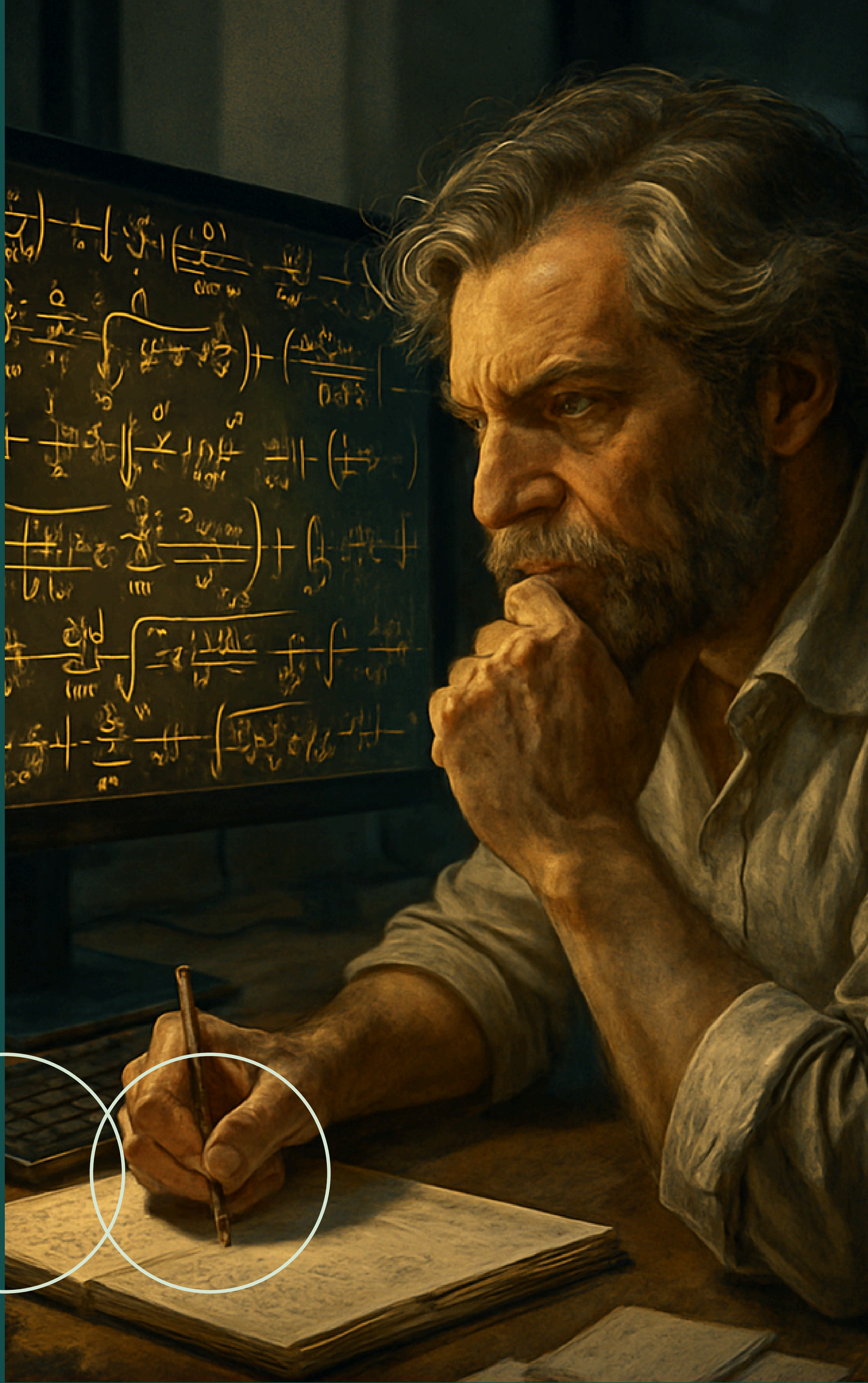
2025 SKKU-NU PHILOSOPHICAL INTERCHANGE

Proof, Computer and Mathematical Agent

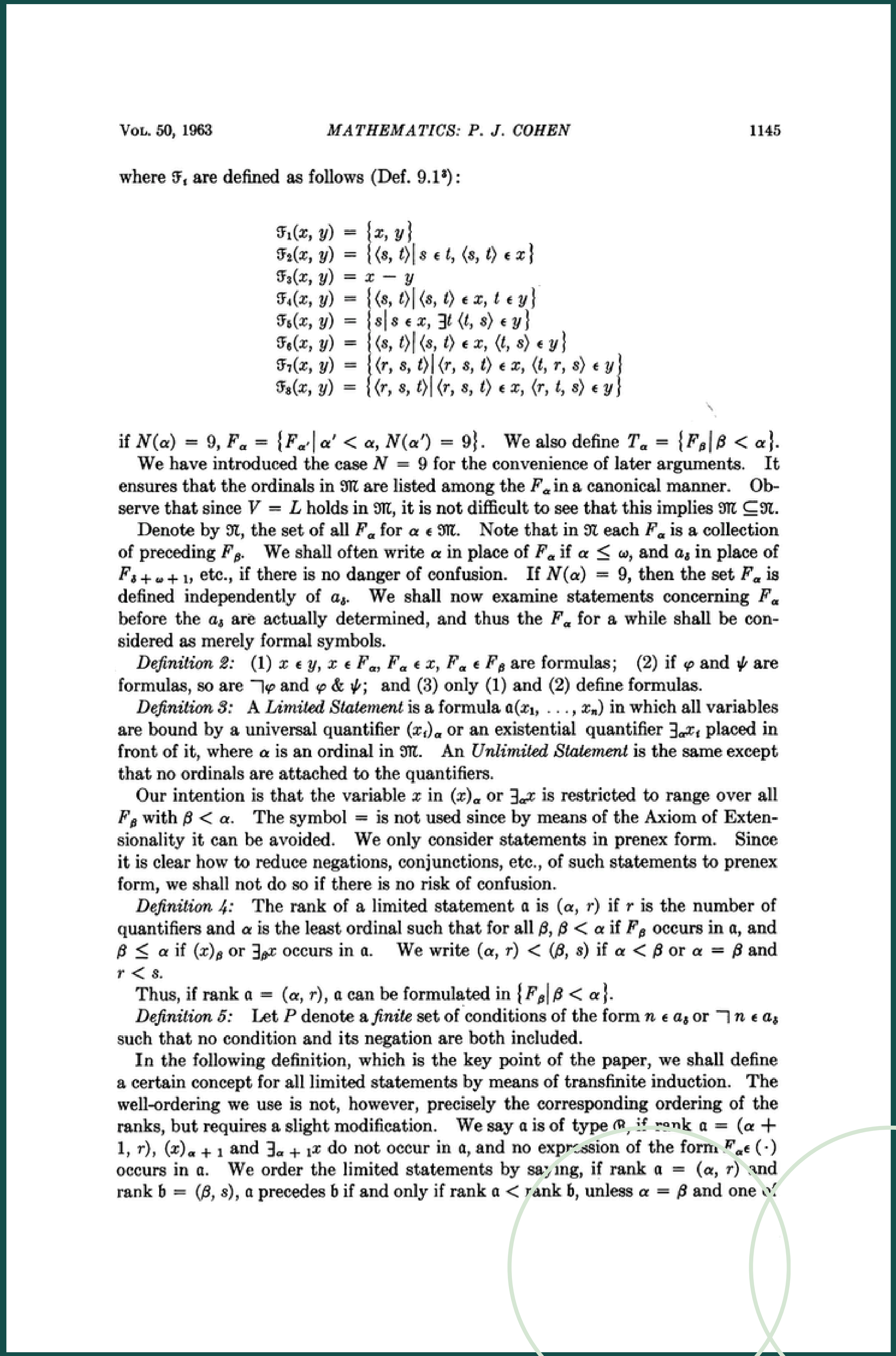


Min Cheol Seo (SKKU)

Framing the Question

What is a mathematical proof?

