PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:

...

Keywords:

...

Author for correspondence:

...

From mathematical axioms to mathematical rules of proof: recent developments in proof analysis

Sara Negri and Jan von Plato

University of Helsinki

A short text in the hand of David Hilbert, discovered in Göttingen a century after it was written, shows that Hilbert had considered adding a 24th problem to his famous list of mathematical problems of the year 1900. The problem he had in mind was to find criteria for the simplicity of proofs and to develop a general theory of methods of proof in mathematics. It is discussed to what extent proof theory has achieved the second of these aims.

....

THE ROYAL SOCIETY PUBLISHING

David Hilbert presented his famous list of open mathematical problems at the international congress in Paris in 1900. First in the list was Cantor's continuum problem, the question of the cardinality of the set of real numbers. The second problem concerned the consistency of the arithmetic of real numbers, i.e., mathematical analysis, and so on until a problem dealing with the calculus of variations, the 23rd and last problem—or so it was for the whole century, when German historian of science Rüdiger Thiele found from old archives in Göttingen some notes in Hilbert's hand:

> As a 24th problem of my Paris talk I wanted to pose the problem: criteria for the simplicity of proofs, or, to show that certain proofs are simpler than any others. In general, to develop a theory of proof methods in mathematics.

Hilbert's text in German, with a picture of the handwritten original, is found in T. Koetsier (2001). We shall here discuss the general part of Hilbert's last problem to which a large part of foundational research in mathematics has been dedicated since the 1920s, namely *proof theory*.

What were the prevailing ideas about mathematical proof around the year 1900? Toward the end of the preceding 19th century, there was a lot of work on the axiomatization of elementary geometry. The traditional big question of elementary geometry had always been the role of Euclid's parallel postulate. Thus, questions of independence of axioms were studied, and such questions are well-posed only relative to a system of mathematical proof. A mere listing of the axioms would not be enough. Perhaps the first one to explicitly give principles of proof in addition to axioms was the German mathematician Gottlob Frege, in his small book titled "Begriffsschrift" (Concept notation) of the year 1879. This book laid the foundations of modern logic: it contained the logic of the connectives and quantifiers together with their rules of inference. The subtitle of Frege's book announced that his "concept notation" was "a formula language for pure thought, built upon the model of arithmetic." Together with such pioneers as Giuseppe Peano, the thought emerged of presenting mathematical statements entirely in formulas.

Frege's notation was forbidding, actually two-dimensional, and his writings were not properly understood. His central discovery, namely what the principles of reasoning with the universal quantifier are, was saved by one single reader around 1902: Bertrand Russell's in his 1903 book *The Principles of Mathematics*. He tells in the preface that he had seen Frege's 1893 *Grundgesetze der Arithmetik* but added that he "failed to grasp its importance or to understand its contents," the reason being "the great difficulty of his symbolism." Upon further study, he wrote a lengthy appendix with the title *The logical and arithmetical doctrines of Frege*, and ended up with the three-volume *Principia Mathematica* in 1910–13 in which Frege's logic is in a central position.

Russell's work, thus, used the notation of Peano and the principles of proof of Frege. A conditional statement, say "A implies B," was written as $A \supset B$ in which the horseshoe is a stylized version of a horizontally inverted "C" used by Peano. The symbol derives from the word "consequence." To express existence, Peano analogously inverted the letter "E" with the formula $\exists x A(x)$, there exists an x such that $A(x)$ holds. (Gerhard Gentzen in 1933 introduced the vertically inverted "A" to express universality, as in $\forall x A(x)$.)

Russell's (and Frege's) formulation of logic started with axioms and included two rules of proof: 1. Whenever a conditional statement $A \supset B$ and its condition A have been proved, it is allowed to conclude B. 2. Whenever the property $A(x)$ has been proved for an arbitrary x, it is allowed to conclude the universally quantified statement $\forall x A(x)$. It is sufficient to consider only implication and universal quantification and their rules, because the rest of the connectives and quantifiers can be defined: Negation $\neg A$ is obtained as a special case of implication, by substituting for B some impossible statement, such as 0 = 1 in arithmetic. Next disjunction can be defined by $A \vee B \equiv \neg A \supset B$, conjunction by $A \& B \equiv \neg(A \supset \neg B)$, and existence by $\exists x A(x) \equiv \neg \forall x \neg A(x)$.

It is common to consider Hilbert's book *Grundlagen der Geometrie* (The Foundations of Geometry) of the year 1899 as the beginning of formalized mathematics. Such a conception prevails only because the book is not read; there one finds no formula language for geometry, not to speak of explicit principles of proof. The axioms and proofs of geometry in Hilbert are verbal explanations not unlike those found in Euclid more than two thousand years earlier.

