

# UNSOUND INFERENCES MAKE PROOFS SHORTER

JUAN P. AGUILERA AND MATTHIAS BAAZ

**ABSTRACT.** We give examples of calculi that extend Gentzen’s sequent calculus LK by unsound quantifier inferences in such a way that (i) derivations lead only to true sequents, and (ii) proofs therein are non-elementarily shorter than LK-proofs.

## 1. INTRODUCTION

Consider the following argument:

- (1) That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
- (2) Hence, that Kurt Gödel is Austrian entails that everyone is Austrian.
- (3) That is, if Kurt Gödel is Austrian, then all people are Austrian.
- (4) Therefore, there exists a person such that, if that person is Austrian, then all people are Austrian.

The argument can be formalized as an instance of the following proof schema:

$$(1) \quad \frac{\frac{\frac{A(a) \vdash A(a)}{A(a) \vdash \forall y A(y)}}{\vdash A(a) \rightarrow \forall y A(y)}}{\vdash \exists x (A(x) \rightarrow \forall y A(y))}$$

The study of formal proofs is motivated by our desire for deductive reasoning to be *correct*, i.e., we wish that it be such that the procedures involved derive only true conclusions. The traditional way of ensuring this involves restricting proofs to those satisfying the two following properties:

- Inferences are *sound*,<sup>1</sup> i.e., only true conclusions result from true premises.
- Derivations are *hereditary*, i.e., initial segments of proofs are proofs themselves.

In particular, one might impose certain *characteristic-variable conditions* that restrict the circumstances under which quantifiers can be inferred. We will use unsound rules based on weak characteristic-variable conditions to define enhanced (correct) logical calculi wherein derivations such as (1) will be allowed.

In our current treatment, the correctness of proofs can only be gauged by considering them in their entirety, as forfeiting soundness while maintaining correctness necessarily violates hereditariness. This is not unlike other aspects of reasoning that are traditionally not reflected in formal derivations. For example, hereditariness is not compatible with proofs by contradiction—one would not take as correct

---

<sup>1</sup>Soundness is usually applied to derivations or logical systems. In this paper, we distinguish ‘soundness’ from ‘correctness.’ See below.

an initial segment of a proof obtained by interrupting it after an assumption directed towards a contradiction had been made, but before the contradiction had been reached.

To illustrate this, we note that the following subproof of the opening example should not be allowed as a proof in the system:

1. That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
2. Hence, that Kurt Gödel is Austrian entails that everyone is Austrian.

The *strong* inference from 1 to 2 is the only unsound one:

1. That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
- 3' That is, if Kurt Gödel is Austrian, then Kurt Gödel is Austrian.
- 4' Therefore, there exists a person such that, if that person is Austrian, then Kurt Gödel is Austrian.

Here, the name ‘Kurt Gödel’ is simply manipulated syntactically, so its meaning plays no role. Moreover, the name does not even appear in sentence 4. It could well have been replaced by any other throughout the derivation with no effect—we call (the result of formalizing) this condition *substitutability*. In our example, it is what allows us to mend the proof after a ‘lie’ has been introduced in 2.

Formal proofs can often be taken without loss of generality to be *regular*, i.e., any variable which is the *characteristic variable*<sup>2</sup> of a quantifier inference can be assumed to appear only before then. Forfeiting this condition allows us to accept the following proof:

$$\begin{array}{c}
 \frac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \quad \frac{A(f(a)) \vdash A(f(a))}{\forall x A(x) \vdash A(f(a))} \\
 \hline
 \frac{A(a) \vdash A(f(a))}{\vdash A(a) \rightarrow A(f(a))} \\
 \hline
 \vdash \exists x (A(x) \rightarrow A(f(x)))
 \end{array}
 \tag{2}$$

In this example,  $a$  is the characteristic variable of only one quantifier inference. We call this *weak regularity*. Our main result is that a sequent calculus augmented with unsound quantifier inferences satisfying substitutability, weak regularity (or even a further weakening thereof), and a technical *side-variable condition* is correct. This means that, although inferences therein may not be sound, the three (global) conditions are enough to guarantee that any lie introduced during the proof is eventually cleared. Moreover, the calculus yields non-elementarily-shorter<sup>3</sup> cut-free proofs than standard proof systems. This is proved in Section 3. Some consequences of the speed-up theorem are explained in Section 4.

The three characteristic-variable conditions are motivated by the rules governing possible inferences in Hilbert’s  $\varepsilon$ -calculus, whose language contains no quantifiers (see Section 5). As a consequence, proofs therein are less restrictive. The weakened conditions thus result from attempting to pull back some extra freedom from the  $\varepsilon$ -calculus to first-order logic. They can also be thought of as a first attempt in tracking down and covering the spots where first-order logic proofs might break down without the usual quantifier-inference restrictions. In particular, it seems plausible that the conditions here can be further weakened, and the corresponding calculi can result even faster (see Section 6).

<sup>2</sup>See Section 2 for preliminaries and definitions.

<sup>3</sup>Size can be measured, e.g., by counting the number of inferences in the proof.

## 2. QUANTIFIER INFERENCES

We consider derivations in sequent calculi. Sequents are expressions of the form  $\Gamma \vdash \Delta$ , where  $\Gamma$  and  $\Delta$  represent collections of formulae. The sequent  $\Gamma \vdash \Delta$  is usually interpreted as ‘if all of  $\Gamma$  hold, then at least one of the formulae in  $\Delta$  holds.’ Equivalently, if  $\Gamma = A_0, A_1, \dots, A_n$  and  $\Delta = B_0, B_1, \dots, B_m$ , then  $\Gamma \vdash \Delta$  is interpreted as:

$$(3) \quad \bigvee_{i \leq n} \neg A_i \vee \bigvee_{i \leq m} B_i$$

Inferences are expressions of the form

$$\frac{\Gamma \vdash \Delta}{\Gamma' \vdash \Delta'}$$

which are to be interpreted as ‘the sequent  $\Gamma' \vdash \Delta'$  follows from the sequent  $\Gamma \vdash \Delta$ .’ Derivations are trees whose leaves are axioms and whose non-leaf nodes are sequents obtained from their predecessors by inferences; they are to be interpreted as *proofs* of the root sequent. In fact, we frequently refer to derivations as ‘proofs.’

If a formula is changed by an inference, we say that the formula is an *auxiliary* formula of the inference. If so, then the resulting formula in the conclusion is called the *critical* formula. An inference is a *quantifier inference* if the (only) critical formula has a quantifier as its outermost logical symbol. Normally, quantifier inferences consist of substituting a formula<sup>4</sup>  $Qx A(x)$  for an instance thereof in a sequent, e.g., as in

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

The *polarity* of a quantifier in a sequent  $\Gamma \vdash \Delta$  is defined as follows: rewrite  $\Gamma \vdash \Delta$  in the form (3) and then rewrite the resulting expression without implication symbols. A quantifier is *positive* if it is universal (resp. existential) and under the scope of an even (resp. odd) number of negation symbols, and *negative* otherwise. A quantifier is *strong* if it is positive and on the right-hand side of a sequent or negative and on the left-hand side of a sequent; it is *weak* otherwise.

If an inference yields a strongly-quantified formula  $Qx A(x)$  from  $A(a)$ , where  $a$  is a free variable, we say that  $a$  is the *characteristic variable* of the inference. We will denote free variables by letters  $a, b, c, \dots$  (in contrast, we denote bound variables by  $x, y, z, \dots$ ). We denote closed terms by  $t, s, r, \dots$  and variants thereof.