A mathematical proof is a *deductive argument* for a mathematical statement, showing that the stated assumptions *logically guarantee the conclusion*. The argument may use other previously established statements, such as theorems; but every proof can, in principle, be constructed using only certain basic or original assumptions known as axioms, along with the accepted rules of inference. [Wikipedia]



where $\mathfrak{F}_i$ are defined as follows (Def. 9.1[3]):

$$
\begin{aligned}
\mathfrak{F}_1(x, y) &= \{x, y\}\\
\mathfrak{F}_2(x, y) &= \{\langle s, t\rangle \mid s \in t, \langle s, t\rangle \in x\}\\
\mathfrak{F}_3(x, y) &= x - y\\
\mathfrak{F}_4(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, t \in y\}\\
\mathfrak{F}_5(x, y) &= \{s \mid s \in x, \exists t\ \langle t, s\rangle \in y\}\\
\mathfrak{F}_6(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, \langle t, s\rangle \in y\}\\
\mathfrak{F}_7(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle t, r, s\rangle \in y\}\\
\mathfrak{F}_8(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle r, t, s\rangle \in y\}
\end{aligned}
$$

if $N(\alpha) = 9$, $F_\alpha = \{F_{\alpha'} \mid \alpha' < \alpha, N(\alpha') = 9\}$. We also define $T_\alpha = \{F_\beta \mid \beta < \alpha\}$.

We have introduced the case $N = 9$ for the convenience of later arguments. It ensures that the ordinals in $\mathfrak{M}$ are listed among the F_α in a canonical manner. Observe that since $V = L$ holds in $\mathfrak{M}$, it is not difficult to see that this implies $\mathfrak{M} \subseteq \mathfrak{N}$.

Denote by $\mathfrak{N}$, the set of all F_α for $\alpha \in \mathfrak{M}$. Note that in $\mathfrak{N}$ each F_α is a collection of preceding F_β. We shall often write α in place of F_α if $\alpha \leq \omega$, and a_δ in place of $F_{\delta+\omega+1}$, etc., if there is no danger of confusion. If $N(\alpha) = 9$, then the set F_α is defined independently of a_δ. We shall now examine statements concerning F_α before the a_δ are actually determined, and thus the F_α for a while shall be considered as merely formal symbols.

Definition 2: (1) $x \in y$, $x \in F_\alpha$, $F_\alpha \in x$, $F_\alpha \in F_\beta$ are formulas; (2) if φ and ψ are formulas, so are $\neg\varphi$ and $\varphi \,\&\, \psi$; and (3) only (1) and (2) define formulas.

Definition 3: A *Limited Statement* is a formula $\mathfrak{a}(x_1, \ldots, x_n)$ in which all variables are bound by a universal quantifier $(x_i)_\alpha$ or an existential quantifier $\exists_\alpha x_i$ placed in front of it, where α is an ordinal in $\mathfrak{M}$. An *Unlimited Statement* is the same except that no ordinals are attached to the quantifiers.

Our intention is that the variable x in $(x)_\alpha$ or $\exists_\alpha x$ is restricted to range over all F_β with $\beta < \alpha$. The symbol $=$ is not used since by means of the Axiom of Extensionality it can be avoided. We only consider statements in prenex form. Since it is clear how to reduce negations, conjunctions, etc., of such statements to prenex form, we shall not do so if there is no risk of confusion.

Definition 4: The rank of a limited statement $\mathfrak{a}$ is (α, r) if r is the number of quantifiers and α is the least ordinal such that for all β, $\beta < \alpha$ if F_β occurs in $\mathfrak{a}$, and $\beta \leq \alpha$ if $(x)_\beta$ or $\exists_\beta x$ occurs in $\mathfrak{a}$. We write $(\alpha, r) < (\beta, s)$ if $\alpha < \beta$ or $\alpha = \beta$ and $r < s$.

Thus, if rank $\mathfrak{a} = (\alpha, r)$, $\mathfrak{a}$ can be formulated in $\{F_\beta \mid \beta < \alpha\}$.

Definition 5: Let P denote a *finite* set of conditions of the form $n \in a_\delta$ or $\neg n \in a_\delta$ such that no condition and its negation are both included.

In the following definition, which is the key point of the paper, we shall define a certain concept for all limited statements by means of transfinite induction. The well-ordering we use is not, however, precisely the corresponding ordering of the ranks, but requires a slight modification. We say $\mathfrak{a}$ is of type $\mathfrak{P}$, if rank $\mathfrak{a} = (\alpha + 1, r)$, $(x)_{\alpha+1}$ and $\exists_{\alpha+1} x$ do not occur in $\mathfrak{a}$, and no expression of the form $F_\alpha \in (\cdot)$ occurs in $\mathfrak{a}$. We order the limited statements by saying, if rank $\mathfrak{a} = (\alpha, r)$ and rank $\mathfrak{b} = (\beta, s)$, $\mathfrak{a}$ precedes $\mathfrak{b}$ if and only if rank $\mathfrak{a} <$ rank $\mathfrak{b}$, unless $\alpha = \beta$ and one of

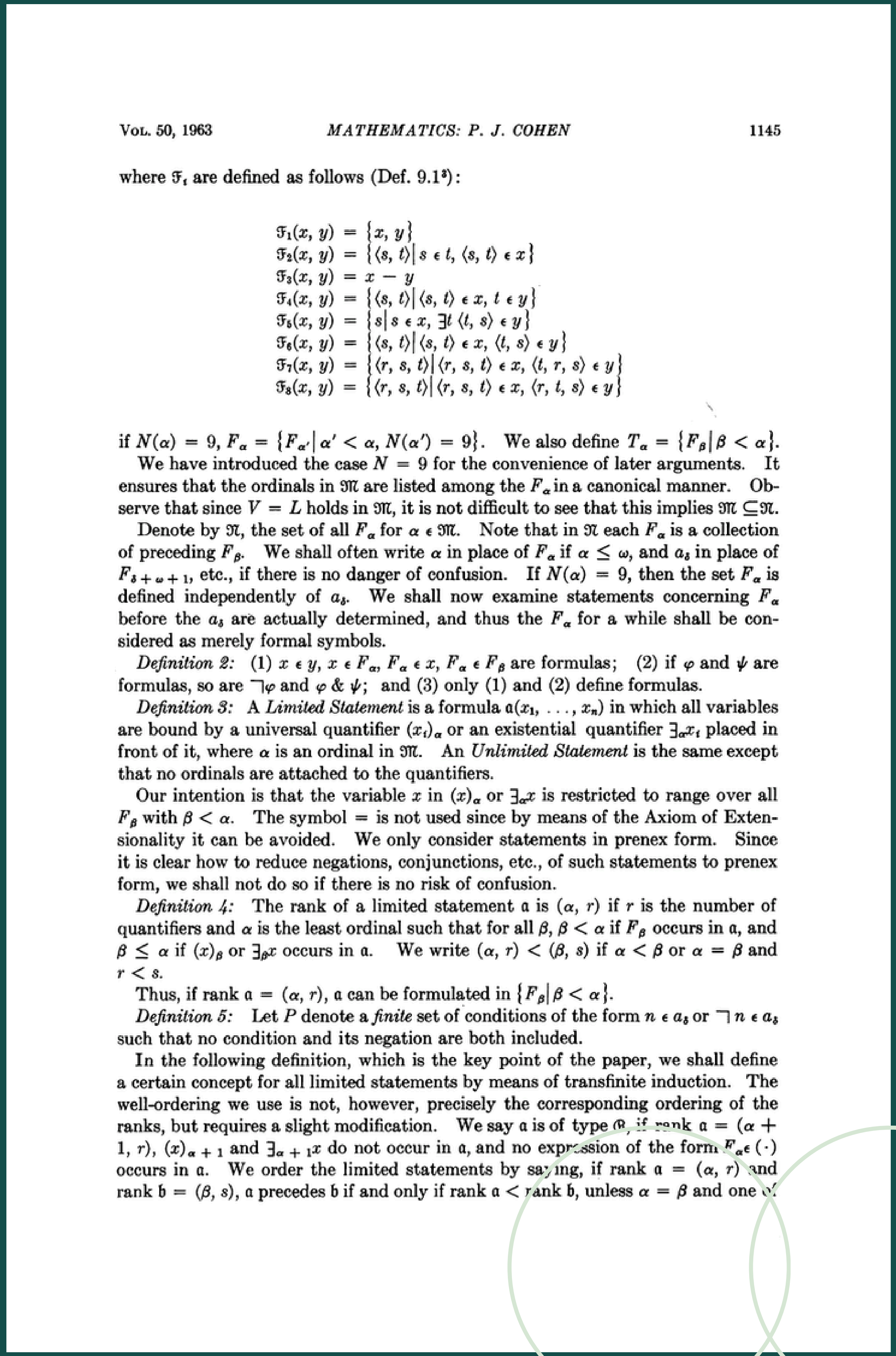
Framing the Question

What is a mathematical proof? (Philosophically)

Formal Sense

- A proof is a finite sequence of formulas derived from axioms by rules of inference.
- It guarantees *truth preservation* in a formal system.