The aim of formalization is that "nothing should be left to guesswork," as Frege expressed it in 1879. The point of departure is a choice of basic concepts, and the method that of trial and error. The lack of formalization and "rigor" in Hilbert's geometry (a word he liked to use) led to errors: For example, the concept of parallel lines was adequately presented only in the seventh edition of the book, in 1930. Another example of "guesswork:" Hilbert states that "any two points that are on different sides of a line in a plane, are distinct," but he gives no reason for why this should follow. To think that it is obvious is precisely to betray the point of formalization. Here is a possible argument: Let the two points be a and b and the line l. Let us call the side on which a is "the left side," formally, $L(a, l)$, and similarly for "the right side" $R(b, l)$. Let now $L(a, l)$ and $R(b, l)$, and assume for the sake of the argument that $a = b$. By *substituting* the expression b for a in the formula $L(a, l)$, we conclude $L(b, l)$. Thus, we need the axiom $\neg(L(b, l) \& R(b, l))$ ("no point is on both sides of a line") to conclude that $a = b$ is impossible if $L(a, l)$ and $R(b, l)$ are assumed.

One essential step of inference in the above was a *principle of substitution of equals*, tacitly used by Hilbert. In modern terms, we state it by requiring that a property such as $L(a, l)$ is a congruence relative to the equality of points. Remarkably, some students of Peano such as the geometer Mario Pieri had understood the importance of substitution principles even before Hilbert published his geometry.

Turning now to the principles of proof of Hilbert's geometry, there is nothing to report. Hilbert's first thoughts for a theory of proofs are from the year 1904. At this stage, the representation of mathematics as a formula language and of mathematical proof as a purely symbolic manipulation of formulas without considering their meaning is clear. The central and almost only aim of the exercise of formalization is to show the consistency and completeness of formalized arithmetic and analysis. When this aim is reached, the problems about the foundations of mathematics can be forgotten, left behind for good, thought Hilbert. Here Hilbert's attitude is completely different from what his statement of the 24th problem only four years earlier suggests. Now it would not matter at all in what way mathematics is formalized, if the two central aims of formalization are reached.

We know since the results of Gödel that Hilbert's original proof-theoretic program failed: By his first incompleteness theorem, there is no complete formalization even for the case of arithmetic. Secondly, by the second incompleteness theorem, the consistency of Peano arithmetic is one of the unprovable "Gödel sentences" and therefore has no strictly finitary, hence no "absolutely reliable," proof. However, it is surprising how soon the study of the foundations of mathematics recovered from this shock. Some, such as von Neumann, thought that Hilbert's aim of a proof of consistency of mathematics is lost forever, but others such as Gerhard Gentzen, a student of Paul Bernays who was working for Hilbert, soon regained hope. It took him only one year from Gödel's paper of 1931 to realize that the consistency of arithmetic is not completely lost, even if it cannot be shown in any Hilbertian strictly finitary way.

Gentzen cleared the way to the proof theory of arithmetic by first recasting the logic of the connectives and quantifiers in a new form. In his magnificent doctoral thesis *Untersuchungen über das logische Schliessen* (Investigations into logical inference) of the year 1933, his objective was to study "the structure of mathematical proofs as they appear in practice." The setting of this task is neutral; it does not commit Gentzen to any specific view on the foundations of mathematics, be it formalism, intuitionism, or Cantorism.

Gentzen presented first a general theory of the structure of mathematical proofs. For each form of proposition, *sufficient conditions* for concluding it are given: A conjunction $A\&B$ can be concluded if the premisses A and B have been separately established, a conditional $A \supset B$ can be concluded if B has been proved from the *assumption* of A, and $\forall x A(x)$ can be concluded if $A(x)$ has been proved for an *arbitrary* x.

The critical point in Gentzen's explanation of conditional statements is to avoid circularity: What if A itself is a conditional statement? Gödel, for example, made this objection in unpublished work of 1941: Gentzen's "sufficient ground" for concluding $A \supset B$ amounts to a method that can convert an arbitrary proof of A into some proof of B. How, then, can one explain what such direct proofs of conditional statements are if it is assumed that the notion of arbitrary proof is understood? This specific problem found its final solution only in the 1970s in the work of Michael Dummett and Per Martin-Löf.

