VICTORIA UNIVERSITY OF WELLINGTON

Department of Computer Science

PO Box 600
Wellington
New Zealand

Tel: +64 4 471 5328
Fax: +64 4 495 5232
Internet: Tech.Reports@comp.vuw.ac.nz

Proof Normalization I: Gentzen's Hauptsatz

Amos R. Omondi

Technical Report CS-TR-93/5
February 1993

Abstract

This paper is the first of a two-part introductory discussion of Gentzen's Hauptsatz, Prawitz's Normalization Theorem, and related results, which are concerned with the elimination of cuts in the calculi of sequents and the calculi of natural deduction. Cuts correspond to detours in proofs and hence there is undoubtedly some philosophical interest in being able to eliminate them. These results also have mathematical interest and applications that are of use in computation and this is the underlying motivation for this paper.

Author Information

Amos Omondi teaches Computer Science at Victoria University

This report has been produced from a set of notes that I wrote some years ago when I took a course in Computational Logic from Prof. Gopalan Nadathur at Duke University. As several other students in a similar situation appear to have found the notes useful, I have decided to reproduce them in this, more accessible, form. I am indebted to Prof. Nadathur for his comments on the first version.

```
  "Talking of Herbert Spencer", he began, "do you really find
no logical difficulty in regarding Nature as a process of
involution, passing from definite coherent homogeneity to
indefinite incoherent heterogeneity?"
.......
  "No physical difficulty," she confidently replied:  "but I
haven't studied Logic much.  Would you state the difficulty?"
  "Well," said Arthur, "do you accept it as self-evident?  Is
it as obvious, for instance, as that 'things that are greater
than the same are greater than one another'?"
  "To my mind," she modestly replied, "it seems quite
obvious.  I grasp both truths by intuition.  But other minds
may need some logical --- I forget the technical terms."
  "For a complete logical argument," Arthur began with
admirable solemnity, "we need two prim Misses ----"
  "Of course!", she interrupted.  I remember that word now.
And they produce ----?"
  "A Delusion," said Arthur.
  "Ye--es?" she said dubiously.  "I don't seem to remember
that so well.  But what is the whole argument called?"
  "A Sillygism."
```

Queer Street, Number Forty. Lewis Carroll.

1 Introduction

This paper is the first of a two-part introductory discussion of Gentzen's Hauptsatz, Prawitz's Normalization Theorem, and related results, which are concerned with the elimination of cuts in the calculi of sequents and the calculi of natural deduction. Cuts correspond to detours [1] in proofs and hence there is undoubtedly some philosophical interest in being able to eliminate them. These results also have mathematical interest and applications that are of use in computation and this is the underlying motivation for this paper. It is useful, at this point, to make a distinction between *normal form* theorems and *normalization* theorems[2][17,43]. The former are results which state a proof not in normal form [3] can be reduced to one in normal form; the latter, in addition to stating this also give an effective procedure for the reduction. Thus, for example, the Hauptsatz as stated in [11] is a normal form theorem, whereas the "official" version (given below) is a normalization theorem; from a computational point of view, the interest in the latter is obvious. A distinction may also be made here between *weak* normalization – some reduction sequence leads to normal form – and *strong* normalization – all reduction sequences lead to normal form.

Some reasons for the interest in these results are:

- *Connection between computation and deduction.* Proof normalization, specifically, certain corollaries[4], is, in a certain sense, the very essence of proof theory (see Girard's [17] discussion on "purity of methods"). Reduction, as in the λ-calculus or term-rewriting is fundamental in computation. Hence a correspondence between the two, which one obtains, for example, via the Curry-Howard isomorphism [21], yields a deep connection between deduction and computation. From this, methods used the one area are frequently useful in the other; [7] is a recent example.

- *Extraction of programs (algorithms) or bounds from proofs* (of $\prod_2^0$ statments). A statement of the form $\forall x \exists y : R(x, y)$ may viewd as a program specification when rendered thus: for every input x, there is some output y such that the condition R holds. From a proof of such a statement one can, by normalization, obtain a recursive function F such that $R(x, F(x))$ (and, more generally, $\forall x \exists y \leq f(x) : R(x, y)$) holds. Alternatively, instead of extracting F from the proof, the proof may itself be "executed" – again, by normalization, to directly yield $F(x)$. This is obviously related to the preceding remark.

- *The formulation of decision procedures for certain deviant logics.* In many of these cases – consider, for example, propositional intuitionistic and modal logics – the usual methods from classical logic are of little help, whereas results such as the Hauptsatz are quite useful.

[1]For example, to prove that $1 + 2 = 2 + 1$, one might do it directly by reduction to $3 = 3$ or first make a detour to prove $\forall x \forall y : x + y = y + x$.

[2]Throughout this paper "normalization" will refer to this classification and "Normalization" to Prawitz's theorem

[3]For the Haputsatz, this means *cut-free*.

[4]Such as the Subformula Property, which is discussed below.

- *Consistency proofs.* Typically, consistency trivially follows from a full cut elimination theorem and the proofs are elementary (finitary) or almost [5] elementary. Because of this, a careful study of such a proof allows *the extraction of precise mathematical information from philosophically shaky results*: As a consequence of Gödel Incompleteness Theorems, it is clear that these proofs, for non-trivial mathematical theories, stand on shaky ground (see [44], for example, for a discussion). Nevertheless, they do have mathematical significance which one obtains most easily from the application of cut-elimination; this, of course, implies that they are not totally devoid of philosophical value.

- *The derivation of other important results.* For example, Herbrand's Theorem (which is of great importance in mechanical theorem proving, Kreisel's No Counterexample Interpretation, as well as other standard results of logic (as obtained via model theory).

In reading the published literature, I have especially relied on the writings of Kreisel and Girard, and if the reader finds any obscure remarks, most can probably be clarified by recourse to these sources.

2 Normalization in the sequent calculi LK and LJ

In what follows uppercase Greek letters denote sequences of formulae of the usual sort of first order language and uppercase Roman letters denote formulae of the language, except in discussions of logical complexity where the latter will have the usual meaning.

Definition 1 A *sequent* is an expression $\Gamma \vdash \Delta$ (read as " Γ yields Δ") where Γ and Δ are finite sequences of formulae of a first order language. Γ is the *antecedent* and Δ is the *succedent.*

The informal understanding is that $A_1, ..., A_m \vdash B_1, ..., B_n$ in the sequent calculus corresponds to $A_1 \wedge ... \wedge A_m \rightarrow B_1 \vee ... \vee B_n$ in the usual calculi for first order logic. In LK (the calculus for classical logic) any number of formulae is allowed in the succedent; in LJ (the calculus for intuitionistic logic) no more than one formula is allowed. Since most of the results below hold for both LK and LJ, no distinction will be made, unless necessary, and a result will simply be stated in terms of LK.

2.1 Rules of inference

We shall refer to rules as *right* rules or *left* rules according to whether they operate on the succedent or antecedent, respectively, of a sequent.

Axioms

[5] "Almost" because they go (slightly) beyond Hilbert's strictly finitist standpoint: *the real mathematical world is* Π^0_1.

$$A \vdash A$$

Structural rules

Thinning

$$\frac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta} \qquad \frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A}$$

Contraction

$$\frac{A, A, \Gamma \vdash \Delta}{A, \Gamma \vdash \Delta} \qquad \frac{\Gamma \vdash \Delta, A, A}{\Gamma \vdash \Delta, A}$$

Exchange

$$\frac{\Gamma, A, B, \Pi \vdash \Delta}{\Gamma, B, A, \Pi \vdash \Delta} \qquad \frac{\Gamma \vdash \Delta, A, B, \Lambda}{\Gamma \vdash \Delta, B, A, \Lambda}$$

Cut rule

$$\frac{\Gamma \vdash \Delta, A \qquad A, \Pi \vdash \Lambda}{\Gamma, \Pi \vdash \Delta, \Lambda} : Cut$$

The resolution rule, that is so central in mechanical theorem proving and logic programming, is easily seen to be a form of the cut rule. For example, the resolution deduction of $P \vee Q$ from $P \vee \neg S$ and $Q \vee S$ is the application of the cut rule to the sequents $\vdash Q, S$ and $S \vdash P$ to obtain the sequent $\vdash P, Q$. So we have the following dichotomy: in logic one tries to eliminate cuts wherever possible, in many computational applications (and practical mathematics) one uses cuts wherever possible; this is discussed further below, although we should hasten to add that the real situation is not so clear cut as this.[6]

Logical rules

Negation

$$\frac{\Gamma \vdash \Delta, A}{\neg A, \Gamma \vdash \Delta} \qquad \frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A}$$

[6] These remarks are only intended to convey a general picture. A precise discussion would be quite lengthy and not necessarily more illuminating at this point. A detailed discussion of the relationship between cuts and resolution will be found in [2].

Conjunction

$$\frac{A, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta} \qquad\qquad \frac{B, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta}$$

$$\frac{\Gamma \vdash \Delta, A \qquad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B}$$

Disjunction

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

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

Implication

$$\frac{\Gamma \vdash \Delta, A \qquad B, \Pi \vdash \Lambda}{A \to B, \Gamma, \Pi \vdash \Delta, \Lambda}$$

$$\frac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \to B}$$

Universal quantification

$$\frac{A(t), \Gamma \vdash \Delta}{\forall x A(x), \Gamma \vdash \Delta} \qquad\qquad \frac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)}$$

where t is an arbitrary term and a (the *eigenvariable*) does not occur in the lower sequent. This will be referred to as the *eigenvariable condition*.

Existential quantification

$$\frac{A(a), \Gamma \vdash \Delta}{\exists x A(x), \Gamma \vdash \Delta} \qquad\qquad \frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, \exists x A(x)}$$

where t is an arbitrary term and a (the *eigenvariable*) does not occur in the lower sequent.

Remark For the simple purposes of providing a proof system for first order logic, some economy and simplification of rules in the system above is possible but this is likely to lead to complications in proving the Hauptsatz [12].

Definition 2 An *LK-proof*, $\mathcal{P}$, of a sequent $\Gamma \vdash \Delta$, is a tree of sequents satisfying:

- All the leaf sequents of $\mathcal{P}$ are axioms.
- Every sequent except $\Gamma \vdash \Delta$ is an upper sequent of an inference whose lower sequent is also in $\mathcal{P}$.
- $\Gamma \vdash \Delta$ is the lowest sequent in $\mathcal{P}$.

We shall use $\mathcal{P}(a)$ to indicate the fact that the variable a occurs in the proof $\mathcal{P}$ and $\mathcal{P}[b/a]$ to indicate the proof $\mathcal{P}$ with every occurrence of a replaced by b.

Remark Evidently, if a sequent is LJ-provable, then it is also LK-provable.

The following simple result will be of great utility in what follows:

Proposition 1 *If a sequent has an LK-proof, then it has an LK-proof in which*

- *All eigenvariables are distinct from one another.*
- *If a free variable* a *occurs as an eigenvariable of some sequent,* $\Gamma \vdash \Delta$, *then* a *occurs only in sequents above* $\Gamma \vdash \Delta$.

Proof Obvious from a suitable renaming of variables.

Such a proof is said to be *regular*. This proposition is neede for the proof of lemma 1; further explanation of its importance of is given in section 2.3. From this point on we shall assume that all proofs are regular; in the proof of the Hauptsatz, we shall indicate where this comes in directly (it is obviously used indirectly wherever lemma 1 is used).

Lemma 1 *Let* $\mathcal{P}(a)$ *be an LK-proof of* $\Gamma(a) \vdash \Delta(a)$, b *some free variable not occurring in* $\mathcal{P}(a)$, *and* t *some term. Then:*

- $\mathcal{P}[b/a]$ *is an LK-proof of* $\Gamma[b/a] \vdash \Delta[b/a]$
- *If every eigenvariable in* $\mathcal{P}(a)$ *is different from* a *and not contained in* t, *then* $\mathcal{P}[t/a]$ *is LK-proof of* $\Gamma[t/a] \vdash \Delta[t/a]$
- *If* $\mathcal{P}'(a)$ *is obtained from* $\mathcal{P}(a)$ *in such a way that every eigenvariable is different from* a *and not contained in* t, *then* $\mathcal{P}'(t)$ *is a proof of* $\Gamma(t) \vdash \Delta(t)$

Proof Straightforward; see [61].