1. **Syntax**: the language and formation rules of the system. Determines what counts as a well-formed formula (WFF).
2. **Axioms**: initial statements taken as given. The basis from which derivations proceed.
3. **Inference Rules**: transformations preserving truth. E.g., Modus Ponens, Universal Instantiation.
4. **Derivation**: the step-by-step application of rules from axioms to theorem.



where $\mathfrak{F}_i$ are defined as follows (Def. 9.1[3]):

$$
\begin{aligned}
\mathfrak{F}_1(x, y) &= \{x, y\}\\
\mathfrak{F}_2(x, y) &= \{\langle s, t\rangle \mid s \in t, \langle s, t\rangle \in x\}\\
\mathfrak{F}_3(x, y) &= x - y\\
\mathfrak{F}_4(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, t \in y\}\\
\mathfrak{F}_5(x, y) &= \{s \mid s \in x, \exists t\, \langle t, s\rangle \in y\}\\
\mathfrak{F}_6(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, \langle t, s\rangle \in y\}\\
\mathfrak{F}_7(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle t, r, s\rangle \in y\}\\
\mathfrak{F}_8(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle r, t, s\rangle \in y\}
\end{aligned}
$$

if $N(\alpha) = 9$, $F_\alpha = \{F_{\alpha'} \mid \alpha' < \alpha, N(\alpha') = 9\}$. We also define $T_\alpha = \{F_\beta \mid \beta < \alpha\}$.

We have introduced the case $N = 9$ for the convenience of later arguments. It ensures that the ordinals in $\mathfrak{M}$ are listed among the F_α in a canonical manner. Observe that since $V = L$ holds in $\mathfrak{M}$, it is not difficult to see that this implies $\mathfrak{M} \subseteq \mathfrak{N}$.

Denote by $\mathfrak{N}$, the set of all F_α for $\alpha \in \mathfrak{M}$. Note that in $\mathfrak{N}$ each F_α is a collection of preceding F_β. We shall often write α in place of F_α if $\alpha \leq \omega$, and a_δ in place of $F_{\delta+\omega+1}$, etc., if there is no danger of confusion. If $N(\alpha) = 9$, then the set F_α is defined independently of a_δ. We shall now examine statements concerning F_α before the a_δ are actually determined, and thus the F_α for a while shall be considered as merely formal symbols.

Definition 2: (1) $x \in y$, $x \in F_\alpha$, $F_\alpha \in x$, $F_\alpha \in F_\beta$ are formulas; (2) if φ and ψ are formulas, so are $\neg\varphi$ and $\varphi \,\&\, \psi$; and (3) only (1) and (2) define formulas.

Definition 3: A *Limited Statement* is a formula $\mathfrak{a}(x_1, \ldots, x_n)$ in which all variables are bound by a universal quantifier $(x_i)_\alpha$ or an existential quantifier $\exists_\alpha x_i$ placed in front of it, where α is an ordinal in $\mathfrak{M}$. An *Unlimited Statement* is the same except that no ordinals are attached to the quantifiers.

Our intention is that the variable x in $(x)_\alpha$ or $\exists_\alpha x$ is restricted to range over all F_β with $\beta < \alpha$. The symbol $=$ is not used since by means of the Axiom of Extensionality it can be avoided. We only consider statements in prenex form. Since it is clear how to reduce negations, conjunctions, etc., of such statements to prenex form, we shall not do so if there is no risk of confusion.

Definition 4: The rank of a limited statement $\mathfrak{a}$ is (α, r) if r is the number of quantifiers and α is the least ordinal such that for all β, $\beta < \alpha$ if F_β occurs in $\mathfrak{a}$, and $\beta \leq \alpha$ if $(x)_\beta$ or $\exists_\beta x$ occurs in $\mathfrak{a}$. We write $(\alpha, r) < (\beta, s)$ if $\alpha < \beta$ or $\alpha = \beta$ and $r < s$.

Thus, if rank $\mathfrak{a} = (\alpha, r)$, $\mathfrak{a}$ can be formulated in $\{F_\beta \mid \beta < \alpha\}$.

Definition 5: Let P denote a *finite* set of conditions of the form $n \in a_\delta$ or $\neg n \in a_\delta$ such that no condition and its negation are both included.

In the following definition, which is the key point of the paper, we shall define a certain concept for all limited statements by means of transfinite induction. The well-ordering we use is not, however, precisely the corresponding ordering of the ranks, but requires a slight modification. We say $\mathfrak{a}$ is of type $\mathfrak{B}$, if rank $\mathfrak{a} = (\alpha + 1, r)$, $(x)_{\alpha+1}$ and $\exists_{\alpha+1} x$ do not occur in $\mathfrak{a}$, and no expression of the form $F_\alpha \in (\cdot)$ occurs in $\mathfrak{a}$. We order the limited statements by saying, if rank $\mathfrak{a} = (\alpha, r)$ and rank $\mathfrak{b} = (\beta, s)$, $\mathfrak{a}$ precedes $\mathfrak{b}$ if and only if rank $\mathfrak{a} <$ rank $\mathfrak{b}$, unless $\alpha = \beta$ and one of

