Extended version of a paper which will appear in *The Bulletin of Symbolic Logic*. In particular, in this version all proofs are spelt out in detail.

Resolution and the Origins of Structural Reasoning: Early Proof-Theoretic Ideas of Hertz and Gentzen

Peter Schroeder-Heister*

Wilhelm-Schickard-Institut, Universität Tübingen

Sand 13, 72076 Tübingen, Germany

psh@informatik.uni-tuebingen.de

In the 1920s, Paul Hertz (1881–1940) developed certain calculi based on structural rules only and established normal form results for proofs. It is shown that he anticipated important techniques and results of general proof theory as well as of resolution theory, if the latter is regarded as a part of structural proof theory. Furthermore, it is shown that Gentzen, in his first paper of 1933, which heavily draws on Hertz, proves a normal form result which corresponds to the completeness of propositional SLD-resolution in logic programming.

1. Introduction: Structural Reasoning

By *structural reasoning* we mean reasoning in a sequent style system using structural rules only. Structural rules do not refer to the internal composition of formulas by means of logical connectives or quantifiers but only affect the way formulas appear within sequents. If sequents are of the form $\Gamma \to A$, prominent structural rules are *Weakening* and *Cut*:

$$\frac{\Gamma \to A}{\Gamma, B \to A} \ (\text{Weakening})$$

$$\frac{\Gamma \to A \qquad A, \Delta \to B}{\Gamma, \Delta \to B} \ (\text{Cut}) \quad .$$

*The results reported here have been presented in part at a seminar at CIS, University of Munich (July 1999), at the Polish-German "Logic and Logical Philosophy" conference, University of Dresden (April 2001), and at the "Natural Deduction" conference, PUC Rio de Janeiro (July 2001). I would like to thank the anonymous referee for helpful comments and suggestions.

If we encode the fact that proofs may start with identities of the form $A \to A$ by means of a rule without premisses, we have as a third prominent structural rule

$$A \to A \quad \text{(Identity)} \quad .$$

Identity, *Weakening* and *Cut* are sufficient as structural rules, if only minimal logic is considered, and if the antecedent of a sequent is a set (rather than a multiset, a sequence or a more sophisticated [e.g., binary] structure) of formulas.

The terminological distinction between structural rules and rules for logical connectives was drawn by Gentzen in his *Investigations into Logical Deduction* (1934/35), where he laid out the idea of a sequent calculus and established the eliminability of *Cut* as its fundamental feature. However, in certain areas structural rules play a role different from their functioning as non-logical inferences within calculi of logic or arithmetic. One such area is the treatment of resolution in terms of a sequent calculus based on structural rules only; another one would be the consideration of rule-based (production) systems in general. The view of structural reasoning as an (albeit essential) *part* of logical reasoning should not conceal the fact that it represents a subject in its own right. This is particularly important in the light of modern developments such as logic programming.

The calculi developed by the Paul Hertz[1] in the 1920s are structural systems in this independent sense. In the following we shall show that the normal form theorems proved by Hertz for these systems anticipate certain ideas and techniques of general proof theory as well as of the theory of resolution, in particular of logic programming (understood in a proof-theoretic setting). We shall also place Gentzen's first publication of 1933, which heavily draws on Hertz's ideas and results, in this context. It turns out that Gentzen proves what in modern terminology can be described as the completeness of propositional SLD-resolution. At the same time, this shows that structural systems stood at the beginning of Gentzen's intellectual development.

Two particular features of structural reasoning which go beyond Gentzen style sequent calculi are the following:

Sequents may occur as assumptions.

and

Cut is an indispensable rule which cannot be eliminated.

The first feature normally implies the second one. Actually, Gentzen (1934/35) called his own sequent calculi, which allow cut elimination, "logistic" calculi (and abbreviated

---

[1]For biographical data on Hertz see Bernays (1969). For an overview of his life and work, placing it in the logical and philosophical context of the time, as well as a bibliography of Hertz's writings, see Legris (1999). A concise sketch of Hertz's contributions to logic is Abrusci (1983).

them by $\boldsymbol{LJ}$ and $\boldsymbol{LK}$) to characterize the fact that they are assumption-free (pp. 184, 190 [Szabo-transl., pp. 75, 81 seq.])[2].

Structural rules may be looked upon as axiomatizing a consequence relation in Tarski's sense (for the finite case, of course). From this point of view, structural systems may be considered as providing a general framework of consequence, in terms of which specific logical systems can be defined. This indicates again that structural reasoning is a powerful tool which can be investigated from a variety of aspects.

2. Propositional Resolution

We recapitulate some notions from resolution theory in a proof-theoretic setting.[3] We consider the fragment of propositional resolution which deals with clauses of the form

$$A_1, \ldots, A_n \to A \ ,$$

because only those are relevant in the context of Hertz's and Gentzen's contributions. In the terminology of resolution, these are clauses with exactly one positive literal. In the terminology of logic programming, they have the form of definite program clauses. In the following we simply speak of *clauses*. The A_i's and A are taken from a finite or denumerably infinite domain $\mathfrak{E}$ of atomic expressions, called *atoms*. The body $A_1, \ldots, A_n$ is considered to be a *set*. Using Γ and Δ for sets of atoms, notations such as $\Gamma, \Delta \to A$ or $\Gamma, B \to A$ are understood in the usual way as standing for $\Gamma \cup \Delta \to A$ or $\Gamma \cup \{B\} \to A$, respectively. Clauses are denoted by $S, S', S_1, S_2, \ldots$ etc.

A *derivation* of S from $S_1, \ldots, S_n$ is a treelike structure of clauses such that (i) top clauses are either identities $A \to A$ or occur among $S_1, \ldots, S_n$, (ii) the bottom clause is S, and (iii) the clauses below the top clauses are generated by the rules of *Cut* and *Weakening*. If there is such a derivation, we write $S_1, \ldots, S_n \vdash S$. In the propositional case, which we are considering here, the resolution rule is just the cut rule. Identities $A \to A$ are only needed to generate trivial clauses of the form $\Gamma, A \to A$.

A derivation of S from $S_1, \ldots, S_n$ is called a *proper resolution derivation* of S from $S_1, \ldots, S_n$, if it only uses *Cut* (and neither *Identity* nor *Weakening*). Moreover, we assume that in applications of *Cut*, A does not occur in Δ, i.e., the cut formula is removed from the body of the right premiss.[4] A derivation of S from $S_1, \ldots, S_n$ is then

[2] As Gentzen emphasizes, they share the property of being assumption-free with the calculi considered in symbolic logic at that time. Gentzen's term is reminiscent of "logistics" ("Logistik"), as symbolic logic was called in Germany (see Carnap 1929).

[3] We do not rely on a particular presentation. An overview of resolution in the standard (disjunction-based) framework is given in Leitsch (1997). The classical reference for the theory of logic programming is Lloyd (1987).

[4] The literature on resolution is not uniform with respect to this requirement. In the presence of *Weakening*, this strict formulation of *Cut* is equivalent to the more relaxed one, where A is allowed

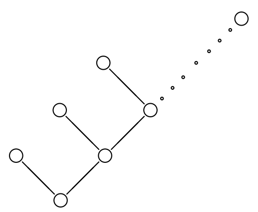
called a *resolution derivation* of S from $S_1, \ldots, S_n$, if it is either (i) an *identity* $A \to A$ or (ii) a *proper resolution derivation* or (iii) a derivation of the form (i) or (ii) followed by one or more applications of *Weakening*. This means, a resolution derivation uses *Weakening* only at the end. If there is such a resolution derivation, we write

$$S_1, \ldots, S_n \vdash_{RES} S \ .$$

Let "$\models$" denote logical consequence in (classical or intuitionistic) propositional logic, where clauses $A_1, \ldots, A_n \to A$ are given their natural reading as formulas $A_1 \& \ldots \& A_n \supset A$. Then the completeness theorem of resolution theory says

$$S_1, \ldots, S_n \vdash_{RES} S \quad \text{iff} \quad S_1, \ldots, S_n \models S \ .$$

A special form of resolution which is of interest in the following is SLD-resolution. A *proper SLD-derivation* is a proper resolution derivation whose tree has the following form[5]:



As for resolution derivations, an SLD-derivation is obtained by (possibly) applying *Weakening* to an identity or to the end clause of a proper SLD-derivation. We write

$$S_1, \ldots, S_n \vdash_{SLD} S \ ,$$

if there is an SLD-derivation of S from $S_1, \ldots, S_n$. Again we have as a theorem:

$$S_1, \ldots, S_n \vdash_{SLD} S \quad \text{iff} \quad S_1, \ldots, S_n \models S \ .$$

This means that for definite clauses, SLD-resolution is as strong as full resolution.

