness theorem and the cut-elimination theorem. Finally, we briefly discuss the error in Herbrand's proof; for a full discussion of this error and its correction, the reader should refer to the papers by Dreben et al and to Goldfarb [7] for an account of Gödel's unpublished work.

References on Herbrand's dissertation include the dissertation itself [8], the translation of its fifth chapter and the accompanying notes by Dreben and van Heijenoort [9]. Discussions of the errors in Herbrand's thesis can be found in the papers by Dreben et al. [3–5] and in Goldfarb [7]. Herbrand's collected works are contained in [10, 11]. Goldfarb [6] has further discussion of the history of Herbrand's theorem and an application to incompleteness.

All proofs in this paper are presented in terms of the sequent calculus; however for space reasons, background material and definitions for the sequent calculus are not included in this paper. A reader unfamiliar with the sequent calculus should either skip all proofs or refer to [2, 17] for definitions.

We are grateful to R. Parikh and W. Goldfarb for comments on an earlier draft of this paper.

2 The weak form of Herbrand's theorem

Herbrand's theorem is one of the fundamental theorems of mathematical logic and allows a certain type of reduction of first-order logic to propositional logic. In its simplest form it states:

Theorem 1. *Let T be a theory axiomatized by purely universal formulas. Suppose that $T \vDash (\forall \boldsymbol{x})(\exists y_1, \ldots, y_k)B(\boldsymbol{x}, \boldsymbol{y})$ with $B(\boldsymbol{x}, \boldsymbol{y})$ a quantifier-free formula. There there is a finite sequence of terms $t_{i,j} = t_{i,j}(\boldsymbol{x})$, with $1 \le i \le r$ and $1 \le j \le k$ so that*

$$T \vdash (\forall \boldsymbol{x}) \left(\bigvee_{i=1}^{r} B(\boldsymbol{x}, t_{i,1}, \ldots, t_{i,k}) \right).$$

It is well-known how to give a model-theoretic proof of Theorem 1; it is also straightforward to give a constructive, proof-theoretic proof based on the cut-elimination theorem as follows:

Proof. Since T is axiomatized by purely universal formulas, it may, without loss of generality, be axiomatized by quantifier-free formulas (obtained by removing the universal quantifiers). Let $\mathfrak{T}$ denote the set of sequents of the form $\longrightarrow A$ with A a (quantifier-free) axiom of T. Define a $LK_{\mathfrak{T}}$ proof to be a sequent calculus proof in Gentzen's system LK, except allowing sequents from $\mathfrak{T}$ in addition to the usual initial sequents.[2] Since $T \vDash (\forall \boldsymbol{x})(\exists \boldsymbol{y})B(\boldsymbol{x}, \boldsymbol{y})$, there is a $LK_{\mathfrak{T}}$-proof of the sequent $\longrightarrow (\exists \boldsymbol{y})B(\boldsymbol{a}, \boldsymbol{y})$.

[2] $LK_{\mathfrak{T}}$ may optionally contain equality axioms as initial sequents.

By the free-cut elimination theorem, there is a free-cut free $LK_{\mathfrak{T}}$-proof P of this sequent, and since the $\mathfrak{T}$-sequents contain only quantifier-free formulas, all cut formulas in P are quantifier-free. Thus, any non-quantifier-free formula in P must be of the form $(\exists y_j)\cdots(\exists y_k)B(\boldsymbol{a}, t_1, \ldots, t_{j-1}, y_j, \ldots, y_k)$ with $1 \leq j < k$. We claim that P can be modified to be a valid proof of a sequent of the form

$$\longrightarrow B(\boldsymbol{a}, t_{1,1}, \ldots, t_{1,k}), \ldots, B(\boldsymbol{a}, t_{r,1}, \ldots, t_{r,k}).$$

The general idea is to remove all $\exists$*:right* inferences in P and remove all existential quantifiers, replacing the bound variables by appropriate terms. Since there may have been contractions on existential formulas that are no longer identical after terms are substituted for variables it will also be necessary to remove contractions and add additional formulas to the sequents. To do this more formally, we know that any sequent in P is of the form $\Gamma \longrightarrow \Delta, \Delta'$ (up to order of the formulas in the sequent), where each formula in Γ and Δ is quantifier-free and where each formula in Δ' is not quantifier-free but is purely existential. We can then prove by induction on the number of lines in the free-cut free proof of $\Gamma \longrightarrow \Delta, \Delta'$ that there is an $r \geq 0$ and a cedent Δ'' of the form

$$B(\boldsymbol{a}, t_{1,1}, \ldots, t_{1,k}), \ldots, B(\boldsymbol{a}, t_{r,1}, \ldots, t_{r,k})$$

such that $\Gamma \longrightarrow \Delta, \Delta''$ is provable. We leave the rest of the details to the reader. $\square$

We define an *instance* of a universal formula $(\forall \boldsymbol{x})A(\boldsymbol{x})$ to be any quantifier-free formula $A(\boldsymbol{t})$. It is not hard to see using cut elimination, that if a quantifier-free formula C is a consequence of a universal theory T, then it is a tautological consequence of some finite set of instances of axioms of T and of equality axioms. In the special case where T is the null theory, we have that C is a consequence of instances of equality axioms (and C is therefore called a *quasitautology*). If, in addition, C does not involve equality, C will be tautologically valid. Thus, Herbrand's theorem reduces provability in first-order logic to generation of (quasi)tautologies.