In addition to sufficient conditions for concluding a proposition of a given form, Gentzen also gave reverse *necessary conditions* that state what we are committed to if we assert a proposition of any of the forms considered. For $A\&B$ these necessary conditions are A and B separately, for $A \supset B$ they are B on the assumption or proof that A holds, and for $\forall x A(x)$ an instance $A(a)$ for any object a. We can collect the sufficient and necessary conditions for propositions, here for the above three forms, into what is known as Gentzen's system of *natural deduction*, with its *introduction* and *elimination rules*:

$$\frac{A \quad B}{A \,\&\, B}\,\&I \qquad \frac{\begin{array}{c}[A]\\ \vdots\\ B\end{array}}{A \supset B}\supset I \qquad \frac{A(x)}{\forall x A(x)}\,\forall I$$

$$\frac{A \,\&\, B}{A}\,\&E \qquad \frac{A \,\&\, B}{B}\,\&E \qquad \frac{A \supset B \quad A}{B}\supset E \qquad \frac{\forall x A(x)}{A(t)}\,\forall E$$

1. GENTZEN'S SYSTEM OF NATURAL DEDUCTION.

In rule $\supset I$, the temporary assumption A has been *closed*, which is indicated by the square brackets. Rule $\forall I$ has a *variable restriction*: the variable x must not occur free (not under the scope of a quantifier) in any assumptions on which the premiss of the rule depends. The condition on the rule guarantees that the property $A(x)$ hold for an arbitrary x.

Two of the rules have two premisses. Formal proofs, called *derivations*, have therefore the form of a tree the leaves of which are assumption formulas. By rule $\&I$, we can combine a derivation of A with any other previously derived result B, and then make away with the latter by one of the rules $\&E$. Thus, there is no upper limit to how complicated derivations of a formula A can be. Gentzen observed that successive applications of introduction and elimination rules are detours in a derivation that can be eliminated. This is obvious with rules $\&I$ and $\&E$. With $\supset I$ and $\supset E$, we observe that a derivation of the second premiss A of rule $\supset E$ can substitute the assumption of A in the preceding rule $\supset I$. Through an opportune combination, we have a derivation of B without the pair of rules $\supset I, \supset E$: Take first the derivation of A, then continue to B as in the derivation of the premiss of rule $\supset I$. We then have the two schematic figures:

$$\begin{array}{c}\vdots \quad \vdots\\ \dfrac{\dfrac{A \quad B}{A \,\&\, B}\,\&I}{A}\,\&E\\ \vdots\end{array} \quad \text{becomes} \quad \begin{array}{c}\vdots\\ A\\ \vdots\end{array} \qquad\qquad \begin{array}{c}\dfrac{\dfrac{\begin{array}{c}[A]\\ \vdots\\ B\end{array}}{A \supset B}\supset I \qquad \begin{array}{c}\vdots\\ A\end{array}}{B}\supset E\\ \vdots\end{array} \quad \text{becomes} \quad \begin{array}{c}\vdots\\ A\\ \vdots\\ B\\ \vdots\end{array}$$

2. CONVERSION OF INTRODUCTION-ELIMINATION PAIRS.

The "conversion to normal form" holds quite generally for the logic of the connectives and quantifiers (cf. e.g. Negri and von Plato, 2001). We can delete all successive "I-E-pairs" to obtain a derivation with the property that all of its formulas are contained, possibly as parts, in the open assumptions or the conclusion of the derivation. This *subformula property* was invented by Gentzen; it gives an upper bound for the complexity of formulas that can appear in a derivation in normal form. It often follows that there is an upper bound also to the number of distinct formulas that can appear. When a logical calculus is built in a right way, it further follows that there can be only a bounded number of different derivations in normal form, which gives a decision method for derivability in the calculus in question. This is the case for the connectives, but with the quantifiers, there is no general decision method. Each instance $A(a)$ is a subformula of both $\forall x A(x)$ and $\exists x A(x)$, but these are infinite in number if the domain of individuals considered $a, b, c, \dots$ is infinite.

Gentzen developed his central results on the structural analysis of proofs for a calculus that differs somewhat from the calculus of natural deduction presented above: In his *sequent calculus*, all open assumptions are collected into a list, let it be Γ, and the derivability of a formula C from assumptions Γ is written as $\Gamma \to C$. Thus, the relation of derivability between a list of assumptions and a formula has been presented in a local way in one line, whereas in natural deduction this relation has to be read from the root and the leaves of the derivation tree. Sequent calculus has rules that correspond to the introduction rules of natural deduction, and they transform the conclusion. Secondly, there are rules that correspond to the elimination rules, and these transform the open assumptions. Both transformations are always such that the formulas in the premisses of a rule are subformulas of the conclusion. Derivations start from sequents of the form $A \to A$ that correspond to the assumption of a formula A, and each rule adds some complexity to the left or right of the arrow.