It should be emphasized that we are dealing here with the resolution calculus as a formal system and not with the *resolution method*, which is a refutation procedure for clauses based on the resolution calculus. Therefore we do not terminologically distinguish goal clauses from other clauses.[6]

---

to occur in Δ. The strict form, which we adopt here, has the advantage of fully separating *Cut* from *Weakening*. Both Hertz and Gentzen adopt the strict form (see the corresponding remarks in the next section).

[5]In treatments of input resolution and SLD-resolution the input clauses usually occur on the right rather than on the left in the derivation tree . This is just a notational variant due to a converse ordering of the premisses of *Cut*, which comes with the writing of program clauses as $A \leftarrow \Gamma$.

[6]We would have to represent them by $A_1, \ldots, A_n \to \bot$, where $\bot$ is a distinguished atom denoting absurdity.

3. Hertz's and Gentzen's Structural Systems

In a series of articles, Hertz proposed structural inference systems and proved results about them. We shall mainly rely on his 1929 paper, which presents the most mature versions of his systems, taking into account certain crucial issues from the 1923 and 1928[7] papers. In his first publication of 1933, Gentzen presents a modified version of Hertz's system.

Besides a propositional version, on which he puts the primary emphasis, Hertz 1929 also presents a system with variables and predicate symbols. This will be dealt with in section 6 below. In the present section we present Hertz's and Gentzen's propositional systems. Sections 4 and 5 deal with the results obtained by Hertz and Gentzen, respectively, for the propositional case. We shall try to make our presentation self-contained so that it can be read without consultation of Hertz's and Gentzen's original papers. This is particular important for Hertz, whose papers have not been translated into English. As for terminology, we shall give the original German terms in parentheses.

Hertz introduces *sentences* ("Sätze") of the form $A_1, \ldots, A_n \to A$ [8], where the capital letters stand for *elements* ("Elemente") from a given *domain* ("Bereich") $\mathfrak{E}$, which is finite or denumerably infinite (Hertz 1929, p. 460). The element A is called the *succedent* ("Sukzedens"), and the *complex* ("Komplex") of elements $A_1, \ldots, A_n$ is called the *antecedent* ("Antezedens") of $A_1, \ldots, A_n \to A$.[9] This complex is understood as a *set* of elements, i.e. order and multiplicity of elements in the antecedent is irrelevant (p. 461). If we identify elements with their names, calling the latter *atoms*, sentences can be identified with sequents or clauses in the terminology used in the previous section. In the following, when presenting Hertz's and Gentzen's systems, we shall use this identification and talk of *elements* $A, A_1, \ldots, A_n$ as synonymous with *atoms*, and of *sentences* of the form $A_1, \ldots, A_n \to A$ as synonymous with *sequents* or *clauses*, respectively. We shall use our previous notation accordingly, writing sentences as $\Gamma \to A$ or $\Gamma, A \to B$ etc.

Hertz's inference system is based on the following rules:

$$\frac{\Gamma_1 \to A_1 \quad \ldots \quad \Gamma_n \to A_n \quad \Delta, A_1, \ldots, A_n \to A}{\Delta, \Gamma_1, \ldots, \Gamma_n \to A} \text{ (Syllogism)}$$

$$\frac{\Gamma \to A}{\Gamma, \Delta \to A} \text{ (Immediate Inference)}$$

[7]This paper was written when the 1929 paper had already been submitted.

[8]He writes $u_1, \ldots, u_n \to v$. For readability, we stick to the notation used so far.

[9]The terms *antecedent* and *consequent* were thus coined by Hertz.

The terms *Syllogism* ("Syllogismus") and *Immediate Inference* ("unmittelbarer Schluß") are due to Hertz.[10] A more modern terminology would be *multicut*[11] for the first inference. The second one is a sort of multiple weakening. In resolution terminology it may be called propositional *subsumption*[12]. Gentzen 1933 introduces the term *Thinning* for it. In the context of the presentation of Hertz's results, we shall use his terminology.

The right premiss of a *Syllogism* is called the *major sentence* ("Obersatz"), the other premisses are called *minor sentences* ("Untersätze"). The $A_1, \ldots, A_n$ are called the *main members* ("Hauptglieder") of the antecedent of the major sentences, the remaining ones its *accessory* ("akzessorische") members.

The schema of *Syllogism* is understood in such a way that the $A_1, \ldots, A_n$ *do not occur in* $\Gamma_1, \ldots, \Gamma_n$. Although this is not clear from Hertz's notation[13], his verbal explanation in the early paper (1923, p. 82) suggests this reading, which we presuppose in the following. This coincides with Gentzen's understanding of *Cut* (see below).

It is very interesting to note that Hertz was the first to consider *tree-like proof structures*, an idea that became essential for Gentzen's later development of natural deduction and of the full sequent calculus. Hertz defines a *proof* ("Beweis") as a linear sequence of inferences (1929, p. 463), and an *inference system* ("Schlußsystem") as a corresponding tree-like structure (p. 464).[14] In the following, we do not make this distinction and always understand proofs as tree-like structures.

If we use this terminology, a *proof* from a *system of sentences* $\mathfrak{T}$ is defined by Hertz as a tree-like structure, whose top sentences ("oberste Sätze") are either *tautological* sentences, i.e., sentences of the form

$$A \to A,$$

[10]Hertz gives detailed explanations of why *Syllogism* is supposed to be the most fundamental inference rule in logic and therefore deserves its name. We shall not discuss this point here.

[11]To my knowledge, Slaney (1989) was the first to introduce this term.

[12]Hertz 1923, p. 82, speaks of the conclusion of an immediate inference as *contained* ("enthalten") in its premiss.

[13]In fact, the literal reading of Hertz 1929, pp. 461 seq., seems to suggest the opposite.

[14]Hertz (1923, pp. 85 seqq.) explicitly shows that proofs can be transformed into inference systems and vice versa. See also 1929, p. 464. It seems to us that Hertz's notion of a (linear) *proof* is not exactly the same as that of a *sequence of sentences* generated by rules, and there are differences between the 1923 and 1929 papers (see 1929, p. 464, footnote 12). We do not want to discuss this issue here as the notion of a tree-like proof (inference system) is clear, and is the only relevant notion in the present context. However, it is interesting to remark that the step from linear to tree-like proofs would have been a very small one already for Hilbert (1923, p. 158), who considers linear proofs in which each formula occurrence is used only once as a premiss, and thus is able to decompose linear proofs into threads which would correspond to branches in tree-like proofs. This is quoted in Hertz (1929, p. 464, footnote).

or sentences taken from $\mathfrak{T}$, and which is generated by means of the rules of *Syllogism* and *Immediate Inference*.[15]

Initially, Hertz considered his systems as directly understandable. A justification by means of an external semantics and a completeness proof was subsequently delivered in Hertz (1928), motivated by comments from Bernays. This issue will be dealt with at the end of the following section.