The weak form of Herbrand's theorem stated above as Theorem 1 has limited applicability since it applies only to $\forall\exists$-consequences of universal theories: fortunately, however, there are several ways to extend Herbrand's theorem to more general situations. In section 3 below, we explain one such generalization; but first we give a simpler method of widening the applicability of Herbrand's theorem, based on the introduction of new function symbols, which we call *Herbrand* and *Skolem* functions, that allow quantifier alternations to be reduced.

For notational simplicity, we will consider only formulas in prenex normal form for the rest of this section; however, the definitions and theorem below can be readily generalized to arbitrary formulas.

Definition 1. *Let $(\exists x)A(x, \boldsymbol{c})$ be a formula with $\boldsymbol{c}$ all of its free variables. The* Skolem function *for $(\exists x)A$ is represented by a function symbol $f_{\exists xA}$ and has the defining axiom:*

$$\textit{Sk-def}(f_{\exists xA}): \quad (\forall \boldsymbol{y})(\forall x)\left(A(x, \boldsymbol{y}) \rightarrow A(f_{\exists xA}(\boldsymbol{y}), \boldsymbol{y})\right).$$

Note that Sk-def($f_{\exists x A}$) implies $(\forall \boldsymbol{y})\,((\exists x)A(x,\boldsymbol{y}) \leftrightarrow A(f_{\exists x A}(\boldsymbol{y}),\boldsymbol{y}))$.

Definition 2. *Let $A(\boldsymbol{c})$ be a formula in prenex form. The* Skolemization, $A^S(\boldsymbol{c})$, *of A is the formula defined inductively by:*

(1) *If $A(\boldsymbol{c})$ is quantifier-free, then $A^S(\boldsymbol{c})$ is $A(\boldsymbol{c})$.*
(2) *If $A(\boldsymbol{c})$ is $(\forall y)B(\boldsymbol{c},y)$, then $A^S(\boldsymbol{c})$ is the formula $(\forall y)B^S(\boldsymbol{c},y)$.*
(3) *If $A(\boldsymbol{c})$ is $(\exists y)B(\boldsymbol{c},y)$, then $A^S(\boldsymbol{c})$ is $B^S(\boldsymbol{c},f_A(\boldsymbol{c}))$, where f_A is the Skolem function for A.*

It is a simple, but important fact that $A^S \vDash A$.

The Skolemization *of a theory T is the theory $T^S = \{A^S : A \in T\}$. Note that T^S is a purely universal theory. Incidentally, the set of Sk-def axioms of the Skolem functions can be equivalently expressed as a set of universal formulas; however, they are not included in theory T^S. From model-theoretic considerations, it is not difficult to see that T^S contains and is conservative over T.*

We next define the concept of 'Herbrandization' which is completely dual to the notion of Skolemization:

Definition 3. *Let $(\forall x)A(x,\boldsymbol{c})$ be a formula with $\boldsymbol{c}$ all of its free variables. The* Herbrand function *for $(\forall x)A$ is represented by a function symbol $h_{\forall x A}$ and has the defining axiom:*

$$(\forall \boldsymbol{y})(\forall x)\,(\neg A(x,\boldsymbol{y}) \rightarrow \neg A(h_{\forall x A}(\boldsymbol{y}),\boldsymbol{y}))\,.$$

Note that this implies $(\forall \boldsymbol{y})\,((\forall x)A(x,\boldsymbol{y}) \leftrightarrow A(h_{\forall x A}(\boldsymbol{y}),\boldsymbol{y}))$. The Herbrand function can also be thought of as a 'counterexample function'; in that $(\forall x)A(x)$ is false if and only if $h_{\forall x A}$ provides a value x which is a counterexample to the truth of $(\forall x)A$.

Definition 4. *Let $A(\boldsymbol{c})$ be a formula in prenex form. The* Herbrandization, *$A^H(\boldsymbol{c})$, of A is the formula defined inductively by:*

(1) *If $A(\boldsymbol{c})$ is quantifier-free, then $A^H(\boldsymbol{c})$ is $A(\boldsymbol{c})$.*
(2) *If $A(\boldsymbol{c})$ is $(\exists y)B(\boldsymbol{c},y)$, then $A^H(\boldsymbol{c})$ is the formula $(\exists y)B^H(\boldsymbol{c},y)$.*
(3) *If $A(\boldsymbol{c})$ is $(\forall y)B(\boldsymbol{c},y)$, then $A^H(\boldsymbol{c})$ is $B^H(\boldsymbol{c},h_A(\boldsymbol{c}))$, where h_A is the Herbrand function for A.*

It is not hard to see that $A \vDash A^H$. Note that A^H is purely existential.