Let  $\pi$  be any derivation. We say  $b$  is a *side variable* of  $a$  in  $\pi$  (written  $a <_\pi b$ ) if  $\pi$  contains a strong-quantifier inference of the form:

$$\frac{\Gamma \vdash \Delta, A(a, b, \vec{c})}{\Gamma \vdash \Delta, \forall x A(x, b, \vec{c})}$$

or of the form:

$$\frac{A(a, b, \vec{c}), \Gamma \vdash \Delta}{\exists x A(x, b, \vec{c}), \Gamma \vdash \Delta}$$

The Skolemization of a first-order formula is defined by replacing every strongly quantified variable  $y$  with a new function symbol  $f_y(x_1, \dots, x_n)$ , where  $x_1, \dots, x_n$  are the weakly quantified variables such that  $Qy$  appears in the scope of their quantifiers, and removing the quantifier  $Qy$ . We write  $sk(A)$  for the Skolemization

<sup>4</sup>We will frequently use  $Q$  to denote any unspecified quantifier.

of  $A$ . Recall that whether a quantifier is weak or strong depends on the side of the sequent on which it is. Thus,  $sk(A)$  means different things depending on whether  $A$  appears in the antecedent or the succedent. The Skolemization of a sequent

$$A_0, \dots, A_m \vdash A_{m+1}, \dots, A_{m+n}$$

is defined as:

$$sk(A_0), \dots, sk(A_m) \vdash sk(A_{m+1}), \dots, sk(A_{m+n}).$$

In classical logic, a sequent is derivable if and only if its Skolemization is.

**Definition 2.1** (Suitable quantifier inference). We say a quantifier inference is *suitable for a proof*  $\pi$  if either it is a weak-quantifier inference, or the following three conditions are satisfied:

- (substitutability) the characteristic variable does not appear in the conclusion of  $\pi$ .
- (side-variable condition) the relation  $<_\pi$  is acyclic.
- (weak regularity) the characteristic variable is not the characteristic variable of another strong-quantifier inference in  $\pi$ .

We will work with various sequent calculi extending *Propositional* LK, the sequent calculus whose only axioms are  $\perp \vdash$  and  $A \vdash A$  (with  $A$  atomic) and whose rules are:

Structural rules:

$$\begin{array}{ll} \frac{\Gamma \vdash \Delta}{\Gamma, A \vdash \Delta} WL & \frac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, A} WR \\ \frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta} CL & \frac{\Gamma \vdash \Delta, A, A}{\Gamma \vdash \Delta, A} CR \\ \frac{\Gamma_1, A, B, \Gamma_2 \vdash \Delta}{\Gamma_1, B, A, \Gamma_2 \vdash \Delta} EL & \frac{\Gamma \vdash \Delta_1, A, B, \Delta_2}{\Gamma \vdash \Delta_1, B, A, \Delta_2} ER \\ \frac{\Gamma \vdash \Delta, A \quad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta} Cut \end{array}$$

Logical rules:

$$\begin{array}{ll} \frac{\Gamma, A \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge_1 L & \frac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, A \vee B} \vee_1 R \\ \frac{\Gamma, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta} \wedge_2 L & \frac{\Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \vee B} \vee_2 R \\ \frac{\Gamma, A \vdash \Delta \quad \Gamma, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta} \vee L & \frac{\Gamma \vdash \Delta, A \quad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B} \wedge R \\ \frac{\Gamma \vdash \Delta, A \quad \Gamma, B \vdash \Delta}{\Gamma, A \rightarrow B \vdash \Delta} \rightarrow L & \frac{\Gamma, A \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B} \rightarrow R \end{array}$$

Negation  $\neg A$  is defined as  $A \rightarrow \perp$ . The names of the structural rules stand, respectively, for ‘weakening,’ ‘contraction,’ and ‘exchange.’ First-order LK is the

extension of Propositional LK obtained by adding quantifier inferences:

$$\begin{array}{c} \frac{\Gamma, A(t) \vdash \Delta}{\Gamma, \forall x A(x) \vdash \Delta} \forall L \qquad \frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, \exists x A(x)} \exists R \\ \frac{\Gamma, A(a) \vdash \Delta}{\Gamma, \exists x A(x) \vdash \Delta} \exists L \qquad \frac{\Gamma \vdash \Delta, A(a)}{\Gamma \vdash \Delta, \forall x A(x)} \forall R \end{array}$$

with the only restriction that in the inferences  $\forall R$  and  $\exists L$ , the variable  $a$  is not allowed to appear in the conclusion. In particular, note that every quantifier inference is suitable for every regular LK-proof. Gentzen's famous *Cut-Elimination Theorem* states that the cut rule is redundant in both propositional and first-order LK.

**Definition 2.2** ( $\text{LK}^+$ ). The calculus  $\text{LK}^+$  is defined like LK, except that we instead allow all weak and strong quantifier inferences with the proviso that they be suitable for the proof.

A further weakening of the characteristic-variable conditions gives rise to the notion of weak suitability:

**Definition 2.3** (Weakly suitable quantifier inference). A quantifier inference is *weakly suitable for a proof  $\pi$*  if either it is a weak-quantifier inference or it satisfies substitutability, the side-variable condition, and:

- (very weak regularity) whenever the characteristic variable is also the characteristic variable of another strong-quantifier inference in  $\pi$ , then it has the same critical formula.

**Definition 2.4** ( $\text{LK}^{++}$ ). The calculus  $\text{LK}^{++}$  is the extension of  $\text{LK}^+$  that results from allowing all weakly suitable quantifier inferences.

It is easy to find examples of sequents that are more easily provable in  $\text{LK}^+$  than in LK:

*Example 2.5.* The sequent

$$\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B)$$

is provable in LK:

$$\begin{array}{c} \frac{A(a) \vdash A(a)}{A(a) \vdash A(a), B} \\ \hline \vdash A(a), A(a) \rightarrow B \\ \hline \vdash A(a), \exists x (A(x) \rightarrow B) \\ \hline \vdash \exists x (A(x) \rightarrow B), A(a) \\ \hline \vdash \exists x (A(x) \rightarrow B), \forall x A(x) \qquad B \vdash B \\ \hline \forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), B \\ \hline \forall x A(x) \rightarrow B, A(b) \vdash \exists x (A(x) \rightarrow B), B \\ \hline \forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), A(b) \rightarrow B \\ \hline \forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B), \exists x (A(x) \rightarrow B) \\ \hline \forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B) \end{array}$$

However, one can find a shorter  $\text{LK}^+$ -proof:

$$\begin{array}{c}
\frac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \quad B \vdash B \\
\hline
\frac{A(a), \forall x A(x) \rightarrow B \vdash B}{\forall x A(x) \rightarrow B, A(a) \vdash B} \\
\hline
\frac{\forall x A(x) \rightarrow B \vdash A(a) \rightarrow B}{\forall x A(x) \rightarrow B \vdash \exists x (A(x) \rightarrow B)}
\end{array}$$

□

Recall that a function on the natural numbers is *elementary* if it can be defined by a quantifier-free formula from  $+$ ,  $\cdot$ , and the function  $x \mapsto 2^x$ . By independent results of R. Statman [8] (which were formalized in [3]) and of V. P. Orevkov [7], the sizes of the smallest cut-free LK-proofs of sequents of length  $n$  are not bounded by any elementary function on  $n$ . Our main theorem is that cut-free LK<sup>+</sup>-proofs are non-elementarily shorter than cut-free LK-proofs:

**Theorem 2.6.** *There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK<sup>+</sup>-proof.*

An immediate consequence is the following:

*Corollary 2.7.* There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK<sup>++</sup>-proof.