This lemma is used in those parts of the proof of the Hauptsatz where quantifiers are involved.

2.2 The cut-elimination theorem

Gentzen's Hauptsatz (also known as the Cut-Elimination Theorem) is undoubtedly the most important result in the proof theory for first order logic and indeed is a cornerstone of all of proof theory. In this section we shall formally state the theorem and present Genzen's proof although not in the exact details in which it was originally given. Given the intended readership, this presentation of the proof is slightly more laborious than is usual, even for the Hauptsatz. A presentation such as Gentzen's [12] is "factored"[7] in several parts: a number

[7] One needs to be awake while reading it ...

of rules that are slightly different, and which therefore require slightly different transformations, are grouped together within a single transformation schema. In understanding such a presentation, one might consider each inference rule in turn, find an appropriate schema, and convince oneself that the right transformation could be obtained. Alternatively, one might imagine that the general idea of the proof (reduction of cut "complexity") is known but no detailed proof is available and then go about formulating a complete proof by considering each rule in turn. What we have done below is to transcribe this "understanding" or "proof re-creation" process.[8] The similarities between the transformations then become apparent and one can then "refactor" this tedious version according to taste, the most natural being by rule structure (see, for example, [61]).

Theorem 1 (Hauptsatz) *If a sequent is LK-provable, then it is cut-free LK-provable.*

Before diving into the proof a few remarks are in order. In particular, since the Cut rule can be eliminated, it is natural to ask why it is there in the first place. There are at least two good reasons. First, it makes helps shorten the length of proofs (see section 2.4); second, it is necessary for an elementary proof of the equivalence[9] of LK and the arbitrary formulations of predicate calculus – consider a system (for the latter) with Modus Ponens. It is also crucial to note here that *the importance of the Hauptsatz is not merely in establishing the existence of cut-free proofs* – as Kreisel [29] points out, this much is obvious[10] simply from the completeness of the cut-free rules – rather, it is to be found in the finitary nature of the "official" proof of the Hauptsatz and in the *Subformula Property* which is a corollary of the Hauptsatz. This makes the Hauptsatz of itself a mathematically[11] shaky result, but one which nonetheless is of unquestionable philosophical significance; see [29] for details.

2.2.1 Proof of the Hauptsatz

In outline the proof consists of showing that in a given proof every cut can be replaced by a cut of smaller "complexity"; induction on the number and "complexity" of cuts then suffices to complete the proof. This "skeleton" proof is visualized easily enough and is explicitly laid out in [17], for example. Nevertheless, the completely "fleshed out" proof is somewhat laborious. For technical reasons, to be discussed (in section 2.3) after the presentation of the proof, the proof is carried out in the system LK* which is defined to be the same as LK but with the *Mix* rule replacing the cut rule.The Mix rule is:

$$\frac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda : mix}$$

where Π and Δ both contain some common formula $\mathcal{M}$, and the sequences $\Pi^{\mathcal{M}}$ and $\Delta^{\mathcal{M}}$ are obtained by deleting all occurrences of $\mathcal{M}$ in, Π and Δ, respectively. $\mathcal{M}$ will be referred to as the *mix formula* and when it is known, will be explicitly indicated. Clearly, a mix is a generalized form of the cut in that a cut may be easily transformed into a mix by the use of thinnings and exchanges. Conversly, a mix may transformed into a cut by the use of

[8] I suspect that a symbol-cruncher will prefer this anyway!

[9] In Gentzen's case it was needed to prove the equivalence of the sequent calculi and the natural deduction calculi.

[10] Actually, only for classical logic. Completeness for intuitionistic logic is quite a delicate business (I got more than my fill of the nature of this "delicacy" by reading [8]). This also shows that cut elimination is not the same as completeness of cut-free rules since the "official" proof of the Hauptsatz applies to both classical logic and intuitionistic logic with equal ease.

[11] It does have some mathematical significance: in the computational aspect of a normalization theorem and from uses of the subformula property.

contractions and exchanges, so LK and LK* are equivalent. Hence what will be proved is the following:

Theorem 2 *If a sequent is LK*–provable, then it is mix-free LK*–provable.*

The proof of this theorem easily follows from the proof of:

Lemma 2 *Suppose a sequent is LK*–provable, and suppose that the proof has only one mix and that this mix is the last inference in the proof, then the sequent is mix-free LK*–provable.*

Having proved this lemma, and this is where the hard work is, by straightforward induction on the number of mixes it is easily proven that any sequent that is LK*-provable is mix-free LK*-provable. From this, by the equivalence of LK and LK*, we shall have obtained a proof of the Hauptsatz.

For the proof of the lemma we need a precise notion of proof complexity[12] ; this is expressed in terms of the complexity introduced by a mix. In what follows we shall assume that $\mathcal{P}$ is a proof satisfying the hypotheses of lemma 1.

Definition 3 A *thread* (of a mix as in lemma 2) is a sequence of sequents:

- that begins with a leaf sequent and ends with the mix formula
- in which every sequent except the last is the upper sequent of an inference and is immediately followed by the lower sequent of the inference

Definition 4 The *right (left) thread* is the thread contains the right (left) upper sequent of the mix.

Definition 5 The *rank* of a right (left) thread is the number of consecutive sequents, counting upward from the right (left) upper sequent of the mix, that contains the mix formula.

Definition 6 The *right (left) rank*, $\rho_R(\mathcal{P})$ ($\rho_L(\mathcal{P})$) of $\mathcal{P}$, is the largest rank over all the right (left) threads.

Definition 7 The rank, $\rho(\mathcal{P})$, of $\mathcal{P}$ is the sum of the right and left ranks.

Remark $\rho(\mathcal{P}) \geq 2$ for all $\mathcal{P}$.

Definition 8 The *degree*, $\lambda(A)$, of a formula, A, is the number of occurrences of logical connectives it contains. The degree, $\lambda(\mathcal{P})$, of $\mathcal{P}$ is the degree of the mix formula (we are assuming there is only one).

The proof of the lemma is by induction on both the rank and degree of $\mathcal{P}$. Specifically, the induction relation, $\prec$, is a lexicographic ordering, defined on finite ordinals as follows: $\langle \lambda_1, \rho_1 \rangle \prec \langle \lambda_2, \rho_2 \rangle$ iff $\lambda_1 < \lambda_2$ or $\lambda_1 = \lambda_2$ and $\rho_1 < \rho_2$.

From the observation above we have two major cases, one corresponding to $\rho(\mathcal{P}) = 2$ and the other to $\rho(\mathcal{P}) > 2$. In the former case the inductive step will be on the degree of $\mathcal{P}$ and in the latter case it will be on the rank. In each case and subcase below, given $\mathcal{P}$ (with only

[12] One of the main difficulties in extending the Hauptsatz (specifically, the type of proof used here) to higher order logics is that there is no "natural", or even useful, measure of complexity; see below.

one mix) a new proof $\mathcal{P}'$ will be constructed, and this proof may have several or no mixes in it. If $\mathcal{P}'$ has any mixes, and each of these is considered in turn in the appropriate order – at each step taking the mix "closest" to the leaves of the proof-tree – then for each subproof $\mathcal{P}''$ with only one mix (this being the last inference) it will be the case that $\langle \lambda_{\mathcal{P}''}, \rho_{\mathcal{P}''} \rangle \prec \langle \lambda_{\mathcal{P}}, \rho_{\mathcal{P}}$. Hence by the induction hypothesis, the mix in $\mathcal{P}''$ – and hence all mixes in $\mathcal{P}'$ and $\mathcal{P}$ – can be eliminated.

Case 1 $\rho(\mathcal{P}) = 2$, which implies that $\rho_R(\mathcal{P}) = 1$ and $\rho_L(\mathcal{P}) = 1$. Then it must be the case, since both threads contain only one sequent, that either the thread is an leaf sequent (axiom) or that the mix formula has been introduced by a thinning, or that it is introduced by a single logical inference. For the first two cases, we consider only the cases of the left thread, the right case being symmetrical.

Case 1.1 The upper left sequent of the mix is an axiom. Then the conclusion of $\mathcal{P}$ has the form:

$$\frac{\mathcal{M} \vdash \mathcal{M} \qquad \Pi \vdash \Lambda}{\mathcal{M}, \Pi^{\mathcal{M}} \vdash \Lambda}$$

We obtain a mix-free proof by transforming to:

$$\cfrac{\cfrac{\cfrac{\Pi \vdash \Lambda}{\mathit{exchanges}}}{\mathcal{M}, \ \ldots, \ \mathcal{M}, \Pi^{\mathcal{M}} \vdash \Lambda}}{\cfrac{\mathit{contractions}}{\mathcal{M}, \Pi^{\mathcal{M}} \vdash \Lambda}}$$

Case 1.2 The mix formula in the upper left sequent comes from a structural rule. As the only structural rule that can introduce a new formula is a thinning, the conclusion of $\mathcal{P}$ has the form:

$$\frac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, \mathcal{M}} \qquad \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta, \Lambda}$$

We obtain a mix-free proof by transforming to:

$$\cfrac{\cfrac{\cfrac{\Gamma \vdash \Delta}{\mathit{thinnings}}}{\Pi^{\mathcal{M}}, \Gamma \vdash \Delta, \Lambda}}{\cfrac{\mathit{exchanges}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta, \Lambda}}$$

Case 1.3 The mix comes from a logical inference. In this case, because $\rho_R(\mathcal{P}) = 1$ and $\rho_L(\mathcal{P}) = 1$, the logical symbol must have been introduced in both the left and right upper sequents of the mix.

Case 1.3.1 The logical symbol is $\wedge$. Then the conclusion of $\mathcal{P}$ has the form:

$$\frac{\dfrac{\Gamma \vdash \Delta, A \qquad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B} \qquad \dfrac{A, \Pi \vdash \Lambda}{A \wedge B, \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda} : \text{mix}$$

Transform this into:

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad A, \Pi \vdash \Lambda}{\Gamma, \Pi^A \vdash \Delta^A, \Lambda}\text{: mix}}{\textit{thinnings and exchanges}}}{\Gamma, \Pi \vdash \Delta, \Lambda}$$

The complexity of the new proof is $\langle \lambda(A), n \rangle$, where $n \geq 2$, which is $\prec$ the complexity, $\langle \lambda(A \wedge B), 2 \rangle$, of the original proof. By the induction hypothesis, there is a mix-free proof of $\Gamma, \Pi^A \vdash \Delta^A, \Lambda$ and hence there is a mix-free proof of $\Gamma, \Pi \vdash \Delta, \Lambda$. A similar argument is easily seen to be applicable to the following subcases and we shall not explicitly state it at every turn.

Case 1.3.2 The logical symbol is $\vee$. Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, A \vee B} \qquad \dfrac{A, \Pi \vdash \Lambda \qquad B, \Pi \vdash \Lambda}{A \vee B, \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda}\text{: mix}$$

Transform this into:

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad A, \Pi \vdash \Lambda}{\Gamma, \Pi^A \vdash \Delta^A, \Lambda}\text{: mix}}{\textit{thinnings and exchanges}}}{\Gamma, \Pi \vdash \Delta, \Lambda}$$

Case 1.3.3 The logical symbol is $\rightarrow$. Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\dfrac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B} \qquad \dfrac{\Pi \vdash \Lambda, A \qquad B, \Sigma \vdash \Theta}{A \rightarrow B, \Pi, \Sigma \vdash \Lambda, \Theta}}{\Gamma, \Pi, \Sigma \vdash \Delta, \Lambda, \Theta}\text{: mix}$$

Transform this into the proof:

$$\dfrac{\dfrac{\dfrac{\Pi \vdash \Lambda, A \qquad \dfrac{A, \Gamma \vdash \Delta, B \qquad B, \Sigma \vdash \Theta}{A, \Gamma, \Sigma^B \vdash \Delta^B, \Theta}\text{: mix}}{\Pi, \Gamma^A, \Sigma^B \vdash \Lambda^A, \Delta^B, \Theta}\text{: mix}}{\textit{thinnings and exchanges}}}{\Gamma, \Pi, \Sigma \vdash \Delta, \Lambda, \Theta}$$

The new proof has two mixes but each has a complexity that is smaller than that of the original proof.