Gentzen (1933) builds on Hertz's approach. In the development of his structural system, he follows Hertz's terminology in many details. The only fundamental deviation from Hertz is that he introduces the notion of *Cut* ("Schnitt") as a variant of *Syllogism* with just a single minor sentence as left premiss. He uses the term *Thinning* ("Verdünnung") for what Hertz calls *Immediate Inference*. So the system Gentzen proposes is based on the following rules

$$\frac{\Gamma \to A \quad \Delta, A \to B}{\Delta, \Gamma \to B} \text{ (Cut)}$$

$$\frac{\Gamma \to A}{\Gamma, \Delta \to A} \text{ (Thinning) },$$

where in *Cut* it is supposed that Δ does not contain A. Corresponding to Hertz's terminology, the left and right premisses of *Cut* are called *minor sentence* and *major sentence*, respectively. The atom A is called the *cut element* ("Schnittelement") (Gentzen 1933, p. 331 [Szabo-transl., p. 31]). We may look upon *Identity* as an (improper) rule, as derivations are allowed to start with tautological sentences of the form $A \to A$. The (obvious) equivalence of Gentzen's system with that of Hertz is explicitly established by Gentzen (p. 332 [Szabo-transl., p. 32]).

4. Hertz's Normal Form and Completeness Proofs

The main goal of both Hertz (1922, 1923, 1929) and Gentzen (1933) was to establish axiomatizability results for systems of sentences. Given a set $\mathfrak{S}$ of sentences which contains all tautological sentences (i.e., sentences of the form $A \to A$) and which is closed under *Immediate Inference / Weakening* and *Syllogism / Cut*, we may ask questions like the following ones: Can $\mathfrak{S}$ be axiomatized? If $\mathfrak{S}$ can be axiomatized, is there an

[15] It might be mentioned that for Hertz (as well as for Gentzen, see Gentzen 1933, p. 331 [Szabo-transl., p. 31]) a formal proof has always to contain at least one proper rule application. For example, a formal proof of $A \to A$ would be

$$\frac{A \to A}{A \to A}$$

rather than the tautological sentence $A \to A$ alone, and analogously for a proof of a non-tautological sentence from itself (see Hertz 1929, p. 463 [footnote]). Therefore it is important that the rule of *Immediate Inference* includes the case where Γ is empty.

independent set of axioms? If $\mathfrak{S}$ can be axiomatized, can axioms be chosen in such a way that they are strongest with respect to the ordering generated by *Immediate Inference / Weakening*? Can a finite set of axioms be found? In which way do results depend on whether the domain of atoms considered is finite or (denumerably) infinite? Which results are obtained if only systems with linear sentences (i.e., sentences of the form $A \to B$) are considered?

Some of the results achieved are highly interesting, particularly from the point of view of inductive logic programming, where one aims at finding programs (i.e., a kind of axiomatization by means of clauses) for facts and clauses given as data.[16] Here we are interested in Hertz's and Gentzen's normal form and completeness results which establish fundamental properties of their standard systems. For both Hertz and Gentzen, these results played an auxiliary role in their treatment of axiomatizability. From a more modern point of view they contain fundamental conceptual insights.

In this section where we describe Hetz's results, we assume that a domain $\mathfrak{E}$ of elements and a set of sentences $\mathfrak{S}$ over $\mathfrak{E}$, called the *axioms*, are fixed. *Axiom* is here synonymous with *assumption* in modern terminology. By a proof we mean a proof from the given axioms using tautological sentences and the rules of *Immediate Inference* and *Syllogism*.

Hertz calls a proof an *Aristotelian normal proof* if each non-tautological major sentence of a syllogism is an axiom. A proof is called a *Goclenian normal proof* if each non-tautological minor sentence of a syllogism is an axiom. This terminology is based on the traditional distinction between Aristotelian and Goclenian[17] chain syllogisms, which lead to Hertz's normal forms when decomposing them into binary proof steps. For example, an Aristotelian chain inference with four premisses

$$\begin{array}{c} A \to B \\ B \to C \\ C \to D \\ \underline{D \to E} \\ A \to E \end{array}$$

leads to the Aristotelian normal proof

$$\dfrac{\dfrac{\dfrac{A \to B \quad B \to C}{A \to C} \qquad C \to D}{A \to D} \qquad D \to E}{A \to E} \quad ,$$

[16] See Nienhuys-Cheng & Wolf (1997).

[17] "Goclenian" after the German logician Rudolph Goclenius [Göckel] (1547–1628).

whereas the Goclenian chain inference

$$\begin{array}{l} D \to E \\ C \to D \\ B \to C \\ \underline{A \to B} \\ A \to E \end{array}$$

leads to the Goclenian normal proof

$$\dfrac{A \to B \qquad \dfrac{B \to C \qquad \dfrac{C \to D \quad D \to E}{C \to E}}{B \to E}}{A \to E} \quad .$$

Obviously, Goclenian chain inferences represent the form of reasoning inherent in SLD-derivations.

Then Hertz obtains the following results, which we formulate as explicit theorems:

Theorem 1. Any proof can be transformed into an Aristotelian normal proof.

Theorem 2. Any proof can be transformed into a Goclenian normal proof.

Theorem 3. It is decidable whether a sentence S is provable from $\mathfrak{S}$*, if* $\mathfrak{S}$ *is finite.*

For *Theorem 1*, Hertz (1923) gives a purely syntactic proof which is highly significant as it uses for the first time the proof-theoretic method of reducing a proof by a syntactic manipulation of inferences, and of justifying the termination of this reduction by an assignment of ordinal numbers. It is obvious that Hertz's proof could provide considerable inspiration to Gentzen, who later made extensive usage of such methods.

For *Theorems 1 and 2*, Hertz (1929) gives proofs in terms of certain fixed point considerations (in modern terminology), which anticipate ideas now standard in logic programming. The fact that a fixed point is reached after finitely many steps is then used to prove *Theorem 3*.

In the following, we present Hertz's proofs in more modern terminology and symbolics.

Hertz's Syntactic Proof of Theorem 1 (Hertz 1923, pp. 88-93)

This proof is given by Hertz for a modified system with

$$\frac{\Gamma_1 \to A_1 \quad \ldots \quad \Gamma_n \to A_n \qquad A_1, \ldots, A_n \to A}{\Gamma_1, \ldots, \Gamma_n \to A} \quad \text{(Pure Syllogism)}$$

("reiner Syllogismus") as a primitive rule instead of *Syllogism*, i.e., *Syllogism* without accessory elements. Obviously, by adding tautologies as minor sentences, any syllogism can be tranformed into a pure syllogism, so the systems based on *Syllogism* and *Pure Syllogism* are equivalent (p. 83).

In order to transform a proof into an Aristotelian normal proof, subproofs ending with

$$\dfrac{\{\Delta_j \to B_j\}_{1\le j\le m} \quad \dfrac{\{\Gamma_i \to A_i\}_{1\le i\le n} \quad A_1,\dots,A_n \to B}{\Gamma_1,\dots,\Gamma_n \to B}}{\Delta_1,\dots,\Delta_m \to B} \quad ,$$

where $\Gamma_1,\dots,\Gamma_n$ represents the same set of atoms as $B_1,\dots,B_m$, are transformed in such a way that the critical major sentence $\Gamma_1,\dots,\Gamma_n \to B$ disappears and the resulting proof is more elementary, where *more elementary* is measured by an assignment of ordinals to proofs (Case 1).

In a similar way, subproofs ending with

$$\dfrac{\{\Delta_j \to A_j\}_{1\le j\le n\le m} \quad \dfrac{A_1,\dots,A_n \to B}{A_1,\dots,A_n,\dots,A_m \to B}}{\Delta_1,\dots,\Delta_m \to B}$$

have to be reduced such that the critical major sentence $A_1,\dots,A_n,\dots,A_m \to B$ disappears.

In order to avoid the heavy use of multiple indices Hertz has to make, we present special examples from which the general method can easily be inferred.