We prove Theorem 2.6 in Section 3. First, we consider the question of the correctness of LK<sup>++</sup>. The proof of Theorem 2.8 that we present here is due to the referee:

**Theorem 2.8.** *If a sequent is LK<sup>++</sup>-derivable, then it is already LK-derivable.*

*Proof.* Let  $\pi$  be an LK<sup>++</sup>-proof. Replace every unsound universal quantifier inference by a  $\rightarrow L$  inference:

$$\frac{\Gamma \vdash \Delta, A(a) \quad \forall x A(x) \vdash \forall x A(x)}{\Gamma, A(a) \rightarrow \forall x A(x) \vdash \Delta, \forall x A(x)}$$

Similarly replace every unsound existential quantifier by an  $\rightarrow L$  inference

$$\frac{\exists x A(x) \vdash \exists x A(x) \quad A(a), \Gamma \vdash \Delta}{\Gamma, \exists x A(x), \exists x A(x) \rightarrow A(a) \vdash \Delta}$$

By doing this, we obtain a proof of the desired sequent, together with many formulae of the form  $A(a) \rightarrow \forall x A(x)$  or  $\exists x A(x) \rightarrow A(a)$  on the left-hand side. However, we can eliminate each of them by adding an existential quantifier inference and cutting with formulae of the form

$$\vdash \exists y (A(y) \rightarrow \forall x A(x))$$

or of the form

$$\vdash \exists y (\exists x A(x) \rightarrow A(y)),$$

both of which are easily derivable. Note that the existential quantifier inferences can be carried out in a way that is permissible by LK because the initial proof satisfied substitutability, weak regularity and the side-variable condition. □

*Corollary 2.9.* If a sequent is derivable in LK<sup>+</sup>, then it is already derivable in LK.

## 3. NON-ELEMENTARY SPEED-UP

Our strategy for proving Theorem 2.6 is to show that  $LK^+$  simulates a strong calculus that is already non-elementarily faster than  $LK$ . It is related to *quantifier shifts*, formulae such as:

$$(\forall x A \rightarrow B) \rightarrow \exists x (A \rightarrow B).$$

The main feature of the calculus is the addition of certain rules for quickly shifting quantifiers. Below, if  $A$  is a subformula of  $B$ , we say that  $B$  is a *context* of  $A$ . We generally write  $\kappa[A]$  to emphasize that  $A$  occurs in a context  $\kappa[\cdot]$ ; they should be thought as syntactical operators  $A \mapsto \kappa[A]$ .

**Definition 3.1.** The calculus  $LK_{\text{shift}}$  is obtained by extending  $LK$  with the following rules:

$$\begin{array}{c} \frac{\Gamma, \kappa[Qx A \triangleleft B] \vdash \Delta}{\Gamma, \kappa[Q'x (A \triangleleft B)] \vdash \Delta} \quad \frac{\Gamma, \kappa[A \triangleleft Qx B] \vdash \Delta}{\Gamma, \kappa[Q'x (A \triangleleft B)] \vdash \Delta} \\ \frac{\Gamma \vdash \Delta, \kappa[Qx A \triangleleft B]}{\Gamma \vdash \Delta, \kappa[Q'x (A \triangleleft B)]} \quad \frac{\Gamma \vdash \Delta, \kappa[A \triangleleft Qx B]}{\Gamma \vdash \Delta, \kappa[Q'x (A \triangleleft B)]} \end{array}$$

where  $\kappa[\cdot]$  is a context,  $\triangleleft \in \{\wedge, \vee, \rightarrow\}$  and  $Q' = Q$ , except if  $\triangleleft$  is  $\rightarrow$  and  $Q$  is taken from the antecedent, in which case  $Q'$  is opposite. We refer to these rules as *deep quantifier shifts*.

Here, recall that, as per our conventions, the syntax of first-order logic includes separate symbols for free and bound variables.

**Proposition 3.2.** *Cut-free  $LK^+$  simulates cut-free  $LK_{\text{shift}}$  double-exponentially, i.e., every  $LK_{\text{shift}}$ -provable sequent is  $LK^+$ -provable and there is a double exponential function that bounds the length of the least cut-free  $LK^+$ -proof of a  $LK^+$ -provable sequent in terms of its least cut-free  $LK_{\text{shift}}$ -proof.*

*Proof.* Let  $\pi$  be a cut-free  $LK_{\text{shift}}$ -proof of size  $n$ . We transform it into a cut-free  $LK^+$ -proof. Assume by induction that there is only one application of a quantifier shift and this is the last inference. We proceed by cases. For simplicity, we assume  $\kappa$  does not change the polarity of the quantifiers. We also assume  $\triangleleft$  is  $\rightarrow$ ; the other connectives are treated similarly.

CASE I. The last inference is:

$$\frac{\Gamma, \kappa[\forall x A(x) \rightarrow B] \vdash \Delta}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}$$

so that the proof has the following structure:

$$\begin{array}{c}
\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'' \vdash \Delta'', A(a)}}{\Gamma'' \vdash \Delta'', \forall x A(x)} (*)}{\frac{\frac{\frac{\vdots \sigma_1}{\Gamma' \vdash \Delta', \forall x A(x)}}{\Gamma', \forall x A(x) \rightarrow B \vdash \Delta'} \quad \frac{\frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \forall x A(x) \rightarrow B \vdash \Delta'}} \\
\frac{\frac{\vdots \sigma_0}{\Gamma, \kappa[\forall x A(x) \rightarrow B] \vdash \Delta}}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}
\end{array}$$

where the  $\sigma_i$  denote subproofs. The subproof  $\sigma_0$  might split into several branches not indicated in the diagram, each of which could potentially include its own copy of  $\forall x A(x)$  being inferred and its own subproof  $\sigma'_3$  with conclusion of the form  $\Pi, B \vdash \Delta$ . Each branch can be dealt with the same way and so we will only focus on the indicated parts of  $\pi$ .

Similarly, the subproof  $\sigma_1$  might split into several branches not indicated in the diagram, each of which could potentially include its own copy of  $\forall x A(x)$  being inferred. This we need to bear in mind. Note that each of those copies of  $\forall x A(x)$  is necessarily inferred from a different characteristic variable. In fact, the variable  $a$  does not appear below  $(*)$ , by the regularity of  $\text{LK}_{\text{shift}}$ .

We modify the proof. Our approach is as follows: we would like to merge the subproofs  $\sigma_1$  and  $\sigma_2$  simply by postponing the inference  $(*)$  as follows:

$$\begin{array}{c}
\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma' \vdash \Delta', A(a)}}{\Gamma', A(a) \rightarrow B \vdash \Delta'} \quad \frac{\frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \exists x (A(x) \rightarrow B) \vdash \Delta'} (**)} \\
\frac{\vdots \sigma_0}{\Gamma, \kappa[\exists x (A(x) \rightarrow B)] \vdash \Delta}
\end{array}$$

The problem that might arise is that some occurrence of  $\forall x A(x)$  that would be contracted in  $\sigma_1$  with the indicated occurrence is unable to be contracted. We describe how to deal with each of them.

Notice that the problematic occurrence of  $\forall x A(x)$  must originate from a strong-quantifier inference. Omit that inference and drag the unquantified formula  $A(b)$  until after  $(**)$ , so that at that point we have a derivation of

$$\Gamma'_0, A(b), \Gamma'_1, \exists x (A(x) \rightarrow B), \vdash \Delta',$$

where  $\Gamma' = \Gamma'_0, \Gamma'_1$ . Add to that some exchanges to obtain:

$$\Gamma', \exists x (A(x) \rightarrow B), A(b) \vdash \Delta',$$

and a subproof

$$(4) \quad \frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}$$



in order to infer

$$\Gamma', \exists x (A(x) \rightarrow B), A(b) \rightarrow B \vdash \Delta'.$$

with an application of  $\exists L$  and a contraction, we are left again with

$$\Gamma', \exists x (A(x) \rightarrow B) \vdash \Delta',$$

to which we can apply  $\sigma_0$  as desired. The proof grows by adding a copy of (4)—of size at most  $n$ —for each problematic occurrence of  $\forall x A(x)$  (of which there are at most  $n$ ). Hence, it grows quadratically.