Theorem 2. *Let T be set of prenex formulas and A any prenex formula. Then the following are equivalent:*

(1) *$T \vDash A$,*
(2) *$T^S \vDash A$,*
(3) *$T \vDash A^H$,*
(4) *$T^S \vDash A^H$,*

This theorem is easily proved from the above definitions and remarks. The importance of Theorem 2 lies in the fact that T^S is a universal theory and that A^H is an existential formula, and that therefore Herbrand's theorem applies to $T^S \vDash A^H$. Thus, Theorem 2 allows Theorem 1 to be applied to an arbitrary logical implication $T \vDash A$, at the cost of converting formulas to prenex form and introducing Herbrand and Skolem functions.

3 A strong form of Herbrand's theorem

Herbrand actually proved a much more general theorem than Theorem 1 which applies directly whenever $\vDash A$, for A a general formula, not necessarily $\forall\exists$. His result also avoids the use of Skolem/Herbrand functions. The theorem we state next is quite similar in spirit and power to the theorem as stated originally by [8].

In this section, we shall consider a first-order formula A such that $\vDash A$. Without loss of generality, we shall suppose that the propositional connectives in A are restricted to be $\wedge$, $\vee$ and $\neg$, and that the $\neg$ connective appears only in front of atomic subformulas of A. (The only reason for this convention is that it avoids having to keep track of whether quantifiers appear positively and negatively in A.)

Definition 5. *Let A satisfy the above convention on negations. An* $\vee$-expansion *of A is any formula that can be obtained from A by a finite number of applications of the following operation:*

(α) *If B is a subformula of an $\vee$-expansion A' of A, replacing B in A' with $B \vee B$ produces another $\vee$-expansion of A.*

A strong $\vee$-expansion *of A is defined similarly, except that now the formula B is restricted to be a subformula with outermost connective an existential quantifier.*

Definition 6. *Let A be a formula. A* prenexification *of A is a formula obtained from A by first renaming bound variables in A so that no variable is quantified more than once in A and then using prenex operations to put the formula in prenex normal form.*

Note that there will generally be more than one prenexification of A since prenex operations may be applied in different orders resulting in a different order of the quantifiers in the prenex normal form formula.

Definition 7. *Let A be a valid first-order formula in prenex normal form, with no variable quantified twice in A. If A has $r \geq 0$ existential quantifiers, then A is of the following form with B quantifier-free:*

$$(\forall x_1 \cdots x_{n_1})(\exists y_1)(\forall x_{n_1+1} \cdots x_{n_2})(\exists y_2) \cdots (\exists y_r)(\forall x_{n_r+1} \cdots x_{n_{r+1}})B(\boldsymbol{x}, \boldsymbol{y})$$

with $0 \leq n_1 \leq n_2 \leq \cdots \leq n_{r+1}$. A witnessing substitution *for A is a sequence of terms (actually, semiterms) $t_1, \ldots t_r$ such that (1) each t_i contains arbitrary free variables but only bound variables from $x_1, \ldots, x_{n_i}$ and (2) the* *formula $B(\boldsymbol{x}, t_1, \ldots, t_r)$ is a quasitautology (i.e., a tautological consequence of instances of equality axioms only). In the case where B does not contain the equality sign, then (2) is equivalent to B being a tautology.*

Let T be a first-order theory. A sequence of terms is said to witness A over T *if the above conditions hold except with condition (2) replaced by the weaker condition that $T \vDash (\forall \boldsymbol{x})B(x, \boldsymbol{t})$.*

Definition 8. *A* Herbrand proof *of a first-order formula A consists of a prenexification A^* of a strong $\vee$-expansion of A plus a witnessing substitution σ for A^*.*

A Herbrand T-proof *of A consists of a prenexification A^* of a strong $\vee$-expansion of A plus a substitution which witnesses A over T.*

We are now in a position to state the general form of Herbrand's theorem:

Theorem 3. *A first-order formula A is valid if and only if A has a Herbrand proof. More generally, if T is a universal theory, then $T \vDash A$ if and only if A has a Herbrand T-proof.*

Proof. We shall sketch a proof of only the first part of the theorem since the proof of the second part is almost identical. Of course it is immediate from the definitions that if A has a Herbrand proof, then A is valid. So suppose A is valid, and therefore has a cut-free LK-proof P. We shall modify P in stages so as to extract a Herbrand proof of P.