Case 1 Given a subproof of the form

$$\dfrac{\begin{matrix}\Pi_4\\ D \to C\end{matrix} \quad \dfrac{\begin{matrix}\Pi_1\\ C \to A_1\end{matrix} \quad \begin{matrix}\Pi_2\\ C \to A_2\end{matrix} \quad \begin{matrix}\Pi_3\\ A_1,A_2 \to B\end{matrix}}{C \to B}}{D \to B} \quad ,$$

where Π_1,Π_2,Π_3 and Π_4 indicate the subproofs with the respective conclusions $C \to A_1, C \to A_2,\ \ A_1,A_2 \to B$ and $D \to C$. Then this subproof is transformed into the following subproof:

$$\dfrac{\dfrac{\begin{matrix}\Pi_4\\ D \to C\end{matrix} \quad \begin{matrix}\Pi_1\\ C \to A_1\end{matrix}}{D \to A_1} \quad \dfrac{\begin{matrix}\Pi_4\\ D \to C\end{matrix} \quad \begin{matrix}\Pi_2\\ C \to A_2\end{matrix}}{D \to A_2} \quad \begin{matrix}\Pi_3\\ A_1,A_2 \to B\end{matrix}}{D \to B}\ ,$$

which means that the applications of *Pure Syllogism* are permuted. The proof complexity is reduced with respect to a certain complexity measure. Hertz defines the *ordinal number* ("Ordnungszahl") of an occurrence of a sentence in a proof as follows:

The ordinal number of a top sentence is 0.
If the premiss of an immediate inference has ordinal number n, then the ordinal number of its conclusion is $n+1$. If the premisses of a syllogism have ordinal numbers $k_1, \ldots, k_{n+1}$, respectively, then the ordinal number of its conclusion is $(max_{1 \leq i \leq n+1} k_i) + 1$

Now the ordinal numbers of major sentences in the subproofs can be computed. If k_1, k_2, k_3 and k_4 are the ordinal numbers of th conclusions of Π_1, Π_2, Π_3 and Π_4, respectively, we have for the original subproof the following assignments:

ordinal number of major sentence $A_1, A_2 \to B$: k_3
ordinal number of major sentence $C \to B$: $max(k_1, k_2, k_3) + 1$

For the modified subproof we obtain the following assignments:

ordinal number of major sentence $A_1, A_2 \to B$: k_3
ordinal number of major sentence $C \to A_1$: k_1.
ordinal number of major sentence $C \to A_2$: k_2.

Hertz argues that "the ordinal number of a major sentence is replaced with a set of lower ordinal numbers of major sentences" (1923, p. 92). In the present example the ordinal number $max(k_1, k_2, k_3) + 1$ is replaced with the set of ordinal numbers $\{k_1, k_2\}$.

Case 2 A subproof of the form

$$\cfrac{\begin{matrix}\Pi_2 \\ D \to C\end{matrix} \quad \begin{matrix}\Pi_3 \\ E \to B\end{matrix} \quad \cfrac{\begin{matrix}\Pi_1 \\ C \to A\end{matrix}}{C, B \to A}}{D, E \to A}$$

is transformed into

$$\cfrac{\cfrac{\begin{matrix}\Pi_2 \\ D \to C\end{matrix} \quad \begin{matrix}\Pi_1 \\ C \to A\end{matrix}}{D \to A}}{D, E \to A} \quad ,$$

again permuting the two inferences. Hertz argues once more that, if the ordinal number of $C, B \to A$ is k, one sentence with the ordinal number k disappears.

Induction argument

Now Hertz proceeds as follows (p. 93). If we start with a subproof such that $C \to B$ (Case 1) or $C, B \to A$ (Case 2) has k as the maximal ordinal number of major sentences occurring in the proof, then after finitely many reductions of the sort indicated we reach a proof with a maximal ordinal number of major sentences less then k. Iterating this procedure finitely many times we reach a proof with all major sentences being of ordinal number 0, i.e. an Aristotelian normal proof. In more modern terminology, Hertz is using ω^2-induction on the pair $\langle \alpha(\Pi), \beta(\Pi) \rangle$, where $\alpha(\Pi)$ is the maximal ordinal number of a major sentence in Π and $\beta(\Pi)$ is the number of major sentences with that ordinal number.

In a footnote in his 1929 paper (p. 474) he remarks that for this procedure to work one has to choose a lowermost subproof of the form considered. Otherwise the ordinal number of $D \to B$ (Case 1) or of $D, E \to A$ (Case 2) and of sentences below them can increase, which is critical if major sentences are among them. This means that the branch starting from $D \to B$ (Case 1) or $D, E \to A$ (Case 2) and proceeding down to the final conclusion of the proof must not pass through major sentences.

Using the terminology coined by Prawitz 1965 for the area of natural deduction, Hertz's normal form demonstration can be described as follows: Call a sentence *maximal* in a proof if it is the conclusion of an application of *Immediate Inference* or *Pure Syllogism* and at the same time the major sentence of an application of *Pure Syllogism*. Call a proof *normal* if it does not contain maximal sentences. Then any proof can be normalized by the procedure described.

Hertz's Fixed Point Proof of Theorem 1 (Hertz 1929, Section 3 = pp. 475–477)

The non-syntactic proofs in Hertz (1929) of *Theorem 1* proceed as follows. We assume that the set of axioms is finite. We start with a finite subset $\mathfrak{E}_0$ of the domain of elements $\mathfrak{E}$. The elements in $\mathfrak{E}$ are called *distinguished ("ausgezeichnet") at level 0.* An element A is *distinguished at level* $n+1$ (with respect to $\mathfrak{E}_0$) if A is not distinguished at any level $\leq n$ (with respect to $\mathfrak{E}_0$) and if there is an axiom $\Gamma \to A$ such that each element in Γ is distinguished at some level $\leq n$ (with respect to $\mathfrak{E}_0$). An atom A is *distinguished* with respect to $\mathfrak{E}_0$, if for some n, A is distinguished at level n (with respect to $\mathfrak{E}_0$). Loosely speaking, A is distinguished with respect to $\mathfrak{E}_0$, if A can be "generated" stepwise from $\mathfrak{E}_0$ by means of axioms (considered as production rules). With this definition Hertz anticipates the definition of the monotonic operator associated with program rules in logic programming (often called T_P) and offers some sort of fixed point semantics.

Based on this definition he proves what we formulate as a Lemma.

Lemma 1 (1) If a sentence $\Gamma \to A$ *is provable from the axioms, then* A *is distinguished with respect to* Γ.
(2) If $\Gamma \to A$*,* A *is distinguished with respect to* Γ*, then there is an Aristotelian normal proof of* $\Gamma \to A$ *from the axioms.*

Hertz emphasizes (1929, p. 475) that this Lemma holds for arbitrary, possibly infinite systems of sentences, whereas the decidability result concluded from it (*Theorem 3* below) holds only for the finite case.

Theorem 1 is an immediate consequence of *Lemma 1.*

Proof of Lemma 1

(1) Suppose $\Gamma \to A$ is provable from the axioms. Define the following property of sentences $\Delta \to B$:

(α) *If each element of* Δ *is distinguished with respect to* Γ*, then so is* B*.*

It is obvious that any axiom or tautological sentence has the property α. Furthermore, it is easy to see that property α carries over from the premisses to the conclusion of *Syllogism* and of *Immediate Inference.* Therefore $\Gamma \to A$ has property α, which implies that A is distinguished with respect to Γ (because Γ is distinguished with respect to itself).[18]

(2) We proceed by induction on the level at which A is distinguished with respect to Γ. If A is distinguished at level 0 with respect to Γ, then A is in Γ. If A is distinguished at level 1 with respect to Γ, then there is an axiom $\Gamma' \to A$ such that $\Gamma' \subseteq \Gamma$. In both cases $\Gamma{\to}A$ can be obtained from a tautological sentence by *Immediate Inference.* Suppose A is distinguished at level $k+1$ with respect to Γ for $k \geq 1$. Then there is an axiom $A_1, \ldots, A_m \to A$ such that each $A_1, \ldots, A_m$ is distinguished at level k with respect to Γ. The $A_1, \ldots, A_m$ must be different from A. By the induction hypothesis, for each i $(1 \leq i \leq m)$ there is an Aristotelian normal proof of $\Gamma \to A_i$, from which, by application of *Syllogism* with $A_1, \ldots, A_m \to A$ as the major sentence, an Aristotelian normal proof of $\Gamma \to A$ is obtained.

Two remarks are appropriate.

