



HAL
open science

Mathematical explanations and axioms as rules*

Elaine Pimentel, Francesca Poggiolesi

► **To cite this version:**

Elaine Pimentel, Francesca Poggiolesi. Mathematical explanations and axioms as rules*. *Topoi*, 2026, Special issue "Logic and explanation". <hal-04993565>

HAL Id: hal-04993565

<https://hal.science/hal-04993565v1>

Submitted on 17 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Mathematical explanations and axioms as rules*

Francesca Poggiolesi¹ and Elaine Pimentel²

¹Institut d'Histoire et de Philosophie des Sciences et des Techniques,
Université Paris 1 Panthéon-Sorbonne, CNRS, IHPST (UMR 8590), 13
Rue du Four, Paris, 75006, France.

²Department of Computer Science, University College London, 66-72
Gower St., London, WC1E 6EA, UK.

Contributing authors: poggiolesi@gmail.com; e.pimentel@ucl.ac.uk;

Abstract

This paper bridges two distinct traditions – philosophy of mathematics and proof theory – to provide a formal framework for explanatory proofs in mathematics. By focusing on explanatory proofs that reveal the grounds of mathematical theorems, we formalize their structure using a novel “axioms-as-rules” approach. Our method ensures the transformation of axioms into inference rules while preserving key logical properties such as soundness, completeness, and cut-admissibility. Two classical examples, the Quadrangle Theorem and Pythagoras’ Theorem, are revisited to show how explanatory steps can be systematically isolated and formalized. This framework not only offers novel proof-theoretic insights but also enhances our understanding of mathematical explanations by bridging informal intuition with rigorous formalism.

Keywords: Mathematical explanations, Proof theory, Axioms-as-rules, Explanatory proofs, Logical frameworks

1 Introduction

This paper is based on and brings together two distinct traditions: one from the philosophy of mathematics, centered on the concept of mathematical explanation, and the other from the proof-theoretical framework, focused on transforming axioms

***Statements and Declarations:** there is no financial or non-financial interest to declare which is directly or indirectly related to this work submitted for publication.

into rules. After briefly presenting both traditions, we will clarify how they can be effectively combined and explore the results that emerge from this combination.

Mathematical explanations. When proving a theorem, mathematicians often rely on defined objects, a mathematical theory, and previously established lemmas. The argumentation is conducted inside the framework of the mathematical theory, through logical reasoning, organizing the proof into a tree structure where such objects and lemmas are employed at specific points. Since Aristotle, two types of proofs have been distinguished, depending on their purpose: those that merely demonstrate the truth of a theorem, and those that go further by explaining why the theorem is true. There are however several ways a proof can explain, *e.g.*, by drawing a diagram or by relying on an analogy. In this paper we will focus on those proofs which explain by providing the (objective) reasons or the grounds why the theorem is true (*e.g.*, see [1]). This latter type of proofs, that we will refer to as *explanatory proofs* or *mathematical explanations*, has long intrigued philosophers, from Leibniz and Bolzano to Frege (*e.g.* see [2] and [3]). In recent decades, the topic has regained attention in contemporary philosophical literature. In particular, philosophers of mathematics have grappled with a central question: what makes them explanatory? Among the several proposals put forward, a recent approach stands out due to its rigorous analysis (see [4–6]). The central idea is to transition from *informal* proofs of mathematical theorems to a *formal* presentation that involves proof trees, with a decomposition of their elements into assumptions, conclusions, and rules. Then the aim is to show that those mathematical proofs that are regarded as having an explanatory power, they all enjoy the same formal presentation where they display an increase of conceptual complexity from the assumptions to the conclusion. In order to find such formalization though, three major constraints need to be respected:

- a. the *grounds* (or reasons why a theorem is true) need to be captured by undischarged assumptions;
- b. the *generalizations*, that typically occur in explanations, link the grounds to their theorems and can be put in an axiomatic form, need to be formalized as rules.
- c. the *rules* that are used in the proof-tree formalizing the explanatory steps of the explanatory proofs meet the desiderata elaborated in [6]. In particular, they are rules that work deep inside formulas and can be seen as introduction rules.

While (a) implies that proof-trees should necessarily be *open*, where the leaves are the grounds, and (c) involves that we will be working with a novel more powerful type of rules, (b) suggests that a *systematic transformation of generalizations into rules* can help on pushing this research plan forward.

One of the aims of this paper is to meet the suggestion contained in (b) and thus to present a significant advancements of the ideas expressed in [4]. This will be achieved by relying on a well-established tradition in proof theory, known as *axioms-as-rules*, which will be presented next.

Axioms as rules. The systematic addition of theories to a base (classical or intuitionistic) logic can be traced back to the work of Gerhard Gentzen [7]. When attempting

to prove the consistency of arithmetic without relying on complete induction in first-order logic, Gentzen proposed an approach where arithmetic axioms were assumed as hypotheses, with the goal of demonstrating that no contradiction (*i.e.*, falsehood) could be derived from these axioms using only the logical rules of the classical sequent calculus. While innovative, this approach proved highly inefficient because the space required for proofs grows exponentially with the number of axioms. Additionally, it diverged from Gentzen’s own proof-theoretical methods by relying on an axiomatic strategy.

This led to several attempts to incorporate inference rules encoding axioms into existing proof systems for (first-order) classical and intuitionistic logics. The main idea is to reflect non-logical axioms (*e.g.*, of the form $P \rightarrow Q$) with non-logical rules of inference, *e.g.*,

$$\frac{\Gamma, Q \Rightarrow \Delta}{\Gamma, P \Rightarrow \Delta} (P \rightarrow Q)_L \quad \frac{\Gamma \Rightarrow \Delta, P}{\Gamma \Rightarrow \Delta, Q} (P \rightarrow Q)_R$$

However bright the insight may be, the question of how to systematically transform axioms such as $P \rightarrow Q$ into the corresponding non-logical inference rules remain. Moreover, it needs to be settled what kind of properties the theory so obtained preserves with respect to the original sequent calculus.

A remarkable step in the direction was the careful investigation of *geometric axioms*. Geometric axioms are first-order formulas that can be converted into (natural deduction/sequent) inference rules having “a certain simple form in which only atomic formulas play a critical part”, as described by Simpson [8]. Other major works in this direction include: Ciabattoni et al. [9], who developed a systematic procedure for transforming a class of (Hilbert) axioms (called \mathcal{N}_2) into equivalent *structural* inference rules; Negri et al. (*e.g.*, see [10]), who proposed methods for transforming axioms into inference rules, without affecting the admissibility of the structural rules; and Viganò [11] who presented a detailed study of extensions of modal systems with Horn relational theories – a sub-class of geometric theories. Finally in [12], Miller and al. showed how *polarities* and *focusing* entails a general approach that encompasses most of the works cited above, providing a final solution for the axioms-as-rules problem for the \mathcal{N}_2 class of axioms over classical and intuitionistic logic.

Putting the two traditions together. In the philosophy of mathematics, a recent approach, which attempts to solve the captivating challenge of characterizing the explanatory power of certain proofs, involves as key ingredient the formalization of the *generalizations*, which typically occur in explanations, into inference rules. This aspect aligns naturally with the proof-theoretical tradition of transforming axioms into rules while preserving important logical properties of the theories so obtained. Motivated by [12],¹ this paper introduces a systematic way of incorporating a class of non-logical generalizations, which correspond to what are called *axioms* in the proof-theoretical tradition, into proof systems, producing classical theories \mathcal{T}_A . We will show that these theories \mathcal{T}_A preserve critical logical features such as soundness, completeness, and cut-admissibility. Moreover, we demonstrate that \mathcal{T}_A -theories allow introduction rules that

¹Unfortunately, the work based on focusing and polarities is not scalable to the general case of transforming generalization into rules. Nevertheless, it served as a source of inspiration for the present work.

operate deeply within formulas, effectively modeling mathematical explanations in a way that reveals their explanatory power.

It is important to underline that the value of these results articulate on three different levels:

Technical Advancement. We introduce a novel method for handling a particular class of axioms in classical logic. As far as we know, this issue has not been addressed in the literature before.

Theoretical Extension. Our results expand the logical framework developed in [6] at the mathematical level. We indeed show the admissibility of mathematical introduction rules, which work deeply within formulas, and naturally possess explanatory force.

Conceptual Integration. By uniting two previously separate traditions, this work provides a systematic formal framework for mathematical explanations, offering a unified structure for understanding their features.

The paper is organized as follows. *Section 2* will serve to introduce the topic of mathematical explanations, the model provided in [5], together with two examples of mathematical explanations that will follow us through the paper and will serve to better illustrate the topic. In *Section 3*, we will focus on the axioms employed in the mathematical explanations presented in Section 2, outlining their logical structure, and showing a method for transforming these axioms into rules while preserving critical properties such as soundness, completeness, and cut-admissibility. *Section 4* will be used to introduce the idea of explanatory rules as developed in [6], and to show how to get mathematical explanatory rules. In *Section 5* we will use the elements so far introduced to properly formalize the two examples of mathematical explanations introduced in Section 2 and show that they are indeed explanatory according to our method, *Section 6* will serve to draw some general conclusions and depict some directions of future work.