CASE II. The last inference is:

$$\frac{\Gamma, \kappa[\exists x A(x) \rightarrow B] \vdash \Delta}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}$$

The proof has the following structure:

$$\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'' \vdash \Delta'', A(t)}}{\Gamma'' \vdash \Delta'', \exists x A(x)}}{\frac{\frac{\vdots \sigma_1}{\Gamma' \vdash \Delta', \exists x A(x)} \quad \frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}}{\Gamma', \exists x A(x) \rightarrow B \vdash \Delta'}} \quad \frac{\vdots \sigma_0}{\Gamma, \kappa[\exists x A(x) \rightarrow B] \vdash \Delta} \quad \frac{\vdots \sigma_0}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}$$

As in Case I, the subproof  $\sigma_0$  could branch off into several subtrees each of which is taken care of in the same way, and so we only focus on the indicated parts of  $\pi$ . We would like to merge the subproofs  $\sigma_1$  and  $\sigma_2$  as follows:

$$\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma' \vdash \Delta', A(t)}}{\Gamma', A(t) \rightarrow B \vdash \Delta'}}{\frac{\vdots \sigma_3}{\Gamma', B \vdash \Delta'}} \quad \frac{\vdots \sigma_0}{\Gamma, \kappa[\forall x (A(x) \rightarrow B)] \vdash \Delta}$$

As before, we face the problem of circumventing a contraction of (possibly several occurrences of)  $\exists x A(x)$  in  $\sigma_1$  and solve it in the same way. The proof grows quadratically again. We note the following: it might happen that the term  $t$  contains some free variable  $a$ . Hence, the proof of this case does not necessarily go through in LK, as  $a$  could be the characteristic variable of a strong-quantifier inference in  $\sigma_1$ . Note that the side-variable condition is satisfied.

CASE III. The last inference is:

$$\frac{\Gamma, \kappa[A \rightarrow \forall x B(x)] \vdash \Delta}{\Gamma, \kappa[\forall x (A \rightarrow B(x))] \vdash \Delta}$$

This is analogous to CASE II.

CASE IV. The last inference is:

$$\frac{\Gamma, \kappa[A \rightarrow \exists x B(x)] \vdash \Delta}{\Gamma, \kappa[\exists x (A \rightarrow B(x))] \vdash \Delta}$$

This is analogous to CASE I.

CASE V. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[\forall x A(x) \rightarrow B]}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}$$

The proof has the following form, modulo qualifications as in Cases I and II:

$$\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'', A(t) \vdash \Delta''}}{\Gamma'', \forall x A(x) \vdash \Delta''} (*)}{\frac{\frac{\vdots \sigma_1}{\Gamma', \forall x A(x) \vdash \Delta', B}}{\Gamma' \vdash \Delta', \forall x A(x) \rightarrow B}} \frac{\frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\forall x A(x) \rightarrow B]}}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}$$

We would like to postpone the inference (\*), thus merging  $\sigma_1$  and  $\sigma_2$ :

$$\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma', A(t) \vdash \Delta', B}}{\Gamma' \vdash \Delta', A(t) \rightarrow B} (**)}{\frac{\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B)}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]} \frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\exists x (A(x) \rightarrow B)]}$$

However, we might have—similarly to the previous cases—an occurrence of  $\forall x A(x)$  that would be contracted on the left-hand side as part of  $\sigma_1$ . For simplicity, assume there is only one. This occurrence is inferred from a formula  $A(s)$ . Postpone this occurrence until after (\*\*) so that at that point we have a derivation of

$$\Gamma'_0, A(s), \Gamma'_1 \vdash \Delta', \exists x (A(x) \rightarrow B),$$

where  $\Gamma' = \Gamma'_0, \Gamma'_1$ . Add to that some exchanges to obtain:

$$\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B),$$

and a weakening:

$$(5) \quad \frac{\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B)}{\Gamma', A(s) \vdash \Delta', \exists x (A(x) \rightarrow B), B} WR$$

in order to infer

$$\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B), A(s) \rightarrow B.$$

with an application of  $\exists R$  and a contraction, we are left again with

$$\Gamma' \vdash \Delta', \exists x (A(x) \rightarrow B),$$

to which we can apply the rest of  $\sigma_0$  as desired. In this case the proof only grows linearly. As in CASE II, the proof of this case would not necessarily work for LK, but it is correct from the point of view of  $LK^+$ , as follows from the regularity of  $LK_{\text{shift}}$  and the fact that the conclusion of the proof is  $\pi$ .

CASE VI. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[\exists x A(x) \rightarrow B]}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]}$$

The proof has the following form, modulo qualifications as in Cases I and II:

$$\frac{\frac{\frac{\frac{\vdots \sigma_2}{\Gamma'', A(a) \vdash \Delta''}}{\Gamma'', \exists x A(x) \vdash \Delta''} (*)}{\frac{\frac{\vdots \sigma_1}{\Gamma', \exists x A(x) \vdash \Delta', B}}{\Gamma' \vdash \Delta', \exists x A(x) \rightarrow B}}}{\frac{\frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\exists x A(x) \rightarrow B]}}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]}}$$

As before, we would like to postpone the inference  $(*)$ , thus merging  $\sigma_1$  and  $\sigma_2$ :

$$\frac{\frac{\frac{\frac{\vdots \sigma_2 + \sigma_1}{\Gamma', A(a) \vdash \Delta', B}}{\Gamma' \vdash \Delta', A(a) \rightarrow B}}{\Gamma' \vdash \Delta', \forall x (A(x) \rightarrow B)}}{\frac{\vdots \sigma_0}{\Gamma \vdash \Delta, \kappa[\forall x (A(x) \rightarrow B)]}}$$

We deal with contractions of  $\exists x A(x)$  on the left-hand side as in the previous case. The proof grows linearly.

CASE VII. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[A \rightarrow \forall x B(x)]}{\Gamma \vdash \Delta, \kappa[\forall x (A \rightarrow B(x))]}$$

This is analogous to CASE VI.

CASE VIII. The last inference is:

$$\frac{\Gamma \vdash \Delta, \kappa[A \rightarrow \exists x B(x)]}{\Gamma \vdash \Delta, \kappa[\exists x (A \rightarrow B(x))]}$$

This is analogous to CASE V.

Hence, we have dealt with all the cases. Note that for each quantifier shifted, the proof grows at most quadratically. Since there are at most  $n$  deep quantifier shifts, the size of the resulting proof is bounded by

$$n^{2^n} = 2^{2^n \cdot \log(n)} \approx 2^{2^n}.$$

This finishes the proof.  $\square$

Theorem 2.6 is a consequence of Proposition 3.2 and the following result:

**Theorem 3.3.** *There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK<sub>shift</sub>-proof.*

*Proof.* We will make use of a very specific family of sequents  $\{S_i\}_{i < \omega}$  described in [3] and due to Statman [8], and specific LK-proofs thereof. The sequents and the proofs themselves are not important for our proof. What is relevant is that they have the following properties:

- (1) the size of  $S_i$  is polynomial in  $i$ ;
- (2) there is no bound on the size of their smallest cut-free LK-proofs that is elementary in  $i$ ;
- (3) the size of these proofs (with cuts), however, is polynomially bounded in  $i$ ;
- (4) all cut formulae are closed; we can also assume they are prenex by, e.g., [3, Theorem 3.3].