1. If we skip the rule of *Immediate Inference* and instead allow a proof to start with trivial sentences of the form $\Gamma, A \to A$, then in the Aristotelian normal proof constructed all minor sentences of syllogisms as well as the end sentence have the same antecedent Γ.[19]

2. It is essential for the construction of the Aristotelian normal form that *Syllogism* (*Multicut*) is used. If we decompose a single multicut into several cuts, then new major sentences arise which are conclusions of *Cut.*[20]

Hertz's Fixed Point Proof of Theorem 2 (Hertz 1929, Section 4 = pp. 478–479)

As for *Theorem 1*, Hertz assumes a finite set of axioms over a domain of elements $\mathfrak{E}$ to be given. Starting with an an element A in $\mathfrak{E}$, a finite set of elements Γ is called

[18] In Hertz's own presentation this argument proceeds indirectly. Hertz shows that the negation of α (which he himself calls α) is carried back from the conclusion of inferences to their premisses.

[19] According to Hertz (1928, p. 277) this observation is due to Bernays.

[20] This was observed by Gentzen (see footnote 23 below).

distinguished at level 0 (with respect to A) if Γ contains A. Γ is called *distinguished at level* $n+1$ (with respect to A) if Γ is not distinguished at any level $\leq n$ (with respect to A), and if there are axioms $\Gamma_1 \to A_1, \ldots, \Gamma_m \to A_m$ such that the following holds:

(i) $\Delta' \cup \{A_1, \ldots, A_m\}$ is distinguished at some level $\leq n$ (with respect to A) for some Δ'

(ii) $\Gamma = \Delta \cup \Delta' \cup \Gamma_1 \cup \ldots \cup \Gamma_m$ for some Δ.

Loosely speaking, Γ is distinguished with respect to A, if, starting with A, Γ can be reached by looking step by step for conditions sufficient to "generate" A by means of axioms taken as production rules. Whereas in "A is distinguished with respect to Γ" (Lemma 1) we were looking for A as a *consequence* of Γ (*forward reasoning*), in "Γ is distinguished with respect to A" we are looking for Γ as a *condition* of A (*backward reasoning*). Formally, this can be described as the construction of a fixed point operator on the power set of $\mathfrak{E}$, which, starting with {A}, associates with each set of elements Γ those sets of elements Δ from which Γ can be generated using axioms.

Based on this definition Hertz proves the following:

Lemma 2 *(1) If a sentence* $\Gamma \to A$ *is provable from the axioms, then* Γ *is distinguished with respect to A.*

(2) If for a sentence $\Gamma \to A$*,* Γ *is distinguished with respect to A, then there is a Goclenian normal proof of* $\Gamma \to A$ *from the axioms.*

Theorem 2 is an immediate consequence of *Lemma 1*.

Proof of Lemma 2

(1) Suppose $\Gamma \to A$ is provable from the axioms. Define the following property of sentences $\Delta \to B$:

(β) *For any* Δ'*, if* $\Delta' \cup \{B\}$ *is distinguished with respect to A, then so is* $\Delta' \cup \Delta$*.*

It is obvious that any axiom or tautological sentence has the property β. Furthermore, it is easy to see that property β carries over from the premisses to the conclusion of *Syllogism* and of *Immediate Inference*. Therefore $\Gamma \to A$ has property β, which implies that Γ is distinguished with respect to A (because A is distinguished with respect to itself).[21]

(2) We proceed by induction on the level at which Γ is distinguished with respect to A. If Γ is distinguished at level 0 with respect to A, then A is in Γ. If Γ is distinguished at level 1 with respect to A, then, since Γ is not distinguished at level 0 with respect to A, there is an axiom $\Gamma' \to A$ such that $\Gamma' \subseteq \Gamma$. In both cases $\Gamma \to A$

[21]As for *Lemma 1*, Hertz's own argument is indirect, showing that the negation of β (which he himself calls β) is carried back from the conclusion of inferences to their premisses.

can be obtained from a tautological sentence by *Immediate Inference*. Suppose Γ is distinguished at level $k+1$ with respect to A for $k \geq 1$. Then by definition there are axioms $\Gamma_1 \to A_1, \ldots, \Gamma_m \to A_m$ and a set of elements Δ' such that the following holds:

(i') $\Delta' \cup \{A_1, \ldots, A_m\}$ is distinguished at some level $\leq n$ with respect to A

(ii') $\Gamma = \Delta \cup \Delta' \cup \Gamma_1 \cup \ldots \cup \Gamma_m$ for some Δ.

By induction hypothesis, from (i') we obtain a Goclenian normal proof Π of $\Delta', A_1, \ldots, A_m \to A$. Using *Syllogism* and *Immediate Inference* we obtain

$$\dfrac{\dfrac{\{\Gamma_i \to A_i\}_{1\leq i\leq m} \quad \begin{matrix}\Pi \\ \Delta', A_1, \ldots, A_m \to A\end{matrix}}{\Delta', \Gamma_1, \ldots, \Gamma_m \to A}}{\Gamma \to A\ ,}$$

which is again a Goclenian normal proof.

The Goclenian normal proof constructed in *Lemma 2* is related to an SLD-derivation in so far as all minor sentences which are used as "input" to antecedents of major sentences are axioms. However, it is not exactly the same. A minor difference is that multiple input in one step (multicut) is possible. The major difference is that *Immediate Inference* (*Thinning*) may be used in between two applications of *Syllogism* and not only as the last step of the proof. This is changed in Gentzen (1933).

Hertz's Proof of Decidability (Theorem 3)

For decidability Hertz (1929) gives two proofs, each for a finite system $\mathfrak{S}$ of sentences. The first one (pp. 474 seq.) argues that the lengths of proofs needed to obtain all sentences of the system is bounded, so that we just have to check all possible proofs from the given axioms step by step. The second one uses the fixed point construction introduced for *Lemma 1*. To check whether $\Gamma \to A$ belongs to $\mathfrak{S}$, we generate, for each n, the elements distinguished with respect to Γ at level n, and determine whether A is among them.

Although these results are trivial from the modern point of view, the way of proving them is original in its proof-theoretic setting. In particular, it is highly significant that in *Lemma 1* Hertz relates the idea of atoms being generated from other atoms by means of sentences (clauses) with the idea of proofs of sentences of a certain form. This means that in principle he is aware of the close relationship betwen *proving* A *from* Γ *by means of axioms* (understood as some sort of inference or production rules) and a formal proof in a sequent style system *of* $\Gamma \to A$ *from the axioms* (see Hertz 1929, middle of p. 476). Today this idea is the basis of relating natural deduction and the sequent calculus with each other. With respect to reasoning with atoms, this sort of reasoning is also basic for logic programming. There a smallest fixed point of an operator defined in terms of clauses can be described by an SLD-derivation using these claues as axioms. Hertz's derivation do not have exactly that form, but in Gentzen (1933), who heavily relies on Hertz, we find what is today called the completeness of SLD-resolution (see section 5 below).

Hertz's Semantics and Completeness Result

Before we turn to Gentzen, we look at Hertz's semantic completeness result which goes beyond *Theorems 1* and *2*. The notions of being *distinguished* used in the proofs of *Lemmas 1* and *2* are not considered a real semantics but rather a technical device. In his 1928 paper (written after Hertz 1929) he explicitly provides such a semantics. Motivated by personal remarks by Bernays (see Hertz 1928, p. 272) he then proves the completeness of his rules with respect to this semantics.

Suppose a finite universal domain ("Grundbereich") $\mathfrak{G}$ *of elements is given. A subset (domain, "Bereich")* $\mathfrak{B}$ *of* $\mathfrak{G}$ *satisfies ("genügt") a sentence* $\Gamma \rightarrow A$*, if either* A *is in* $\mathfrak{B}$ *or not every element of* Γ *is in* $\mathfrak{B}$*, i.e., if* $\Gamma \subseteq \mathfrak{B}$*, then* $A \in \mathfrak{B}$*.*

This definition of a domain $\mathfrak{B}$ satisfying a sentence $\Gamma \rightarrow A$ can be read as a classical truth definition for a sentence under a valuation (represented by the domain $\mathfrak{B}$) as well as a constructive interpretation of $\Gamma \rightarrow A$ as a production rule under which $\mathfrak{B}$ is closed. Hertz presents several philosophical interpretations of $\Gamma \rightarrow A$, which we do not want to discuss here in detail. It is interesting that he mentions the possibility of identifying a sentence with the set of the domains satisfying it (p. 273[22]), which corresponds to identifying a proposition with the set of worlds in which it is valid, again anticipating a modern idea.