2 Mathematical explanations

Mathematical explanations are proofs that seem to explain the theorem they prove. Typically, the distinction between proofs that explain and proofs that merely show the theorem to be true comes from the mathematical practice itself where working mathematicians struggle to find new proofs for theorems that have already been demonstrated; often they claim to do so to satisfy the exigence of showing *why* the theorem is true. Hence in order to get a better grip on the topic, here we propose two different theorems and related proofs, which have been judged explanatory by mathematicians, and are by now standardly accepted as such in the contemporary literature. The first proof is a case study coming from Bernard Bolzano, a great if underestimated philosopher and mathematician, who can be seen as one of the pioneers in the study of mathematical explanations.² Amongst the numerous examples that can be

²*E.g.*, see [13].

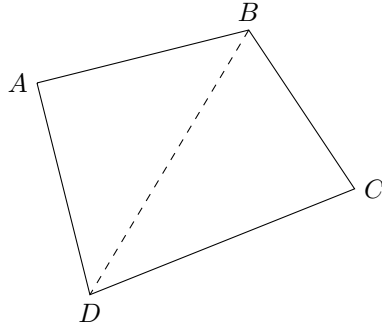


Fig. 1: A quadrangle divided into two triangles

found in his monumental *Theory of Science* (see [14]), one which is at the same time simple and illustrative for our purposes, concerns a property of quadrangles, and for this reason we call it the *Quadrangle Theorem*.

Quadrangle Theorem. The sum of the angles of any quadrangle is equal to 360° .

Explanatory proof. Consider the quadrangle $ABCD$. By the theorem which says that any quadrangle can be divided into two triangles that have a common side, we can consider $ABCD$ as the sum of the two triangles ABD and BCD .³ But the sum of the angles of ABD is 180° , as it is the case for the sum of the angles of the triangle BCD . Hence the sum of the angles of the quadrangle $ABCD$ is $180^\circ + 180^\circ$, namely 360° .

According to Bolzano's analysis, this proof is explanatory, in that it shows that the reason why the sum of the internal angles of quadrangles is 360° is the fact that the sum of the internal angles of triangles is 180° . In other terms, a property of triangles stands as the ground or the reason why another property, concerning quadrangles, holds; hence the proof which relies on a property of triangles to prove a property of quadrangles can be considered explanatory.

Our second example concerns the famous Pythagoras' Theorem. In the literature on mathematical explanation, there is a proof that is based on the similarity of triangles that is often considered as a legitimate explanation of Pythagoras' Theorem.

Pythagoras' Theorem. For any right triangle ABC with the right angle on A , the square BC^2 on the hypotenuse BC equals the sum of the squares AB^2 and AC^2 on the other sides AB and AC of the triangle.

Explanatory proof. Consider any right triangle ABC with the right angle on A and let AH be its height with respect to the hypotenuse BC , as shown in Figure 2. First, we notice that the triangle AHC contained in ABC is similar to ABC itself since they both have a right angle and they both share the angle on C .⁴ The similarity between

³See Figure 1.

⁴Since by fixing two angles of a triangle, we also fix the third, the equality of the two right angles of the two triangles and of their angle on C implies that the angle on B of ABC and the angle on A of AHC are equal, which, by the definition of similarity, implies that the two triangles are similar.

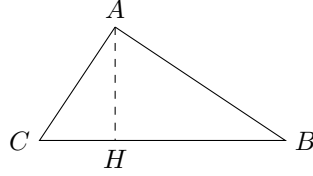


Fig. 2: A right triangle with height AH

the triangles ABC and AHC implies the following equality of ratios:

$$\frac{AC}{BC} = \frac{CH}{AC}$$

which can in its turn be expressed in the following way: $AC^2 = CH \cdot BC$.

Symmetrically, the triangle ABH contained in ABC is similar to ABC itself since they both have a right angle and they both share the angle on B . The similarity between the triangles ABC and ABH implies the equality of ratios

$$\frac{AB}{BC} = \frac{BH}{AB}$$

which can in its turn be expressed in the following way: $AB^2 = BH \cdot BC$. Since we have established that $AC^2 = CH \cdot BC$ and $AB^2 = BH \cdot BC$, we can put them together obtaining $AC^2 + AB^2 = CH \cdot BC + BH \cdot BC$, which is equivalent to $AC^2 + AB^2 = (CH + BH)(BC)$. This, in turn, precisely gives us the conclusion $AC^2 + AB^2 = BC^2$, since by construction $CH + BH = BC$.

The first mathematician who argued that this proof is explanatory is Bouligand [15]. Accordingly, whilst other proofs of Pythagoras' theorem might convince us of the truth of the theorem but are not explanatory, the one we presented above explains why the theorem is true. Both Bouligand and Steiner [16] claim that the explanatory character of the proof is due to its more general nature: indeed, the proof relies on a general property of similar figures to prove that the theorem is true. In other words, the explanatory proof explains a property of right-angled triangles by means of another property, that of similar figures: actually the latter stands as the reason why the former is true.

We have thus presented two proofs that mathematicians and philosophers standardly consider as explanatory. The issue is to identify those elements of such proofs which are the sign of their explanatory power. Following a recent account proposed in [4], in order to check whether a proof is explanatory one needs to check whether there exists a formalization - in terms of tree-structure - of the proof which displays an increase of complexity from the undischarged assumptions to the conclusion. Although original and so far successful, the account proposed in [4] encounters the following limitations: (i) the formalization only concerns the tree-structure of the proof, but does not involve any introduction of a formal language and formal theory where the proof is developed; (ii) the primary tool employed is a natural deduction calculus,

which makes it more challenging to prove general results, such as cut-admissibility; (iii) finally, the method for checking the increase of conceptual complexity, although revealing, is tedious and long. Here we improve on each of these aspects. Hence, given a mathematical explanatory proof, we will first identify the non-logical generalizations on which the proof relies and formalizes them in a formal language. We will show that there exists a method to transform these non-logical generalizations into rules so that if these rules are added to the classical sequent calculus, one obtains a mathematical theory which is sound and complete with respect to the axiomatic correspondent, and has cut-admissibility. Moreover, in line with more recent research developed in [6] where it has been shown that in the classical sequent calculus one can isolate a class of admissible rules that can be considered as explanatory and whose main characteristics consists in being introduction rules and working deep inside formulas, we will show that in these new mathematical theories, analogous mathematical rules can be shown to be admissible. We will use these new mathematical rules to formalize our examples of explanatory proofs, thus being able to precisely isolate each explanatory step of the proofs.

3 On axioms and rules' admissibility

We will work with first-order language which is standardly defined in the following way.

Definition 1. *The formulas of first-order classical logic are built from atomic predicates along with \neg (negation), \wedge and \vee (conjunction and disjunction), and \forall , \exists (universal and existential quantification). The connectives \rightarrow (implication), \top and \perp (true and false) are defined as usual. Finally, we will use the symbols \circ and \odot as follows: $\circ \in \{\wedge, \vee\}$ and $\odot \in \{\forall, \exists\}$.*

We will consider sequent systems based on sequents of the form $\Gamma \Rightarrow \Delta$, with Γ, Δ sets of formulas. We will work with the classical sequent calculus for first order logic G3c, introduced in Figure 3.

Definition 2. *Let $H(\bar{x})$ be a formula and $G(\bar{x})$ be an atomic predicate. A conditional axiom is a formula of the form*

$$\forall \bar{x}. H(\bar{x}) \rightarrow (F(\bar{x}) \leftrightarrow \mathbf{G}(\bar{x}))$$

where the free variables in F are in \bar{x} . H is called the condition and G , highlighted in bold, is the atomic definition of the axiom.

We will generally omit variables and the outer quantifiers conditional axioms to simplify the notation. Observe that, although $H \rightarrow (F \leftrightarrow G)$ is logically equivalent to $H \rightarrow (G \leftrightarrow F)$, in this work it is important to highlight the atomic predicate that will be used in non-logical inference rules. For this reason, we will consider $H \rightarrow (F \leftrightarrow \mathbf{G})$ and $H \rightarrow (G \leftrightarrow \mathbf{F})$ as *different* conditional axioms. Note also that every axiom of the

form $F \leftrightarrow \mathbf{G}$ is equivalent to a conditional axiom of the form

$$\top \rightarrow (F \leftrightarrow \mathbf{G})$$

where the free variables of F form a subset of the free variables of G . Hence conditional axioms generalize the concept of *definition clauses* [17–19]. Finally, the *condition* can also be viewed as a *guard*, functioning as a scheduler that controls the activation of a definition, similar to *guarded Horn clauses* [20].

Similarly to the work on definition clauses, not all conditional axioms are admissible. Restrictions may need to be imposed to ensure properties such as consistency, termination, or well-definedness of the system. These limitations reflect the need for careful design and validation of conditional axioms to maintain the correctness of the underlying framework.

Definition 3. Let \mathcal{T}_A be a set of conditional axioms of the form $\forall \bar{x}. H(\bar{x}) \rightarrow (F(\bar{x}) \leftrightarrow \mathbf{G}(\bar{x}))$.