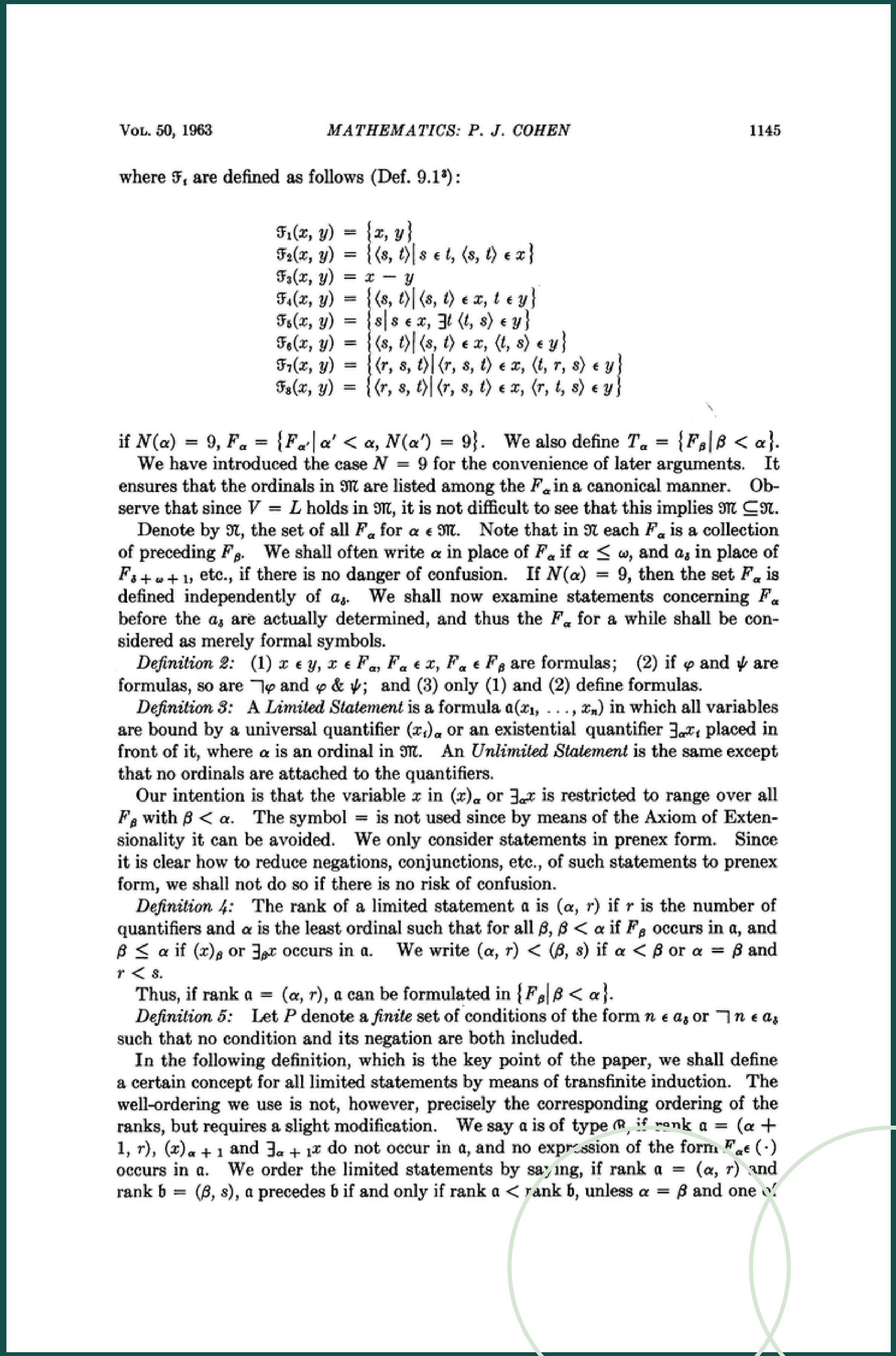
Framing the Question

What is a mathematical proof? (Philosophically)

Epistemic sense

- A proof is an *instrument of knowledge*: it's what entitles us to believe a theorem is true.
- It is not just a string of formulas but a means of showing why something must be true.

1. **Justification**: A proof shows not only *that* a theorem is true, but *why*.
2. **Reasons**: Proofs gives reasons to transform belief into knowledge.



Vol. 50, 1963 *MATHEMATICS: P. J. COHEN* 1145

where $\mathfrak{F}_i$ are defined as follows (Def. 9.1[3]):

$$\begin{aligned}
\mathfrak{F}_1(x, y) &= \{x, y\}\\
\mathfrak{F}_2(x, y) &= \{\langle s, t\rangle \mid s \in t, \langle s, t\rangle \in x\}\\
\mathfrak{F}_3(x, y) &= x - y\\
\mathfrak{F}_4(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, t \in y\}\\
\mathfrak{F}_5(x, y) &= \{s \mid s \in x, \exists t\, \langle t, s\rangle \in y\}\\
\mathfrak{F}_6(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, \langle t, s\rangle \in y\}\\
\mathfrak{F}_7(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle t, r, s\rangle \in y\}\\
\mathfrak{F}_8(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle r, t, s\rangle \in y\}
\end{aligned}$$

if $N(\alpha) = 9$, $F_\alpha = \{F_{\alpha'} \mid \alpha' < \alpha, N(\alpha') = 9\}$. We also define $T_\alpha = \{F_\beta \mid \beta < \alpha\}$.

We have introduced the case $N = 9$ for the convenience of later arguments. It ensures that the ordinals in $\mathfrak{M}$ are listed among the F_α in a canonical manner. Observe that since $V = L$ holds in $\mathfrak{M}$, it is not difficult to see that this implies $\mathfrak{M} \subseteq \mathfrak{N}$.

Denote by $\mathfrak{N}$, the set of all F_α for $\alpha \in \mathfrak{M}$. Note that in $\mathfrak{N}$ each F_α is a collection of preceding F_β. We shall often write α in place of F_α if $\alpha \leq \omega$, and a_δ in place of $F_{\delta+\omega+1}$, etc., if there is no danger of confusion. If $N(\alpha) = 9$, then the set F_α is defined independently of a_δ. We shall now examine statements concerning F_α before the a_δ are actually determined, and thus the F_α for a while shall be considered as merely formal symbols.

Definition 2: (1) $x \in y$, $x \in F_\alpha$, $F_\alpha \in x$, $F_\alpha \in F_\beta$ are formulas; (2) if φ and ψ are formulas, so are $\neg\varphi$ and $\varphi \,\&\, \psi$; and (3) only (1) and (2) define formulas.

Definition 3: A *Limited Statement* is a formula $\mathfrak{a}(x_1, \ldots, x_n)$ in which all variables are bound by a universal quantifier $(x_i)_\alpha$ or an existential quantifier $\exists_\alpha x_i$ placed in front of it, where α is an ordinal in $\mathfrak{M}$. An *Unlimited Statement* is the same except that no ordinals are attached to the quantifiers.

Our intention is that the variable x in $(x)_\alpha$ or $\exists_\alpha x$ is restricted to range over all F_β with $\beta < \alpha$. The symbol $=$ is not used since by means of the Axiom of Extensionality it can be avoided. We only consider statements in prenex form. Since it is clear how to reduce negations, conjunctions, etc., of such statements to prenex form, we shall not do so if there is no risk of confusion.

Definition 4: The rank of a limited statement $\mathfrak{a}$ is (α, r) if r is the number of quantifiers and α is the least ordinal such that for all β, $\beta < \alpha$ if F_β occurs in $\mathfrak{a}$, and $\beta \leq \alpha$ if $(x)_\beta$ or $\exists_\beta x$ occurs in $\mathfrak{a}$. We write $(\alpha, r) < (\beta, s)$ if $\alpha < \beta$ or $\alpha = \beta$ and $r < s$.

Thus, if rank $\mathfrak{a} = (\alpha, r)$, $\mathfrak{a}$ can be formulated in $\{F_\beta \mid \beta < \alpha\}$.

Definition 5: Let P denote a *finite* set of conditions of the form $n \in a_\delta$ or $\neg n \in a_\delta$ such that no condition and its negation are both included.

In the following definition, which is the key point of the paper, we shall define a certain concept for all limited statements by means of transfinite induction. The well-ordering we use is not, however, precisely the corresponding ordering of the ranks, but requires a slight modification. We say $\mathfrak{a}$ is of type $\mathfrak{B}$, if rank $\mathfrak{a} = (\alpha + 1, r)$, $(x)_{\alpha+1}$ and $\exists_{\alpha+1} x$ do not occur in $\mathfrak{a}$, and no expression of the form $F_\alpha \in (\cdot)$ occurs in $\mathfrak{a}$. We order the limited statements by saying, if rank $\mathfrak{a} = (\alpha, r)$ and rank $\mathfrak{b} = (\beta, s)$, $\mathfrak{a}$ precedes $\mathfrak{b}$ if and only if rank $\mathfrak{a} <$ rank $\mathfrak{b}$, unless $\alpha = \beta$ and one of