Case 1.3.4 The logical symbol is $\forall$. Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)} \qquad \dfrac{A(t), \Pi \vdash \Lambda}{\forall x A(x), \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda}\text{: mix}$$

Transform this into;

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A(t) \qquad A(t), \Pi \vdash \Lambda}{\Gamma, \Pi^{A(t)} \vdash \Delta^{A(t)}, \Lambda}\text{: mix}}{\textit{thinnings and exchanges}}}{\Gamma, \Pi \vdash \Delta, \Lambda}$$

Here the proof of $\Gamma \vdash \Delta, A(t)$ has been obtained from the proof of $\Gamma \vdash \Delta, A(a)$ by replacing all free occurrences of a by t; see lemma 1.

Case 1.3.5 The logical symbol is $\exists$. Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, \exists x A(x)} \quad \dfrac{A(a), \Pi \vdash \Lambda}{\exists x A(x), \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda} : \text{mix}$$

Transform this into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A(t) \qquad A(t), \Pi \vdash \Lambda}{\Gamma, \Pi^{A(t)} \vdash \Delta^{A(t)}, \Lambda} : \text{mix}}{\dfrac{\textit{thinnings and exchanges}}{\Gamma, \Pi \vdash \Delta, \Lambda}}$$

Here the proof of $A(t), \Pi \vdash \Lambda$ has been obtained from the proof of $A(a), \Pi \vdash \Lambda$ by replacing all free occurrences of a by t; see lemma 1.

Case 1.3.6 The logical symbol is $\neg$. Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\dfrac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A} \quad \dfrac{\Pi \vdash \Lambda, A}{\neg A, \Pi \vdash \Lambda}}{\Gamma, \Pi \vdash \Delta, \Lambda} : \text{mix}$$

Transform this into:

$$\dfrac{\dfrac{\Pi \vdash \Lambda, A \qquad A, \Gamma \vdash \Delta}{\Pi, \Gamma^{A} \vdash \Lambda^{A}, \Delta} : \text{mix}}{\dfrac{\textit{thinnings and exchanges}}{\Gamma, \Pi \vdash \Delta, \Lambda}}$$

Case 2 $\rho(\mathcal{P}) > 2$. There are two main possibilities to consider, depending on the magnitude of the left or right rank. These are $\rho_R(\mathcal{P}) > 1$ and $\rho_L(\mathcal{P}) \geq 1$ or $\rho_R(\mathcal{P}) \geq 1$ and $\rho_L(\mathcal{P}) > 1$. Because of the obvious symmetry, we will go into the details of only the former case. The three subcases considered below correspond roughly to the three considered above. However, there is one major difference: In the former cases, new proofs were constructed in such a way that the degree was lowered; in the following cases new proofs will be constructed in a manner that lowers the rank. For the case considered below ($\rho_R(\mathcal{P}) > 1$), the right rank will be reduced (in most cases by 1) whereas the left rank will be unchanged (the other, symmetric, half of the proof covers this). In a few cases a new mix is introduced by the transformation; in such a case the degree of each new mix will also be lower than that of the original mix.

Case 2.1 Γ or Λ (see the definition of mix) contains $\mathcal{M}$. Then a mix-free proof can be constructed by transforming the mix as follows:

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta^{\mathcal{M}}, \mathcal{M}}\ \textit{contractions and exchanges}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\ \textit{thinnings and exchanges} \qquad \dfrac{\dfrac{\Pi \vdash \Lambda}{\mathcal{M}, \Pi^{\mathcal{M}} \vdash \Lambda}\ \textit{contractions and exchanges}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\ \textit{thinnings and exchanges}$$

Case 2.2 The inference, I, ending with the right upper sequent of the mix is an application of a structural rule.

This case demonstrates quite well the general situation in the remaining subcases and provides a rather clear way to visualize[13] *the general process of mix elimination: the mix is "pushed upwards" through the proof (in the new proof the mix appears above the inference); in this process, it may break into other mixes (inferences involving two upper sequents) which are in turn pushed upwards; the process ends with all mixes "falling off" the top of the proof or "disappearing" within the proof. Evidently, each time a mix is pushed up, the rank is reduced.*

(a) *Left Rules*

• The mix formula is the same one that is introduced by a thinning ($\Sigma = \Sigma_{\mathcal{M}}$) or contracted ($\mathcal{M} \in \Sigma$), or exchanged ($\mathcal{M} \in \Sigma$). Then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Sigma \vdash \Lambda}{\Pi \vdash \Lambda}\text{: I}}{\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}$$

This is transformed into:

$$\dfrac{\Gamma \vdash \Delta \qquad \Sigma \vdash \Lambda}{\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}$$

• If the mix formula is not the same one that is introduced (in the case of a thinning), or contracted, or exchanged, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Sigma \vdash \Lambda}{\Pi \vdash \Lambda}\text{: I}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Sigma \vdash \Lambda}{\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}}{\textit{exchanges}}}{\Sigma^{\mathcal{M}}, \Gamma \vdash \Delta^{\mathcal{M}}, \Lambda}}{\Pi^{\mathcal{M}}, \Gamma \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: I}}{\dfrac{\textit{exchanges}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}$$

The mix which concludes the proof, $\mathcal{P}'$, of $\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda$ occurs above the inference I and, therefore, can clearly be seen to have a smaller rank: specifically, $\rho_R(\mathcal{P}') = \rho_R(\mathcal{P}) - 1$.

(b) *Right Rules*

The conclusion of $\mathcal{P}$ has the following form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Theta}{\Pi \vdash \Lambda}\text{: I}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}$$

[13] cf. Kreisel's remark that a proof is elementary if one can visualize it.

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Theta}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta}\text{: mix}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: I}$$

Case 2.3 The inference concluding with the upper right hand sequent of the mix is an application of a logical rule. The various cases are considered according to the outermost logical symbol.

Case 2.3.1 The outermost logical symbol is $\wedge$.

(a) *Left Rules.* Because of the symmetry we consider only one of the two rules.

- If $A \wedge B$ is the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{B, \Pi \vdash \Lambda}{A \wedge B, \Pi \vdash \Lambda}\text{: }\wedge}{\Gamma, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda}\text{: mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, B, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda}\text{: mix}}{B, \Gamma, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda}\textit{ exchanges}}{A \wedge B, \Gamma, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda}\text{: }\wedge}{\Gamma, \Gamma, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Delta^{A \wedge B}, \Lambda}\text{: mix}}{\Gamma, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda}\textit{ contractions and exchanges}$$

The upper mix has right rank $\rho_R(\mathcal{P}) - 1$; the lower mix has right rank 1. Therefore, by the induction hypothesis, $\Gamma, B, \Pi^{A \wedge B} \vdash \Delta^{A \wedge B}, \Lambda$ has a mix-free proof. Having obtained this mix-free proof, the lower mix can now be eliminated (induction hypothesis) to obtain a mix-free proof equivalent to $\mathcal{P}$. (In the rest of the proof we shall simply state the ranks rather than repeat this entire, and obvious, argument).

- If $A \wedge B$ is not the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{B, \Pi \vdash \Lambda}{A \wedge B, \Pi \vdash \Lambda}\text{: }\wedge}{\Gamma, A \wedge B, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}$$

If B is not the mix formula, then this is transformed into:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, B, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}}{B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\textit{ exchanges}}{A \wedge B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: }\wedge}{\Gamma, A \wedge B, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\textit{ exchanges}}{}$$

If B is the mix formula, then the transformation is the same except that the proof of $B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda$ is replaced by

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

In either case, the mix has right rank $\rho_R(\mathcal{P}) - 1$.

(b) *Right Rule.*

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Lambda, A \qquad \Pi \vdash \Lambda, B}{\Pi \vdash \Lambda, A \wedge B} : \wedge}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A \wedge B} : \text{mix}$$

this is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, A}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A} : \text{mix} \qquad \dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, B}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, B} : \text{mix}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A \wedge B} : \wedge$$

Case 2.3.2 The outermost logical symbol is $\vee$.

(a) *Left Rule.*

• If $A \vee B$ is the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A, \Pi \vdash \Lambda \qquad B, \Pi \vdash \Lambda}{A \vee B, \Pi \vdash \Lambda} : \vee}{\Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda} : \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A, \Pi \vdash \Lambda}{\Gamma, A, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda} : \text{mix}}{\dfrac{\textit{exchanges}}{A, \Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda}} \qquad \dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, B, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda} : \text{mix}}{\dfrac{\textit{exchanges}}{B, \Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda}}}{A \vee B, \Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda} : \vee}{\Gamma, \Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Delta^{A \vee B}, \Lambda} : \text{mix}}{\dfrac{\textit{contractions and exchanges}}{\Gamma, \Pi^{A \vee B} \vdash \Delta^{A \vee B}, \Lambda}}$$

If $A \vee B$ is not the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A, \Pi \vdash \Lambda \qquad B, \Pi \vdash \Lambda}{A \vee B, \Pi \vdash \Lambda, B} : \vee}{\Gamma, A \vee B, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda} : \text{mix}$$

• If neither A nor B is the mix formula, then this is transformed into:

$$\dfrac{\dfrac{\dfrac{\Gamma\vdash\Delta \qquad A,\Pi\vdash\Lambda}{\Gamma,A,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}\text{: mix}}{A,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}\,\textit{exchanges} \qquad \dfrac{\dfrac{\Gamma\vdash\Delta \qquad B,\Pi\vdash\Lambda}{\Gamma,B,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}\text{: mix}}{B,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}\,\textit{exchanges}}{\dfrac{A\vee B,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}{\Gamma,A\vee B,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}\,\textit{exchanges}}\,:\vee$$

If A or B is the mix formula then the proof of $A,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda$ or $B,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda$ is replaced by:

$$\dfrac{\dfrac{\Gamma\vdash\Delta \qquad A,\Pi\vdash\Lambda}{\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}}{A,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda} \qquad\qquad \dfrac{\dfrac{\Gamma\vdash\Delta \qquad B,\Pi\vdash\Lambda}{\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}}{B,\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda}$$

(b) *Right Rules.* Because of the obvious symmetry, only one rule is considered.
The conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma\vdash\Delta \quad \dfrac{\Pi\vdash\Lambda,A}{\Pi\vdash\Lambda,A\vee B}:\vee}{\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda,A\vee B}\text{: mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma\vdash\Delta \qquad \Pi\vdash\Lambda,A}{\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda,A}\text{: mix}}{\Gamma,\Pi^{\mathcal{M}}\vdash\Delta^{\mathcal{M}},\Lambda,A\vee B}:\vee$$

Case 2.3.3 The outermost logical symbol is $\rightarrow$.
(a) *Left Rule.*
- If $A\rightarrow B$ is the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma\vdash\Delta \quad \dfrac{\Sigma\vdash\Theta,A \qquad B,\Pi\vdash\Lambda}{A\rightarrow B,\Sigma,\Pi\vdash\Theta,\Lambda}:\rightarrow}{\Gamma,\Sigma^{A\rightarrow B},\Pi^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Theta,\Lambda}\text{: mix}$$