The normal form of natural deduction, or the eliminability of "I-E-pairs," is expressed in sequent calculus through Gentzen's famous cut rule. In a typical case, we have a derivation of some result of interest C from assumptions A and Δ, and in a second stage we succeed in establishing a lemma that shows A superfluous in the sense that it follows from some assumptions Γ that we are prepared to accept, say axioms or previously established theorems. By the lemma we have $\Gamma \to A$, and by the main result we have $A, \Delta \to C$. The cut rule combines these:

$$\frac{\Gamma \to A \quad A, \Delta \to C}{\Gamma, \Delta \to C}\,\mathit{Cut}$$

Formula A has disappeared from the list of assumptions, and there is no a priori bound on the complexity of possible cut formulas that might be needed in order to derive $\Gamma, \Delta \to C$. Gentzen's main result is a procedure by which applications of the rule of cut can be eliminated from derivations. The subformula property is then straightforward, with again an upper bound on the complexity of formulas in a derivation.

Gentzen's structural proof theory first covered logic, that is, the structure of mathematical argument in general. However, his initial objective was to extend the theory to arithmetic and analysis. Work on the former was completed in 1935, only two years after the doctoral dissertation. The derivations of formalized Peano arithmetic are represented as trees constructed by suitable rules of inference. Next these trees are ordered linearly in a certain way according to their complexity and it is shown that if a contradiction is derivable, this ordering has an infinite descending chain. The ordering Gentzen used is a "transfinite induction up to the first ε-number." Such a generalized induction characterizes Peano arithmetic: The induction principle itself can be expressed in arithmetic, but it follows from Gödel's results that it cannot be proved in arithmetic. Thus, Gentzen's transfinite induction is the first example of an "ordinary" arithmetical statement that is unprovable in Peano arithmetic, whereas Gödel's unprovable formula was obtained through an arithmetic coding of the provability relation, a concept of the metamathematics of formalised systems. Gentzen also found a direct proof of the unprovability of his transfinite induction principle, in his paper of 1939, published in 1943, that marks the beginning of what is known as *ordinal proof theory*.

In conclusion, Gentzen's work solved the general part of Hilbert's 24th problem, "to develop a theory of proof methods in mathematics," as far as the general principles of proof and proofs in elementary arithmetic are concerned. After this success, Gentzen's next natural aim was the proof theory of analysis, but this work was in its beginnings when he died in tragic circumstances in August 1945 (for extensive and carefully detailed documentation on Gentzen's personal and

rsta.royalsocietypublishing.org Phil. Trans. R. Soc0000000

scientific life the reader is referred to Menzler-Trott, 2007). Others have continued the program, even if the difficulties have been great. The strength of different subsystems of analysis is measured by determining what transfinite principles are needed in order to prove these systems consistent.

Arithmetic and analysis were the first natural objectives of the proof theory that Hilbert envisioned, brought to success in arithmetic, and with an ongoing work in analysis. Other parts of mathematics have been left relatively untouched by proof theorists. In our own work, we have found that the structure of proofs can often be analyzed without encountering at all the kind of limitations that follow from Gödel's incompleteness results. Such limitations are met if the theory in question has the natural numbers and their arithmetic built in.

It is possible to study, for example, elementary theories of order, lattice theory, and elementary geometry by purely proof-theoretical means, by converting their axioms into suitable systems of rules in what we have called *proof analysis*, a methodology extensively studied in our monograph Negri and von Plato (2011) and in a series of articles. Let us look at the theory of linear order as a clear and concise example:

We have, as usual, a reflexive and transitive partial order relation $a \leqslant b$ to which the linearity postulate $a \leqslant b \vee b \leqslant a$ is added. Our proof analysis of order relations uses Gentzen's sequent calculus, but we can describe it here at least approximately in terms of natural deduction. Reflexivity is taken into account by letting derivation trees have formulas $a \leqslant a$ as leaves. Transitivity of order is the two-premiss rule by which $a \leqslant c$ can be concluded whenever the premisses $a \leqslant b$ and $b \leqslant c$ are at hand. The linearity postulate is translated into a rule of inference as follows: Let a formula C be derived from the assumption $a \leqslant b$, and let it also be derived from the assumption $b \leqslant a$. Both cases of the linearity postulate have thus led to C, so the rule of linearity concludes C with the assumptions $a \leqslant b$ and $b \leqslant a$ deleted from the list of open assumptions.