The first stage will involve restricting the formulas which can be combined by a contraction inference. In order to properly keep track of contractions of formulas in a sequent calculus proof, we must be careful to formulate inference rules with two hypotheses in a "multiplicative" fashion so as to avoid the problem of having implicit contractions on side formulas in inferences with two hypothesis such as $\vee$*:left* and $\wedge$*:right*. For example, we want to formulate the $\vee$*:left* inference rule in the multiplicative form

$$\frac{A, \Gamma \longrightarrow \Delta \qquad B, \Gamma' \longrightarrow \Delta'}{A \vee B, \Gamma, \Gamma' \longrightarrow \Delta, \Delta'}$$

rather than in the "additive" form

$$\frac{A, \Gamma \longrightarrow \Delta \qquad B, \Gamma \longrightarrow \Delta}{A \vee B, \Gamma \longrightarrow \Delta}$$

since the additive form contains implicit contractions on side formulas in Γ and Δ, whereas the multiplicative formulation does not contain implicit contractions. We also use analogous multiplicative formulations of the $\wedge$*:right* and cut rule. Of course, using multiplicative formulations rules instead of additive formulation does not change the strength of the sequent calculus, since either form may be derived from the other with the use of weak structural inferences. Furthermore, the cut-elimination and free-cut elimination theorems hold with either formulation. We therefore henceforth use the multiplicative formulation of the rules of inference for the sequent calculus.

A contraction inference is said to be a *propositional contraction* (resp., an $\exists$-*contraction* provided that the principal formula of the contraction is quantifier-free (resp., its outermost connective is an existential quantifier). The first step in modifying P is to form a cut-free proof P_1, also with endsequent $\longrightarrow A$ such that all contraction inferences in P_1 are propositional or $\exists$-contractions. The construction of P_1 from P is done by a "contraction-elimination" procedure. For this purpose, we define the *E-depth* of a formula by letting the E-depth of a quantifier-free formula or a formula which begins with an existential quantifier be equal to zero, and defining the E-depth of other formulas inductively by letting the E-depth of $\neg\varphi$ equal the E-depth of φ plus one and by letting the E-depths of $\varphi \vee \psi$ and $\varphi \wedge \psi$ equal one plus the maximum of the E-depths of φ and ψ. Then we prove, by double induction on the maximum E-depth d of contraction formulas and the number of contractions of formulas of this maximum E-depth, that P_1 can be transformed into a proof in which all contractions are on formulas of E-depth zero. The induction step consists of removing a topmost contraction inference of the maximum E-depth d. For example, suppose that the following inference is a topmost contraction with principal formula of E-depth d;

$$\dfrac{\begin{array}{c}\ddots\vdots\cdot^{\cdot} R\\ \Gamma \longrightarrow \Delta, (\forall x)B, (\forall x)B\end{array}}{\Gamma \longrightarrow \Delta, (\forall x)B}$$

Since P_1 is w.l.o.g. in free variable normal form and since this is a topmost contraction of E-depth d, we can modify the subproof R of P_1 by removing at most two $\forall$*:right* inferences and/or changing some *Weakening:right* inferences to get a proof of $\Gamma \longrightarrow \Delta, B(a), B(a')$, where a and a' are free variables not appearing in the endsequent of R. Further replacing a' everywhere by a gives a proof of $\Gamma \longrightarrow \Delta, B(a), B(a)$: we use this get to a proof ending:

$$\dfrac{\dfrac{\begin{array}{c}\ddots\vdots\cdot^{\cdot}\\ \Gamma \longrightarrow \Delta, B(a), B(a)\end{array}}{\Gamma \longrightarrow \Delta, B(a)}}{\Gamma \longrightarrow \Delta, (\forall x)B}$$

Thus we have reduced the E-depth of the contraction inference. A similar procedure works for contractions of E-depth d with outermost connective a propositional connective—we leave the details to the reader. Note that the construction of P_1 depends on the fact that propositional inferences and $\forall$*:right* inferences can be pushed downward in the proof. It is not generally possible to push $\exists$*:right* inferences downward in a proof without violating eigenvariable conditions.

The second step in modifying P is to convert P_1 into a cut-free proof P_2 of some strong $\vee$-expansion A' of A such that every contraction in P_2 is propositional. This is done by the simple expedient of replacing every $\exists$-contraction in P_1 with an $\vee$*:right* inference, and then making the induced changes to all descendents of the principal formula of the inference. More precisely, starting with a lowermost $\exists$-contraction in P_1, say

$$\frac{\Gamma \longrightarrow \Delta, (\exists x)B, (\exists x)B}{\Gamma \longrightarrow \Delta, (\exists x)B}$$

replace this with an $\vee$*:left* inference

$$\frac{\Gamma \longrightarrow \Delta, (\exists x)B, (\exists x)B}{\Gamma \longrightarrow \Delta, (\exists x)B \vee (\exists x)B}$$

and then, in order to get a syntactically correct proof, replace, as necessary, subformulas $(\exists x)B'$ of formulas in P with $(\exists x)B' \vee (\exists x)B'$ (we use the notation B' since terms in B may be different in its descendents). Iterating this process yields the desired proof P_2 of a strong $\vee$-expansion A' of A. By renaming bound variables in P_2 we can assume w.l.o.g. that no variable is quantified twice in any single sequent in P_2.