1. \mathcal{T}_A is a conditional mathematical theory if
 - a. if $H \rightarrow (F \leftrightarrow \mathbf{G}) \in \mathcal{T}_A$, then $H' \rightarrow (T \leftrightarrow \mathbf{G}) \notin \mathcal{T}_A$ for any $T \neq F$;
 - b. if $H \rightarrow (F \leftrightarrow \mathbf{G}) \in \mathcal{T}_A$, then $G \rightarrow (P \leftrightarrow \mathbf{T}) \notin \mathcal{T}_A$;
2. Consider the following rules

$$\frac{\Gamma, H(\bar{t}), G(\bar{t}), F(\bar{t}) \Rightarrow \Delta}{\Gamma, H(\bar{t}), G(\bar{t}) \Rightarrow \Delta} \text{R1} \quad \frac{\Gamma, H(\bar{t}) \Rightarrow \Delta, G(\bar{t}), F(\bar{t})}{\Gamma, H(\bar{t}) \Rightarrow \Delta, G(\bar{t})} \text{R2}$$

where \bar{t} is a list of terms of the same arity as \bar{x} . We define $G3c(\mathcal{T}_A)$ to be the extension of $G3c$ (Figure 3) with the non-logical inference rules $R1, R2$ regarding each conditional mathematical theory in \mathcal{T}_A .⁵

The rules $R1, R2$ are the natural extension of the rules for definition clauses for intuitionistic logic presented in [17–19] both in the presence of conditionals and in the classical setting. Observe the obvious symmetry between $R1$ and $R2$, which is analogous to the symmetry between the right-introduction and left-introduction rules for logical operators. Hence, whilst the rule $R1$ can be seen as the rule which eliminates the predicate G , the rule $R2$ introduces it and its introduction increases the complexity of the proof that contains its application. The rules $R1$ and $R2$ determine, *proof-theoretically*, the meaning of the predicate G under the condition H .

Regarding the restrictions 1, note that 1a implies that, in mathematical theories, atomic predicates are uniquely determined, excluding *non-deterministic definitions*; and 1b forbids that defined atoms can act as conditionals since this introduces *non-deterministic conditionals*. Both forms of non-determinism break cut-admissibility, as discussed in Remark 2.

It is well known that, in order to obtain cut-admissibility of sequent systems with definitions, it is necessary to have *stratified theories* (see Remark 1). To that end we follow [17, 19, 21] and assume that each predicate symbol P in the language is

⁵Here we decided to explicitly copy the formula G from the conclusion to premises for didactic purposes only. Remember that all rules have implicit contraction.

INTRODUCTION RULES

$$\begin{array}{c}
\frac{A, B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta} \wedge L \qquad \frac{\Gamma \Rightarrow \Delta, A \quad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B} \wedge R \\
\\
\frac{A, \Gamma \Rightarrow \Delta \quad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta} \vee L \qquad \frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B} \vee R \\
\\
\frac{\Gamma \Rightarrow \Delta, A}{\neg A, \Gamma \Rightarrow \Delta} \neg L \qquad \frac{A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A} \neg R \\
\\
\frac{[t/x]B, \Gamma \Rightarrow \Delta}{\forall x.B, \Gamma \Rightarrow \Delta} \forall L \qquad \frac{\Gamma \Rightarrow \Delta, [y/x]B}{\Gamma \Rightarrow \Delta, \forall x.B} \forall R \\
\\
\frac{[y/x]B, \Gamma \Rightarrow \Delta}{\exists x.B, \Gamma \Rightarrow \Delta} \exists L \qquad \frac{\Gamma \Rightarrow \Delta, [t/x]B}{\Gamma \Rightarrow \Delta, \exists x.B} \exists R
\end{array}$$

IDENTITY RULE

$$\frac{}{A, \Gamma \Rightarrow \Delta, A} \text{init}$$

Fig. 3: The classical sequent calculus $G3c$. Here, A and B are arbitrary formulas. Contexts are sets, hence implicit contraction is assumed in all the rules. In the $\forall R, \exists L$ the eigenvariable y does not occur free in any formula of the conclusion. The rules for implication were made explicit for clarity.

associated with it a natural number $\text{lvl}(P)$, the level of the predicate. We then extend the notion of level to formulas and derivations as follows.

Definition 4. Given a formula F , its level $\text{lvl}(F)$ is defined as follows:

1. $\text{lvl}(P(\bar{t})) = \text{lvl}(P)$
2. $\text{lvl}(\neg F) = \text{lvl}(F) + 1$
3. $\text{lvl}(F \wedge G) = \text{lvl}(F \vee G) = \max\{\text{lvl}(F), \text{lvl}(G)\}$
4. $\text{lvl}(\forall x.F) = \text{lvl}(\exists x.F) = \text{lvl}(F)$.

A conditional mathematical theory \mathcal{T}_A is stratified if there exists an assignment of level definitions to atomic predicates such that, for all $H \rightarrow (F \leftrightarrow G) \in \mathcal{T}_A$, $\text{lvl}(F) < \text{lvl}(G)$.

We note that this definition slightly differs from those in [17–19]. Specifically, in the intuitionistic setting, the level ordering does not need to be strict: for defined atoms, it is sufficient to have $\text{lvl}(F) \leq \text{lvl}(G)$. As will be clarified in the proof of Theorem 3, this does not hold in the classical setting.

Example 1. Let G_1, G_2 be different 4-ary predicates and H, F be 3-ary predicates. Consider \mathcal{T}_A be composed of the following axioms

$$\begin{array}{l}
\forall xyzw. H(xyw) \wedge H(yzw) \rightarrow F(xyw) \wedge F(yzw) \leftrightarrow \mathbf{G}_1(xyzw) \\
\forall xyzw. H(xyw) \wedge H(yzw) \leftrightarrow \mathbf{G}_2(xyzw)
\end{array}$$

Then \mathcal{T}_A is stratified since we can define, e.g.,

$$\text{lvl}(H) = \text{lvl}(F) = 0 \text{ and } \text{lvl}(G_1) = \text{lvl}(G_2) = 1$$

All theories considered in this text will be determined by the terms in Definitions 2, 3 and 4. In particular, \mathcal{T}_A will always refer to a conditional mathematical stratified theory.

In the following, we will present standard definitions and structural lemmas adapted to $G3c\langle\mathcal{T}_A\rangle$, starting with the definitions of derivation and height of derivations, and admissibility and invertibility of rules.

Definition 5. A derivation in $G3c\langle\mathcal{T}_A\rangle$ is a finite rooted tree with nodes labeled by sequents, axioms at the top nodes, and where each node is connected with the (immediate) successor nodes (if any) according to its inference rules. A sequent S is derivable in $G3c\langle\mathcal{T}_A\rangle$ iff there is a derivation of S in $G3c\langle\mathcal{T}_A\rangle$.

Definition 6. The height of a derivation π , denoted by $\text{ht}(\pi)$, is the greatest number of successive applications of rules in it, where an axiom have height 0.

Definition 7. An inference rule

$$\frac{S_1 \quad \cdots \quad S_k}{S}$$

is called:

- i. admissible in $G3c\langle\mathcal{T}_A\rangle$ if S is derivable in $G3c\langle\mathcal{T}_A\rangle$ whenever S_1, \dots, S_k are derivable in $G3c\langle\mathcal{T}_A\rangle$.
- ii. invertible in $G3c\langle\mathcal{T}_A\rangle$ if the rules

$$\frac{S}{S_1}, \dots, \frac{S}{S_k}$$

are admissible in $G3c\langle\mathcal{T}_A\rangle$.

The following lemmas have been mechanized in the L-framework tool [22]. The L-Framework uses rewrite-based reasoning for proving crucial properties of sequent systems such as admissibility of structural rules, invertibility of rules and cut-elimination.

Lemma 1. The weakening rules

$$\frac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta} \text{WL} \quad \frac{\Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, A} \text{WR}$$

are height-preserving admissible in $G3c\langle\mathcal{T}_A\rangle$.

It is worth noting that the invertibility of classical rules is generally lost when conditional rules are introduced. However, as shown in the next lemma, by treating contexts as sets of formulas invertibility is restored.

Lemma 2. *All the logical rules in $G3c\langle\mathcal{T}_A\rangle$ are height-preserving invertible.*

Proof. The only interesting cases are the ones involving the rules $R1, R2$ and their interaction with the logical rules. We will illustrate the invertibility of the logical rules for the cases $R1, \wedge L, \exists L$. The other are similar or simpler. In the following, assume that $H(\bar{t}), G(\bar{t}) \in \Gamma$.