We can represent the application of the linearity postulate in a schematic form:

$$\dfrac{\begin{array}{c}\overset{1}{[a \leqslant b]} \\ \vdots \\ C\end{array} \qquad \begin{array}{c}\overset{1}{[b \leqslant a]} \\ \vdots \\ C\end{array}}{C}\,{\scriptstyle Lin,1}$$

The two cases in linearity, here temporary assumptions, have been closed at the inference.

The above inference seems to go against the subformula property: two formulas seemed to disappear entirely from the list of open assumptions. However, the following can be proved: If the derivation of C is of minimum size, meaning that it cannot be locally shortened from here or there by the deletion of a superfluous rule, say a two-premiss rule in which one premiss is identical to the conclusion, then all the elements in the overall derivation, such as a and b above, appear somewhere in the open assumptions or the conclusion. This can be called the *subterm property* of minimum-size derivations, in analogy to Gentzen's subformula property of cut-free derivations. As a consequence of the subterm property, the linearity postulate has only a bounded number of distinct instances, with results such as the decidability of derivability for quantifier-free formulas in a linear order as a consequence. In particular, the word problem for a linear order receives a positive solution by a clear-cut analysis of the structure of minimal derivations. The actual proof of the subterm property is far from trivial: we consider an uppermost instance of the linearity rule, and above it, we have instances of transitivity. It is shown that for any term, such as a and b, there is at least one path up the derivation tree in which that term has to appear in a formula, and it has to be a term in an open assumption at some point, or the derivation grows to infinity.

The conversion of mathematical axioms into rules added to a logical calculus can be extended to other theories, such as the first-order theory of *linear Heyting algebras*. More specifically, one can show how rules in a sequent calculus extended by mathematical rules correspond to rewrite rules for expressions of the form $a \leqslant b$ where a and b are terms in the algebra. Linearity allows one to decompose such terms with rewrite rules for operations on both sides of the order relation

rsta.royalsocietypublishing.org Phil. Trans. R. Soc0000000

Consider the example (with the rewrite operation indicated by $\rightarrow$):

$$c \leqslant a{\vee}b \rightarrow c \leqslant a \vee c \leqslant b$$

As in linearity, there are two cases, but only under the assumption $c \leqslant a{\vee}b$. In terms of natural deduction, we have the rule:

$$\dfrac{c \leqslant a{\vee}b \qquad \begin{matrix}[c \leqslant a] \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[c \leqslant b] \\ \vdots \\ C\end{matrix}}{C}$$

Whenever C follows from each of the cases, it follows from the condition $c \leqslant a{\vee}b$ that gives the cases. The rule in a formulation in sequent calculus is:[1]

$$\dfrac{c \leqslant a, \Gamma \rightarrow \Delta \quad c \leqslant b, \Gamma \rightarrow \Delta}{c \leqslant a{\vee}b, \Gamma \rightarrow \Delta}$$

Similar rules are found for rewrite conditions of the "exponentiation" operation of Heyting algebras. This operation is characterized by the equivalence

$$c \leqslant b^a \supset\subset c{\wedge}a \leqslant b$$

In the presence of linearity, one obtains the rewrite conditions (with 1 denoting the top element of the algebra)

$$c \leqslant b^a \rightarrow c \leqslant b \vee a \leqslant b \qquad c^a \leqslant b \rightarrow c \leqslant b \,\&\, (a \leqslant c \supset 1 \leqslant b)$$

and corresponding sequent calculus rules. Dyckhoff and Negri (2006) developed this approach and gave a simple decision method, based on terminating proof search in a suitable sequent calculus, for the fragment of positively quantified formulas of the first-order theory of linearly ordered Heyting algebras. (Positively quantified formulas are those in which the universal quantifier occurs only in positive position, and the existential quantifier only in negative position.)

A natural question concerns the potential range of applications of the methodology of "axioms-as-rules." Many interesting mathematical theories can be expressed by means of coherent/geometric[2] implications: these are universal closures of formulas of the form $C \supset D$ where C and D are first-order formulas built up from atoms using conjunction, disjunction and existential quantification. They include all algebraic theories, such as group theory and ring theory, all essentially algebraic theories, such as category theory, the theory of fields, the theory of local rings, lattice theory, projective geometry, the theory of separably closed local rings, and the infinitary theory of torsion abelian groups (see Dyckhoff and Negri 2015 for references). A useful normal form for such axiomatizations has C that consists of conjunctions of atoms and D of a disjunction of existentially quantified conjunctions of atoms, i.e., any geometric implication can be reduced to the form

$$\forall \overline{x}(\& P_i \supset \exists \overline{y}_1 M_1 \vee \cdots \vee \exists \overline{y}_n M_n)$$

The P_i range over a finite set of atomic formulas and all the M_j are conjunctions of atomic formulas Q_{j_i} and the variables in the vector $\overline{y}_j$ are not free in the P_i.