Thirdly, from P_2 we can construct a prenexification A^* of A' together with a witnessing substitution, thereby obtaining a Herbrand proof of A. To do this, we iterate the following procedure for pulling quantifiers to the front of the proved formula. Find any lowest quantifier inference in P_2 which has not already been handled: this quantifier inference corresponds to a unique quantifier, (Qx), appearing in the endsequent of the proof (and conversely, each quantifier in the endsequent of the proof corresponds to a unique quantifier inference, since all contraction formulas are quantifier-free). Use prenex operations to pull (Qx) as far to the front of the endsequent formula as possible (but not past the quantifiers that have already been moved to the front of the endsequent formula). Also, push the quantifier inference downward in the proof until it reaches the group of quantifier inferences that have already been pushed downward in the proof. It is straightforward to check that this procedure preserves the property of having a syntactically valid proof. When we are done iterating this procedure, we obtain a proof P_3 of a prenexification $\longrightarrow A^*$ of A. It remains to define a witnessing substitution for A^*: this is now easy, for each existential quantifier $(\exists y_i)$ in A^*, find the corresponding $\exists$*:right* inference

$$\frac{\Gamma \longrightarrow \Delta, B(t_i)}{\Gamma \longrightarrow \Delta, (\exists y_i)B(y_i)}$$

and let the term t_i be from this inference. That this is a witnessing substitution for A^* is easily proved by noting that by removing the $\exists$*:right* inference from P_3, a proof of $A^*_M(\boldsymbol{x}, \boldsymbol{t})$ is obtained where A^*_M is the quantifier-free portion of A^*. □

The above theorem can be used to obtain the following 'no-counterexample interpretation' which has been very useful recently in the study of bounded arithmetic (see [12, 1, 18]).[3]

Corollary 1. *Let T be a universal theory and suppose $T \vDash (\exists x)(\forall y)A(x, y, \boldsymbol{c})$ with A a quantifier-free formula. There is a $k > 0$ and terms $t_1(\boldsymbol{c})$, $t_2(\boldsymbol{c}, y_1)$,*

[3] This corollary is named after the more sophisticated no-counterexample interpretations of [13, 14].

$t_3(\boldsymbol{c}, y_1, y_2), \ldots, t_k(\boldsymbol{c}, y_1, \ldots y_{k-1})$ *such that*

$$
\begin{aligned}
T \vDash & (\forall y_1)[A(t_1(\boldsymbol{c}), y_1, \boldsymbol{c}) \\
& \vee(\forall y_2)[A(t_2(\boldsymbol{c}, y_1), y_2, \boldsymbol{c}) \\
& \vee(\forall y_3)[A(t_3(\boldsymbol{c}, y_1, y_2), y_3, \boldsymbol{c}) \\
& \vee \cdots \vee (\forall y_k)[A(t_k(\boldsymbol{c}, y_1, \ldots, y_{k-1}), y_k, \boldsymbol{c}))] \cdots]]]
\end{aligned}
$$

To prove the corollary, note that the only strong $\vee$-expansions of A are formulas of the form $\bigvee(\exists x)(\forall y)A(x, y, \boldsymbol{c})$ and apply the previous theorem.

4 No recursive bounds on number of terms

It is interesting to ask whether it is possible to bound the value of r in Theorem 1. For this, consider the special case where the theory T is empty, so that we have an LK-proof P of $(\exists x_1, \ldots, x_k)B(\boldsymbol{a}, \boldsymbol{x})$ where B is quantifier-free. There are two ways in which one might wish to bound the number r needed for Herbrand's theorem: as a function of the size of P, or alternatively, as a function of the size of the formula $(\exists \boldsymbol{x})B$. For the first approach, it follows immediately from the proof of the cut-elimination theorem in [2] and the proof of Herbrand's theorem, that $r \leq 2^{||P||}_{2||P||}$, where 2^x_i is defined inductively by $2^x_0 = x$ and $2^x_{i+1} = 2^{2^x_i}$ and where $||P||$ equals the number of strong inferences in P. For the second approach, we shall sketch a proof below that r can not be recursively bounded as a function of the formula $(\exists \boldsymbol{x})B$. The proof is based on the unification algorithm contained in Herbrand [9, para. 2.4]

To show that r cannot be recursively bounded as a function of $(\exists \boldsymbol{x})B$, we shall prove that having a recursive bound on r would give a decision procedure for determining if a given existential formula is valid. Since it is well known that validity of existential first-order formulas is undecidable; this implies that r cannot be recursively bounded in terms of the formula size.

What we shall show is that, given a formula B as in Theorem 1 and given an $r > 0$, it is decidable whether there are terms $t_{1,1}, \ldots, t_{r,k}$ which make the formula

$$
\bigvee_{i=1}^{r} B(\boldsymbol{a}, t_{i,1}, \ldots, t_{i,k}) \tag{1}
$$

a tautology. (This fact was first proved by Herbrand by the same argument that we sketch here.) This will suffice to show that r cannot be recursively bounded. The quantifier-free formula B is expressible as a Boolean combination $C(D_1, \ldots, D_\ell)$ where each D_j is an atomic formula and $C(\cdots)$ is a propositional formula. If the formula (1) is a tautology, it is by virtue of certain formulas $D_j(\boldsymbol{a}, t_{i,1}, \ldots, t_{i,k})$ being identical. That is to say there is a finite set X of equalities of the form

$$
D_j(\boldsymbol{a}, t_{i,1}, \ldots, t_{i,k}) \;=\; D_{j'}(\boldsymbol{a}, t_{i',1}, \ldots, t_{i',k})
$$

such that, any set of terms $t_{1,1}, \ldots, t_{r,k}$ which makes all the equalities in X true will make (1) a tautology.