- Case $R1$. Let π be a proof of $\Gamma \Rightarrow \Delta$. We want to show that $\Gamma, F(\bar{t}) \Rightarrow \Delta$ is also provable with the same height. But this is trivial from Lemma 1. The same for rule $R2$ (considering $H(\bar{t}) \in \Gamma, G(\bar{t}) \in \Delta$).
- Case $\wedge L$. Assume that $H = H_1 \wedge H_2 \in \Gamma$ and consider the derivation

$$\frac{\Gamma, F(\bar{t}) \Rightarrow \Delta}{\Gamma \Rightarrow \Delta} \pi \quad R1$$

By IH, there is a proof π' of $\Gamma, H_1, H_2, F(\bar{t}) \Rightarrow \Delta$ with $\text{ht}(\pi) = \text{ht}(\pi')$. Thus

$$\frac{\Gamma, H_1, H_2, F(\bar{t}) \Rightarrow \Delta}{\Gamma, H_1, H_2 \Rightarrow \Delta} \pi' \quad R1$$

Recall that $H \in \Gamma$, which allows the application of $R1$. When working with multisets, the application of logical rules on the conditional introduces some complicated extra steps and conditions, avoided with the use of sets.

- Case $\exists L$. Let $H = \exists x.H'(x) \in \Gamma$ and consider the derivation

$$\frac{\Gamma, F(\bar{t}) \Rightarrow \Delta}{\Gamma \Rightarrow \Delta} \pi \quad R1$$

Let y be a variable not in Γ, Δ . In particular, observe that y cannot be in \bar{t} . By IH, there is a proof π' of $\Gamma, H'[y/x], F(\bar{t}) \Rightarrow \Delta$ with $\text{ht}(\pi) = \text{ht}(\pi')$. Hence

$$\frac{\Gamma, H'[y/x], F(\bar{t}) \Rightarrow \Delta}{\Gamma, H'[y/x] \Rightarrow \Delta} \pi' \quad R1$$

□

Theorem 3. *$G3c\langle\mathcal{T}_A\rangle$ is sound and complete w.r.t. the first order logic augmented with the conditional axioms in \mathcal{T}_A . Moreover, the following cut rule*

$$\frac{\Gamma \Rightarrow \Delta, A \quad A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta} \text{cut}$$

is admissible in $G3c(\mathcal{T}_A)$.

Proof. We present below the proofs of soundness (Sound.), completeness (Comp.) and cut admissibility (Cut.).

Sound. First of all, let π be a proof of the sequent $\Gamma \Rightarrow \Delta$ in $G3c(\mathcal{T}_A)$. We will show that $\bigwedge \Gamma \rightarrow \bigvee \Delta$ is valid in $\text{FOL} + \mathcal{T}_A$.

If the last rule applied in π is in $G3c$, then the result follows from soundness of $G3c$ w.r.t. FOL. The case where π has one of the the shapes

$$\frac{\Gamma, H, G, F \Rightarrow \Delta}{\Gamma, H, G \Rightarrow \Delta} R1 \quad \frac{\Gamma, H \Rightarrow \Delta, G, F}{\Gamma, H \Rightarrow \Delta, G} R2$$

follows from the following facts:

- the validity of $\bigwedge \Gamma \wedge H$ implies the validity of H ;
- the validity of H implies that $F \leftrightarrow G$ holds;
- if $F \leftrightarrow G$ holds then
 - * $\Gamma \wedge H \wedge F$ is valid iff $\Gamma \wedge H \wedge G$ is valid;
 - * $\bigvee \Delta \vee F$ is valid iff $\bigvee \Delta \vee G$ is valid;

Comp. Every theorem of the logic $\text{FOL} + \mathcal{T}_A$ is provable in $G3c(\mathcal{T}_A)$. In fact, all axioms of first-order logic are provable in $G3c(\mathcal{T}_A)$. The axioms of the form $H \rightarrow (F \leftrightarrow G)$ are provable in $G3c(\mathcal{T}_A)$, as shown below.

$$\frac{\frac{\overline{H, F \Rightarrow G, F}}{H, F \Rightarrow G} \text{init} \quad \frac{\overline{H, G, F \Rightarrow F}}{H, G \Rightarrow F} \text{init}}{\frac{H \Rightarrow F \rightarrow G}{\Rightarrow H \rightarrow (F \leftrightarrow G)} \rightarrow R} \rightarrow R \quad \frac{\frac{H, G, F \Rightarrow F}{H, G \Rightarrow F} \text{init} \quad \frac{H, G \Rightarrow F}{H \Rightarrow G \rightarrow F} \rightarrow R}{\Rightarrow H \rightarrow (F \leftrightarrow G)} \rightarrow R, \wedge R$$

Finally, *modus ponens* can be simulated by the rule *cut*, which is admissible by the next result.

Cut. We will only illustrate the admissibility of the cut rule for the case where the rules $R1, R2$ are involved, all the others are standard. The proof is by induction on the tuple (n, m, k) where n is the level of the cut formula, m is the number of occurrences of logical connectives in the cut formula A (denoted by $|A|$) and k is the cut-height of the derivation.

In the following, assume that $H \rightarrow (F \leftrightarrow \mathbf{G}) \in \mathcal{T}_A$.

1. Consider the derivation

$$\frac{\frac{\Gamma, H \Rightarrow \Delta, G, F}{\Gamma, H \Rightarrow \Delta, G} R2 \quad \Gamma, H, G \Rightarrow \Delta}{\Gamma, H \Rightarrow \Delta} \text{cut}$$

and let's consider the possible cases for the last rule in π_2 . If G, H are not principal, a rule over Γ, Δ is applied, in which case the cut rule moves up the proof.

In the case that either G or H is principal, there are three possibilities (since G is atomic):

- i. H is principal in π_2 w.r.t. a logical rule;
- ii. H is principal in π_2 w.r.t. a non-logical rule, it is a condition of an axiom in \mathcal{T}_A with atomic definition different from G ;
- iii. H is principal in π_2 w.r.t. a non-logical rule, it is a condition of an axiom in \mathcal{T}_A with atomic definition G . Observe that, by condition 1a in Definition 3, G is uniquely defined, that is, the non-logical rule is determined by $H \rightarrow (F \leftrightarrow \mathbf{G})$;

Note also that 1b, which basically says that conditions cannot be defined atoms, *rules out* the following possibilities:

- iv. H is principal in π_2 w.r.t. a non-logical rule and H is an atomic definition;
- v. G is principal in π_2 and it is the condition of an axiom in \mathcal{T}_A .

Please refer to Remarks 1 and 2 for a discussion about stratification, non-determinism and cut-elimination.

We will describe next the reduction steps related to the cases above.

- i. Suppose that *e.g.*, $H = H_1 \wedge H_2$

$$\frac{\Gamma, H, H_1, \overset{\pi'_2}{H_2}, G \Rightarrow \Delta}{\Gamma, H, G \Rightarrow \Delta} \wedge R$$

This can be reduced to

$$\frac{\frac{\Gamma, H, H_1, \overset{\pi_1^w}{H_2} \Rightarrow \Delta, G, F}{\Gamma, H, H_1, H_2 \Rightarrow \Delta, G} R2 \quad \Gamma, H, H_1, \overset{\pi_2'^w}{H_2}, G \Rightarrow \Delta}{\frac{\Gamma, H, H_1, H_2 \Rightarrow \Delta}{\Gamma, H \Rightarrow \Delta} \wedge} \text{cut}$$

where π^w is the weakened version of π . Since $\text{ht}(\pi_2'^w) = \text{ht}(\pi_2') < \text{ht}(\pi_2)$, the result follows by induction.

- ii. Suppose that $H \rightarrow (T \leftrightarrow \mathbf{S}) \in \mathcal{T}_A$ and $S \in \Delta$. Hence

$$\frac{\Gamma, H, G \Rightarrow \Delta, T}{\Gamma, H, G \Rightarrow \Delta} \overset{\pi'_2}{R2}$$

and the cut permutes up

$$\frac{\frac{\Gamma, H \Rightarrow \overset{\pi_1^w}{\Delta}, G, T, F}{\Gamma, H \Rightarrow \Delta, G, T} R2 \quad \Gamma, H, G \Rightarrow \Delta, T}{\frac{\Gamma, H \Rightarrow \Delta, T}{\Gamma, H \Rightarrow \Delta} R2} \text{cut}$$

iii. π_2 has necessarily the shape

$$\frac{\Gamma, H, G, F \Rightarrow \Delta}{\Gamma, H, G \Rightarrow \Delta} \overset{\pi'_2}{R1}$$

This reduces to

$$\frac{\frac{\Gamma, H \Rightarrow \Delta, F, G}{\Gamma, H \Rightarrow \Delta, F} \overset{\pi_1}{\text{cut}} \frac{\overline{F, \Gamma, H \Rightarrow \Delta, F}}{\overline{G, \Gamma, H \Rightarrow \Delta, F}} \overset{\text{init}}{R1}}{\Gamma, H \Rightarrow \Delta} \text{cut} \frac{\overline{\Gamma, H, F \Rightarrow \Delta, F}}{\overline{\Gamma, H, F \Rightarrow \Delta, G}} \overset{\text{init}}{R2} \frac{G, \Gamma, H, F \Rightarrow \Delta}{\Gamma, H, F \Rightarrow \Delta} \overset{\pi'_2}{\text{cut}} \text{cut}$$

Hence the result follows by the IH since the cuts either strictly reduces n or k . In fact, $\text{vl}(F) < \text{vl}(G)$ and the proofs ending with the init rule have height 1, the original cut has height $\text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$ and

- $\text{ht}(\pi_1) + 1 < \text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$;
- $\text{ht}(\pi'_2) + 1 = \text{ht}(\pi_2) < \text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$.