The general rule scheme that corresponds to the geometric axiom above is

$$\dfrac{\overline{Q}_1(z_1/y_1), \overline{P}, \Gamma \rightarrow \Delta \quad \ldots \quad \overline{Q}_k(z_k/y_k), \overline{P}, \Gamma \rightarrow \Delta}{\overline{P}, \Gamma \rightarrow \Delta}$$

The scheme has the condition that its eigenvariables z_i are not free in $\overline{P}, \Gamma, \Delta$. Adding such a rule scheme to a suitable sequent calculus does not affect its properties, most importantly that

[1]In this rule, Δ can represent a finite number of alternative cases rather than a single conclusion. The possibility to have more than one formulas to the right of the sequent arrow gives what is called a *multisuccedent* sequent calculus, an approach that gives a greater generality.

[2]The terminology in the literature is not uniform and the name *geometric* is sometimes reserved for axiomatizations of the mentioned form, sometimes for the more general case that allows infinitary disjunctions. We shall follow here the first, finitary, sense.

the rule of cut is not needed (a result proved in Negri 2003), thus a coherent/geometric theory can be presented directly as a sequent calculus rule system, with the mathematical rules each corresponding to the conjuncts of the normal form of its axioms. Important examples of theories that become in this way amenable to a cut-free sequent calculus treatment are the theories of *Robinson arithmetic* and of *real closed fields*.

The conversion of geometric axiomatic theories into inference rules results in a uniform treatment of theories based on classical and intuitionistic logic: with multisuccedent sequent calculi the difference between the two logical systems is just in the rules for implication and the universal quantifier, whereas the additional part corresponding to the geometric rules for the theory under consideration is identical. By this uniformity, a striking application is obtained of the method of conversion of geometric axioms into rules, namely an almost immediate proof of the first-order Barr's theorem, a central result of constructive mathematics by which a classical proof of a geometric implication in a geometric theory can be transformed into a constructive proof. It turned out that in a cut-free system of sequent calculus with any collection of geometric rules, a classical proof of a geometric implication is already a constructive proof (cf. Negri 2003 and Negri and von Plato 2011). Although the term "geometric" for these axiomatizations does not originate from geometry but from category theory, geometric theories and their proof-theoretic treatment through the geometric rule scheme have been employed for a formalization of Euclidean geometry in Avigad et al. (2009) and for projective and affine geometry in Negri and von Plato (ch. 10, 2011).

Not all first-order axioms are reducible to geometric implications. A generalization of geometric implications, originally motivated by problems in the proof-theoretic study of certain epistemic logics (cf. Maffezioli, Naibo, Negri, 2013), was introduced by Negri (2016), where a class $\mathbf{GGI}$ of first-order sentences is defined recursively as follows: $\mathbf{GA_0}$ is the class of *special coherent implications*[3] $\forall \mathbf{x}.H_0 \supset \exists \mathbf{y}_1 H_1 \vee \cdots \vee \exists \mathbf{y}_m H_m$, and, for $n \geqslant 0$, $\mathbf{GA_{n+1}}$ is the class of sentences $\forall \mathbf{x}.H \supset (\exists \mathbf{y}_1 G_1 \vee \cdots \vee \exists \mathbf{y}_m G_m)$, where each H_i is a (possibly empty) conjunction of atoms, $m \geqslant 0$ and each G_i is a conjunction (with free variables in $\mathbf{x}, \mathbf{y}_i$) of sentences from $\mathbf{GA_k}$ for $k \leqslant n$. With $m = 0$, the succedent of the implication is just $\bot$; likewise, sentences where H_0 is the empty conjunction $\top$ are identified with sentences from which the implication symbol and its trivial antecedent are omitted. Sentences from $\mathbf{GA_n}$ are implicitly also sentences of $\mathbf{GA_{n+1}}$. $\mathbf{GGI}$ is then the union of the classes $\mathbf{GA_n}$ for $n \geqslant 0$. The subformulae G_i will be called **GGI**-*subformulae*.

It then follows that (i) a sentence in **GGI** has no negative occurrences of implication (hence has no negative occurrences of negation, since $\neg A$ is just an abbreviation for $A \supset \bot$) or universal quantification; (ii), conversely, a sentence with this property is intuitionistically equivalent to a conjunction of sentences in **GGI**.