But now the question of whether there exist terms $t_{1,1}, \ldots, t_{r,k}$ which satisfy such a finite set X of equations is easily seen to be a first-order unification problem. The algorithm for solving first-order unification problems is given in Herbrand's thesis and is now-a-days well-known; Robinson [16] gives a method of getting a most general solution, and Paterson-Wegman [15] give a linear-time algorithm for unification. This algorithm either determines that no choice of terms will satisfy all the equations in X or will find a (most general) set of terms that satisfy the equations of X.

Since, for a fixed $r > 0$, there are only finitely many possible sets X of equalities, we have the following algorithm for determining if there are terms which make (1) a tautology: for each possible set X of equalities, check if it has a solution (i.e., a most general unifier), and if so, check if the equalities are sufficient to make (1) a tautology. $\square$

5 The actual theorem of Herbrand

In this final section, we discuss the results contained in chapter 5 of Herbrand's Ph.D. thesis. The fundamental theorem of this chapter is very similar to Theorem 3 but differs in some details. We also describe the two proof systems, now called Q_H and Q'_H, that Herbrand used. The results stated by Herbrand include a version of the cut-elimination theorem and his proof methods give (or nearly give) a version of the completeness theorem. There was also a fairly serious error in Herbrand's proof, which was first described in published material by Dreben et al.; this error was apparently also recognized by Bernays in the 1930's and was discovered and corrected by Gödel in unpublished notes (see [7]). These errors in no way detract from the importance of Herbrand's work, since alternative proofs could be given. In any event, although there are some false lemmas in Herbrand's work, his main theorems are all fully correct.

5.1 Herbrand's fundamental theorem.

Herbrand's fundamental theorem applied to arbitrary first-order formulas A; in particular, A need not be in prenex normal form. By renaming variables, one can assume that no variable is quantified more than once in A and that no variable occurs both free and bound in A. Herbrand took prenex operations as fundamental in his proof theory (see the definitions of Q_H and Q'_H below). His formal system allowed prenex operations to be applied not only in a 'forward' direction which brings quantifiers to the front of a formula, but also in a 'reverse' direction pushing quantifiers further inward in a formula. Herbrand noted that for every formula A there is a unique formula, called the *canonical form* of A, which is obtained by applying prenex operations to subformulas of A to push quantifiers as far inward as possible. Let M be obtained from A by erasing all quantifiers from A. Since we use only connectives $\neg$, $\vee$ and $\wedge$ (Herbrand used only the first two) and because of our conventions on not reusing variables, it is easy to see that any prenex formula obtained from A by prenex operations consists of M preceded by a string of quantifiers. Thus all prenexifications of A differ only in the order of the quantifiers.

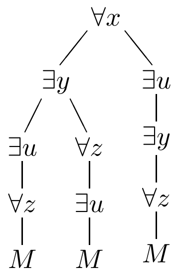
We now describe a *tree expansion of* A to consist of a finite set (also called a forest) of labeled trees: each tree has its leaves labeled with the formula M and has its internal nodes labeled with quantifiers $(\exists x)$ or $(\forall x)$ which already appear in A. Furthermore, the following properties should hold:

1. For any simple path from a root of a tree to a leaf, if the labels on the path are concatenated, then one obtains a formula which is equivalent to A and is obtainable from A by prenex operations only.
2. The trees are finite in that each internal node has only finitely many children.
3. If a node has more than one child, then none of its children are labeled with universal quantifiers.[4]

To given an example, consider a formula of the form

$$(\forall x)[(\exists y)(\forall z)A(x,y,z) \vee (\exists u)B(x,u)].$$

One possible set of trees associated to this formula is:



where M is $A(x,y,z) \vee B(x,u)$. Herbrand used a tabular notation to represent this situation; namely, for this example, he would write

$$\begin{array}{cccc} +x & -u & -y & +z \\ +x & -y & -u & +z \\ +x & -y & +z & -u \end{array}$$

using $+x$ to mean $(\forall x)$ and $-u$ to mean $(\exists y)$, etc. Note that each line in the table corresponds to a path in the tree. To make the tree structure clearer, Herbrand then rewrites the table above as:

$$+x \left\{ \begin{array}{l} -u \quad -y \quad +z \\ \\ -y \left\{ \begin{array}{ll} -u & +z \\ +z & -u \end{array} \right. \end{array} \right.$$

[4] This optional condition is given in Herbrand's paragraph 2.33 of chapter 5 of his thesis. It is the analogue of our use of *strong* $\vee$-expansions in place of $\vee$-expansions.