Let  $\Gamma_i \vdash \Delta_i$  be one of the sequents. We modify the proof as follows: first, replace each cut

$$\frac{\Gamma \vdash \Delta, A \quad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta} \text{Cut}$$

with an application of  $\rightarrow L$ :

$$\frac{\Gamma \vdash \Delta, A \quad \Gamma, A \vdash \Delta}{\Gamma, A \rightarrow A \vdash \Delta} \rightarrow L$$

We are left with a cut free proof  $\pi_0$  whose end sequent is of the form:

$$(6) \quad A_0 \rightarrow A_0, \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i.$$

Choose any occurrence of a quantifier in  $A_0$  that is not in the scope of another quantifier. Since it appears once in the antecedent of  $A_0 \rightarrow A_0$  and once in the consequent, it appears once with each polarity. Because  $A_0$  is assumed to be prenex,

we can apply two deep quantifier shifts, so that we are left with a proof  $\pi'_0$  of the sequent:

$$(7) \quad \forall x_0^0 \exists x_1^0 (A'_0 \rightarrow A'_0), \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i,$$

Continue choosing any occurrence of a quantifier in  $A'_0$  that is not in the scope of another quantifier (within  $A'_0$ ) and applying deep quantifier shifts in pairs until we obtain a proof  $\pi_1$  of the sequent:

$$\forall x_0^0 \exists x_1^0 \dots (\hat{A}_0 \rightarrow \hat{A}_0), \dots, A_m \rightarrow A_m, \Gamma_i \vdash \Delta_i,$$

where  $\hat{A}_0$  is quantifier-free and the prefix of  $\hat{A}_0 \rightarrow \hat{A}_0$  consists of alternating quantifiers. Repeat this procedure for each of the  $A_i$  to obtain a proof  $\pi_m$  where each formula  $A_i \rightarrow A_i$  is replaced by an expression of the form:

$$\forall x_0^i \exists x_1^i \dots (\hat{A}_i \rightarrow \hat{A}_i),$$

where  $\hat{A}_i$  is quantifier free and the quantifier prefix is alternating. Let  $\hat{\Gamma}^i \vdash \hat{\Delta}^i$  be the sequent:

$$(8) \quad \forall x_0^0 \exists x_1^0 \dots (\hat{A}_0 \rightarrow \hat{A}_0), \dots, \forall x_0^m \exists x_1^m \dots (\hat{A}_m \rightarrow \hat{A}_m), \Gamma_i \vdash \Delta_i.$$

Note that, since the size of the initial proof was polynomial in  $i$ , there were polynomially many quantifiers in the proof; hence, we added only polynomially many deep quantifier shifts, and so the size of the resulting proof  $\pi_m$  of (8) is bounded polynomially in  $i$ . Moreover, it is cut-free. Consequently, it suffices to show:

**Claim 3.4.** *There is no elementary function bounding the size of the smallest cut-free LK-proofs of (8).*

*Proof.* Let  $\{\sigma_i\}_{i < \omega}$  be a sequence of such proofs. First, we transform it into a sequence of proofs of the Skolemizations of the sequents (8), so that for each sequent, each of the implications  $\hat{A}_i \rightarrow \hat{A}_i$  remains only universally quantified. Then, by Herbrand's theorem, there are propositional proofs  $\{\theta_i\}_{i < \omega}$  of the Herbrand sequents of (8), each of which is of the form:

$$(9) \quad \bigwedge_k B_k^0, \dots, \bigwedge_k B_k^m, \Gamma' \vdash \Delta'.$$

Moreover, the lengths of the Herbrand sequents are bounded exponentially by the lengths of  $\{\theta_i\}_{i < \omega}$  (see [3, Theorem 4.3]). Each conjunct  $B_k^p$  is a quantifier-free implication of the form:

$$(10) \quad A(t_1, \dots, t_l) \rightarrow A(s_1, \dots, s_l),$$

for some terms  $t_1, \dots, t_l, s_1, \dots, s_l$ . The key point is that, in the process of Skolemizing the sequents (8), exactly one term in each pair  $(t_j, s_j)$  was weakly quantified and the other (which was under the scope of the strong quantifier) was replaced by a Skolem function. Hence, either  $t_j$  is of the form  $f(s_j, \vec{r})$  or  $s_j$  is of the form  $f(t_j, \vec{r})$ , for some tuple of terms  $\vec{r}$ .

Transform the sequent (9) as follows: pick a pair  $(t_j, s_j)$  of some disjunct  $B_k^p$  such that the element thereof that was replaced by a Skolem term—say,  $f(s_j, \vec{r}) = t_j$ —does not have any other of  $t_1, \dots, t_l, s_1, \dots, s_l$  as an argument. Such a term of course exists—it is the first term whose quantifier was shifted in (7). Substitute  $s_j$  for  $t_j$  throughout the sequent.

By repeating this process sufficiently-many times, each formula (10) is transformed into a propositional tautology of the form  $A(\vec{t}) \rightarrow A(\vec{t})$ . This means that the sequent

$$\vdash \bigwedge_k B_k^0, \dots, \bigwedge_k B_k^m.$$

is also a propositional tautology, whereby so too is  $\Gamma' \vdash \Delta'$ , from which follows that it is LK-provable. By [2, Theorem 2], there is an LK-proof of the unskolemized sequent of length exponential in that of  $\Gamma' \vdash \Delta'$ . But this is impossible, since—by assumption—there exist no short proofs of the unskolemized sequent.  $\square$

This establishes Theorem 3.3 and thus Theorem 2.6.  $\square$

#### 4. PROPERTIES OF THE CALCULI

We discuss some properties of the calculi augmented with unsound inferences. Some of the arguments are only sketched—we leave the details to the interested reader.

**4.1.  $\text{LJ}^+$  and  $\text{LJ}^{++}$ .** Consider  $\text{LJ}^+$  and  $\text{LJ}^{++}$ , the analogs of the extended calculi  $\text{LK}^+$  and  $\text{LK}^{++}$  for intuitionistic logic. Specifically, these are the calculi obtained by restricting possible sequents to those with at most one formula on the right-hand side. The calculi are not sound for intuitionistic logic, as shown by any of the examples in the introduction. A consequence of this that they do not admit cut elimination. The argument makes use of an important consequence of cut elimination, *subformula property*—every formula appearing in a cut-free proof of a sequent  $S$  is a subformula of a formula in  $S$ .

**Proposition 4.1.**  *$\text{LJ}^+$  and  $\text{LJ}^{++}$  do not admit cut elimination.*

*Proof.* Consider example (2). Since only one formula appears on the right-hand side of each sequent, this is an  $\text{LJ}^+$ -proof. Suppose towards a contradiction that there were a cut-free proof of

$$(11) \quad \vdash \exists x (A(x) \rightarrow A(f(x))).$$

By the subformula property, this derivation would consist entirely of subformulae of (11). In particular, it would contain no strong quantifier inferences. Therefore, it would already be an LJ-proof. But this is impossible, as (11) is not intuitionistically valid.  $\square$

Proposition 4.1 has the consequence that  $\text{LK}^+$  and  $\text{LK}^{++}$  do not admit cut elimination by an algorithm that resembles Gentzen's. We give a rough definition of what we mean by this, but the reader may consult the appendix of [1] for further details.