Negri (2016) introduced "systems of rules" for the class **GGI**: alternations of quantifiers are coded up through a prescribed order in which the added rules and their eigenvariables may occur in a derivation, so that the rules require extra book-keeping of the variable dependences.

As an example of an axiom in the class of generalized geometric implications, consider the axiom of *join semi-lattices*:

$$\forall xy \exists z ((x \leqslant z \,\&\, y \leqslant z) \,\&\, \forall w (x \leqslant w \,\&\, y \leqslant w \supset z \leqslant w)) \qquad \textit{lub-A}$$

The system of rules consists of the following two rules of existence and uniqueness of the least upper bound:

$$\frac{x \leqslant z, y \leqslant z, \Gamma \to \Delta}{\Gamma \to \Delta}\,\textit{lub-E} \qquad \frac{z \leqslant w, x \leqslant w, y \leqslant w, \Gamma \to \Delta}{x \leqslant w, y \leqslant w, \Gamma \to \Delta}\,\textit{lub-U}$$

Rule *lub-E* has the condition that z is fresh (i.e., not in the conclusion of the rule), whereas rule *lub-U* has the condition that in a derivation it should be applied only above (but not necessarily immediately above) rule *lub-E* (and we use the same x, y, z). This means that any derivation that uses rule *lub-U* must have a branch of the following form, with the condition is that z is not free

[3]Special coherent implications are a normal form for coherent implications, cf. def. 2.5 in Dyckhoff and Negri, 2015.

in Γ, Δ:

$$\dfrac{z \leqslant w, x \leqslant w, y \leqslant w, \Gamma' \to \Delta'}{x \leqslant w, y \leqslant w, \Gamma' \to \Delta'}\ {\scriptstyle lub\text{-}U}$$

$$\vdots\ \mathcal{D}\ \vdots$$

$$\dfrac{x \leqslant z, y \leqslant z, \Gamma \to \Delta}{\Gamma \to \Delta}\ {\scriptstyle lub\text{-}E}$$

As an application of systems of rules, a generalisation of the first-order version of Barr's theorem, and in the spirit of Orevkov's work (1968) on Glivenko classes is proved: if the antecedent formulas of a single-succedent sequent are **GGI** and the succedent formula is a coherent implication, then any classical proof of the sequent can be transformed to an intuitionistic proof (see also Negri 2016a for these classes).

The method of axioms-as-rules can be extended further to any first-order axiomatization, namely one can prove that any first-order axiom can be replaced by a series of geometric rules which is built starting from either the conjunctive or the disjunctive normal form of the axiom. The conversion to normal form is not even necessary; in fact, one of the main results of Dyckhoff and Negri (2015) is an algorithm of "coherentization" that preserves as much as possible of the formula structure. This is detailed out in section 17 of the mentioned paper. Compared to the approach of system of rules, this latter method is more expressive, as it covers any first-order axiom, not just those in the class **GGI**, and codifies the variable dependences through the addition of new predicate symbols, with no need of external conditions on the order of application of the rules. For example, to continue with the example of join semilattices, one adds a new primitive ternary predicate $J(x, y, z)$ (with the intended meaning that z is the least upper bound of x and y) with the following rules (in the first, z is a fresh variable):

$$\dfrac{J(x,y,z), \Gamma \to \Delta}{\Gamma \to \Delta} \qquad \dfrac{x \leqslant z, y \leqslant z, \Gamma \to \Delta}{J(x,y,z), \Gamma \to \Delta} \qquad \dfrac{z \leqslant w, \Gamma \to \Delta}{J(x,y,z), x \leqslant w, y \leqslant w, \Gamma \to \Delta}$$

It turns out that the above is a sequent calculus version of the fragment for least upper bound of Skolem's rules for relational lattice theory (cf. sec. 5.3 of Negri and von Plato, 2011). The conversion of axioms into rules, presented in a systematic way in Dyckhoff and Negri (2015), is a procedure that works in full generality for all first-order theories. It can also be used to give an alternative proof, purely based on proof analysis, of the result stating that any first-order theory has a coherent conservative extension (cf. sec. 10 of Dyckhoff and Negri, 2015).

As a final example, and closer to Hilbert's concerns around 1900 when he formulated his last problem, the question of the independence of the parallel postulate in plane affine geometry can be answered through proof analysis: No derivation can end with the postulate in a system of geometric rules that corresponds to the rest of the geometric axioms. The standard proof of independence is based on models of non-Euclidean geometry, and these in turn are based on real numbers and thereby on arithmetic. The proof-theoretic result instead comes from a thoroughly elementary combinatorial property of minimal derivation trees, namely the subterm property. The procedure by which such minimal derivations are established is of extreme complexity.