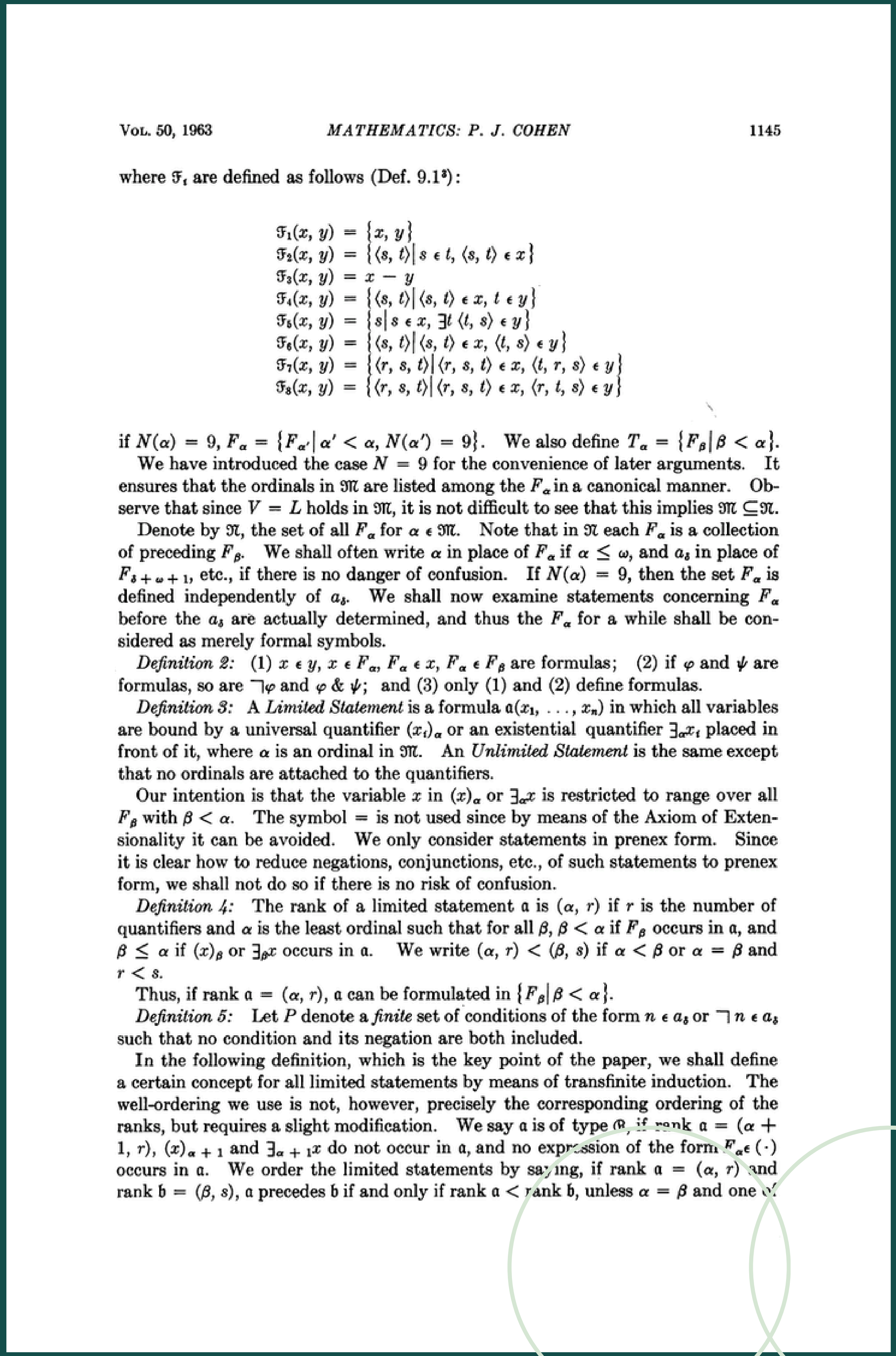
Framing the Question

What does it mean that we *understand* proof?

The Gap

- It is not just a string of formulas but a means of showing why something must be true.
- We can *know* a theorem is true *without* understanding its proof.

- Understanding proof requires something more than just knowing that the proof (theorem) is true!



where $\mathfrak{F}_i$ are defined as follows (Def. 9.1[3]):

$$
\begin{aligned}
\mathfrak{F}_1(x, y) &= \{x, y\} \\
\mathfrak{F}_2(x, y) &= \{\langle s, t\rangle \mid s \in t, \langle s, t\rangle \in x\} \\
\mathfrak{F}_3(x, y) &= x - y \\
\mathfrak{F}_4(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, t \in y\} \\
\mathfrak{F}_5(x, y) &= \{s \mid s \in x, \exists t\, \langle t, s\rangle \in y\} \\
\mathfrak{F}_6(x, y) &= \{\langle s, t\rangle \mid \langle s, t\rangle \in x, \langle t, s\rangle \in y\} \\
\mathfrak{F}_7(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle t, r, s\rangle \in y\} \\
\mathfrak{F}_8(x, y) &= \{\langle r, s, t\rangle \mid \langle r, s, t\rangle \in x, \langle r, t, s\rangle \in y\}
\end{aligned}
$$

if $N(\alpha) = 9$, $F_\alpha = \{F_{\alpha'} \mid \alpha' < \alpha, N(\alpha') = 9\}$. We also define $T_\alpha = \{F_\beta \mid \beta < \alpha\}$.

We have introduced the case $N = 9$ for the convenience of later arguments. It ensures that the ordinals in $\mathfrak{M}$ are listed among the F_α in a canonical manner. Observe that since $V = L$ holds in $\mathfrak{M}$, it is not difficult to see that this implies $\mathfrak{M} \subseteq \mathfrak{N}$.

Denote by $\mathfrak{N}$, the set of all F_α for $\alpha \in \mathfrak{M}$. Note that in $\mathfrak{N}$ each F_α is a collection of preceding F_β. We shall often write α in place of F_α if $\alpha \leq \omega$, and a_δ in place of $F_{\delta + \omega + 1}$, etc., if there is no danger of confusion. If $N(\alpha) = 9$, then the set F_α is defined independently of a_δ. We shall now examine statements concerning F_α before the a_δ are actually determined, and thus the F_α for a while shall be considered as merely formal symbols.

Definition 2: (1) $x \in y$, $x \in F_\alpha$, $F_\alpha \in x$, $F_\alpha \in F_\beta$ are formulas; (2) if φ and ψ are formulas, so are $\neg\varphi$ and $\varphi \,\&\, \psi$; and (3) only (1) and (2) define formulas.

Definition 3: A *Limited Statement* is a formula $\mathfrak{a}(x_1, \ldots, x_n)$ in which all variables are bound by a universal quantifier $(x_i)_\alpha$ or an existential quantifier $\exists_\alpha x_i$ placed in front of it, where α is an ordinal in $\mathfrak{M}$. An *Unlimited Statement* is the same except that no ordinals are attached to the quantifiers.

Our intention is that the variable x in $(x)_\alpha$ or $\exists_\alpha x$ is restricted to range over all F_β with $\beta < \alpha$. The symbol $=$ is not used since by means of the Axiom of Extensionality it can be avoided. We only consider statements in prenex form. Since it is clear how to reduce negations, conjunctions, etc., of such statements to prenex form, we shall not do so if there is no risk of confusion.

Definition 4: The rank of a limited statement $\mathfrak{a}$ is (α, r) if r is the number of quantifiers and α is the least ordinal such that for all β, $\beta < \alpha$ if F_β occurs in $\mathfrak{a}$, and $\beta \leq \alpha$ if $(x)_\beta$ or $\exists_\beta x$ occurs in $\mathfrak{a}$. We write $(\alpha, r) < (\beta, s)$ if $\alpha < \beta$ or $\alpha = \beta$ and $r < s$.

Thus, if rank $\mathfrak{a} = (\alpha, r)$, $\mathfrak{a}$ can be formulated in $\{F_\beta \mid \beta < \alpha\}$.

Definition 5: Let P denote a *finite* set of conditions of the form $n \in a_\delta$ or $\neg n \in a_\delta$ such that no condition and its negation are both included.