**Definition 4.2.** We say that a cut-eliminating procedure is *Gentzen-style* if it is a transformation of proofs consisting of permutation of rules, substitution of free variables, reduction of cuts of a formula to cuts of its outermost subformulae, and absorption of axioms, i.e., elimination of cuts

$$\frac{\Gamma \vdash \Delta, A \quad A \vdash A}{\Gamma \vdash \Delta, A} \text{Cut}$$

by deleting the indicated occurrence of  $A \vdash A$ .

The point in considering Gentzen-style cut elimination algorithms is that they transform intuitionistic proofs into cut-free intuitionistic proofs. Thus, we obtain:

*Corollary 4.3.*  $LK^{++}$  and  $LK^+$  do not admit any Gentzen-style cut elimination.

As is well known, the following three formulae are the only quantifier shifts not derivable in LJ:

- (1)  $\forall x (A \vee B(x)) \vdash A \vee \forall x B(x)$ ;
- (2)  $(\forall x A(x) \rightarrow B) \vdash \exists x (A(x) \rightarrow B)$ ;
- (3)  $(A \rightarrow \exists x B(x)) \vdash \exists x (A \rightarrow B(x))$ .

They give rise to a characterization of  $LJ^{++}$ :

**Proposition 4.4.** *A sequent is provable in  $LJ^{++}$  if and only if it is provable in LJ with all quantifier shifts added as axioms.*

*Proof.* All three quantifier shifts 1–3 are provable in  $LJ^{++}$  (see, e.g., Example 2.5). We show the converse. Clearly, the sequents  $\vdash \forall y A(y) \rightarrow \forall x A(x)$  and  $\vdash \exists y A(y) \rightarrow \exists x A(x)$  are provable. Hence, using the quantifier shifts 2 and 3 one derives  $\vdash \exists y (A(y) \rightarrow \forall x A(x))$  and  $\vdash \exists y (\exists x A(x) \rightarrow A(y))$ . Therefore, we can apply the proof of Theorem 2.8 to obtain the desired result.  $\square$

**4.2. Skolemization.** The following are applications of (the proof of) Theorem 2.6:

**Proposition 4.5.** *If  $\Gamma \vdash \Delta$  is  $LK^+$ -derivable (resp.  $LK^{++}$ -derivable), then its Skolemization is  $LK^+$ -derivable (resp.  $LK^{++}$ -derivable) in quadratically many steps using additional cuts.*

*Proof.* Let  $\Gamma \vdash \Delta$  be any  $LK^{++}$ -derivable sequent. For each formula  $A$  in  $\Delta$ , the sequent  $A \vdash sk(A)$  is already  $LK$ -derivable, it follows that, by using one cut per formula, we can derive  $\Gamma \vdash sk(\Delta)$  in  $LK^{++}$ . Similarly,  $sk(A) \vdash A$  is already  $LK$ -derivable. Hence, by also using one cut for each formula in  $\Gamma$ , we can derive  $sk(\Gamma) \vdash sk(\Delta)$ .  $\square$

However, as we see, the additional cuts are necessary:

**Proposition 4.6.** *Let  $S$  be a sequent. Then, there is no elementary bound on the length of the smallest cut-free  $LK^{++}$ -proof of the Skolemization of  $S$  in terms of the smallest cut-free  $LK^+$ -proof of  $S$ .*

*Proof.* We argue as in the proof of Claim 3.4: if the Skolemization of  $\Gamma \vdash \Delta$  were cut-free  $LK^{++}$ -derivable in elementary many steps, the cut-free  $LK^{++}$ -proof would already be an  $LK$ -proof, whence by [2, Theorem 2], there would be an  $LK$ -proof of  $\Gamma \vdash \Delta$  of elementary length with respect to the  $LK^+$ -proof.  $\square$

Note, in contrast, that a cut-free proof can be Skolemized in  $LK$  at no extra cost.

**Proposition 4.7.** *There is no elementary bound on the length of the Herbrand sequent of a sequent in terms of its smallest cut-free  $LK^+$ -proof (resp.  $LK^{++}$ -proof).*

*Proof.* This follows from Proposition 4.6, as a cut-free  $LK^{++}$ -proof of the Skolemization of a sequent contains no strong quantifier and is hence already an  $LK$ -proof, whence we can apply Herbrand's theorem.  $\square$

Although deskolemization of cut-free proofs is exponential in  $LK$ , it is linear in  $LK^{++}$ :

**Proposition 4.8.** *If the Skolemization of  $\Gamma \vdash \Delta$  is cut-free  $\text{LK}^{++}$ -derivable, then  $\Gamma \vdash \Delta$  is cut-free  $\text{LK}^{++}$ -derivable in linearly many steps.*

*Proof.* Let  $\pi$  be a cut-free  $\text{LK}^{++}$ -proof of the Skolemization of  $\Gamma \vdash \Delta$ . By definition, the sets of Skolem functions assigned to each formula in  $\Gamma \vdash \Delta$  are disjoint. Modify  $\pi$  into a proof of  $\Gamma \vdash \Delta$  by replacing each Skolem term  $t$  by a new free variable  $a_t$  and quantifying it as soon as it is introduced. This might not have been possible in  $\text{LK}$ , but it is in  $\text{LK}^{++}$ : substitutability clearly holds, as does weak regularity. The side-variable condition must also hold, since every Skolem function has only finite arity.  $\square$

*Remark 4.9.* The proof of Proposition 4.8 also yields the intuitionistic analog.

## 5. CONNECTIONS TO THE EPSILON CALCULUS

We start by recalling the sequent-calculus reformulation of Hilbert's  $\varepsilon$ -calculus (see [5]).

**Definition 5.1.** The quantifier-free language  $\mathcal{L}_\varepsilon$  is obtained by removing symbols  $\exists$  and  $\forall$  from the language of first-order logic and adding symbols  $\varepsilon$  and  $\tau$ . Formulae and terms are simultaneously defined by induction in a way that the following clauses are satisfied. We leave the precise recursion to the reader.

- (1) constants and free variables are terms;
- (2) if  $t_1, \dots, t_n$  are terms and  $f$  is an  $n$ -ary function symbol, then  $f(t_1, \dots, t_n)$  is a term;
- (3) if  $A(t)$  is a formula,  $t$  is a term, and  $x$  is a bound variable, then  $\varepsilon_x A(x)$  and  $\tau_x A(x)$  are terms.
- (4) if  $t_0, \dots, t_n$  are terms and  $A$  is an  $n$ -ary predicate symbol, then  $A(t_0, \dots, t_n)$  is a formula;
- (5) if  $A$  and  $B$  are formulae, then  $A \wedge B$ ,  $A \vee B$ ,  $\neg A$ , and  $A \rightarrow B$  are formulae.