2. Consider the derivation

$$\frac{\Gamma, G \Rightarrow \Delta, H}{\Gamma, G \Rightarrow \Delta} \overset{\pi_1}{\text{cut}} \frac{H, \Gamma, G, F \Rightarrow \Delta}{H, \Gamma, G \Rightarrow \Delta} \overset{\pi_2}{R1}$$

Since H cannot be neither a condition nor a defined atom in π_1 , the only non-trivial possible case is that H is principal in π_1 with a logical rule.

As an example, suppose that $H = H_1 \vee H_2$ and consider the derivation

$$\frac{\Gamma, G \Rightarrow \Delta, H, H_1, H_2}{\Gamma, G \Rightarrow \Delta, H} \overset{\pi'_1}{\wedge R}$$

This can be reduced to

$$\frac{\frac{\Gamma, G \Rightarrow \Delta, H_1, H_2, H}{\Gamma, G \Rightarrow \Delta, H_1, H_2} \overset{\pi'_1}{\text{cut}} \frac{\overline{H, \Gamma, G \Rightarrow \Delta, H_1, H_2}}{\overline{H_2, \Gamma, G \Rightarrow \Delta, H_1, H}} \overset{\text{cut}}{\text{cut}} \frac{\overline{H, H_2, \Gamma, G, F \Rightarrow \Delta, H_1}}{\overline{H, H_2, \Gamma, G \Rightarrow \Delta, H_1}} \overset{\pi'_2}{R1}}{\Gamma, G \Rightarrow \Delta, H_1} \overset{\text{cut}}{\text{cut}} \overset{\pi_3}{\text{cut}} \text{cut}$$

where π_3 is

$$\frac{\overline{H_1, \Gamma, G \Rightarrow \Delta, H_2, H} \overset{\text{cut}}{\text{cut}} \frac{\overline{H, H_1, \Gamma, G, F \Rightarrow \Delta, H_2}}{\overline{H, H_1, \Gamma, G \Rightarrow \Delta, H_2}} \overset{\pi'_2}{R1}}{H_1, \Gamma, G \Rightarrow \Delta, H_2} \text{cut} \frac{\overline{H_2, H_1, \Gamma, G \Rightarrow \Delta, H} \overset{\text{cut}}{\text{cut}} \frac{\overline{H, H_2, H_1, \Gamma, G, F \Rightarrow \Delta}}{\overline{H, H_2, H_1, \Gamma, G \Rightarrow \Delta}} \overset{\pi'_2}{R1}}{H_2, H_1, \Gamma, G \Rightarrow \Delta} \text{cut}$$

Observe that all cuts either strictly reduces m or k , since $|H_1|, |H_2| < |H|$, the proofs with double bar have height 1, the original cut has height $\text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$ and

- $\text{ht}(\pi'_1) + 1 = \text{ht}(\pi_1) < \text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$;
- $\text{ht}(\pi_2) + 1 < \text{ht}(\pi_1) + \text{ht}(\pi_2) + 1$.

□

We will now briefly examine how non-stratification and non-determinism in conditional axioms can compromise cut-elimination.

Remark 1 (Non-stratification breaks cut-admissibility). *It is well known that cut-elimination does not hold for theories in general [17]. In fact, let G be any atomic predicate and consider \mathcal{T}_A consisting of the following conditional axiom*

$$\top \rightarrow (\neg G \leftrightarrow G)$$

with the corresponding rules

$$\frac{\Gamma, G, \neg G \Rightarrow \Delta}{\Gamma, G \Rightarrow \Delta} R1 \qquad \frac{\Gamma \Rightarrow \Delta, G, \neg G}{\Gamma \Rightarrow \Delta, G} R2$$

The following is a derivation in $G3c\langle\mathcal{T}_A\rangle$ plus the cut rule, where the cut instance cannot be eliminated

$$\frac{\frac{\frac{\overline{G \Rightarrow G}}{\Rightarrow G, \neg G} \text{init} \quad \frac{\overline{G \Rightarrow G}}{\Rightarrow G} \text{init}}{\Rightarrow G} \neg R \quad \frac{\frac{\overline{G \Rightarrow G}}{G, \neg G \Rightarrow \perp} \text{init} \quad \frac{\overline{G \Rightarrow G}}{G \Rightarrow \perp} \text{init}}{G \Rightarrow \perp} \neg L}{\Rightarrow \perp} R1}{\Rightarrow \perp} R2 \quad \text{cut}$$

The key issue here is that no definition of level for G can render \mathcal{T}_A stratified.

Remark 2 (Non-determinism breaks cut-admissibility). *Observe that mathematical theories are very strict w.r.t. the use of conditionals and defined atoms. In particular, atomic predicates cannot be multi-defined and conditions cannot be defined. Allowing this kind of non-determinism can occur in losing cut-admissibility.*

Indeed, suppose that $\{H \rightarrow (F \leftrightarrow \mathbf{G}), H \rightarrow (T \leftrightarrow \mathbf{G})\} \subset \mathcal{T}_A$. Hence the following is a valid derivation with no cut-free correspondent.

$$\frac{\frac{\frac{\overline{H, T \Rightarrow F, G, \overline{T}}}{H, T \Rightarrow F, G} \text{init} \quad \frac{\overline{H, T \Rightarrow F, G}}{H, T \Rightarrow F, G} \text{init}}{H, T \Rightarrow F, G} R2 \quad \frac{\frac{\overline{F, G, H, T \Rightarrow F}}{G, H, T \Rightarrow F} \text{init} \quad \frac{\overline{F, G, H, T \Rightarrow F}}{G, H, T \Rightarrow F} \text{init}}{G, H, T \Rightarrow F} R1}{H, T \Rightarrow F} \text{cut}$$

This justifies, proof-theoretically, the condition 1a in Definition 3.

Regarding condition 1b, suppose that $\{H \rightarrow (F \leftrightarrow \mathbf{G}), G \rightarrow (F \leftrightarrow \mathbf{T}), H' \rightarrow (F \leftrightarrow \mathbf{H})\} \subset \mathcal{T}_A$ and consider the derivation

$$\frac{\frac{\frac{F, H, H' \Rightarrow T, G, F}{F, H, H' \Rightarrow T, G} \text{ init}}{F, H, H' \Rightarrow T, G} \text{ R2} \quad \frac{\frac{F, H, H' \Rightarrow T, F}{G, F, H, H' \Rightarrow T} \text{ init}}{G, F, H, H' \Rightarrow T} \text{ R2}}{\frac{F, H, H' \Rightarrow T}{H, H' \Rightarrow T} \text{ cut}} \text{ R1}$$

There is no cut-free derivation of the end-sequent.

4 Explanatory rules

Our main goal is to formalize mathematical proofs so to investigate their explanatory power. Any mathematical proof involves a logical and a mathematical level. In [6] the explanatory power of a proof at the logical level has been thoroughly investigated: in particular, explanatory rules, that are rules such that not only is the conclusion inferable from the premisses, but also the premisses represent the reasons why the conclusion is true, have been introduced. Any application of these rules at the logical level is the countersign of an explanatory step in the proof.

Amongst the several characteristics of these rules, the one that it is of interest for us here is that these rules may work with the main connective of a formula - as standard inferential rules do - but also with connectives that lie deep inside the formulas themselves.

To capture both possibilities, the notions of *context* and *formula in a context* are introduced. Informally, a context represents the portion of a formula not being directly analyzed and is denoted by $C[\cdot]$. For instance, consider the first-order formula

$$\exists x.(S(x) \wedge T(x)) \vee \forall xy.(P(x) \rightarrow Q(x) \wedge R(y)).$$

If we focus on $Q(x) \wedge R(y)$ while setting aside the rest, this is represented as $C[\cdot]$, where $C[\cdot]$ corresponds to the context $\exists x(S(x) \wedge T(x)) \vee \forall xy(\neg P(x) \vee [\cdot])$. Focusing on $Q(x) \wedge R(y)$ in this context is then written as $C[Q(x) \wedge R(y)]$. As clarified in [6], when working in an explanatory framework, negation needs to be handled with particular care (although this will not be visible in the examples considered in the next sections as in none of them negations occur). This is also true for the notion of contexts, which are thus introduced in the following way.

Definition 8. *The set of contexts is inductively defined in the following way:*

- $[\cdot]$ is a context;
- if $C[\cdot]$ is a context, then $\neg\neg C[\cdot]$, $D \circ C[\cdot]$, $C[\cdot] \circ D$, $\odot x C[\cdot]$ are contexts;
- if $C[\cdot]$ is a context and $C[\cdot] \neq \overbrace{\neg \dots \neg}^{2n} [\cdot]$, where $n \geq 0$, then $\neg C[\cdot]$ is a context.