In the following definition, which is the key point of the paper, we shall define a certain concept for all limited statements by means of transfinite induction. The well-ordering we use is not, however, precisely the corresponding ordering of the ranks, but requires a slight modification. We say $\mathfrak{a}$ is of type $\mathfrak{B}$, if rank $\mathfrak{a} = (\alpha + 1, r)$, $(x)_{\alpha+1}$ and $\exists_{\alpha+1} x$ do not occur in $\mathfrak{a}$, and no expression of the form $F_\alpha \in (\cdot)$ occurs in $\mathfrak{a}$. We order the limited statements by saying, if rank $\mathfrak{a} = (\alpha, r)$ and rank $\mathfrak{b} = (\beta, s)$, $\mathfrak{a}$ precedes $\mathfrak{b}$ if and only if rank $\mathfrak{a} <$ rank $\mathfrak{b}$, unless $\alpha = \beta$ and one of

The Four-Colour Theorem

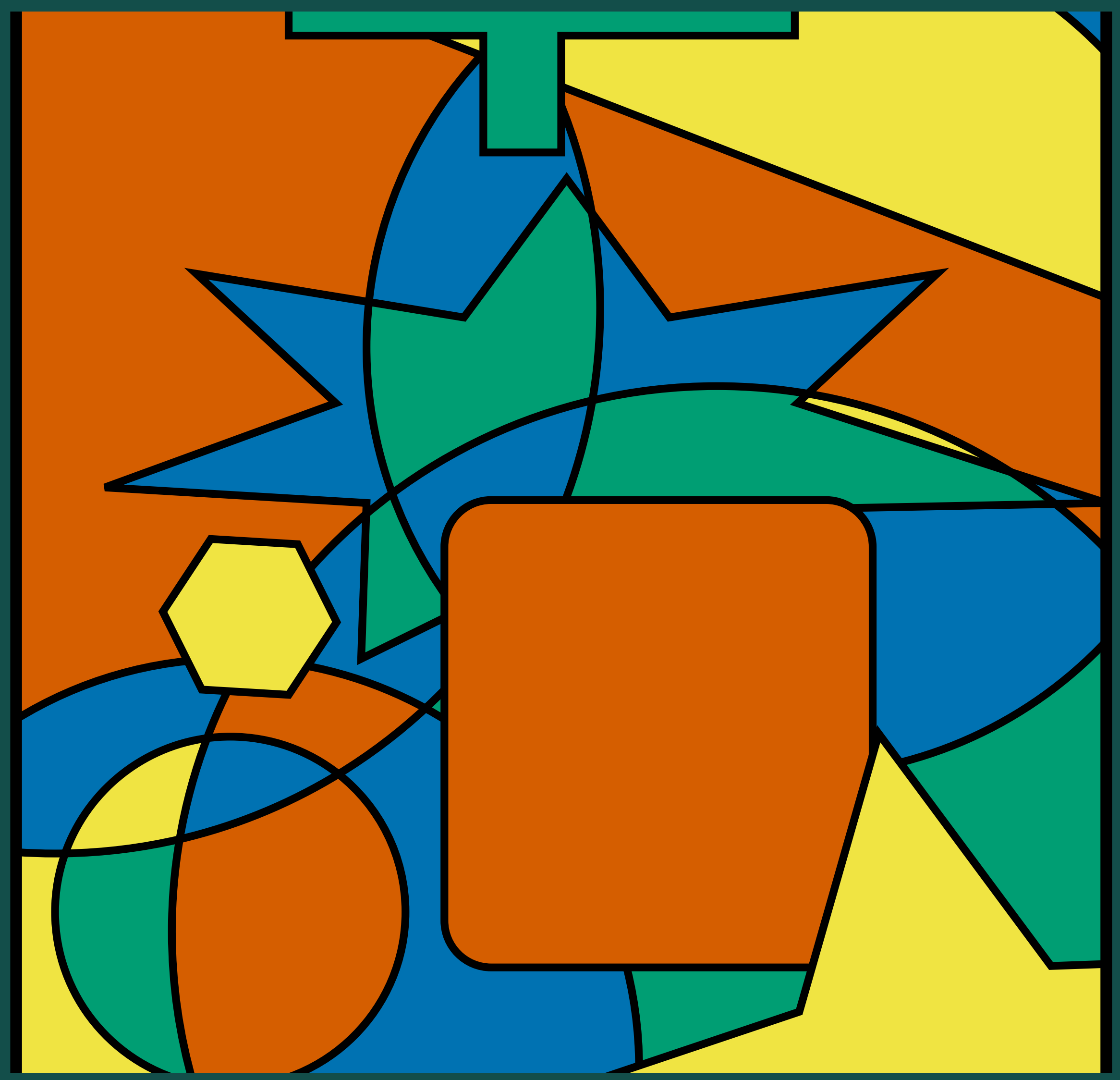
Where the epistemic gap appears.

The Four-Colour Theorem

- Every planar map can be coloured with no more than four colours, so that no adjacent regions share the same colour.

To prove it, we need to show that there is no minimal counterexample (five coloured planar map) can exists. Which then turns on the following two core notions:

1. **Reducibility**: a configuration is reducible if it cannot appear in a minimal counterexample.
2. **Unavoidability**: the proof identifies a finite set of reducible configurations (the unavoidable set) such that every map contains at least one of them.



The Four-Colour Theorem

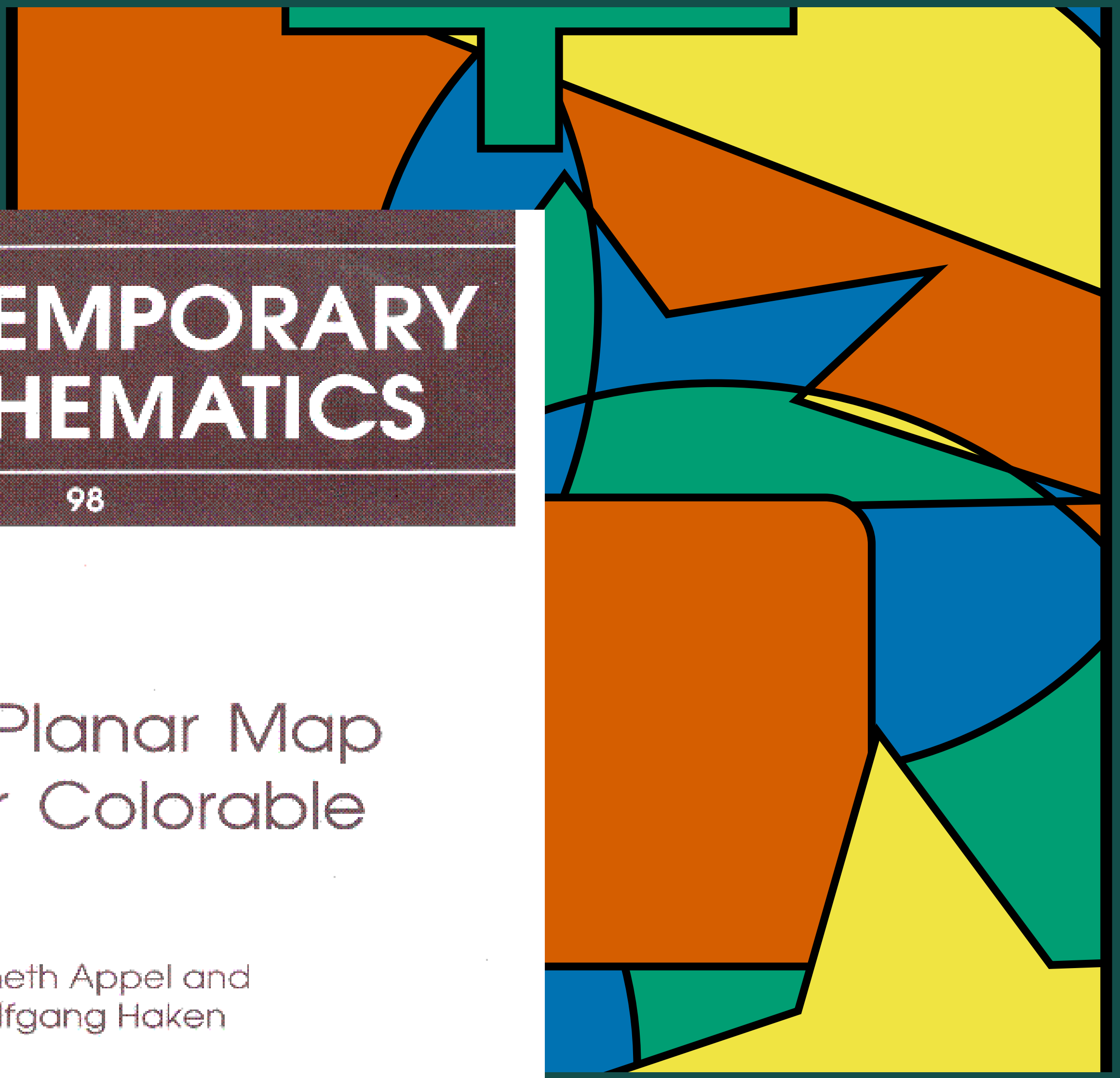
Where the epistemic gap appears.