The  $\varepsilon$ -sequent calculus  $\text{LK}^\varepsilon$  is obtained by adding to Propositional  $\text{LK}$  the rules:

$$\frac{\Gamma, A(t) \vdash \Delta}{\Gamma, A(\tau_x A(x)) \vdash \Delta} \tau \qquad \frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, A(\varepsilon_x A(x))} \varepsilon$$

The term  $\varepsilon_x A$  is to be understood as a ‘generic’ object satisfying property  $A$ . The dual term  $\tau_x A$  is added for symmetry; it is in fact redundant—one can define  $A(\tau_x A) = A(\varepsilon_x (\neg A))$ . The *standard translation* of first-order logic into epsilon calculus is defined recursively by mapping atomic formulae to themselves, respecting Booleans, and dealing with quantified formulae as follows: suppose  $A'$  is the standard translation of  $A$ , then we map:

$$\exists x A(x) \mapsto A'(\varepsilon_x A'(x)), \qquad \forall x A(x) \mapsto A'(\tau_x A'(x)).$$

**Proposition 5.2.** *A sequent is  $\text{LK}^\varepsilon$ -derivable if and only if the standard translation of its Skolemization is  $\text{LK}^\varepsilon$ -derivable.*

In the  $\text{LK}^\varepsilon$ -proof of the standard translation of a sequent,  $\varepsilon$ - and  $\tau$ -rules play the role of weak quantifier inferences, while the substitution of variables for  $\varepsilon$ - and  $\tau$ -terms throughout the proof plays the role of strong quantifier inferences. The proof of Proposition 5.2 relies on the *Extended First Epsilon Theorem* (see [6, Theorem 16]). It states the following: let

$$(12) \qquad A_0, \dots, A_m \vdash A_{m+1}, \dots, A_{m+n}$$



be an  $\text{LK}^\varepsilon$ -provable sequent all of whose occurrences of  $\varepsilon$  and  $\tau$  are *weak* (this is defined analogously as for quantifiers in first-order logic) and let  $\vec{t}^i$  be all the terms of the form  $\tau_x A(x)$  or  $\varepsilon_x A(x)$  in  $A_i$ . Then there exist finitely-many tuples of terms  $\vec{t}^i_j$  without any occurrence of the symbols  $\tau$  or  $\varepsilon$  such that if  $A_j^i$  is the result of substituting the terms  $\vec{t}^i_j$  for  $\vec{t}^i$  in  $A_i$ , then the Herbrand sequent

$$\bigwedge_i A_0^i, \dots, \bigwedge_i A_m^i \vdash \bigvee_i A_{m+1}^i, \dots, \bigvee_i A_{m+n}^i$$

is a propositional tautology. If so, then the sequent resulting from ‘shifting’ the disjunctions and conjunctions inside each formula to the subformula where the  $\varepsilon$ - and  $\tau$ -terms appeared originally is also a propositional tautology. Call this sequent  $S$ . Assuming (12) was the standard translation of the Skolemization of a sequent  $\Gamma \vdash \Delta$ , one can then use  $S$  to obtain an LK-proof of  $\Gamma \vdash \Delta$ . Let us illustrate this with an example:

*Example 5.3.* Consider the following proof of a sequent whose only occurrence of  $\tau$  is weak:

$$\begin{array}{c} \frac{A(\tau_x(A(x) \rightarrow B)) \vdash A(\tau_x(A(x) \rightarrow B))}{\frac{A(\tau_x A(x)) \vdash A(\tau_x(A(x) \rightarrow B))}{\frac{A(\tau_x A(x)) \vdash B, A(\tau_x(A(x) \rightarrow B))}{\vdash A(\tau_x A(x)) \rightarrow B, A(\tau_x(A(x) \rightarrow B))} \quad B \vdash B} \\ \frac{A(\tau_x(A(x) \rightarrow B)) \rightarrow B \vdash A(\tau_x A(x)) \rightarrow B, B}{\frac{A(\tau_x(A(x) \rightarrow B)) \rightarrow B, A(\tau_x A(x)) \vdash A(\tau_x A(x)) \rightarrow B, B}{\frac{A(\tau_x(A(x) \rightarrow B)) \rightarrow B \vdash A(\tau_x A(x)) \rightarrow B, A(\tau_x A(x)) \rightarrow B}{A(\tau_x(A(x) \rightarrow B)) \rightarrow B \vdash A(\tau_x A(x)) \rightarrow B}} \end{array}$$

The corresponding Herbrand sequent is

$$(13) \quad A(a) \rightarrow B \vdash (A(a) \rightarrow B) \vee (A(b) \rightarrow B),$$

which is a propositional tautology. Moreover, the result of shifting the disjunction to where the  $\varepsilon$ -term originally appeared, namely,

$$A(a) \rightarrow B \vdash A(a) \wedge A(b) \rightarrow B,$$

is also a propositional tautology. Here, the disjunction is replaced by a conjunction, as the  $\varepsilon$ -term appeared in the antecedent of an implication. The conclusion of the proof is the translation of the Skolemized sequent

$$\forall x (A(x) \rightarrow B) \vdash \forall x A(x) \rightarrow B,$$

and so this is LK-provable by Proposition 5.2. Compare this with

$$\forall x (A(x) \rightarrow B) \vdash \exists x (A(x) \rightarrow B),$$

which is the Skolemized sequent suggested by (13).

There is a very intimate connection between  $\text{LK}^{++}$  and the  $\varepsilon$ -calculus. In fact, the characteristic-variable conditions of  $\text{LK}^{++}$  are precisely chosen so as to ensure that we have the following result:

**Proposition 5.4.** *If a sequent has an  $\text{LK}^{++}$ -proof of length  $k$ , then its standard translation has an  $\text{LK}^\varepsilon$ -proof of length  $\leq k$ .*

The proof proceeds by replacing weak-quantifier inferences with  $\varepsilon$ - or  $\tau$ -rules and omitting strong quantifier inferences—instead, we substitute the appropriate  $\varepsilon$ - or  $\tau$ -term for the characteristic variable throughout the proof.

Proposition 5.4 readily yields an alternative proof of Theorem 2.8: let  $\Gamma \vdash \Delta$  be an  $\text{LK}^{++}$ -provable sequent. Then its Skolemization is  $\text{LK}^{++}$ -provable by Proposition 4.5. But the Skolemization is common to  $\text{LK}$  and preserves validity, so by Propositions 5.2 and 5.4,  $\Gamma \vdash \Delta$  is  $\text{LK}$ -provable.

One can state the relationship between the  $\varepsilon$ -calculus and  $\text{LJ}^{++}$ .

**Proposition 5.5.** *A sequent is  $\text{LJ}^{++}$ -provable if and only if its standard translation is  $\text{LJ}^\varepsilon$ -provable.*

*Proof.* Suppose  $\Gamma \vdash D$  is  $\text{LJ}^{++}$ -provable. By Proposition 4.4 it is then  $\text{LJ}$ -provable by adding quantifier shifts as axioms. As in Proposition 5.4, the standard translation of  $\Gamma \vdash D$  is  $\text{LJ}^\varepsilon$ -provable by adding the standard translations of quantifier shifts as axioms. However, the standard translations of quantifier shifts are already  $\text{LJ}^\varepsilon$ -provable.

Conversely, suppose the standard translation of  $\Gamma \vdash D$  is  $\text{LJ}^\varepsilon$ -provable, say by a proof  $\pi$ . We want to directly translate  $\pi$  into an  $\text{LJ}^{++}$ -proof. The problem is that  $\pi$  might have inferences arranged the following way:

$$\frac{\frac{\vdots}{\Pi \vdash B(s)}}{\Pi \vdash B(\varepsilon_x B(x))} \quad \frac{\frac{\vdots}{\Pi' \vdash A(\varepsilon_x B(x))}}{\Pi' \vdash A(\varepsilon_x A(x))}$$

To rid ourselves of the problem, we will transform  $\pi$  into a cut-free proof where no inferred  $\varepsilon$ - or  $\tau$ -term is modified by an  $\varepsilon$ - or  $\tau$ -inference. We do this by replacing occurrences of  $A(\varepsilon_x A)$  by  $A(f_A(x))$ , for Skolem functions  $f_A$ . Suppose  $A(t)$  is a cut formula in  $\pi$  and  $t$  is a term of the form  $\varepsilon_x B$  that is inferred in  $\pi$ . For example:

$$\frac{\frac{\frac{\vdots}{\Pi \vdash B(s)}}{\Pi \vdash B(\varepsilon_x B(x))} \quad \frac{\frac{\vdots}{\Pi' \vdash A(\varepsilon_x B(x))} \quad \frac{\pi_0 \vdots}{A(\varepsilon_x B(x)), \Pi' \vdash E'}}{\Pi' \vdash E'}$$