A hole in a context can be replaced by any formula A : $C[A]$ represents the formula that is obtained from $C[\cdot]$ by replacing the hole by A . Once formulas are considered

in contexts, they will naturally have an *occurrence polarity* which is either positive or negative, defined as standard (see e.g., [23]).

Definition 9. We define the set of contexts with positive \mathcal{P} and negative polarities \mathcal{N} simultaneously by the following inductive definition.

(i) $[\cdot] \in \mathcal{P}$;

if $C[\cdot]^+ \in \mathcal{P}$, $C[\cdot]^- \in \mathcal{N}$, and A is any formula, then:

(ii) $\neg C[\cdot]^-, A \wedge C[\cdot]^+, C[\cdot]^+ \wedge A, A \vee C[\cdot]^+, C[\cdot]^+ \vee A, \forall x C[\cdot]^+, \exists x C[\cdot]^+ \in \mathcal{P}$;

(iii) $\neg C[\cdot]^+, A \wedge C[\cdot]^-, C[\cdot]^- \wedge A, A \vee C[\cdot]^-, C[\cdot]^- \vee A, \forall x C[\cdot]^-, \exists x C[\cdot]^- \in \mathcal{N}$.

We say that a formula A is positive (resp. negative) in a context $C[A]$ if $C[\cdot] \in \mathcal{P}$ (resp. $C[\cdot] \in \mathcal{N}$).

Definition 10. If $C[\cdot]$ is a context, then the scope of a context, $SC(C)$, and the inverse scope $SC^{inv}(C)$, are defined inductively in the following way:

- if $C[\cdot] = [\cdot]$, then $SC(C) = SC^{inv}(C) = \emptyset$;
- if $C[\cdot] = \neg\neg D[\cdot]$, then $SC(C) = SC(D)$ and $SC^{inv}(C) = SC^{inv}(D)$;
- if $C[\cdot] = D' \circ D[\cdot]$ or $D[\cdot] \circ D'$, then $SC(C) = SC(D)$ and $SC^{inv}(C) = SC^{inv}(D)$;
- if $C[\cdot] = \forall x D[\cdot]$, then $SC(C) = \forall x.(SC(D))$ and $SC^{inv}(C) = \exists x.(SC^{inv}(D))$;
- if $C[\cdot] = \exists x D[\cdot]$, then $SC(C) = \exists x.(SC(D))$ and $SC^{inv}(C) = \forall x.(SC^{inv}(D))$;
- if $C[\cdot] = \neg D[\cdot]$, then $SC(C) = SC^{inv}(D)$ and $SC^{inv}(C) = SC(D)$.

Note that scopes of contexts are defined in such a way that an existential that occurs in a context with a negative polarity is transformed into an universal in a context with a positive polarity, whilst an universal in a context with a negative polarity is transformed into an existential in a context with a positive polarity. For instance, consider a context $C[\cdot] = \forall x \neg \exists y \neg [\cdot]$, then $SC(C) = \forall x \forall y$. On the other hand, consider a context $C[\cdot] = \exists x \neg \forall y \neg [\cdot]$, then $SC(C) = \exists x \exists y$.

Given a formula in a context $C[F]$, we can restrict the scope of context $SC(C)$ to the *scope of context relative to the formula F* , $SC_F(C)$, depending on the variables that the quantifiers in $SC(C)$ bound in F .

Definition 11. Let F be a formula of \mathcal{L} , then the free variables of F , $FV(F)$, are standardly defined as those variables occurring in F which are not bound by any quantifier. We define the restricted free variables of a formula F that is of the form $G \circ G'$ as the set $FV^+(F)$, in the following way:

- if F is of the form $G \circ G'$, then $FV^+(F) = FV(G) \cap FV(G')$.

Definition 12. Let $C[F]$ be a formula in a context, and let $FV^+(F) = x_1, \dots, x_n$. The scope of a context $C[\cdot]$ relative to the formula F , $SC_F(C)$, is the result of removing from $SC(C)$ any quantifier that is not of the form $\odot x_1, \dots, \odot x_n$.

Amongst the explanatory propositional rules introduced in [6], the one that will be useful in the next Section is the following⁶

$$\frac{\Gamma \Rightarrow \Delta, C[F] \quad \Gamma \Rightarrow \Delta, C[G]}{\Gamma \Rightarrow \Delta, C[F \circ G]} \circ_1$$

In order to be applied the rule obeys the following restrictions. If $\circ = \wedge$ and $SC_{F \circ G}(C) \neq \emptyset$, then the rule \circ_1 can be applied if, only if, $C[\cdot] \in \mathcal{P}$ and $SC_F(C) = \forall x_1, \dots, \forall x_n$ or $C[\cdot] \in \mathcal{N}$ and $SC_F(C) = \exists x_1, \dots, \exists x_n$. If $\circ = \wedge$ and $SC_{F \circ G}(C) = \emptyset$, then the rule \circ_1 can always be applied. If $\circ = \vee$ and $SC_{F \circ G}(C) \neq \emptyset$, then the rule \circ_1 can be applied if, only if, $C[\cdot] \in \mathcal{N}$ and $SC_F(C) = \forall x_1, \dots, \forall x_n$, or $C[\cdot] \in \mathcal{P}$ and $SC_F(C) = \exists x_1, \dots, \exists x_n$. If $\circ = \vee$ and $SC_{F \circ G}(C) = \emptyset$, then the rule \circ_1 can always be applied. The double line indicates that the rule \circ_1 is not simply an inferential rule, but it can also be seen as explaining its conclusion. In other words, not only in \circ_1 the conclusion follows from the premisses, but also the premisses can be seen as the reasons why the conclusion is true.⁷

In [6] it is proved that the rule \circ_1 is admissible in $G3c$. Here we will obtain an analogous result for mathematical rules. Indeed we will first introduce rules that stand as the explanatory counterpart of the rules $R1$ and $R2$ in that they only operate on the right side of the sequent - they are indeed introduction rules - and they work deep inside formulas. Secondly we will show that they are admissible in all theories $G3c(\mathcal{T}_A)$ which contain the rules $R1$ and $R2$. Thirdly we will use them to formalize mathematical explanations: each of their application in the formalization of a mathematical proof correspond to an explanatory step.

Theorem 4. *Let $C[\cdot]$ be a context, G a predicate, M, N sets of formulas and F, H be first-order formulas such that $H \rightarrow (F \leftrightarrow G) \in \mathcal{T}_A$. Then the following rules, where $C[\cdot] \in \mathcal{P}$, are admissible in $G3c(\mathcal{T}_A)$:*

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge F \rightarrow \vee N]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge G \rightarrow \vee N]} R1c \quad \frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \rightarrow \vee N \vee F]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \rightarrow \vee N \vee G]} R2c$$

where $\wedge M$ ($\vee N$) stands for the conjunction (disjunction) of formulas in M (N). When $H = \top$, then the following rule is admissible

$$\frac{\Gamma \Rightarrow \Delta, C[F]}{\Gamma \Rightarrow \Delta, C[G]} Rc$$

for $C[\cdot] \in \mathcal{P}$ or $C[\cdot] \in \mathcal{N}$.

Proof. We examine the case of the rule $R2c$; the rule $R1c$ and Rc can be treated analogously. We consider the premise of the rule, namely the sequent $\Gamma \Rightarrow \Delta, C[\wedge M \wedge$

⁶We do not present the other rules as they would need other technical details to be properly introduced.

⁷In [6], conditions for some formulas to count as the reasons why another formula is true are set. Moreover, it is proved that the premisses of the rule \circ_1 meets these conditions, see Theorems 5.7 and 5.14 in [6]. Analogous conditions apply in the mathematical case and are satisfied by the rules $R1c$, $R2c$ and Rc . However, for the sake of clarity and brevity, we will not dwell on the proof.

$H \rightarrow \bigvee N \vee F]$. We apply on the context $C[\cdot]$ as many rules \overline{R} as necessary to unfold the context itself and reach the formula $\bigwedge M \wedge H \rightarrow \bigvee N \vee F$. Once arrived to $\bigwedge M \wedge H \rightarrow \bigvee N \vee F$, we apply the inverse of the appropriate logical rules to get H on the left side of the sequent, and F on the right side of the sequent, both separated from other formulas. We then apply the rule $R2$ to F , obtaining the predicate G . We then apply the standard logic rules to get the desired conclusion. \square

5 Formalizing mathematical proofs and displaying their explanatory power

We now have all the elements to formalize the proofs of the theorems introduced in Section 2 and show whether they are explanatory according to our perspective. We start by analyzing the proof of the Quadrangle's theorem and then move to that of Pythagoras' theorem.

Example 2 (Formalization of the proof of the Quadrangles' theorem). *This is an instance of Example 1. First of all we isolate the conditional axioms that occur in the proof. Adopting the following notation*

- $Tr(xyz)$ for xyz is a triangle,
- $Tr(xyzw)$ for $xyzw$ is a quadrangle,
- $St180(xyz)$ for the sum of the internal angles of the triangle xyz is 180° ,
- $Sq360(xyzw)$ for the sum of the internal angles of the quadrangle $xyzw$ is 360° .

the conditional axioms of the theory \mathcal{T}_A are

- (i) $\forall xyzw.Tr(xyz) \wedge Tr(yzw) \rightarrow St180(xyz) \wedge St180(yzw) \leftrightarrow \mathbf{Sq360}(xyzw)$,
- (ii) $\forall xyzw.Tr(xyz) \wedge Tr(yzw) \leftrightarrow \mathbf{Qu}(xyzw)$.