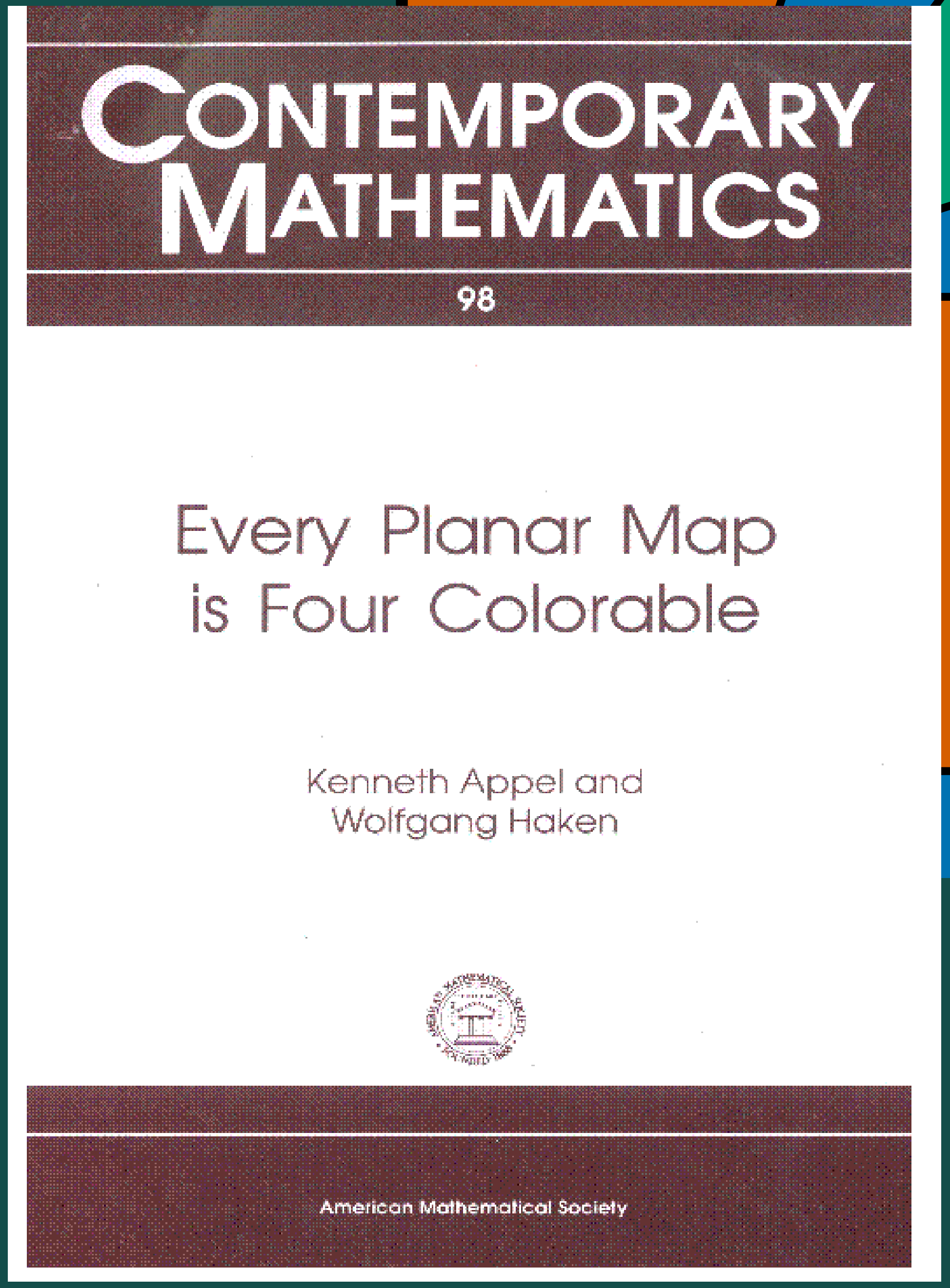
Appel and Haken's Proof of 4CT

- Appel and Haken identifies that there are 1,834 configurations in the unavoidable set!

But 1,834 configurations is to big! No human can calculate that!

- Appel and Haken used computer to calculate and verify, that each of 1,834 cases are all indeed reducible!
- How can we verify that A&H's proof is indeed correct?





The Four-Colour Theorem

Where the epistemic gap appears.

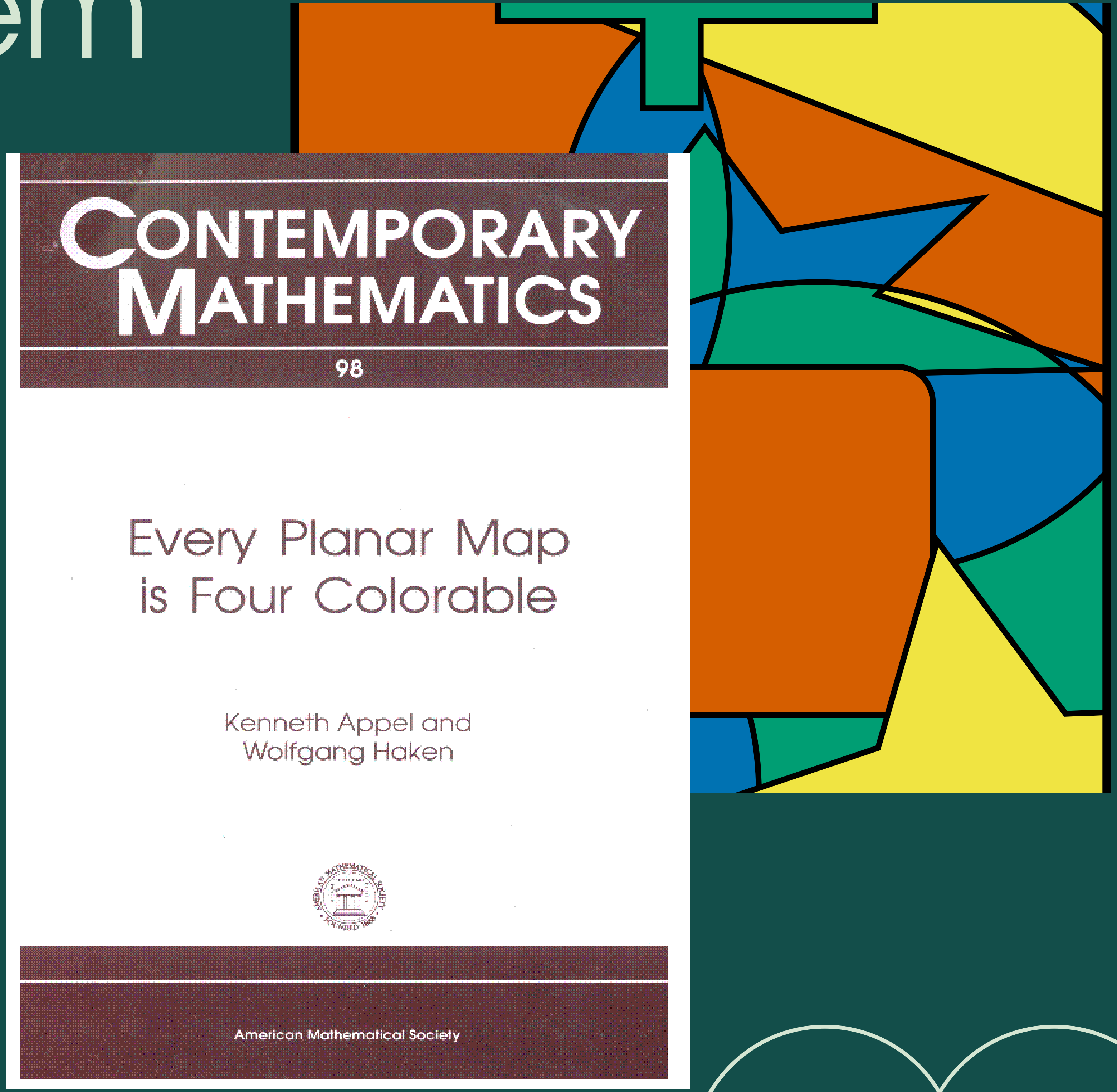
Gonthier’s Proof [2005]

- We now know that the proof is indeed correct, and that the configuration can be further reduced to 633.
- And this result is also achieved by computer (Coq).
- But even so, this requires the full trust of Coq kernel.