If $A\rightarrow B$ is in Σ and Π, then this is transformed into:

$$\dfrac{\dfrac{\Gamma\vdash\Delta \quad \dfrac{\dfrac{\Gamma\vdash\Delta \qquad \Sigma\vdash\Theta,A}{\Gamma,\Sigma^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Theta,A}\text{: mix} \quad \dfrac{\dfrac{\Gamma\vdash\Delta \qquad B,\Pi\vdash\Lambda}{\Gamma,B,\Pi^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Lambda}\text{: mix}}{B,\Gamma,\Pi^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Lambda}\,\textit{exchanges}}{A\rightarrow B,\Gamma,\Sigma^{A\rightarrow B},\Gamma,\Pi^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Theta,\Delta^{A\rightarrow B},\Lambda}:\rightarrow}{\Gamma,\Gamma,\Sigma^{A\rightarrow B},\Gamma,\Pi^{A\rightarrow B}\vdash,\Delta^{A\rightarrow B},\Delta^{A\rightarrow B},\Theta,\Delta^{A\rightarrow B},\Lambda}\text{: mix}}{\Gamma,\Sigma^{A\rightarrow B},\Pi^{A\rightarrow B}\vdash\Delta^{A\rightarrow B},\Theta,\Lambda}\,\textit{contractions and exchanges}$$

If $A \to B$ is not in Σ or not in Θ (it must occur in one since $\rho_R(\mathcal{P}) > 1$) , then the proof of $\Gamma, \Sigma^{A\to B} \vdash \Delta^{A \to B}, \Theta, A$ or of $B, \Gamma, \Pi^{A\to B} \vdash \Delta^{A\to B}, \Lambda$ is replaced by:

$$\dfrac{\dfrac{\Sigma \vdash \Theta, A}{\textit{thinnings and exchanges}}}{\Gamma, \Sigma \vdash \Delta, \Theta, A} \qquad \dfrac{\dfrac{B, \Pi \vdash \Lambda}{\textit{thinnings and exchanges}}}{B, \Gamma, \Pi \vdash \Delta, \Lambda}$$

- If $A \to B$ is not the mix formula, then the conclusion of $\mathcal{P}$ is:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Sigma \vdash \Theta, A \qquad B, \Pi \vdash \Lambda}{A \to B, \Sigma, \Pi \vdash \Theta, \Lambda}{:\to}}{\Gamma, A \to B, \Sigma^{\mathcal{M}}, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta, \Lambda}{: \text{mix}}$$

If neither A nor B is the mix formula, then this is transformed into:

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Sigma \vdash \Theta, A}{\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta, A}{: \text{mix}} \qquad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, B, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}{: \text{mix}}}{\textit{exchanges}}}{B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{\dfrac{A \to B, \Gamma, \Sigma^{\mathcal{M}}, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta, \Delta^{\mathcal{M}}, \Lambda}{\textit{contractions and exchanges}}}{:\to}}{\Gamma, A \to B, \Sigma^{\mathcal{M}}, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

If A or B is the mix formula then the proof of $\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta, A$ or $B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda$ is replaced by:

$$\dfrac{\Gamma \vdash \Delta \qquad \Sigma \vdash \Theta, A}{\Gamma, \Sigma^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Theta, A} \qquad \dfrac{\dfrac{\Gamma \vdash \Delta \qquad B, \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{B, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

(b) *Right Rule.*
The conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A, \Pi \vdash \Lambda, B}{\Pi \vdash \Lambda, A \to B}{:\to}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A \to B}{: \text{mix}}$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A, \Pi \vdash \Lambda, B}{\Gamma, A, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, B}{: \text{mix}}}{\textit{exchanges}}}{A, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, B}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A \to B}{:\to}$$

Case 2.3.4 The outermost logical symbol is $\neg$.
(a) *Left Rule.*
- If $\neg A$ is the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Lambda, A}{\neg A, \Pi \vdash \Lambda}: \neg}{\Gamma, \Pi^{\neg A} \vdash \Delta^{\neg A}, \Lambda}: \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, A}{\Gamma, \Pi^{\neg A} \vdash \Delta^{\neg A}, \Lambda, A}: \text{mix}}{\neg A, \Gamma, \Pi^{\neg A} \vdash \Delta^{\neg A}, \Lambda}: \neg}{\Gamma, \Gamma, \Pi^{\neg A} \vdash \Delta^{\neg A}, \Delta^{\neg A}, \Lambda}: \text{mix}}{\dfrac{\textit{contractions and exchanges}}{\Gamma, \Pi^{\neg A} \vdash \Delta^{\neg A}, \Lambda}}$$

- If $\neg A$ is not the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Lambda, A}{\neg A, \Pi \vdash \Lambda}: \neg}{\Gamma, \neg A, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}: \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, A}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A}: \text{mix}}{\neg A, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}: \neg}{\dfrac{\textit{exchanges}}{\Gamma, \neg A, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}$$

(b) *Right Rule.*
The conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A, \Pi \vdash \Lambda}{\Pi \vdash \Lambda, \neg A}: \neg}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \neg A}: \text{mix}$$

If A is not the mix formula, then this is transformed into:

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A, \Pi \vdash \Lambda}{\Gamma, A, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}: \text{mix}}{\textit{exchanges}}}{\dfrac{A, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \neg A}: \neg}$$

If A is the mix formula then the proof of $A, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda$ is replaced by:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A, \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{A, \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

Case 2.3.5 The outermost logical symbol is $\forall$.

(a) *Left Rule.*

• If $\forall x A(x)$ is the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A(t), \Pi \vdash \Lambda}{\forall x A(x), \Pi \vdash \Lambda} : \forall}{\Gamma, \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Lambda,} : \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A(t), \Pi \vdash \Lambda}{\Gamma, A(t), \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Lambda} : \text{mix}}{\textit{exchanges}}}{A(t), \Gamma, \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Lambda}}{\forall x A(x), \Gamma, \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Lambda} : \forall}{\dfrac{\Gamma, \Gamma, \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Delta^{\forall x A(x)}, \Lambda}{\textit{contractions and exchanges}}} : \text{mix}}{\Gamma, \Pi^{\forall x A(x)} \vdash \Delta^{\forall x A(x)}, \Lambda}$$

• If $\forall x A(x)$ is not the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A(t), \Pi \vdash \Lambda}{\forall x A(x), \Pi \vdash \Lambda} : \forall}{\Gamma, \forall x A(x), \Pi \vdash \Delta^{\mathcal{M}}, \Lambda,} : \text{mix}$$

If $A(t)$ is not the mix formula, then this is transformed into:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A(t), \Pi \vdash \Lambda}{\Gamma, A(t), \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda} : \text{mix}}{\textit{exchanges}}}{A(t), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{\forall x A(x), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda} : \forall}{\textit{exchanges}}}{\Gamma, \forall x A(x), \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

If $A(t)$ is the mix formula then the proof of $A(t), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda$ is replaced by:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A(t), \Pi \vdash \Lambda}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}}{A(t), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

(b) *Right Rule.*

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Lambda, A(a)}{\Pi \vdash \Lambda, \forall x A(x)} : \forall}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \forall x A(x)} : \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, A(a)}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A(a)} : \text{mix}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \forall x A(x)} : \forall$$

Note that the regularity condition is tacitly being invoked here. In the absence of this, a may occur, for example, in $\Delta^{\mathcal{M}}$ and the new proof would violate the eigenvariable condition. Also, $A(a)$ cannot be the mix formula since this too would violate the same condition.

Case 2.3.6 The outermost logical symbol is $\exists$.
Left Rule.

• If the $\exists x A(x)$ is the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A(a), \Pi \vdash \Lambda}{\exists x A(x), \Pi \vdash \Lambda} : \exists}{\Gamma, \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Lambda} : \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A(a), \Pi \vdash \Lambda}{\Gamma, A(a), \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Lambda} : \text{mix}}{\textit{exchanges}}}{A(a), \Gamma, \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Lambda}}{\exists x A(x), \Gamma, \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Lambda} : \exists}{\Gamma, \Gamma, \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Delta^{\exists x A(x)}, \Lambda} : \text{mix}}{\dfrac{\textit{contractions and exchanges}}{\Gamma, \Pi^{\exists x A(x)} \vdash \Delta^{\exists x A(x)}, \Lambda}}$$

• If the $\exists x A(x)$ is not the mix formula, then the conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{A(a), \Pi \vdash \Lambda}{\exists x A(x), \Pi \vdash \Lambda} : \exists}{\Gamma, \exists x A(x), \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda} : \text{mix}$$

This is transformed into:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta \qquad A(a), \Pi \vdash \Lambda}{\Gamma, A(a), \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: mix}}{\textit{exchanges}}}{\dfrac{A(a), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}{\exists x A(x), \Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}\text{: }\exists}}{\textit{exchanges}}}{\Gamma, \exists x A(x), \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda}$$

Again, as above, the regularity condition is tacitly being invoked; similarly, $A(a)$ cannot be the mix formula.

(b) *Right Rules.*
The conclusion of $\mathcal{P}$ has the form:

$$\dfrac{\Gamma \vdash \Delta \qquad \dfrac{\Pi \vdash \Lambda, A(t)}{\Pi \vdash \Lambda, \exists x A(x)}\text{: }\exists}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \exists x A(x)}\text{: mix}$$

This is transformed into:

$$\dfrac{\dfrac{\Gamma \vdash \Delta \qquad \Pi \vdash \Lambda, A(t)}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, A(t)}\text{: mix}}{\Gamma, \Pi^{\mathcal{M}} \vdash \Delta^{\mathcal{M}}, \Lambda, \exists x A(x)}\text{: }\exists$$

End of Proof (Hauptsatz)

2.2.2 Discussion of the proof

Typically the first question one asks about the proof is: *why the mix instead of the cut?* The answer is that the mix simplifies the proof. To see the type of difficulty that might arise with the cut, consider the following proof:

$$\dfrac{\Gamma \vdash \Delta, A \qquad \dfrac{A, A, \Pi \vdash \Lambda}{A, \Pi \vdash \Lambda}\text{: contraction}}{\Gamma, \Pi \vdash \Delta, \Lambda}\text{: I}$$

If I is taken to be a mix, then we can easily reduce the rank of the proof by eliminating the contraction and by applying the mix to the original upper sequent of the contraction – thus:

$$\dfrac{\Gamma \vdash \Delta, A \qquad A, A, \Pi \vdash \Lambda}{\Gamma, \Pi \vdash \Delta, \Lambda}\text{: mix}$$

On the other hand, if I is taken to be a cut and we try to apply this simple (and rather natural) transformation with, we get:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad A, A, \Pi \vdash \Lambda}{\Gamma, A, \Pi \vdash \Delta, \Lambda}\text{: cut}}{\textit{exchanges}}}{A, \Gamma, \Pi \vdash \Delta, \Lambda}}{\Gamma, \Gamma, \Pi \vdash \Delta, \Delta, \Lambda}\text{: cut}}{\Gamma, \Pi \vdash \Delta, \Lambda}$$

It is apparent that no progress has been made since, although the upper cut has a smaller complexity than the original, the complexity of the proof has been increased by the lower cut! And should we simple-mindedly (try to remedy the situation by chosing to) apply the procedure of thinning to re-introduce a formula deleted by a "less complex" cut and having the original inference end up between the two cuts, we would end up with a non-terminating sequence of transformations, starting with:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad \dfrac{\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad A, A, \Pi \vdash \Lambda}{\Gamma, A, \Pi \vdash \Delta, \Lambda}: \text{cut}}{\dfrac{\textit{exchanges}}{A, \Gamma, \Pi \vdash \Delta, \Lambda}}}{A, A, \Gamma, \Pi \vdash \Delta, \Lambda}}{A, \Gamma, \Pi \vdash \Delta, \Lambda}}{\Gamma, \Gamma, \Pi \vdash \Delta, \Delta, \Lambda}: \text{cut}}{\Gamma, \Pi \vdash \Delta, \Lambda}$$

It is of course possible to work with the Cut instead of the Mix – in Szabo's [55] category-theoretic formulation of proof theory, for example, it is not possible to use the Mix; likewise in Zucker's [64] work on the correspondence between the Hauptsatz and the Normalization Theorem – but only, it seems, at the cost of additional complications in the proof.