\mathcal{T}_A is stratified since we can define

$$\text{vl}(Tr) = \text{vl}(St180) = 0 \text{ and } \text{vl}(Qu) = \text{vl}(Sq360) = 1$$

According to Definition 3, we can form the following two pairs of rules:

$$\frac{Tr(xyz) \wedge Tr(yzw), St180(xyz) \wedge St180(yzw), \Gamma \Rightarrow \Delta}{Tr(xyz) \wedge Tr(yzw), Sq360(xyzw), \Gamma \Rightarrow \Delta} R^i L$$

$$\frac{Tr(xyz) \wedge Tr(yzw), \Gamma \Rightarrow \Delta, St180(xyz) \wedge St180(yzw)}{Tr(xyz) \wedge Tr(yzw), \Gamma \Rightarrow \Delta, Sq360(xyzw)} R^i R$$

which correspond to the conditional axiom (i), and

$$\frac{Tr(xyz) \wedge Tr(yzw), \Gamma \Rightarrow \Delta}{Qu(xyzw), \Gamma \Rightarrow \Delta} R^{ii} L \quad \frac{\Gamma \Rightarrow \Delta, Tr(xyz) \wedge Tr(yzw)}{\Gamma \Rightarrow \Delta, Qu(xyzw)} R^{ii} R$$

which correspond to the conditional axiom (ii).

By adding these rules to the classical sequent calculus we obtain the calculus $G3c(\mathcal{T}_A)$ which, by Theorem 3, is sound and cut-free complete to the theory \mathcal{T}_A . Moreover, according to Lemma 4, the following explanatory rules:

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge Tr(xyz) \wedge Tr(yzw) \wedge St180(xyz) \wedge St180(yzw) \rightarrow \vee N]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge Tr(xyz) \wedge Tr(yzw) \wedge Sq360(xyzw) \rightarrow \vee N]} R^{i1c}$$

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge Tr(xyz) \wedge Tr(yzw) \rightarrow \vee N \vee St180(xyz) \wedge St180(yzw)]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge Tr(xyz) \wedge Tr(yzw) \rightarrow \vee N \vee Sq360(xyzw)]} R^{i2c}$$

where $C[\cdot] \in \mathcal{P}$, are admissible. The same with the explanatory rule

$$\frac{\Gamma \Rightarrow \Delta, C[Tr(xyz) \wedge Tr(yzw)]}{\Gamma \Rightarrow \Delta, C[Qu(xyzw)]} R^{ii_c}$$

where $C[\cdot] \in \mathcal{P}$ or $C[\cdot] \in \mathcal{N}$. We will use these rules plus the explanatory logical rule \circ_1 to formalize the proof of the Quadrangles' theorem in the following way.

$$\frac{\frac{\frac{\forall xyzw(Tr(xyz) \wedge Tr(yzw) \rightarrow St180(xyz)) \quad \forall x, y, z, w(Tr(xyz) \wedge Tr(yzw) \rightarrow St180(yzw))}{\forall xyzw(Tr(xyz) \wedge Tr(yzw) \rightarrow St180(xyz) \wedge St180(yzw))} \circ_1}{\frac{\forall xyzw(Tr(xyz) \wedge Tr(yzw) \rightarrow Sq360(xyzw))}{\forall xyzw(Qu(xyzw) \rightarrow Sq360(xyzw))} R^{ii_c}} R^{i2c}}{\forall xyzw(Qu(xyzw) \rightarrow Sq360(xyzw))} R^{i1c}$$

We have thus found a formalization that is faithful to the informal proof and is composed of explanatory steps: in each step the complexity's increase - as in each step we use an introduction rules that introduce in the proof a new concept. At each step we can see the premisses as providing the reason(s) why their conclusion is true. Thus, within our account, not only can we confirm that the intuitively explanatory proof is indeed explanatory as it contains applications of explanatory rules, but we can also identify any of its single explanatory step.

Example 3 (Formalization of the proof of the Pythagoras' theorem). *First, we isolate the conditional axioms present in the proof. Adopting the following notation:*

- $Tr(xyz)$: xyz is a triangle,
- $R(x)$: x is a right angle,
- $Ob(xyz \sim xyw / xzw)$: the triangles xyw and xzw are obtained from the triangle xyz by drawing the height xw ,
- $Sim(xyz, xyw)$: xyz and xyw are similar triangles,
- $Sm(xywz)$, represented by $xy^2 = (yw)(yz)$: the measure of xy squared corresponds to the multiplication of yw and yz ,
- $Ss(xyz)$, represented by $xy^2 + xz^2 = yz^2$: the measure of yz squared corresponds to the sum of xy squared and xz squared.

the axioms of the theory \mathcal{T}_A are:

- (i) $\forall xyzw(Tr(xyz) \wedge Tr(xyw) \wedge Tr(xzw) \wedge Ob(xyz \sim xyw//xzw) \rightarrow (Sim(xyz, xyw) \wedge Sim(xyz, xzw) \leftrightarrow \mathbf{R}(x))),$
(ii) $\forall xyzw(Tr(xyz) \wedge Tr(xyw) \wedge Tr(xzw) \wedge Ob(xyz \sim xyw//xzw) \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw) \rightarrow (xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz) \leftrightarrow \mathbf{xy}^2 + \mathbf{xz}^2 = \mathbf{yz}^2)).$

\mathcal{T}_A is stratified since we can define

$$\text{lvl}(Tr) = \text{lvl}(Ob) = \text{lvl}(Sym) = \text{lvl}(Sm) = 0 \text{ and } \text{lvl}(R) = \text{lvl}(Ss) = 1$$

According to Definition 3, we can form the following two pairs of rules (where by H we denote $Tr(xyz) \wedge Tr(xyw) \wedge Tr(xzw) \wedge Ob(xyz \sim xyw//xzw)$)

$$\frac{H, Sim(xyz, xyw) \wedge Sim(xyz, xzw), \Gamma \Rightarrow \Delta}{H, R(x), \Gamma \Rightarrow \Delta} R^i L$$

$$\frac{H, \Gamma \Rightarrow \Delta, Sim(xyz, xyw) \wedge Sim(xyz, xzw)}{H, \Gamma \Rightarrow \Delta, R(x)} R^i R$$

which correspond to the conditional axiom (i), and,

$$\frac{H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw), xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz), \Gamma \Rightarrow \Delta}{H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw), xy^2 + xz^2 = yz^2, \Gamma \Rightarrow \Delta} R^i L$$

$$\frac{H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw), \Gamma \Rightarrow \Delta, xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz)}{H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw), \Gamma \Rightarrow \Delta, xy^2 + xz^2 = yz^2} R^i R$$

which correspond to the conditional axiom (ii).

By adding these rules to the classical sequent calculus we obtain the calculus $G3c(\mathcal{T}_A)$ which, by Theorem 3, is sound and cut-free complete to the theory \mathcal{T}_A . Moreover, according to Lemma 4, the following explanatory rules:

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge X \wedge Sim(xyz, xwy) \wedge Sim(xyz, xzw) \rightarrow \vee N]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge X \wedge R(x) \rightarrow \vee N]} R^i 1c$$

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge X \rightarrow \vee N \vee Sim(xyz, xwy) \wedge Sim(xyz, xzw)]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge X \rightarrow \vee N \vee R(x)]} R^i 2c$$

where $C[\cdot] \in \mathcal{P}$, are admissible in $G3c(\mathcal{T}_A)$. The explanatory rules

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw) \wedge xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz) \rightarrow \vee N]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge Sim(xyz, xyw) \wedge Sim(xyz, xzw) \wedge xy^2 + xz^2 = yz^2 \rightarrow \vee N]} R^i 1c$$

$$\frac{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \rightarrow \vee N \vee xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz)]}{\Gamma \Rightarrow \Delta, C[\wedge M \wedge H \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \rightarrow \vee N \vee xy^2 + xz^2 = yz^2]} R^{i2c}$$

where $C[\cdot] \in \mathcal{P}$, are also admissible. We will use these rules plus the explanatory logical rule \circ_1 to formalize the proof of the Pythagoras' theorem in the following way.

$$\frac{\frac{\frac{\forall xywz(\text{Tr}(xyz) \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \wedge Y \rightarrow xy^2 = (yw)(yz)) \quad \forall xywz(\text{Tr}(xyz) \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \wedge Y \rightarrow xy^2 = (yw)(yz))}{\forall xywz(\text{Tr}(xyz) \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \wedge Y \rightarrow xy^2 = (yw)(yz) \wedge xz^2 = (wz)(yz))} \circ_1}{\forall xywz(\text{Tr}(xyz) \wedge \text{Sim}(xyz, xyw) \wedge \text{Sim}(xyz, xzw) \wedge Y \rightarrow xy^2 + xz^2 = yz^2)} R^{i2c}}{\forall xywz(\text{Tr}(xyz) \wedge R(x) \wedge Y \rightarrow xy^2 + xz^2 = yz^2)} R^{i4c}$$

where Y denotes the formula $\text{Tr}(xyw) \wedge \text{Tr}(xwz) \wedge \text{Ob}(xyz \sim xyw // xzw)$.⁸

We have thus found a formalization that is faithful to the informal proof and is composed of explanatory steps: in each step the complexity's increase - as in each step we use an explanatory rule that introduces in the proof a new concept. At each step we can see the premisses as providing the reason(s) why their conclusion is true. Thus, within our account, not only can we confirm that intuitively explanatory proof is indeed explanatory as it involves applications of explanatory rules, but we can also identify any single explanatory step of the proof.