According to Hertz, a finite set of sentences $\mathfrak{S}$ implies ("impliziert") a sentence S, if any domain $\mathfrak{B}$ satisfying all sentences of $\mathfrak{S}$ satisfies S as well. If we write $\mathfrak{B} \models S$ for "$\mathfrak{B}$ satisfies S", $\mathfrak{B} \models \mathfrak{S}$ for "$\mathfrak{B}$ satisfies every element of $\mathfrak{S}$" and $\mathfrak{S} \models S$ for "$\mathfrak{S}$ implies S", this can be expressed as

$$\mathfrak{S} \models S \textit{ iff for every } \mathfrak{B}, \textit{ if } \mathfrak{B} \models \mathfrak{S} \textit{ then } \mathfrak{B} \models S \quad .$$

The completeness result Hertz proves, can then be formulated as follows:
Theorem 4 A sentence S *is provable from a finite set* $\mathfrak{S}$ *of axioms iff* $\mathfrak{S} \models S$*.*

The validity of the soundness direction is obvious. Due to the definition of $\mathfrak{S} \models \Gamma \rightarrow A$, satisfaction of a sentence is carried over from the premisses to the conclusion of *Syllogism* and *Immediate Inference.*

For the completeness direction Hertz relies on *Lemma 1 (2)* which says that the fact that A is distinguished with respect to Γ is sufficient for the provability of $\Gamma \rightarrow A$

[22] See also 1929a, p. 188

from the axioms. It remains to show that $\mathfrak{S} \models \Gamma \rightarrow A$ implies that A is distinguished with respect to Γ given the elements of $\mathfrak{S}$ as axioms.

In order to establish this fact, Hertz considers the domain Γ^*, which consists of all elements distinguished with respect to Γ. This means that Γ^* is the closure of Γ under the sentences of $\mathfrak{S}$ (taken as production rules). By definition, Γ^* satisfies all sentences of $\mathfrak{S}$, formally $\Gamma^* \models \mathfrak{S}$. By the assumption $\mathfrak{S} \models \Gamma \rightarrow A$, this implies $\Gamma^* \models \Gamma \rightarrow A$. Since $\Gamma \subseteq \Gamma^*$, this means that $A \in \Gamma^*$, i.e., A is distinguished with respect to Γ.

This is exactly as a modern proof would proceed. However, for applications such as logic programming, an Aristotelian normal proof is not as relevant as a special Goclenian one with single input (*Cut*) and *Immediate Inference* (*Thinning*) only at the end of the proof. Results for this system were achieved by Gentzen (1933).

5. Gentzen's Normal Form and Completeness Proofs

After presenting his system (see section 3 above), Gentzen (1933, pp. 333 seqq. [Szabo-transl., pp. 33 seqq.]) provides the same semantics as Hertz (1928). So we can use the notation introduced at the end of the previous section. We shall write $S_1, \ldots, S_n \models S$, if $\{S_1, \ldots, S_n\}$ implies S. Terminologically differing from Hertz, Gentzen prefers to speak of S as a *consequence* ("Folgerung") of $S_1, \ldots, S_n$, if $S_1, \ldots, S_n \models S$. In the present section, we adopt this terminology. Furthermore, Gentzen makes it explicitly clear that, when considering the question of whether $S_1, \ldots, S_n \models S$ holds, the underlying domain of elements comprises just the elements occurring in $S_1, \ldots, S_n, S$. In the terminology of logic programming, this corresponds to considering the (propositional) *Herbrand universe* as a basis.

Gentzen first proves soundness with respect to this semantics.

Theorem 5 If there is a proof of S from $S_1, \ldots, S_n$, then S is a consequence of $S_1, \ldots, S_n$.

Gentzen (p. 333 seq. [Szabo-transl., p. 33 seq.]) first shows that tautological sentences are consequences of any sentence and that the conclusions of *Cut* and *Thinning* are consequences of their respective premisses. Therefore, if a sentence is obtained from consequences of $S_1, \ldots, S_n$ by means of *Cut* or *Thinning*, it is itself a consequence of $S_1, \ldots, S_n$. Hence, since all tautological sentences as well as $S_1, \ldots, S_n$ are consequences of $S_1, \ldots, S_n$, the sentence S is a consequence of $S_1, \ldots, S_n$. The last step implicitly contains the induction argument.

For completeness, Gentzen proves a stronger result which yields a normal form theorem at the same time. He explicitly refers to Hertz, mentioning that he is aiming at normal proofs different from those considered by Hertz.[23]

Call a sentence of the form $\Gamma, A \rightarrow A$ *trivial*[24]. A *normal proof* ("Normalbeweis") of a non-trivial sentence S from the sentences $S_1, \ldots, S_n$ is defined as a proof of the following form:

$$\cfrac{R_{m-1} \qquad \cfrac{\vdots}{\cfrac{R_1 \qquad \cfrac{R_0 \quad Q_0}{Q_1}\,\text{Cut}}{\vdots}\,\text{Cut}} \; Q_{m-1}}{\cfrac{Q_m}{S}\,\text{Thinning}}\,\text{Cut} \; ,$$

where $m \geq 0$, and the R_i and Q_i are sentences such that the following holds:

(i) All *initial sentences* ("Anfangssätze") $Q_0, R_0, \ldots, R_{m-1}$ occur among $S_1, \ldots, S_n$.

(ii) No trivial sentence occurs in the proof.

Then the completeness theorem is formulated by Gentzen as follows (p. 336 [Szabo-transl., p. 36]):

Theorem 6 If a non-trivial sentence S is a consequence of $S_1, \ldots, S_n$, then there is a normal proof of S from $S_1, \ldots, S_n$.

Gentzen does not explicitly define the notion of a normal proof for trivial sentences. Of course, this would be just a proof of the form

$$\frac{A \rightarrow A}{\Gamma, A \rightarrow A}\,\text{Thinning} \quad .$$

It is obvious that a normal proof is an SLD-derivation in the sense of section 2. Therefore the proof of *Theorem 6* establishes the completeness of SLD-resolution. Due to condition (ii), a normal proof is slightly more restricted than SLD-derivations in general, so *Theorem 6* proves more than needed for the completeness of SLD-resolution. In an SLD-derivation of a non-trivial sentence we would admit trivial clauses as assumptions, if they belong to the program considered. We would also admit trivial clauses if

---

[23]See Gentzen 1933, p. 334 (footnote) and p. 335 (footnote) [Szabo-transl., p. 312, notes 6 and 7]. Gentzen remarks that his own normal proofs are related to Hertz's *Goclenian normal proofs*, because the *minor* premiss of each application of *Cut* is a sentence of $S_1, \ldots, S_n$ (an *axiom* in Hertz's terminology). Gentzen also gives the following simple counterexample showing that there is no analogue to Hertz's *Aristotelian normal proofs* in his system based on *Cut* rather than *Multicut*:

$$\cfrac{E \rightarrow A \qquad \cfrac{D \rightarrow B \quad A, B \rightarrow C}{A, D \rightarrow C}}{E, D \rightarrow C}$$

[24]This term ("trivialer Satz") is due to Hertz. See his 1929, p. 463.

they happen to occur as conclusions of *Cut* (e.g., the trivial clause $A \to A$ as a consequence of the non-trivial clauses $A \to B$ and $B \to A$). As an immediate consequence of *Theorems 5* and *6* Gentzen obtains the following normal form theorem.

Theorem 7 If a non-trivial sentence S is provable from $S_1, \ldots, S_n$, then there is a normal proof of S from $S_1, \ldots, S_n$.

Gentzen remarks that a direct syntactic proof of *Theorem 7* is possible, but that he prefers the deviation via the soundness and completeness theorems, as they present important additional insights into the system.[25]

Proof of Theorem 6 (Gentzen 1933, pp. 336 seqq. [Szabo-transl., pp. 36 seqq.])