Now to the classification of the Hauptsatz. As given above, it is clearly a normalization result; the proof gives a definite procedure (a set of rewrite rules) for eliminating cuts (mixes). It is, however, a weak normalization result due to the restriction that in any given proof cuts (mixes) must be eliminated from the "top" of the proof downwards, as indicated by the main lemma above. This restriction is necessary since there are instances in which permitting the permutation of mixes can lead to non-terminating transformation sequences. Let us suppose that the transformation rules permit pushing mixes up past other mixes. Then consider the following example, due to Zucker [64]:

Starting with the proof whose conclusion is:

$$\dfrac{A \vdash B \qquad \dfrac{B \vdash A \qquad A, B \vdash C}{B, B \vdash C}}{A \vdash C}$$

the first two transformations produce:

$$\dfrac{\dfrac{A \vdash B \qquad B \vdash A}{A \vdash A} \qquad \dfrac{A \vdash B \qquad A, B \vdash C}{A, A \vdash C}}{A \vdash C}$$

and

$$\dfrac{A \vdash B \qquad \dfrac{B \vdash A \qquad \dfrac{A \vdash B \qquad A, B \vdash C}{A, A \vdash C}}{B \vdash C}}{A \vdash C}$$

Now the original proof is clearly isomorphic to part of the last proof, and it is apparent that no progress has been made.

For the classification according to logic complexity, it is easily seen that the Hauptsatz is a $\prod_2^0$ statement. What is says is that for every proof x with cuts there is an equivalent cut-free proof y. That is, in $\forall x \exists y : R(x, y)$ x and y are variables for codes of finite proofs and R states that x encodes a proof with cuts, y encodes a cut-free proof, y is equivalent to x, and so forth; evidently, R is recursive.

Lastly, a point which is of relevance later: the termination of the cut elimination procedure has been proved by induction up to ω. Superficially, this is transfinite induction up to ω^3 since the proof is carried out using three nested inductions (on the number of mixes, the degree, and the rank) but this is evidently reducible, by suitable encoding, to mathematical induction – encode every triplet as a single ordinal.

2.3 Size of cut-free proofs

An issue that is undoubtedly of interest in computation but less so in logic is that of the relation between the size of a proof with cuts and the corresponding cut-free proof. For the main result in this section we shall need a notion of *depth* (analogous to that of degree) and of *height* (analogous to rank).

Definition 9 The *depth, d,* of a formula is defined inductively as follows:

- If A is atomic then $d(A) = 0$
- $d(A \vee B) = d(A \wedge B) = d(A \rightarrow B) = \max\{d(A) + 1, d(B) + 1\}$
- $d(\neg A) = d(\exists x A) = d(\forall x A) = d(A) + 1$

The *depth of a cut* is the depth of its cut-formula and *the depth of a proof* is $\max\{d(C){+}1\}$, over all cuts, C, in the proof. The depth of a cut-free proof is 0.

Definition 10 The *height, h,* of a proof, $\mathcal{P}$, is defined inductively as follows:

- if $\mathcal{P}$ consists of one axiom, then $h(\mathcal{P}){=}0$,
- if $\mathcal{P}$ is obtained from $\mathcal{P}'$ by an application of a structural rule then $h(\mathcal{P}) = h(\mathcal{P}')$.
- if $\mathcal{P}$ is obtained from $\mathcal{P}'$ by an application of a logical rule with one premise, then $h(\mathcal{P}) = h(\mathcal{P}') + 1$.
- if $\mathcal{P}$ is obtained from $\mathcal{P}'$ and $\mathcal{P}''$ by an application of a logical rule with two premises or a cut , then $h(\mathcal{P}) = \max\{h(\mathcal{P}') + 1, h(\mathcal{P}'') + 1\}$.

It is easy to see that the above proof of the Hauptsatz can be reformulated with depth and height instead of degree and rank; see [11, 17, 48], for example. From such a proof[14] one readily obtains:

Theorem 3 *If a sequent has an LK-proof of depth n and height k, then it has a cut-free proof of height $\leq H(k, n)$, where $H(k, 0) = k$ and $H(k, m + 1) = 4^{H(k,m)}$.*

This potentially[15] astronomical blow-up in the size of the proof is the main reason why one uses the cut-rule in computation (and, in general, in practical mathematics). Logicians on the other hand strive for cut-free proofs. So one sees here something of a supporting case for Kreisel's [35] warning about the irrelevance of logical proof theory to mechanical theorem proving (and, in general, computer science). Nevertheless, it may be admitted that cut-free proofs are easy to find (both for human and machine) and is the sort of proof that one naturally looks for when working backwards (which is also natural) from the theorem[16].

---

[14]Note that this is an elementary proof and see the discussions in section 3.3.2.

[15]Statman [52] shows that there are cases in which the blow-up is *always* as large as the upper bound

[16]This may be just the way I do things and may not be "natural" to others!

2.4 A calculus with equality

The relevant calculus, $LK_=$, is obtained by adding (as initial axioms) to LK all sequents of the following form:

- $\vdash s = s$
- $s_1 = t_1, ..., s_n = t_n \vdash f(s_1, ..., s_n) = f(t_1, ..., t_n)$
  for every n-ary function letter f
- $s_1 = t_1, ..., s_n = t_n, P(s_1, ..., s_n) \vdash P(t_1, ..., t_n)$
  for every n-ary predicate letter P

where $s, s_1, ..., s_n, t_1, ..., t_n$ are arbitrary terms.

The main point here is that in this new calculus it is not possible to eliminate every cut[17]. The best that one can do by way of cut elimination is:

Theorem 4 *If a sequent has an $LK_=$-proof, then it has an $LK_=$-proof in which in which all cuts except those of the form $s = t$ have been eliminated.*

A proof of this may be found in [61] and also in [11] where calculi with equality are discussed in great detail.

2.5 The Hauptsatz and higher-order logics

The Hauptsatz for second order logic was first conjectured by Takeuti [60] and, consequently, the result is also popularly known as "Takeuti's Conjecture". Based on work by Schutte, Tait [56] gave a semantic proof. Girard produced additional developments, which included showing how to obtain the applicable cut-elimination procedure [14, 17]. Related discussions include [41, 58, 61] on cut elimination in type theory.

The second order sequent calculus, L^2K, is defined as consisting of the the obvious rules as in LK plus the following set quantifier rules (ϕ and φ are set variables and T is a set or predicate constant):

universal

$$\frac{\Gamma \vdash A(\varphi), \Delta}{\Gamma \vdash \forall\phi A(\phi), \Delta} \qquad \frac{\Gamma, A(T) \vdash \Delta}{\Gamma, \forall\phi A(\phi) \vdash \Delta}$$

existential

$$\frac{\Gamma \vdash A(T), \Delta}{\Gamma \vdash \exists\phi A(\phi), \Delta} \qquad \frac{\Gamma, A(\varphi) \vdash \Delta}{\Gamma, \exists\phi A(\phi) \vdash \Delta}$$

with the usual proviso that φ does not occur free in Γ or Δ.

The main point in this section is to briefly highlight the difficulties of finding an elementary proof of a cut-elimination theorem in higher-order logics; specifically, why the approach used by Gentzen will not work here. the problematic aspect is easily explained: If one considers a rule such as right $\exists$, one cannot really say, in any reasonable fashion, that $\exists\phi A(\phi)$ is more complex than $A(\varphi)$ if φ is abstract [10, 39]. For example, if $\varphi = \{x \mid B(x)\}$, it might very

[17] Presumably this is the sort of thing Girard [17] has in mind when he says that equalitarian theories are "not nice".

well be the case that B is more complex than A; more on this below (section 3.1). Another point is that a study of Takeuti's Conjecture gives insights into the limits of the usefulness of cut elimination[18] and the importance of semantic notions (specifically, 3-valued semantics) [19] in proof theory; these points are discussed in great detail in [14, 17].

3 Corollaries and applications

3.1 The subformula property

We now turn to *the* important corollary of the Hauptsatz.

Definition 11 Let A be a wff. The *subformulae* of A, $Subf(A)$ are given by:

- if A is atomic then $Subf(A) = \{A\}$
- if A is $B \wedge C, B \vee C$ or $B \to C$, then $Subf(A) = \{A\} \cup Subf(B) \cup Subf(C)$
- if A is $\neg B$, then $Subf(A) = \{A\} \cup Subf(B)$
- if A is $\exists x B(x)$ or $\forall x B(x)$, then $Subf(A) = \{A\} \cup Subf(B(t))$, t is an arbitrary term.

Corollary 1 (The Subformula Property) *Let $\mathcal{P}$ be a cut-free proof of $\Gamma \vdash \Delta$. Then each formula in $\mathcal{P}$ is a subformula of some formula in $\Gamma \vdash \Lambda$.*

This is easily seen from the fact that all formulae that appear in the premises of the logical rules also appear in the conclusion (the proof is by induction on the cut-free proofs). From a philosophical standpoint, the subformula property is satisfying since it is a more precise statement of Hilbert's "purity of methods"[20].

The subformula property has a number of important applications; typical example is in obtaining consistency proofs. For consistency one proves that the empty sequent $\vdash$ is not provable; this is equivalent to the usual forms of statement: For example, suppose we have a proof of $\vdash A$ and a proof of $\vdash \neg A$. Obtain a proof of $\vdash$ as follows:

$$\dfrac{\vdash \neg A \qquad \dfrac{\vdash A}{\neg A \vdash}}{\vdash}$$

And one readily sees that from a proof of $\vdash$ thinnings will give a proof of any and every sequent.

It is easy to see that any calculus with a subformula property is consistent. The subformula property also has uses in formulating decision procedures; see below.

Another important consequence is this: although Tarski's famous result on the undefinability of truth for arithmetic states that unrestricted truth definitions are an impossibility, we do have partial truth definitions for subformula of a cut-free provable formula [29]. Stated another way: the subformula property allows us to bound the logical complexity (number of quantifiers) of the formulae used in a proof (these are exactly the subformulae) and truth definitions do exist for bounded logical complexity; this is important for, among other things, proving **reflection principles** (see below). This leaves the significance of the subformula property in no doubt.

[18] In Girard's words: "Cut elimination theorems (if one forgets pretentious more or less <<philosophical>> consequences) can easily become trifling."

[19] Semantic notions are needed for the analysis of cut since syntactic notions clearly do not suffice: *provability* has the same meaning for both proofs with cuts and cut-free proofs.

[20] Roughly: "All methods used and all intermediate results must be directly related to the conclusion".

Also, to go back to the higher-order logic case, we can now summarise the situation as follows: there is no meaningful/useful notion of a subformula in these logics. If we look back at the proof of the Hauptsatz, it is immediately obvious that the subformula property is being implicitly invoked: new mixes introduced by the transformations apply (whenever the degree is being reduced) to a subformula of the original mix formula. In other words, "thinking backwards", it is the subformula property that allows us to define (one part of) proof complexity in terms of the number of logical connectives.

Indeed, in proof theory one generally regards[21] a cut elimination theorem without a subformula property as being practically useless; see earlier comments on the significance of cut elimination. At the very least, if a full cut elimination is not possible then one tries to get a subformula property for formulae of a restricted logical complexity.

3.1.1 The Hauptsatz and logical complexity

A major premise in Girard's[22] [17] work is that progress in proof theory goes hand in hand with an increase in logical complexity. Let us see how this relates to the Hauptsatz:

Cut elimination with a subformula property holds for the $\sum_1^0$ subset of Peano Arithmetic (which subset can be developed using predicate calculus); since we have the subformula property, we also have completeness for $\sum_1^0$ statements. Also provability here is $\sum_1^0$. In order to make progress we must consider a system (ω-logic) in which mathematical induction can be proved; this yields similar results for $\prod_1^1$ statements. Further progress is made by considering a system (β-logic) in which transfinite induction can be proven; this yields similar results for $\prod_2^1$ statements. And so on, and so forth ... The general idea is that the logical complexity of what we work with must equal to at least that of what we want to analyse; thus, for example, in section 3.3.1 we shall see the Hauptsatz (of 2.2) used most naturally in the analysis of $\prod_2^0$ statements.

Since the concepts (e.g., finite trees, well-founded trees, dilators, etc.) involved in these logics are of increasing complexity, we see that to get the subformula property for increasingly bigger systems, one is naturally led to an increase in logical complexity. A quick summary of all this may be found in (Table 1 of) [15].