The concept of a *proposition derived from A* is defined as follows: for each node in the tree assign a formula as follows: assign the matrix M to every leaf node, and assign to an internal node α labels with (Qv) the formula

$$(Qv)[P_1 \vee ... \vee P_n]$$

where $P_1, \ldots, P_m$ are the formulas assigned to the n children of α. Finally, take the disjunction of the formulas assigned to all the roots of trees in the forest, then rename variables so that no variable is used twice in this disjunction and form an arbitrary prenexification of this disjunction; the result is called a *proposition derived from A*.

It is clear that a proposition derived from A is equivalent to A, since it is obtained by using only the following types of operations: (a) prenex operations, (b) variable renamings, and (c) replacing subformulas Z with $Z \vee Z$ (i.e., $\vee$-expansion steps). Herbrand's fundamental theorem can now be stated as follows (the theory Q_H is described below; since it is sound and complete, A is Q_H-provable iff A is valid):

Theorem 4. *A is provable in the theory Q_H iff there is a proposition derived from A which has a witnessing substitution.*

5.2. Herbrand's proof systems

Herbrand's thesis primarily used a proof system which we shall denote Q_H; he also used a modified version, Q'_H, and his fundamental theorem states that provability in Q_H is equivalent to provability in Q'_H. Formulas in these proof systems involve the logical connectives $\neg$, $\vee$, $\forall$ and $\exists$; other symbols, such as $\rightarrow$ are abbreviations for more complex formulas. It is not permitted for a variable to be quantified twice in a formula, or to appear both free and bound in a formula. The system Q_H has all tautologies as axioms and has the following rules of inference:

1. Modus Ponens; from A and $A \rightarrow B$, infer B.
2. Rule of Simplification: If Z' is an alphabetic variant of Z, then Z may be inferred from $Z \vee Z'$.
3. Universal Generalization: from Φ, infer $(\forall x)\Phi$.
4. Existential Instantiation: from $\Phi(t)$, infer $(\exists x)\Phi(x)$.
5. The Rules of Passage: consider the following six pairs of logically equivalent formulas:

$$\begin{aligned}
\neg\forall x\Phi &\Leftrightarrow \exists x(\neg\Phi)\\
\neg\exists x\Phi &\Leftrightarrow \forall x(\neg\Phi)\\
(\forall x\Phi) \vee Z &\Leftrightarrow \forall x(\Phi \vee Z)\\
Z \vee (\forall x\Phi) &\Leftrightarrow \forall x(Z \vee \Phi)\\
(\exists x\Phi) \vee Z &\Leftrightarrow \exists x(\Phi \vee Z)\\
Z \vee (\exists x\Phi) &\Leftrightarrow \exists x(Z \vee \Phi)
\end{aligned}$$

There are twelve *rules of passage*; these allow a formula B to be inferred from the formula A provided B is obtained from A by replacing an occurrence of a subformula in A which is in one of the above twelve forms with the equivalent subformula given in the above table. (Note that the conventions on variable usage imply that x does not appear in Z.)

Herbrand's second proof system, which we call Q'_H, is obtained from Q_H by disallowing the rule of modus ponens, and replacing the rule of simplification by the *generalized rule of simplification* which permits B to be inferred from A when B is obtained from A by replacing a subformula of the form $Z \vee Z'$ with the subformula Z, provided Z' is an alphabetic variant of Z.

A corollary of Herbrand's fundamental theorem is the statement that a formula is Q_H-provable if and only if it is Q'_H-provable. This is a very intriguing fact, since it is evident that Q'_H is very similar to a cut-free sequent calculus proof system; in particular, there is an analogue of the subformula property of the sequent calculus which holds for Q'_H; namely, if one measures the complexity of formula in terms of the depth of quantifier nesting in the canonical form of a formula, then it is evident that all the formulas which appear in a Q'_H-proof of a formula A have complexity no greater than the complexity of A. Gentzen's paper on LK and cut-elimination appeared only four years later in 1934. However, we are reluctant to ascribe much of the credit for the cut-elimination theorem to Herbrand for two reasons: firstly, Q'_H does not have the elegance of the sequent calculus LK, and secondly, the errors in Herbrand's proof impinge directly on the proof of the equivalence of Q_H and Q'_H.

Indeed, it is precisely at the step of "elimination of modus ponens", which is the analogue of cut-elimination, that the errors in Herbrand's proof occur (see paragraph 5.3, lemma 3, chapter 5 of Herbrand's thesis). It is well-known that the process of cut-elimination in first-order logic leads to superexponential growth rates; however, in his erroneous proof, Herbrand claimed that much lower growth rates sufficed. The corrected versions of Herbrand's proof, given by Gödel (see [7]) and by Dreben et al. [3–5] do give superexponential growth rates that are similar to the growth rates known to hold for the cut-elimination theorem; and these growth rates are (nearly) optimal.