Modify  $\pi_0$  into a proof  $\pi_1$  as follows: the indicated occurrence of  $\varepsilon_x B(x)$  must have been either introduced by weakening—in which case we modify the weakening so that  $f_B(s)$  is introduced instead—or by an axiom:

$$\frac{C(\varepsilon_x B(x)) \vdash C(\varepsilon_x B(x))}{\vdots} \quad \frac{\vdots}{A(\varepsilon_x B(x)), \Pi' \vdash E'}$$

in which case we replace it by:

$$\frac{\frac{C(f_B(s)) \vdash C(f_B(s)) \quad C(\varepsilon_x B(x)) \vdash C(\varepsilon_x B(x))}{C(f_B(s)) \rightarrow C(\varepsilon_x B(x)), C(f_B(s)) \vdash C(\varepsilon_x B(x))}}{\vdots} \frac{}{A(f_B(s)), \Pi'' \vdash E'}$$

where  $\Pi''$  has the form  $C(f_B(s)) \rightarrow C(\varepsilon_x B(x)), \Pi'$ . We modify  $\pi$ —say, in the latter case—as follows:

$$\frac{\frac{\frac{\vdots}{\Pi \vdash B(s)} \quad \frac{\vdots}{B(f_B(s)) \vdash B(f_B(s))}}{B(s) \rightarrow B(f_B(s)), \Pi \vdash B(f_B(s))}}{\vdots} \frac{\frac{\vdots}{B(s) \rightarrow B(f_B(s)), \Pi' \vdash A(f_B(s))} \quad \frac{\pi_1 \vdots}{A(f_B(s)), \Pi'' \vdash E'}}{C(f_B(s)) \rightarrow C(\varepsilon_x B(x)), B(s) \rightarrow B(f_B(s)), \Pi' \vdash E'} \vdots$$

Repeat this for every weak  $\varepsilon$ - and  $\tau$ -term inferred in each cut formula. We then make a similar modification to the proof for each inferred  $\varepsilon$ - or  $\tau$ -term that is modified by an  $\varepsilon$ - or  $\tau$ -inference. For example,

$$\frac{\frac{\vdots}{\Pi, \vdash B(s)}}{\Pi, \vdash B(\varepsilon_x A(x))} \frac{\vdots}{\Pi', \vdash A(\varepsilon_x B(x))} \frac{}{\Pi', \vdash A(\varepsilon_x A(x))}$$

becomes

$$\frac{\frac{\vdots}{\Pi, \vdash B(s)} \quad B(f_B(s)) \vdash B(f_B(s))}{B(s) \rightarrow B(f_B(s)), \Pi, \vdash B(f_B(s))} \frac{\vdots}{B(s) \rightarrow B(f_B(s)), \Pi', \vdash A(f_B(s))} \frac{}{B(s) \rightarrow B(f_B(s)), \Pi', \vdash A(\varepsilon_x A(x))}$$

After this, replace all cuts with  $\rightarrow L$  rules as in the proof of Theorem 3.3 and add a  $\tau$ -rule for each Skolem term and each  $\varepsilon$ - and  $\tau$ -term in each of the introduced implications.

This yields a proof of a sequent  $A'_1, \dots, A'_n, \Gamma' \vdash D'$  such that:

- (1)  $\Gamma'$  and  $D'$  are respectively the standard translations of  $\Gamma$  and  $D$ ;

- (2) Each  $A'_i$  is an implication of the form  $A \rightarrow A$  with only  $\tau$ -terms, and so the standard translation of an implication of the form  $A \rightarrow A$  preceded by a block of universal quantifiers (in particular,  $A'_1, \dots, A'_n, \Gamma' \vdash D'$  is the standard translation of a sequent  $A_1, \dots, A_n, \Gamma \vdash D$ );
- (3) No inferred  $\varepsilon$ - or  $\tau$ -term is modified by an  $\varepsilon$ - or  $\tau$ -inference in the proof.

Using this, one can translate back the proof to an  $\text{LJ}^{++}$ -proof of

$$A_1, \dots, A_n, \Gamma \vdash D$$

sequent by sequent. This is done by replacing  $\varepsilon$ - and  $\tau$ -rules with weak quantifier inferences and introducing strong quantifier inferences whenever a formula is the  $\varepsilon$ -translation of a strongly quantified formula. The point is that the  $\varepsilon$ - and  $\tau$ -terms that were not inferred in the proof can be replaced by free variables. Finally, we obtain a proof of  $\Gamma \rightarrow D$  by cutting the  $A_i$ . We leave the details to the reader.  $\square$

## 6. CONCLUDING REMARKS

Many related questions remain open. Perhaps one should more thoroughly investigate the role that (individual) inferences (should) play in proofs, both from a philosophical and from a technical standpoint. We believe that it is not straightforward, as there are grounds for challenging the view of inferences simply as one-step subproofs of proofs, or—conversely—of proofs (only) as arbitrary concatenations of inferences.

Another problem is to explore the addition of different unsound rules to proof systems. Here, we gave two examples, but many more are possible. It is unknown exactly how faster the calculus  $\text{LK}^+$  really is—the weakening of the usual characteristic-variable condition into substitutability, weak regularity, and the side-variable condition can perhaps be exploited further.

It is not clear whether  $\text{LK}^+$  is any (or significantly) slower than  $\text{LK}^{++}$ . It is also not clear in what precise relation they both stand to  $\text{LK}_{\text{shift}}$  and  $\text{LK}^\varepsilon$ . All these questions seem to promise fruitful lines of future research.

**Acknowledgements.** We would like to thank the anonymous referee for several comments that certainly improved the paper. This work was partially supported by FWF grants I-2671-N35 and W1255-N23.

## REFERENCES

- [1] M. Baaz and S. Hetzl, On the Non-Confluence of Cut-Elimination, *J. Symb. Logic*, 76 (2011), 313–340.
- [2] M. Baaz, S. Hetzl, and D. Weller, On the Complexity of Proof Deskolemization, *J. Symb. Logic*, 77 (2012), 669–686.
- [3] M. Baaz and A. Leitsch, On Skolemization and Proof Complexity, *Fund. Inform.*, 20 (1994), 353–379.
- [4] M. Baaz and A. Leitsch, Cut Normal Forms and Proof Complexity, *Ann. Pure Appl. Logic*, 97 (2000), 127–177.
- [5] D. Hilbert and P. Bernays, *Grundlagen der Mathematik II* (1970), Springer-Verlag,
- [6] G. Moser and R. Zach, The Epsilon Calculus and Herbrand Complexity, *Stud. Logica*, 82 (2006), 133–155.
- [7] V. P. Orevkov, Lower Bounds for Increasing Complexity of Derivations after Cut Elimination (*in Russian*), *J. Soviet Math.* (2006), 2337–2350.
- [8] R. Statman, Lower Bounds on Herbrand’s Theorem, *Proc. Amer. Math. Soc.* 75 (1979), 104–107.

INSTITUTE OF DISCRETE MATHEMATICS AND GEOMETRY, VIENNA UNIVERSITY OF TECHNOLOGY,  
WIEDNER HAUPTSTRASSE 8–10, 1040 VIENNA, AUSTRIA

INSTITUTE OF DISCRETE MATHEMATICS AND GEOMETRY, VIENNA UNIVERSITY OF TECHNOLOGY,  
WIEDNER HAUPTSTRASSE 8–10, 1040 VIENNA, AUSTRIA