3.1.2 A Comment on the regularity condition

To conclude this section, we are now in a position to to give another explanation of why the regularity condition (see Proposition 1 above) is required. First, observe that since only the $\rightarrow$-rules and the $\neg$-rules move a sequent from antecedent to succedent, or vice versa, the following also obviously follows from the Hauptsatz:

Corollary 2 *Let $\mathcal{P}$ be a cut-free, $\rightarrow$-free, and $\neg$-free proof of $\Gamma \vdash \Delta$. Let $\Pi \vdash \Lambda$ be any sequent in $\mathcal{P}$. Then each formula in Π is a subformula of a formula in Γ and each formula in Λ is a subformula of one in Δ.*

Now consider the following proof (this is due to Kleene [24]):

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A(x) \vdash A(x)}{B(y) \wedge A(x) \vdash A(x)}}{\forall y(B(y) \wedge A(x)) \vdash A(x)}}{\forall x \forall y(B(y) \wedge A(x)) \vdash A(x)}}{\forall x \forall y(B(y) \wedge A(x)) \vdash \forall x A(x)} \qquad \dfrac{A(y) \vdash A(y)}{\forall x A(x) \vdash A(y)}}{\forall x \forall y(B(y) \wedge A(x)) \vdash A(y)}$$

[21] This, however, is not the case in some important applications in Computer Science.

[22] In view of the title of his rather long book, it would be silly of me to attempt more than a sketchy explanation of all this.

This is not regular since x is the eigenvariable of a $\forall$-rule but it also occurs free in a sequent below the $\forall$-application. Now the subformulae of $\forall x \forall y(B(y) \wedge A(x))$ are: $\forall x \forall y(B(y) \wedge A(x)), \forall y(B(y) \wedge A(t)), B(u) \wedge A(t), B(u),$ and $A(t)$, for every term u and every term t not containing y free. The only subformula of $A(y)$ is $A(y)$. The cut in this proof cannot (without suitable variable renaming) be eliminated: consider corollary 2 and the fact that $A(x)$ is not a subformula of $A(y)$.

3.2 The Sharpened Hauptsatz and Herbrand's Theorem

Lemma 3 *If a sequent $\Gamma \vdash \Delta$ has an LK-proof, then it has an LK-proof in which*

- *the formulae in the axioms are all atomic and*
- *the formula introduced in all the thinnings are atomic.*

Proof [23] By induction on the degree of formulae.
(i) Assume we have cut-free proofs of $A \vdash A$ and $B \vdash B$ from atomic formulae. Then:

- *Conjunction*: obtain a cut-free proof of $A \wedge B \vdash A \wedge B$ as follows

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A \qquad B \vdash B}{A, B \vdash A \wedge B}}{A \wedge B, B \vdash A \wedge B}}{B, A \wedge B \vdash A \wedge B}}{A \wedge B, A \wedge B \vdash A \wedge B}}{A \wedge B \vdash A \wedge B}$$

- *Disjunction*: obtain a cut-free proof of $A \vee B \vdash A \vee B$ as follows

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A \vdash A}{A \vdash A, B} \qquad \dfrac{B \vdash B}{B \vdash A, B}}{A \vee B \vdash A, B}}{A \vee B \vdash A, A \vee B}}{A \vee B \vdash A \vee B, A}}{A \vee B \vdash A \vee B, A \vee B}}{A \vee B \vdash A \vee B}$$

- *Implication*: obtain a cut-free proof of $A \rightarrow B \vdash A \rightarrow B$ as follows

$$\dfrac{\dfrac{A \vdash A \qquad B \vdash B}{A \rightarrow B, A \vdash B}}{A \rightarrow B \vdash A \rightarrow B}$$

- *Negation*: obtain a cut-free proof of $\neg A \vdash \neg A$ as follows

$$\dfrac{\dfrac{A \vdash A}{\vdash A, \neg A}}{\neg A \vdash \neg A}$$

- *Quantification*: from a cut-free proof of $A(a) \vdash A(a)$ obtain cut-free proofs of $\forall x A(x) \vdash \forall x A(x)$ and $\exists x A(x) \vdash \exists x A(x)$ as follows

$$\dfrac{\dfrac{A(a) \vdash A(a)}{\forall x A(x) \vdash A(a)}}{\forall x A(x) \vdash \forall x A(x)} \qquad \dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \exists x A(x)}}{\exists x A(x) \vdash \exists x A(x)}$$

[23][17], exercise 2.1.7-(ii).

(ii) Assume we have a proof of $\Gamma \vdash \Delta$

- *Conjunction*: obtain proofs of $\Gamma \vdash \Delta, A \wedge B$ and $A \wedge B, \Gamma \vdash \Delta$ as follows

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A} \qquad \dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, B}}{\Gamma \vdash \Delta, A \wedge B} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}}{A \wedge B, \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{B, \Gamma \vdash \Delta}}{A \wedge B, \Gamma \vdash \Delta}$$

- *Disjunction*: obtain proofs of $A \vee B, \Gamma \vdash \Delta$ and $\Gamma \vdash \Delta, A \vee B$ as follows

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta} \qquad \dfrac{\Gamma \vdash \Delta}{B, \Gamma \vdash \Delta}}{A \vee B, \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A}}{\Gamma \vdash \Delta, A \vee B} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, B}}{\Gamma \vdash \Delta, A \vee B}$$

- *Implication*: obtain proofs of $A \to B, \Gamma \vdash \Delta$ and $\Gamma \vdash \Delta, A \to B$ as follows

$$\dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A} \qquad \dfrac{\Gamma \vdash \Delta}{B, \Gamma \vdash \Delta}}{A \to B, \Gamma, \Gamma \vdash \Delta, \Delta}}{A \to B, \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\dfrac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}}{A, \Gamma \vdash \Delta, B}}{\Gamma \vdash \Delta, A \to B}$$

- *Negation*: obtain proofs of $\neg A, \Gamma \vdash \Delta$ and $\Gamma \vdash \Delta, \neg A$ as follows

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A}}{\neg A, \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{A, \Gamma \vdash \Delta}}{\Gamma \vdash \Delta, \neg A}$$

- *Quantification*: obtain proofs of $\forall x A(x), \Gamma \vdash \Delta$ and $\Gamma \vdash \Delta, \forall x A(x)$ and of $\forall x A(x), \Gamma \vdash \Delta$ and $\Gamma \vdash \Delta, \forall x A(x)$ as follows

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{A(t), \Gamma \vdash \Delta}}{\forall x A(x), \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A(a)}}{\Gamma \vdash \Delta, \forall x A(x)}$$

and

$$\dfrac{\dfrac{\Gamma \vdash \Delta}{A(a), \Gamma \vdash \Delta}}{\exists x A(x), \Gamma \vdash \Delta} \quad \dfrac{\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A(t)}}{\Gamma \vdash \Delta, \exists x A(x)}$$

Lemma 4 *Let $\mathcal{P}$ be a cut-free proof of a sequent $\Gamma \vdash \Delta$ consisting only of prenex formulae. Then we can construct another proof, $\mathcal{P}^m$, of $\Gamma \vdash \Delta$ such that in $\mathcal{P}^m$:*

- *if a rule, R, is applied below an application of a quantifier rule, then R is either a structural rule, a right ∀-rule, or a ∃-rule.*
- *all uses of initial axioms are on atomic formulae.*
- *the formula introduced in each thinning is an atomic formula.*

Proof By lemma 3, we can obtain a proof, $\mathcal{P}'$ satisfying the last two conditions. From $\mathcal{P}'$ we obtain $\mathcal{P}^m$ by appropriately interchanging the positions of the inferences; see [11, 12, 17, 61] for this last part of the proof.

Theorem 5 (The Sharpened Hauptsatz) [24] *If a sequent $\Gamma \vdash \Delta$ consisting only of prenex formulae is LK-provable, then there is a cut-free proof, $\mathcal{P}$ of the sequent such that $\mathcal{P}$ satisfys the conditions of lemma 4 and there is a sequent (the* midsequent*), $\Pi \vdash \Lambda$, of $\mathcal{P}$ such that:*

- $\Pi \vdash \Lambda$ *is quantifier-free.*
- *no quantifier rules are used above* $\Pi \vdash \Lambda$
- *only quantifier and structural rules are used below* $\Pi \vdash \Lambda$

Proof Immediate from the Hauptsatz and lemma 5.

Corollary 3 *If $\exists x_1, ..., \exists x_n : R(x_1, ..., x_n)$, R quantifier-free, is provable in predicate calculus then there exist terms $t_1^1, ..., t_n^1, ..., t_1^m, ..., t_n^m$ such that $R(t_1^1, ..., t_n^1) \vee ... \vee R(t_1^m, ..., t_n^m)$ is provable in propositional calculus.*

Proof Immediate from the Sharpened Hauptsatz (the terms in the midsequent are exactly the terms which make up the conjunction) and the completeness of LK.

The preceding theorem and corollary apply only to LK; the sharpened Hauptsatz does not hold for LJ unless applications of left $\vee$-rules are disallowed. Here is a counterexample (due to Kleene [24]):

$$\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \exists x A(x)} \quad \dfrac{A(b) \vdash A(b)}{A(b) \vdash \exists x A(x)}}{A(a) \vee A(b) \vdash \exists x A(x)}$$

If the sharpened Hauptsatz were applicable for proofs of $A(a) \vee A(b) \vdash \exists x A(x)$, then the midsequent would be of the form $\Pi \vdash \Lambda$ where Π consists if at least one occurrence of $A(a) \vee A(b)$ and Λ is $A(t)$ for some term t. Suppose then that $\Pi \vdash \Lambda$ is $A(a) \vee A(b) \vdash A(t)$. Then t may be either a or b or neither. Consider the case when t is a: Then $A(a) \vee A(b) \rightarrow A(a)$, and hence $A \vee B \rightarrow A$, would be provable in propositional calculus; clearly this is impossible.

Theorem 6 (Herbrand's Theorem) [25] *Let F be a formula in prenex normal form. Then F is provable in predicated calculus iff there is a sequence of formulae $F_1, ..., F_n$, in which bound variables have been replaced by certain*[26] *terms, such that the disjunction $F_1 \vee ... \vee F_n$ is provable in propositional calculus. Furthermore, $F_1, ..., F_n$ can be obtained from F, and F from $F_1, ..., F_n$ by primitive recursive operations.*

Proofs of the theorem (from the sharpened Hauptsatz) may be found in [11, 17]; the original proof of the theorem is in [19] but it is difficult to follow and it is also known to be flawed.

The first half of the proof consists of showing that from the midsequent one can get a sequent that essentially consists instances of the skolemized formula; the second half, that from the latter one can get a sequent to which right $\forall$-rules and right $\exists$-rules can be applied to obtain the original formula. Note also that corollary 3 is already quite close to what needs to be proved; the only information missing is on the detailed nature of the terms involved.

For other proofs of Herbrand's Theorem, not involving the Hauptsatz, see [1, 46].

[24] Also known as the Midsequent Theorem or the Miteilsatz.
[25] The generalization, in LK and LJ, to formulae other than prenex is given in [4].
[26] Skolem.

3.2.1 Comparison of the Sharpened Hauptsatz and Herbrand's Theorem

The similarities between these two results are quite apparent from the corollary to the Sharpened Hauptsatz: one can readily see that after ("jumping ahead" and) skolemizing the universal quantifiers, the midsequent consists (essentially) of the terms in the Herbrand disjunction. Indeed, the system Herbrand worked with was the first example of what we now call a cut-free system [62]. The Sharpened Hauptsatz appears to be sharper result than Herbrand's theorem since the latter may be obtained from the former, but not the former from the latter. We summarise some of the main differences:

- From one point of view, the Sharpened Hauptsatz tells us more than Herbrand's Theorem: that the midsequent has a cut-free proof.
- From another point of view, Herbrand's Theorem tells us more than the Sharpened Hauptsatz: we know what the terms in the midsequent look like.
- The Sharpened Hauptsatz applies to both classical logic and intuitionistic logic, with certain restrictions in the latter case, whereas Herbrand's Theorem applies only to classical logic.
- Herbrand's Theorem applies also to formulae not in prenex normal form while the Sharpened Hauptsatz does not.
- Herbrand's Theorem can be extended to apply to ω-logic, this gives Kreisel's No Counterexample Interpretation, but in the case of the Sharpened Hauptsatz this is not possible since the notion of a midsequent is problematic in the presence of infinitary rules (see below).
- The (proof of the) Sharpened Hauptsatz is more "proof-theoretic" than (that of) Herbrand's Theorem which smacks of semantics: it is[27] natural in trying to understand the *champs finis* (which tells us how to obtain the F_is) to think of models.
- It seems impossible, or, more accurately, nobody has tried, to extend Herbrand's Theorem to modal logics as has been done for the Sharpened Hauptsatz.