Due to the specific requirements of the normal form, this proof differs from the argument by Hertz discussed in the previous section. However, as in Hertz, it uses the fixed point construction which is now standard in the theory of logic programming.

Let S be non-trivial. Suppose there is no normal proof of S from $S_1, \ldots, S_n$. Gentzen shows that S is not a consequence of $S_1, \ldots, S_n$ by constructing a domain Γ^* of elements such that

$$\Gamma^* \models S_i \text{ for all } i \ (1 \leq i \leq n), \text{ but } \Gamma^* \not\models S.$$

Suppose S has the form $\Gamma \to A$. Then Γ^* is constructed (nondeterministically) by expanding Γ step by step as follows.

$$\begin{array}{lcll} \Gamma_1 & = & \Gamma & \\ \Gamma_{j+1} & = & \Gamma_j \cup \{B_j\} & \text{if } \Delta_j \to B_j \text{ belongs to } S_1, \ldots, S_n \\ & & & \text{such that } \Gamma_j \not\models \Delta_j \to B_j \\ & & & \text{(i.e., } \Delta_j \subseteq \Gamma_j, \text{ but } B_j \notin \Gamma_j \text{).} \end{array}$$

If there is no such $\Delta_j \to B_j$, then the procedure terminates with Γ_j, and Γ^* is set as Γ_j. So we obtain a sequence of domains, for which the following holds:

$$\Gamma = \Gamma_1 \subsetneq \ldots \subsetneq \Gamma_m = \Gamma^* .$$

Obviously, this sequence is finite as there are only finitely many elements in the basic domain (which is the set of elements occurring in $S_1, \ldots, S_n, S$).

Γ^* is the closure of Γ under $S_1, \ldots, S_n$, if $S_1, \ldots, S_n$ are considered as production rules, i.e., in the terminology of logic programming, Γ^* is the fixed point of the operator T_P characteristic of the *program* $\{S_1, \ldots, S_n\}$.

Due to the construction of Γ^*, we have $\Gamma^* \models S_i$ for any i $(1 \leq i \leq n)$. Furthermore, $\Gamma \subseteq \Gamma^*$ means that $\Gamma^* \models \Gamma$. To show that $\Gamma^* \not\models S$ we just have to show $\Gamma^* \not\models A$, i.e., $A \notin \Gamma^*$.

---

[25]See p. 337 [Szabo-transl., p. 37 seq.]. What Gentzen obviously has in mind is the method of shifting down *Thinning* by permuting applications of *Cut* with *Thinning*.

For that purpose, Gentzen proves the following stronger assertion by induction on the construction of Γ^*.

Lemma *Let* $\mathfrak{S}$ *be the set of those non-trivial sentences of the form* $\Delta \to A$*, for which there is a normal proof from* $S_1, \ldots, S_n$ *without Thinning at the end. Then for any* Γ_k *reached in the construction of* Γ^**, we have*

$$(1)\ \Gamma_k \models \mathfrak{S},\ \textit{but}\ \ (2)\ \Gamma_k \not\models A\ .$$

The proof of the *Lemma* proceeds as follows.

k = 1: (1) Suppose $\Gamma \not\models \Delta \to A$ for some $\Delta \to A$ in $\mathfrak{S}$. Then $\Gamma \models \Delta$, i.e., $\Delta \subseteq \Gamma$. Since $\Gamma \to A$ can be obtained from $\Delta \to A$ by *Thinning*, this would mean that there is a normal proof of $\Gamma \to A$ from $S_1, \ldots, S_n$, contradicting the main assumption of the completeness proof.

(2) We have $\Gamma \not\models A$, since $\Gamma \to A$ is assumed to be non-trivial.

$k = j + 1$*:* Suppose as the induction hypothesis $\Gamma_j \models \mathfrak{S}$ and $\Gamma_j \not\models A$. Suppose furthermore that Γ_{j+1} has been obtained from Γ_j by adding B_j using $\Delta_j \to B_j$.

(1) Suppose $\Gamma_{j+1} \not\models S'$ for some S' in $\mathfrak{S}$. Since $\Gamma_j \models S'$, S' must have the form $\Delta, B_j \to A$. We can assume that $B_j \notin \Delta$. Now consider the following cut:

$$\frac{\Delta_j \to B_j \quad \Delta, B_j \to A}{\Delta, \Delta_j \to A}\ .$$

$\Delta, \Delta_j \to A$ is non-trivial. Since $\Delta, B_j \to A$ is in $\mathfrak{S}$, the sentence $\Delta, \Delta_j \to A$ must be in $\mathfrak{S}$ as well, as it is the conclusion of a *Cut* applied to an axiom as its minor sentence and a member of $\mathfrak{S}$ as its major sentence. However, since $\Delta \subseteq \Gamma_j$ and $\Delta_j \subseteq \Gamma_j$, but $A \notin \Gamma_j$, we have $\Gamma_j \not\models \Delta, \Delta_j \to A$, which contradicts $\Gamma_j \models \mathfrak{S}$.

(2) We have $\Delta_j \subseteq \Gamma_j$, hence $\Delta_j \not\models A$. Since $\Gamma_j \not\models \Delta_j \to B_j$ and $\Gamma_j \models \mathfrak{S}$, B_j differs from A. Therefore $\Gamma_{j+1} \not\models A$.

This is a full-fledged completeness proof for propositional SLD-resolution, invalidating $\Gamma \to A$ by constructing the closure Γ^* of Γ under the clauses given. The special case of SLD-refutations as SLD-derivations of the empty clause, which is normally considered in logic programming, can be obtained from Gentzen's results as follows. Since the succedent of the right premiss and of the conclusion of a *Cut* are identical, Gentzen normal proofs can be written as

$$\cfrac{\cfrac{\cfrac{R_1 \qquad \cfrac{R_0 \quad \Gamma_0 \to A}{\Gamma_1 \to A}\,\text{Cut}}{\vdots}\,\text{Cut}}{\cfrac{R_{m-1} \qquad \Gamma_{m-1} \to A}{\Gamma_m \to A}\,\text{Cut}}}{\Gamma \to A}\,\text{Thinning}\ .$$

If we represent absurdity by a special atom denoted by the empty succedent, we obtain

$$\cfrac{\cfrac{\cfrac{R_{m-1} \quad \cfrac{\vdots}{\cfrac{R_1 \quad \cfrac{R_0 \quad \Gamma_0 \to}{\Gamma_1 \to}\text{Cut}}{}\text{Cut}}}{\Gamma_{m-1} \to}}{\Gamma_m \to}\text{Cut}}{\Gamma \to}\text{Thinning} \quad .$$

Clauses with empty succedent are usually called *goal clauses.* A Gentzen normal proof of the empty goal would then have the form

$$\cfrac{\to\ B \qquad \cfrac{\vdots}{\cfrac{R_1 \quad \cfrac{R_0 \quad \Gamma_0 \to}{\Gamma_1 \to}\text{Cut}}{}\text{Cut}}\ B \to}{\to}\text{Cut}$$

without *Thinning* at the end. If we know that for a set $\mathfrak{P}$ of sentences without empty succedents (the *program*),

$$\mathfrak{P},\ \Gamma_0 \to \quad \models \quad \to$$

holds, then the derivation resulting from *Theorem 7* has the form required, as $\Gamma_0 \to$ can only occur as a major sentence of a cut.

6. Hertz's Structural Logic with Variables

From the point of view of modern resolution theory, propositional resolution based on the cut rule is a "trivial" discipline, as the unification of variables, which is the basic ingredient of the resolution rule, is not involved. Although the resolution calculus based on unification was first presented by Robinson (1963), it is interesting that Hertz sketches a calculus with variables, which for definite clauses has the deductive power of the resolution calculus. It is obvious and well-known that *Cut* together with the substitution rule

$$\frac{S}{S[t/x]}\ \text{(Substitution)},$$

is strong enough to replace the resolution rule[26]. Hertz's system with variables, which is developed in his 1929 paper[27], contains principles corresponding to *Substitution.* He even sketches a proof that, modulo substitution, his normal form theorems extend to

[26] but not the resolution method as a method of constructing proofs. For that purpose resolution with unification is indispensable. (See the last paragraph of section 2 above.)