The proof analysis that leads to the independence of the parallel postulate shows, with the notation $a \in l$ for the incidence of a point a on a line l and $par(l, a)$ for the parallel line construction, the underivability of the sequent $b \in l, b \in par(l, a) \to a \in l$: in words, if point b is incident on line l and on the parallel to l through point a, then also point a is incident on line l. The contrapositive gives a traditional formulation: "If a given point is outside a given line, then no point is on the line and its parallel through the given point." The system of rules of affine geometry obeys the subterm property. It follows that there are only two rules that can be used to conclude the above sequent, and that no rule can give the premisses of these rules. The consistency of this system of geometry also follows at once from the subterm property. The result is in stark contrast with Hilbert's own statement of 1900, in the second of his Paris problems:

> In geometry, the proof of the consistency of the axioms can be effected by constructing a suitable field of numbers, such that analogous relations between the numbers of this field correspond to the geometrical axioms. Any contradiction in the deductions from the geometrical axioms must thereupon be recognizable in the arithmetic of this field of numbers. In this way the desired proof for the consistency of the geometrical axioms is made to depend on the theorem of the consistency of the arithmetical axioms

Hilbert obviously found this reduction of the consistency of geometry to that of real arithmetic to be a step ahead, but Gödel's second incompleteness theorem of 1931, about the impossibility to prove even the consistency of elementary arithmetic by finitistic means, teaches us differently: Direct finitary consistency proofs for systems of geometry are also direct progress in what Hilbert called "metamathematics." Finally, we remark that the "local" shortening of derivations, as in the subterm property, is different from a shortest or simplest possible proof in a global sense. For example, in the theory of linear order, we could apply transitivity to the premisses $a \leqslant a$ and $a \leqslant b$, with the conclusion $a \leqslant b$ identical to the second premiss. This situation is called a "loop" in the derivation. More complicated loops can have any number of steps of derivation between them. Our minimal derivations need not be shortest or simplest in any global sense: the aim is just to have any finite bound on their size. Hilbert instead in the first part of his last problem was asking for criteria of the simplicity of proofs in a general sense. What these criteria could be is a question studied in the theory of the complexity of proofs, in which general methods and results of complexity theory are applied to derivation trees. Complexity theory has received a lot of attention lately, especially through the popularity of the "P=NP" problem. Our purpose here has been to discuss developments that answer to the general part of Hilbert's last problem.

References

Dyckhoff, R. and S. Negri (2006) Decision methods for linearly ordered Heyting algebras. *Archive for Mathematical Logic*, vol 45, pp. 411–422.

Dyckhoff, R. and S. Negri (2015) Geometrization of first-order logic. *The Bulletin of Symbolic Logic*, vol. 21, pp. 123–163.

Gentzen, G. (1969) *The Collected Papers of Gerhard Gentzen*, ed. M. Szabo, North-Holland.

Koetsier, T. (2001) Hilberts 24ste probleem. *Nieuw Archief voor Wiskunde*, vol. 5, pp. 65–67.

Maffezioli, P., A. Naibo, and S. Negri (2003) The Church-Fitch knowability paradox in the light of structural proof theory. *Synthese*, vol. 190, pp. 2677–2716.

Menzler-Trott, E. (2007) *Logic's Lost Genius: The Life of Gerhard Gentzen*. AMS.

Negri, S. (2003) Contraction-free sequent calculi for geometric theories, with an application to Barr's theorem. *Archive for Mathematical Logic*, vol. 42, pp. 389–401.

Negri, S. (2016) Proof analysis beyond geometric theories: from rule systems to systems of rules. *Journal of Logic and Computation*, vol. 27, pp. 513–537.

Negri, S. (2016a) Glivenko sequent classes in the light of structural proof theory. *Archive for Mathematical Logic*, vol. 55, pp. 461–473.

Negri, S. and J. von Plato (2001) *Structural Proof Theory*. Cambridge University Press.

Negri, S. and J. von Plato (2011) *Proof Analysis: A Contribution to HIlbert's Last Problem*. Cambridge University Press.

Orevkov, V. P. (1968) On Glivenko sequent classes. *Proc. Steklov Inst. Math.*, vol. 98, pp. 147–173 (translated from the Russian original V. P. Orevkov, Glivenko's sequence classes, Logical and logical-mathematical calculus. Part I, *Trudy Mat. Inst. Steklov*, vol. 98, 1968, pp. 131–154).