This reasoning behind these examples has been implemented in the Profound-Intuitionistic tool, aka *Profint*, accessible at <https://chaudhuri.info/research/profint/>. Profint is an interactive theorem prover for first-order intuitionistic logic that is designed around expressive direct manipulation based interactions. Although our setting is classical, constructive reasoning was enough for the examples in this work. The basic idea is to use deep-reasoning in order to automatize the derivation steps, where different appearances of the atomic predicates are matched in the formula structure, simulating the deep application of rules.

The code for following part of the proof explanation in Example 2

$$\frac{\forall xyzw(\text{Tr}(xyz) \wedge \text{Tr}(yzw) \rightarrow \text{Sq360}(xyzw))}{\forall xyzw(\text{Qu}(xyzw) \rightarrow \text{Sq360}(xyzw))} R^{iic}$$

is shown below.

```

Typing:
i : \type.
tr : i -> i -> i -> \o.
qu : i -> i -> i -> i -> \o.
sq360 : i -> i -> i -> i -> \o.

```

Program :

⁸Note that the formula $\forall xywz(\text{Tr}(xyz) \wedge R(x) \rightarrow xy^2 + xz^2 = yz^2)$, which represents a proper formalization of Pythagoras' theorem, can be proved to be equivalent to the formula $\forall xywz(\text{Tr}(xyz) \wedge R(x) \wedge Y \rightarrow xy^2 + xz^2 = yz^2)$.

$$(\forall [x, y, z, w] \text{qu } x y z w \rightarrow (\text{tr } x y z \wedge \text{tr } y z w)) \rightarrow$$

$$(\forall [x, y, z, w] (\text{tr } x y z \wedge \text{tr } y z w \rightarrow \text{sq360 } x y z w) \rightarrow$$

$$\text{qu } x y z w \rightarrow \text{sq360 } x y z w)$$

6 Conclusion

This paper shows the viability and utility of combining the philosophy of mathematical explanations with advanced proof-theoretical techniques. By adopting the “axioms-as-rules” paradigm, we established a systematic method to transform generalizations into inference rules, enabling the formalization of explanatory proofs. Our approach preserves essential logical properties and offers a robust framework for analyzing explanatory power in proofs.

Through the formalization of classic examples, we showcased how this method clarifies the structure and conceptual complexity of mathematical explanations. This work not only connects two previously disparate traditions but also paves the way for future research into the formal and conceptual aspects of mathematical reasoning and explanations.

References

- [1] Betti, A.: Explanation in metaphysics and Bolzano’s theory of ground and consequence. *Logique et analyse* **53**(211), 281–316 (2010)
- [2] Detlefsen, M.: Fregean hierarchies and mathematical explanation. *International Studies in the Philosophy of Science* **3**, 97–116 (1988)
- [3] Mancosu, P., Poggiolesi, F., Pincock, C.: Mathematical explanation. In: Zalta, E. (ed.) *Stanford Encyclopedia of Philosophy*, pp. 1–44. Stanford, Stanford, USA (2022)
- [4] Poggiolesi, F.: Mathematical explanations: An analysis via formal proofs and conceptual complexity. *Philosophia Mathematica* **32**(2), 145–176 (2023) <https://doi.org/10.1093/phimat/nkad023> <https://academic.oup.com/phimat/article-pdf/32/2/145/58997913/nkad023.pdf>
- [5] Poggiolesi, F., Genco, F.: Conceptual (and hence mathematical) explanation, conceptual grounding and proof. *Erkenntnis* **88**(4), 1481–1507 (2023) <https://doi.org/10.1007/s10670-021-00412-x>
- [6] Poggiolesi, F.: (Conceptual) explanations in logic. *Journal of Logic and Computation*, 1–35 (2025) <https://doi.org/10.1093/logcom/exae064>
- [7] Gentzen, G.: Investigations into logical deduction. In: Szabo, M.E. (ed.) *The Collected Papers of Gerhard Gentzen*, pp. 68–131. North-Holland, Amsterdam (1935). <https://doi.org/10.1007/BF01201353>

- [8] Simpson, A.K.: The proof theory and semantics of intuitionistic modal logic. PhD thesis, College of Science and Engineering, School of Informatics, University of Edinburgh (1994)
- [9] Ciabattoni, A., Galatos, N., Terui, K.: From axioms to analytic rules in nonclassical logics. In: Proceedings of the Twenty-Third Annual IEEE Symposium on Logic in Computer Science, LICS 2008, 24-27 June 2008, Pittsburgh, PA, USA, pp. 229–240 (2008). <https://doi.org/10.1109/LICS.2008.39>
- [10] Negri, S.: Contraction-free sequent calculi for geometric theories with an application to Barr’s theorem. *Archives of Mathematical Logic* **42**(4), 389–401 (2003) <https://doi.org/10.1007/s001530100124>
- [11] Viganò, L.: *Labelled Non-Classical Logics*. Kluwer Academic Publishers, New York, USA (2000). <https://doi.org/10.1007/978-1-4757-3208-5>
- [12] Marin, S., Miller, D., Pimentel, E., Volpe, M.: From axioms to synthetic inference rules via focusing. *Ann. Pure Appl. Log.* **173**(5), 103091 (2022) <https://doi.org/10.1016/j.apal.2022.103091>
- [13] Kitcher, P.: Bolzano’s ideal of algebraic analysis. *Studies in History and Philosophy of Science Part A* **6**(3), 229–269 (1975) [https://doi.org/10.1016/0039-3681\(75\)90024-2](https://doi.org/10.1016/0039-3681(75)90024-2)
- [14] Bolzano, B.: *Theory of Science vol. 1*. Oxford University Press, Oxford, UK (2014). <https://doi.org/10.1093/actrade/9780199684397.book.1>
- [15] Bouligand, G.: *Premières Leçons sur la Théorie Générale des groupes*. Vuibert, Paris, France (1932)
- [16] Steiner, M.: Mathematical explanation. *Philosophical Studies* **34**(2), 135–151 (1978) <https://doi.org/10.1007/BF00354494> . Accessed 2012-12-27
- [17] Schroeder-Heister, P.: Rules of definitional reflection. In: Proceedings of the Eighth Annual Symposium on Logic in Computer Science (LICS ’93), Montreal, Canada, June 19-23, 1993, pp. 222–232. IEEE Computer Society, ??? (1993). <https://doi.org/10.1109/LICS.1993.287585> . <https://doi.org/10.1109/LICS.1993.287585>
- [18] McDowell, R., Miller, D.: Cut-elimination for a logic with definitions and induction. *Theor. Comput. Sci.* **232**(1-2), 91–119 (2000) [https://doi.org/10.1016/S0304-3975\(99\)00171-1](https://doi.org/10.1016/S0304-3975(99)00171-1)
- [19] Miller, D., Tiu, A.: A proof theory for generic judgments. *ACM Trans. Comput. Log.* **6**(4), 749–783 (2005) <https://doi.org/10.1145/1094622.1094628>
- [20] Ueda, K.: Guarded horn clauses. In: Wada, E. (ed.) *Logic Programming ’85*,

Proceedings of the 4th Conference, Tokyo, Japan, July 1-3, 1985. Lecture Notes in Computer Science, vol. 221, pp. 168–179. Springer, Berlin, Germany (1985). https://doi.org/10.1007/3-540-16479-0_17 . https://doi.org/10.1007/3-540-16479-0_17

- [21] Miller, D., Tiu, A.F.: A proof theory for generic judgments: An extended abstract. In: 18th IEEE Symposium on Logic in Computer Science, 22-25 June 2003, Ottawa, Canada, Proceedings, pp. 118–127. IEEE Computer Society, LICS (2003). <https://doi.org/10.1109/LICS.2003.1210051> . <https://doi.org/10.1109/LICS.2003.1210051>
- [22] Olarte, C., Pimentel, E., Rocha, C.: A rewriting logic approach to specification, proof-search, and meta-proofs in sequent systems. *J. Log. Algebraic Methods Program.* **130**, 100827 (2023) <https://doi.org/10.1016/J.JLAMP.2022.100827>
- [23] Troelstra, A.S., Schwichtenberg, H.: *Basic Proof Theory*, 2nd edn. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, UK (2000)