[27] Second Section: "Sentences with Variables", pp. 485 – end of paper

the case with variables, yielding a result which corresponds to the completeness of SLD-resolution with variables.

Hertz defines atoms of the form $R(x_1, \ldots, x_n)$ for n-ary predicates and variables $x_1, \ldots, x_n$. As before, sentences have the form $A_1, \ldots, A_n \rightarrow B$ for atoms $A_1, \ldots, A_n, B$. For philosophical reasons, Hertz does not permit any mixture of variables and constants in the same sentence, not even in the same proof.[28] He distinguishes between a *macrosentence* ("Makrosatz") with variables and a *microsentence* ("Mikrosatz") with constants. So

$$R(x_1, x_2), R(x_2, x_3) \rightarrow R(x_1, x_3)$$

is a macrosentence, whereas

$$R(a, b), R(b, c) \rightarrow R(a, c)$$

is a microsentence.

In Hertz's writings the situation is even more complicated as macrosentences are abstracted from sets of microsentences being their instances. Thus macrosentences are identified if they have the same set of instances, i.e. if they result from each other by renaming of variables. In the terminology of logic programming they are *variants* of each other.[29]

The proof system for macrosentences Hertz proposes, results from the propositional system dealt with in Section 3 above by adding the following two inference rules:

$$\frac{S}{S[y/x]}\ \text{(Binding)} \qquad\qquad \frac{S}{S'}\ \text{(Formal Inference)}$$

$$(x \text{ and } y \text{ variables in } S) \qquad\qquad (S' \text{ variant of } S)$$

Binding ("Bindung") allows one to identify variables, i.e. to properly specialize a clause. By *Formal Inference* ("Formaler Schluß") we can change the names of variables. The term *Formal Inference* results from the fact that in Hertz's framework this inference passes from one representation of a macrosentence to another one of the same sentence, i.e. the premiss and the conclusion of the inference denote the same sentence. This means that for Hertz the difference between premiss and conclusion is just a "formal" difference (a matter of symbols, not of what is symbolized).

From the modern point of view, both rules taken together have the strength of *Substitution* for systems with mixed variables and constants. We just have to let certain distinguished variables play the role of constants. Then *Formal Inference* accounts for the substitution of a term which does not occur in S, whereas *Binding* accounts for the

[28]He does not even admit extra variables in the succedent of a sentence, i.e. variables not already occurring in its antecedent (pp. 486 seq.).

[29]We do not follow Hertz's construction of sentences and to his terminology in detail, as we are mainly interested in normal forms for proofs.

substitution of a term already occurring in S. It should be noted that, since *Formal Inference* is just a matter of rewriting a clause, it can be omitted in the "official" definition of a proof.

Hertz remarks that due to the presence of *Binding*, normal forms in the original sense cannot be achieved, as the following example shows, for which there is no Aristotelian normal form:

$$\dfrac{P(x) \to Q(x,x) \quad \dfrac{Q(x,y) \to R(x,y)}{Q(x,x) \to R(x,x)}\text{ Binding}}{P(x) \to R(x,x)}\text{ Syllogism}$$

However, Hertz also remarks that both Aristotelian and Goclenian normal forms can be obtained, when restrictions concerning normal forms are relaxed in the following way: Instead of requiring that major or minor premisses, respectively, of *Syllogism* be axioms, they may now be axioms followed by applications of *Binding*[30].

The argument Hertz gives (1929, pp. 498 seq.) is closely connected to modern demonstrations of the completeness of SLD-resolution with variables. He first replaces macrosentences with corresponding microsentences, which yields propositional proofs. Then he applies his normal form theorems for the propositional case to obtain normal proofs containing only microsentences. Finally he observes that these normal proofs are of such a form that normal proofs consisting of macrosentences can be extracted from them. In modern terminology, this observation corresponds to the *Lifting Lemma*[31], which says that substitution and resolution can be interchanged, so that from a ground resolution proof, a proof for clauses with variables can be obtained.

Hertz structural system with variables and his normal form theorem for this system was not taken up by Gentzen, who never considered structural reasoning with non-ground atoms. Inspired by Hertz's investigations of the propositional case, Gentzen passed on to his systems for first-order logic and arithmetic. It is the more modern background of resolution and logic programming which enables us to fully appreciate Hertz's achievements.

References

Abrusci, V. Michele (1983). Paul Hertz's logical works: Contents and relevance. In: *Atti del Convegno Internazionale di Storia della Logica, San Gimignano, 4–8 dicembre 1982*, Bologna: CLUEB, pp. 369–374.

[30] or by applications of *Binding* and *Formal Inference*, if *Formal Inference* is considered a proper rule.

[31] See e.g. Lloyd (1987), pp. 47 seq.

Bernays, Paul (1969). Hertz, Paul. Biographical entry in: *Neue Deutsche Biographie* (ed. Historische Kommission bei der Bayerischen Akademie der Wissenschaften), Vol. 8, Berlin: Duncker & Humblot, pp. 711–712.

Carnap, Rudolf (1929). *Abriß der Logistik. Mit besonderer Berücksichtigung der Relationstheorie und ihrer Anwendungen.* Wien: Springer.

Gentzen, Gerhard (1933). Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen. *Mathematische Annalen, 107,* 329–350. English translation in: *The Collected Papers of Gerhard Gentzen* (ed. M. E. Szabo), Amsterdam: North Holland (1969), pp. 29–52.

Gentzen, Gerhard (1934/35). Untersuchungen über das logische Schließen. *Mathematische Zeitschrift, 39*, pp. 176–210, 405–431. English translation in: *The Collected Papers of Gerhard Gentzen* (ed. M. E. Szabo), Amsterdam: North Holland (1969), pp. 68–131.

Hertz, Paul (1922). Über Axiomensysteme für beliebige Satzsysteme. I. Teil. Sätze ersten Grades. (Über die Axiomensysteme von der kleinsten Satzzahl und den Begriff des idealen Elementes.) *Mathematische Annalen, 87,* pp. 246–269.

Hertz, Paul (1923). Über Axiomensysteme für beliebige Satzsysteme. II. Teil. Sätze höheren Grades. *Mathematische Annalen, 89*, pp. 76–100. 246–269.

Hertz, Paul (1928). Reichen die üblichen syllogistischen Regeln für das Schließen in der positiven Logik elementarer Sätze aus? *Annalen der Philosophie und philosophischen Kritik, 7,* pp. 272–277.

Hertz, Paul (1929). Über Axiomensysteme für beliebige Satzsysteme. *Mathematische Annalen, 101,* pp. 457–514. 246–269.

Hertz, Paul (1929a). Über Axiomensysteme beliebiger Satzsysteme. *Annalen der Philosophie und philosophischen Kritik, 8,* pp. 178–204.

Hilbert, David (1923). Die logischen Grundlagen der Mathematik. *Mathematische Annalen, 88,* pp. 151–165.

Legris, Javier (1999). Paul Hertz's Proof-Theoretical Conception of Logical Consequence. Unpublished Manuscript.

Leitsch, Alexander (1997). *The Resolution Calculus.* Berlin etc.: Springer.

Lloyd, John W. (1987). *Foundations of Logic Programming.* Berlin etc.: Springer (2nd edition.)

Nienhuys,Cheng, Shan-Hwei & Wolf, Ronald de (1997). *Foundations of Inductive Logic Programming.* Berlin etc.: Springer (Lecture Notes in Computer Science, Vol. 1228.)

Robinson, John A. (1963). A machine-oriented first-order logic. Abstract. *Journal of Symbolic Logic. 28,* p. 302. Full paper 1965: A machine-oriented logic based on the resolution principle, *Journal of the Association for Computing Machinery, 12,* pp. 23–41.

Slaney, John (1989). Solution to a Problem of Ono and Komori. *Journal of Philosophical Logic, 18,* pp. 103–111.