5.3. The completeness theorem.

Herbrand's thesis also includes a construction that is very close to the completeness theorem. (Recall that the completeness theorem was first proved by Gödel in 1930, in the same year that Herbrand's thesis was completed.) In his thesis, Herbrand discusses that fact that if there is no witnessing substitution for a proposition derived from A (as in Theorem 4), then it is possible to construct an sequence of finite domains where appropriate translations of A are false. Herbrand also discusses the possibility of having an infinite domain where A would be false in the usual sense; had he actually done this, he would have proved the completeness theorem. Somewhat surprisingly, Herbrand evidently knew that such an infinite domain could be obtained, but because of his constructive outlook, he declined to carry out the proof that such an infinite domain existed. Indeed he says "but only a 'principle of choice' could lead us to take a fixed system of values in an infinite domain."[5]

By this he means that it would be necessary to use the axiom of choice to obtain an infinite model in which A is false under the usual Tarskian semantics.

It is interesting to speculate why Herbrand chose not to state the completeness theorem. Firstly, Herbrand took a very strong constructive, formalist point of view, and he would have rejected non-constructive arguments on philosophical grounds. Indeed, Herbrand defined "true" to mean "provable in Q_H" rather than "true in all possible structures". Secondly, it seems that Herbrand felt that his fundamental theorem was of greater interest than a model-theoretic completeness theorem.

The issue of the completeness theorem has also some bearing on the status of the errors in Herbrand's thesis. The errors in his proof affected only the proof-theoretic results, and the completeness theorem, which Herbrand *could* have stated and proved, would not have been affected by these errors. Therefore, Herbrand could have obtained a alternative and correct proof of his fundamental theorem by using the following argument: suppose A is a formula and there is no proposition derived from A which has a witnessing substitution; then by the completeness theorem, there is an infinite domain (i.e., structure) where A is false; therefore, since the proof system Q_H is sound, there is no Q_H-proof of A. This argument proves the contrapositive of Theorem 4 and is thereby an error-free proof of Herbrand's fundamental theorem. Of course, this proof uses non-constructive methods and presumably would not have been attractive to Herbrand.

References

1. S. R. Buss, *Relating the bounded arithmetic and polynomial-time hierarchies*, Annals of Pure and Applied Logic, 75 (1995), pp. 67–77.
2. ——, *An introduction to proof theory*, in Handbook of Proof Theory, S. R. Buss, ed., North-Holland, 1998, pp. 1–78.
3. B. Dreben and S. Aanderaa, *Herbrand analyzing functions*, Bulletin of the American Mathematical Society, 70 (1964), pp. 697–698.
4. B. Dreben, P. Andrews, and S. Aanderaa, *False lemmas in Herbrand*, Bulletin of the American Mathematical Society, 69 (1963), pp. 699–706.
5. B. Dreben and J. Denton, *A supplement to Herbrand*, Journal of Symbolic Logic, 31 (1966), pp. 393–398.
6. W. D. Goldfarb, *Herbrand's theorem and the incompleteness of arithmetic*, Iyyun, A Jerusalem Philosophical Quarterly, 39 (1990), pp. 45–64.
7. ——, *Herbrand's error and Gödel's correction*, Modern Logic, 3 (1993), pp. 103–118.
8. J. Herbrand, *Recherches sur la théorie de la démonstration*, PhD thesis, University of Paris, 1930.

[5] Herbrand [9, p.552]

9. ———, *Investigations in proof theory: The properties of true propositions*, in From Frege to Gödel: A Source Book in Mathematical Logic, 1978-1931, J. van Heijenoort, ed., Harvard University Press, Cambridge, Massachusetts, 1967, pp. 525–581. Translation of chapter 5 of [8], with commentary and notes, by J. van Heijenoort and B. Dreben.
10. ———, *Écrits logique*, Presses Universitaires de France, Paris, 1968. Ed. by J. van Heijenoort.
11. ———, *Logical Writings*, D. Reidel, Dordrecht-Holland, 1971. Ed. by W. Goldfarb, Translation of [10].
12. J. Krajíček, P. Pudlák, and G. Takeuti, *Bounded arithmetic and the polynomial hierarchy*, Annals of Pure and Applied Logic, 52 (1991), pp. 143–153.
13. G. Kreisel, *On the interpretation of non-finitist proofs–part I*, Journal of Symbolic Logic, 16 (1951), pp. 241–267.
14. ———, *On the interpretation of non-finitist proofs, part II. interpretation of number theory, applications*, Journal of Symbolic Logic, 17 (1952), pp. 43–58.
15. M. S. Paterson and M. N. Wegman, *Linear unification*, J. Comput. System Sci., 16 (1978), pp. 158–167.
16. J. A. Robinson, *A machine-oriented logic based on the resolution principle*, J. Assoc. Comput. Mach., 12 (1965), pp. 23–41.
17. G. Takeuti, *Proof Theory*, North-Holland, Amsterdam, 2nd ed., 1987.
18. D. Zambella, *Notes on polynomially bounded arithmetic*, Journal of Symbolic Logic, 61 (1996), pp. 942–966.