More on the subject of the differences and similarities of the two results may be found in [15, 17, 62].

3.3 Applications of the Hauptsatz

3.3.1 Extraction of bounds and algorithms

Given a proof of $\forall x \exists y : R(x, y)$, we want to find a recursive function f such that $\forall x \exists y \leq f(x) : R(x, y)$. The connection with computation (programming) is obvious: render $\forall x \exists y : R(x, y)$ as "for all inputs x, there is an output y that satisfies specification R". Hence from a proof we can, in principle, extract an algorithm. The same techniques used also apply to the practice of everyday mathematics where the problem of extracting bounds ("unwinding theorems") is common.

In what follows, we are greatly simplifying the story; for a more complete presentation see [48]. Suppose we have a proof of $\vdash \forall x \exists y : R(x, y)$. The conclusion of this proof is

$$\frac{\vdash \exists y R(a, y)}{\vdash \forall x \exists y : R(x, y)}$$

[27] This may simply be something of a rather personal reaction.

Now consider the (sub)proof of $\vdash \exists y R(a,y)$. By lemma 1, we may suppose that from this we have a proof of[28] $\vdash \exists y R(\overline{n},y)$ for any $\overline{n}$. If we normalize this proof then by corollary 2 there are terms $t_1, ..., t_m$ such that $R(\overline{n},t_1) \vee ... R(\overline{n},t_m)$ is true. This means that the set $\{t_j \mid R(\overline{n},t_j) = \underline{true}\}$ is non-empty. From this set pick the term with the least numerical value; this is the value for $f(\overline{n})$. Here one clearly sees the link between logic and computation: we plug in some input value, $\overline{n}$, run through some algorithm (normalization), and read off an output value, $f(\overline{n})$. Although it is possible to extract f from a normalized proof (the process just outlined above will, generally, give a definition for f), for purely mechanical work cut elimination seems [29] more suited to *executing proofs* rather than *extracting programs.*

Completely worked out examples for extraction of bounds can be quite tedious for anything but the simplest cases; recall the nasty bounds for the size of cut-free proofs. Several mathematically non-trivial examples are discussed in [25, 32, 17]; these omit some details but would still be too long to present and explain here. Takasu [59] also presents (in all detail) a small example of how to use cut elimination to mechanically extract programs from proofs in system a that implements Gödel's functional interpretation for intuitionistic number theory.

An important point to note: straightforward normalization will remove even those cuts which have nothing to do with the computational content of the proof. Since this is a computationally expensive business, in practice one needs to be careful about the choice of the cuts to be eliminated [33, 34]. [Kreisel 1981a, 1981b].

3.3.2 Consistency of elementary number theory

Gentzen's proof of the consistency of elementary number theory is generally regarded as (one of) the most outstanding result(s) in proof theory: essentially, it saved proof theory after savage blow of Gödel's Incompleteness Theorems[30]. We shall sketch the main points of the proof below. This is based on Schutte's work using ω-logic [39, 47]; technically, this is more satisfying[31] than the exact proof of Gentzen who rejected such rules as the ω-rule[32]:

$$\frac{\Gamma \vdash A(\overline{0}), \Delta \quad \Gamma \vdash A(\overline{1}), \Delta \quad \Gamma \vdash A(\overline{2}), \Delta \quad ...}{\Gamma \vdash \forall x A(x), \Delta}$$

on ideological grounds connected with Hilbert's program.

Let PA_∞ be a system, with the ω-rule, for Peano arithmetic and PA_∞^{CF}, the corresponding system without the cut rule; see [Schwichtenburg 1977] for appropriate examples of PA_∞ and PA_∞^{CF}. Then it is provable that if a formula has a proof in PA, then it has a proof in PA_∞; this latter proof has height $< \omega \cdot 2$ and depth $< \omega$. Now Schutte's work shows that ω-logic enjoys full cut elimination (with a subformula property) and what has been done above is to embed Peano arithmetic in ω-logic. So by applying this cut elimination procedure, we obtain a corresponding proof in PA_∞^{CF}; this proof has height[33] $< \varepsilon_0$ and, furthermore, *this bound is minimal.* From the subformula property, it is easy to see that the empty sequent $\vdash$ is not provable in PA_∞^{CF} and the consistency of PA is thus established. One also sees where transfinite induction up to ε_0 comes in: it is used to prove the termination of the cut elimination process. Furthermore, this is quantifier-free transfinite induction. The question of whether or not Gentzen's proof is elementary or not then boils down to whether or not the use transfinite induction is elementary; see [6] for an elementary (no pun intended) discussion of this, as well as [17], the introduction of [54], and also [13].

[28] From here on $\overline{n}$ will denote the nth numeral, i.e., $Succ^n(0)$.
[29] To me at least ...
[30] Recall that the whole purpose of early proof theory was to vindicate Hilbert.
[31] One sees quite clearly how the transfinite ordinals get into the picture. It is certainly easier to follow!
[32] The notion of a midsequent is problematic here since a unary rule is required for that.
[33] ε_0 is the first ordinal α such that $\omega^\alpha = \alpha$.

Let us now examine what is special about this proof. This is discussed in detail in the works of Kreisel [28, 29, 30], who is largely responsible for redirecting the focus of attention in these proofs, and also, to some extent, by Girard [17]. A summary: no one has the slightest doubts about the consistency of this theory: we call it "elementary" because we have absolute faith in it. Therefore, the value of an (almost?) elementary consistency proof is not in boosting our confidence; it lies elsewhere. The correct question to ask then is not to what extent the proof makes us feel good (we already feel quite good about the theory) but, as Kreisel points out: *what more do we know when we have proved a result [by limited means] than if we merely know it to be true [or have proved it by more abstract methods]?* The answer is that in general we know more because we must prove more; in the example at hand, we know that our methods of proof must include at least quantifier-free[34] transfinite induction (not full transfinite induction!) up to ε_0 (not just any old well order); i.e., ε_0 is *the* ordinal of arithmetic. (Correspondingly larger ordinals appear with increasing logical complexity). Since this use of transfinite induction is not provable in elementary number theory, it easy to see that this can be regarded as a sharpening of Gödel's second Incompleteness Theorem; in fact this result says something quite close to a truly finitary consistency proof of number theory [17]. Additional details of what one gets from carrying out this proof may be found in [48]; these have to do with the nature of certain functions/functionals in PA/extensions of PA.

In addition to the references named above, relevant material is in [61] where Gentzen's original proof is analysed in great detail.

Note The system that is obtained by a straightforward addition to LK of sequents corresponding to Peano's axioms and using the induction rule:

$$\frac{A(a), \Gamma \vdash \Delta, A(Succ(a))}{A(0), \Gamma \vdash \Delta, A(t)}$$

where the variable a may occur in the term t.

will not do since one does not have a full cut elimination theorem here, this would imply a complete and consistent formalization of PA. Gentzen's original proof uses reductions (similar to those used to prove the Haupsatz) which add up to something weaker than cut elimination.

3.3.3 Miscellaneous

In this section, we briefly mention other situations where the Hauptsatz is of importance; typically, in proving some fundamental result.

3.3.3.1 The reflection schema

[15, 17]: It is trifling (mathematically although perhaps not philosophically) to say (soundness) that all theorems of PA are true, simply because we have agreed beforehand on the axioms we believe to be true and on the rules of inference we believe preserve truth; what the reflection schema do is to allow us to salvage something from this trifle by formalizing it [actually, slightly diluted] and applying it.

Let $Prov_{\mathcal{T}}(u, v)$ denote the fact that u is the Gödelnumber of a proof, in the theory $\mathcal{T}$ (which we assume includes PA), of the formula with Gödelnumber v, and $Tr_{\mathcal{T}}(w)$, that the formula with Gödelnumber w is true. Then a first attempt at reflection is:

Global Reflection Principle

$$\forall u \forall v [Prov_{\mathcal{T}}(u, v) \rightarrow Tr_{\mathcal{T}}(v)]$$

[34] This is particularly important; as Kreisel points out, it is this that removes (what would be) the silly part of the joke: "Gentzen is the guy who proved mathematical induction by using transfinite induction".

Now from Tarski's theorem, Tr cannot possibly exist so, instead, one usually works with the following ($\ulcorner ... \urcorner$ denotes the Gödelnumber of ...):

Local Reflection Principle

$$\forall u[Prov_{\mathcal{T}}(u, \ulcorner A \urcorner) \rightarrow A]$$

1st Uniform Reflection Principle

$$\forall u[Prov_{\mathcal{T}}(u, \ulcorner A(\overline{x}) \urcorner) \rightarrow A(x)]$$

2nd Uniform Reflection Principle

$$\forall u \forall x[Prov_{\mathcal{T}}(u, \ulcorner A(\overline{x}) \urcorner) \rightarrow A(x)]$$

An immediate explanation of the interest in reflection principles is that consistency is equivalent to a restricted reflection principle; for example, Gödel's Incompletenesss Theorems are easily stated in terms of reflection [49]. Also Hilbert's concerns can now be summed as follows: prove the local reflection principle for finitary statements.

Reflection principles form some of the fundamental tools of proof theory. Typical applications are [36]:

- for sound systems, they provide methods for constructing stronger systems
- they provide methods for comparing the strength of given formal systems

Typical results are:

Theorem 7 PA+*reflection schema* $\equiv$ PA+ε_0*-transfinite induction.*

Theorem 8 *The uniform reflection principle is provable in* PA *if* $\mathcal{T}$ *is predicate calculus or a finitely axiomatizable subsystem of* PA*.*

And following from this:

Theorem 9 PA *is not finitely axiomatizable.*

Theorem 10 PA *proves the consistency of all its finitely axiomatizable subsystems.*

Where does cut elimination come in? If we consider theorem 8, for example, and attempt to formalize the proof in a naive fashion, we immediately run into the limitations of Tarski's theorem. But by using the subformula property, a proof may be carried out by bounding the logical complexity of the formulae considered in a proof of $A(\overline{x})$.

3.3.3.2 No conterexample interpretation

No Counterexample Interpretation is the extension of Herbrand's Theorem to the case of ω-logic. We explain by considering a specific formula, $\exists x \forall y \exists z \forall t : R(x, y, z, t)$, without loss of generality. Herbrand's Theorem tells us that this formula is provable in predicate calculus iff there are terms, Us and Ws, such that $R(U_1, f(U_1), W_1, g(U_1, W_1)) \vee ... \vee R(U_n, f(U_n), W_n, g(U_n, W_n))$ is provable in propositional calculus. No Counterexample Interpretation tells us that it is provable iff there are functionals U and W such that $R(U(f, g), f(U(f, g)), W(f, g), g(U(f, g), W(f, g))$ is provable for every choice of f and g.

A fairly straightforward proof of No Counterexample Interpretation (from the Hauptsatz) can be found in [48]. Like Herbrand's Theorem and the Hauptsatz, the No Counterexample

Interpretation is also useful for the extraction of bounds/algorithms; a number of examples may be found in [25].

3.3.3.3 Decision procedures

Another application of the subformula property is in decision procedures. For propositional calculi, this is straightforward: each formula only has a finite number of subformulae to consider. This is of interest in formulating decision procedures for uneven calculi, such as intuitionistic and modal calculi, for which the standard truth table approach does apply; see [13, 17] for the outlines of such procedures for intuitionistic logic, and [63] for sequent-style formulations, together with cut elimination theorems, of several modal calculi. For a quantified formula, one has an infinite of subformulae but all is not lost for provable formulae (see [11]).

3.3.3.4 Craig's interpolation lemma

References here are [11, 17, 61].

Theorem 11 (Interpolation) *If $A \to B$ is LK-provable and A and B have at least one predicate in common, then there exists a formula I such that: $A \to I$ and $I \to B$ are LK-provable, and I contains only those predicates, constants, and variables that occur in both A and B.*

A constructive proof of this can be given from the Hauptsatz. The result is useful for proving such results as Beth's Definability Theorem and Robinson's Joint Consistency Theorem; the importance for a proof theorist here is (I think) the satisfaction of being able to prove standard results of logic (model theory) within a purely proof-theoretic framework. Also, for computation it is well known that the latter result can be used to combine decision procedures.

4 Bibliography

1. Andrews, P.B. 1986. *An Introduction to Mathematical Logic and Type Theory: To Truth through Proof.* Academic Press, Orlando, Florida.

2. Avron, A. 1990. Gentzen-type systems, resolution, and tableaux. Department of Computer Science, Tel Aviv University.

3. Beth, E.W. 1959. *The Foundations of Mathematics.* North-Holland, Amsterdam, The Netherlands.

4. Beeson, M.J. 1986. Proving programs and programming proofs. In: *Logic, Methodology and Philosophy of Science, VII*, B. Marcus et al, editors, North-Holland, Amsterdam, The Netherlands, pp. 51–82.

5. Bellini, G. 1990. Herbrand's theorem for sequents of LK and LJ. In: *Proceedings, 5th Scandinavian Logic Symposium* (Springer-Verlag, New York), pp 285–300.

6. DeLong, H. 1970. *A Profile of Mathematical Logic.* Addison-Wesley, Reading, Massachusetts.

7. Dershowitz, N., and M. Okada. 1988. Proof-theoretic techniques for term rewriting theory. In: *Proceedings, Logic in Computer Science Symposium*, pp. 104–111.

8. Dummett, M. 1977. *Elements of Intuitionism.* Oxford University Press, London, England.

9. Feferman, S. 1977. Review of PROOF THEORY by G. Takeuti. *Bulletin of the American Mathematical Society*, 83, 3:351–361.

10. Feferman, S. and W. Sieg. 1981. (Uses of) theories of inductive definitions. In: *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, W. Bucholz et al, editors, Springer-Verlag, New York, pp. 16–77.

11. Gallier, J.H. 1986. *Logic for Computer Science: Foundations of Automatic Theorem Proving.* Harper and Row, New York.

12. Gentzen, G. 1935. Investigations into logical deduction. In: [Szabo 1969], pp. 68–131.

13. Gentzen, G. 1936. The consistency of elementary number theory. In: [Szabo 1969], pp. 132–213.

14. Girard, J.-V. 1976. Three-valued logic and cut-elimination: the actual meaning of Takeuti's conjecture. *Dissertationes Mathematicae*, 136, Polish Academic Publishers, Warsaw, Poland.

15. Girard, J.-V. 1982. Herbrand's theorem and proof theory. In: *Proceedings of the Herbrand Symposium*, J. Stern, editor, North-Holland, Amsterdam, The Netherlands, pp. 29–38.

16. Girard, J.-V. 1982b. A survey of $\prod_2^1$ logic. In: *Logic, Methodology and Philosophy of Science, VII*, L.J. Cohen et al, editors, North-Holland, Amsterdam, The Netherlands, pp. 89–107.

17. Girard, J.-V. 1987. *Proof Theory and Logical Complexity.* Bibliopolis, Naples, Italy.

18. Goad, C.A. 1980. Proofs as descriptions of computation. In: *5th Conference on Automated Deduction*, Lecture Notes in Computer Science, no. 87, W. Bibel and R. Kowalski, editors, Springer-Verlag, New York, pp. 39–52.

19. Herbrand, J. 1930. Investigations in proof theory: the properties of true propositions. In: *From Frege to Gödel: A Source Book in Mathematical Logic*, Harvard University Press, J. van Heijenoort, editor, 1967, Cambridge, Massachusetts, pp. 525–582.

20. Hindley, J.R. and J.P. Seldin. 1986. *Introduction to Combinators and the λ-Calculus.* Cambridge University Press, Cambridge, England.

21. Howard, W.A. 1980. The formulae-as-types notion of construction. In: *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J.P. Seldin and J.R. Hindley, editors, Academic Press, Orlando, Florida, pp. 479–502.

22. Jervell, H.J. 1971. A normalform in first order arithmetic. In: *Proceedings, Second Scandinavian Logic Symposium*, J.E. Fenstad, editor, North-Holland, Amsterdam, The Netherlands, pp. 93–108.

23. Körner, S. 1960. *The Philosophy of Mathematics.* Dover Publications, New York.

24. Kleene, S.C. 1952. *Introduction to Metamathematics.* North-Holland, Amsterdam, The Netherlands.

25. Kreisel, G. 1952. On the interpretation of non-finitist proofs II. *Journal of Symbolic Logic*, 17:43–58.

26. Kreisel, G. 1958. Mathematical significance of consistency proofs. *Journal of Symbolic Logic*, 23:155–182.

27. Kreisel, G. 1964. Hilbert's programme. In: *Philosophy of Mathematics*, P. Benaceraf and H. Putnam, editors, Prentice-hall, Englewood Cliffs, New Jersey, pp. 157–180.

28. Kreisel, G. 1965. Mathematical logic. In: *Lectures on Modern Mathematics III*, T.L. Saaty, editor, Wiley, New York, 95–195.

29. Kreisel,G. 1968. A survey of proof theory. *Journal of Symbolic Logic*, 33, 3:321–288.

30. Kreisel,G. 1971. A survey of proof theory II. In: *Proceedings, Second Scandinavian Logic Symposium*, J.E. Fenstad, editor, North-Holland, Amsterdam, The Netherlands, pp. 109–170

31. Kreisel, G. 1976. What have we learned from Hilbert's second problem? In: *Proceedings of the Symposia in Pure Mathematics*, 28, American Mathematical Society, pp. 93–130.

32. Kreisel, G. 1977. On the kind of data needed for a theory of proofs. In: *Logic Colloquium 76*, R. Gandy and M. Hyland, editors, North-Holland, Amsterdam, The Netherlands, pp. 111–128.

33. Kreisel, G. 1981a. Neglected possibilities of processing assertions and proofs mechanically: choice of problems and data. In: *University-level Computer-assisted Instruction at Stanford: 1968–1980*, P. Suppes, editor, Stanford University. pp. 131–147.

34. Kreisel, G. 1981b. Extraction of bounds: interpreting some tricks of the trade. In: *University-level Computer-assisted Instruction at Stanford: 1968–1980*, P. Suppes, editor, Stanford University. pp. 149–163

35. Kreisel, G. 1985. Proof theory and the synthesis of programs: potential and limitations. In: *Proceedings, European Conference on Computer Algebra*, Lecture Notes in Computer Science, no. 203, B. Buchberger, editor, Springer-Verlag, New York, pp. 136–150.

36. Kreisel, G. and A. Levy. 1968. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift fur math. Logik und Grundlagen der Mathematik*, 14:97–142.

37. Martin-Löf, P. 1971. Hauptsatz for the theory of species. In: *Proceedings, Second Scandinavian Logic Symposium*, J.E. Fenstad, editor, North-Holland, Amsterdam, The Netherlands, pp. 217–233.

38. Martin-Löf, P. 1971. Hauptsatz for the intuitionistic theory of iterated inductive definitions. In: *Proceedings, Second Scandinavian Logic Symposium*, J.E. Fenstad, editor, North-Holland, Amsterdam, The Netherlands, pp. 179–216.

39. Pohlers, W. 1987. Contributions of the Schutte school in Munich to Proof Theory. In: [Takeuti 1987], pp. 406–431.

40. Prawitz, D. 1965. *Natural Deduction: A Proof-Theoretical Study.* Almqvist and Wiskell, Stockholm, Sweden.

41. Prawitz, D. 1968. Hauptsatz for higher order logic. *Journal of Symbolic Logic*, 33:452–457.

42. Prawitz, D. 1970. Some results for intuitionistic logic with second order quantification rules. In: *Intuitionism and Proof Theory*, A. Kino, J. Myhill, and R.E. Vesley, editors, North-Holland, Amsterdam, The Netherlands, 259–269.

43. Prawitz, D. 1971. Ideas and results in proof theory. In: *Proceedings, Second Scandinavian Logic Symposium*, J.E. Fenstad, editor, North-Holland, Amsterdam, The Netherlands, pp. 235–307.

44. Resnik, M.D. 1988. On the philosophical significance of consistency proofs. In: *Gödel's Theorem in Focus*, S.G. Shanker, editor, Croom Helm, London, England, pp. 115–131.

45. Scarpellini, B. 1970. On cut elimination in intuitionistic systems of analysis. In: *Intuitionism and Proof Theory*, A. Kino, J. Myhill, and R.E. Vesley, editors, North-Holland, Amsterdam, The Netherlands, 271–286.

46. Shoenfield, J.R. 1967. *Mathematical Logic.* Addison-Wesley, Reading, Massachusetts.

47. Schutte, K. 1977. *Proof Theory.* Springer-Verlag, New York.

48. Schwichtenburg, H. 1977. Proof theory: some applications of cut-elimination. In: *Handbook of Mathematical Logic*, J. Barwise, editor, North-Holland, Amsterdam, The Netherlands, pp. 867–896.

49. Smorynski, C. 1977. The incompleteness theorems. In: *Handbook of Mathematical Logic*, J. Barwise, editor, North-Holland, Amsterdam, The Netherlands, pp. 821–865.

50. Smullyan, R.M. 1968. Analytic cut. *Journal of Symbolic Logic*, 33, 4:560–564.

51. Statman, R. 1977. Herbrand's theorem and Gentzen's notion of a direct proof. In: *Handbook of Mathematical Logic*, J. Barwise, editor, North-Holland, Amsterdam, The Netherlands, pp. 897–912.

52. Statman, R. 1979. Lower bounds on Herbrand's Theorem. *Proceedings of the American Mathematical Society*, 75(1):104–107.

53. Stenlund, S. 1972. *Combinators, λ-Terms, and Proof Theory.* D. Reidel, Dordrecht, The Netherlands.

54. Szabo, M.E., editor. 1969. *The Collected Papers of Gerhard Gentzen.* North-Holland, Amsterdam, The Netherlands.

55. Szabo, M.E. 1978. *Algebra of Proofs.* North-Holland, Amsterdam, The Netherlands.

56. Tait, W. 1966. A non-constructive proof of Gentzen's Hauptsatz for second order predicate logic. *Bulletin of the American Mathematical Society*, 72:980–983.

57. Tait, W. 1970. Applications of the cut elimination theorem to some subsystems of analysis. In: *Intuitionism and Proof Theory*, North-Holland, Amsterdam, The A. Kino, J. Myhill, and R.E. Vesley, editors, Netherlands, pp. 475–488.

58. Takahashi, M. 1967. A proof of cut elimination in simple type theory. *Journal of the Mathematical Society of Japan*, 19:399-410.

59. Takasu, S. 1978. Programs and proofs. In: *Proceedings, Third IBM Symposium on Mathematical Foundations of Computer Science*, Kansai, Japan.

60. Takeuti, G. 1953. On a generalized logical calculus. *Japanese Journal of Mathematics*, 23:39–96.

61. Takeuti, G. 1987. *Proof Theory.* North-Holland, Amsterdam, The Netherlands.

62. van Heijenoort, J. 1985. Jacques Herbrand's work in logic and its historical context. In: *Selected Essays.* Bibliopolis, Naples, Italy, 88–121.

63. Zeman, J.J. 1973. *Modal Logic: The Lewis Modal Systems.* Oxford University Press, London, England.

64. Zucker, J. 1974. The correspondence between cut-elimination and normalization. *Annals of Mathematical Logic*, 7:1–155.