Still, mathematicians think that they *understand* the proof.

Question:

What kind of cognitive or epistemic achievement does such understanding amount to?



Proof, Computer and Mathematical Agent

Reconsidering mathematical understanding in the age of AI

Plan-based account of Proof Understanding

We examine a novel, plan-based account of proof understanding, where understanding proof requires the rational reconstruction of proof's underlying plan.

The Limit Case: Robbins Conjecture

We consider the proof of Robbins conjecture, and shows that the adequacy of the plan-based account can be challenged.

Extending the plan-based account: HPA

Extending the plan-based account, the Hierarchical Proof Architecture (HPA) framework suggests proofs as multi-level architectures—human intentions at the top, computational procedures below, bound by ordered dependence.

Plan-Based Account

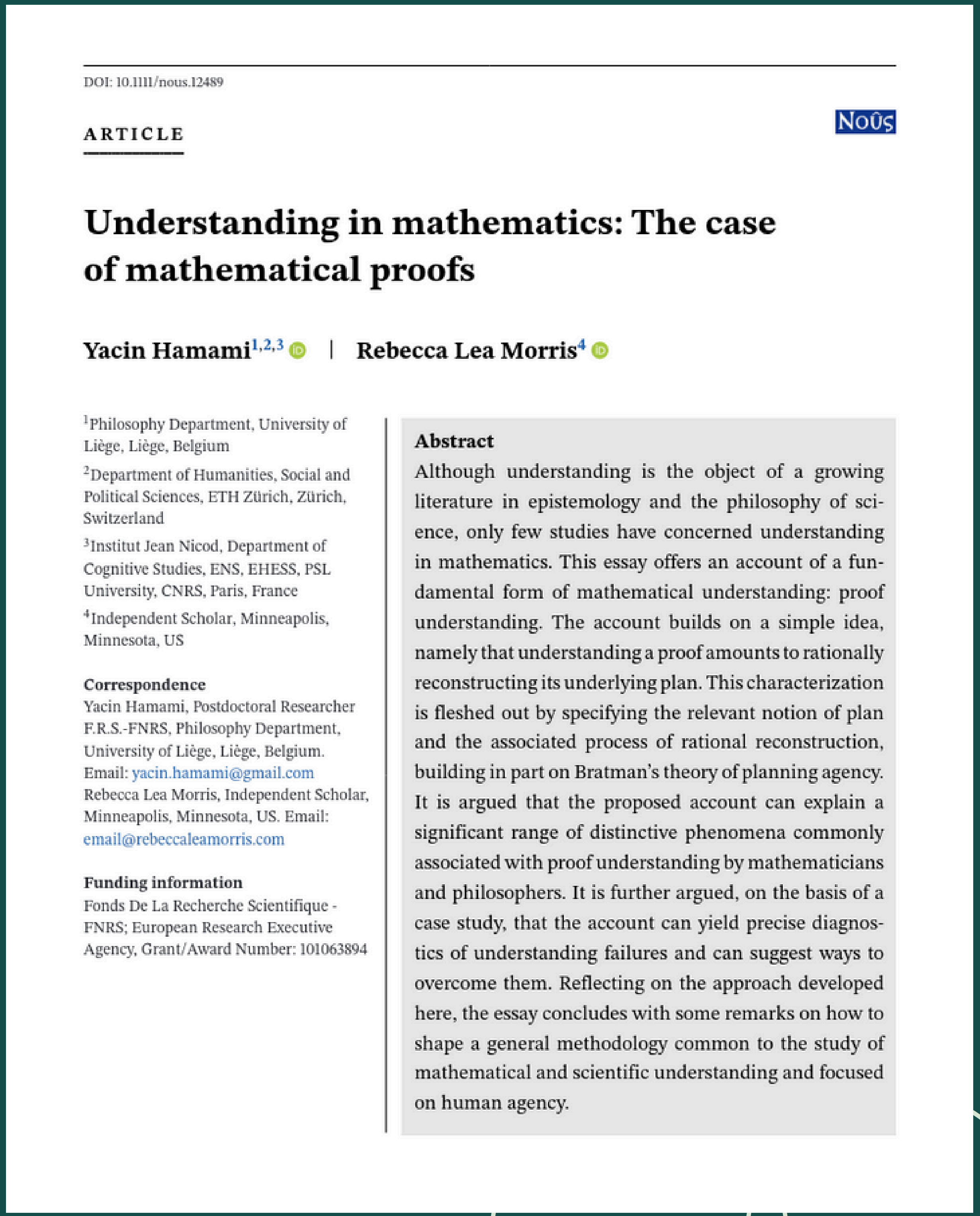
Understanding in mathematics: The case of mathematical proofs

The Plan-Based Account

- "An agent *understands* a proof *P* iff she has *rationally reconstructed* the plan underlying *P*". [2024]

Key detail:

1. Proofs = *activities*, not just texts.
2. Each proof encodes a *hierarchy of intentions*.
3. Understanding = *reconstructing* those intentions.



DOI: 10.1111/nous.12489

Noûs

ARTICLE

Understanding in mathematics: The case of mathematical proofs

Yacin Hamami[1,2,3] | Rebecca Lea Morris[4]

[1]Philosophy Department, University of Liège, Liège, Belgium
[2]Department of Humanities, Social and Political Sciences, ETH Zürich, Zürich, Switzerland
[3]Institut Jean Nicod, Department of Cognitive Studies, ENS, EHESS, PSL University, CNRS, Paris, France
[4]Independent Scholar, Minneapolis, Minnesota, US

Correspondence
Yacin Hamami, Postdoctoral Researcher F.R.S.-FNRS, Philosophy Department, University of Liège, Liège, Belgium. Email: yacin.hamami@gmail.com
Rebecca Lea Morris, Independent Scholar, Minneapolis, Minnesota, US. Email: email@rebeccaleamorris.com

Funding information
Fonds De La Recherche Scientifique - FNRS; European Research Executive Agency, Grant/Award Number: 101063894

Abstract
Although understanding is the object of a growing literature in epistemology and the philosophy of science, only few studies have concerned understanding in mathematics. This essay offers an account of a fundamental form of mathematical understanding: proof understanding. The account builds on a simple idea, namely that understanding a proof amounts to rationally reconstructing its underlying plan. This characterization is fleshed out by specifying the relevant notion of plan and the associated process of rational reconstruction, building in part on Bratman's theory of planning agency. It is argued that the proposed account can explain a significant range of distinctive phenomena commonly associated with proof understanding by mathematicians and philosophers. It is further argued, on the basis of a case study, that the account can yield precise diagnostics of understanding failures and can suggest ways to overcome them. Reflecting on the approach developed here, the essay concludes with some remarks on how to shape a general methodology common to the study of mathematical and scientific understanding and focused on human agency.

Plan-Based Account

Understanding in mathematics: The case of mathematical proofs

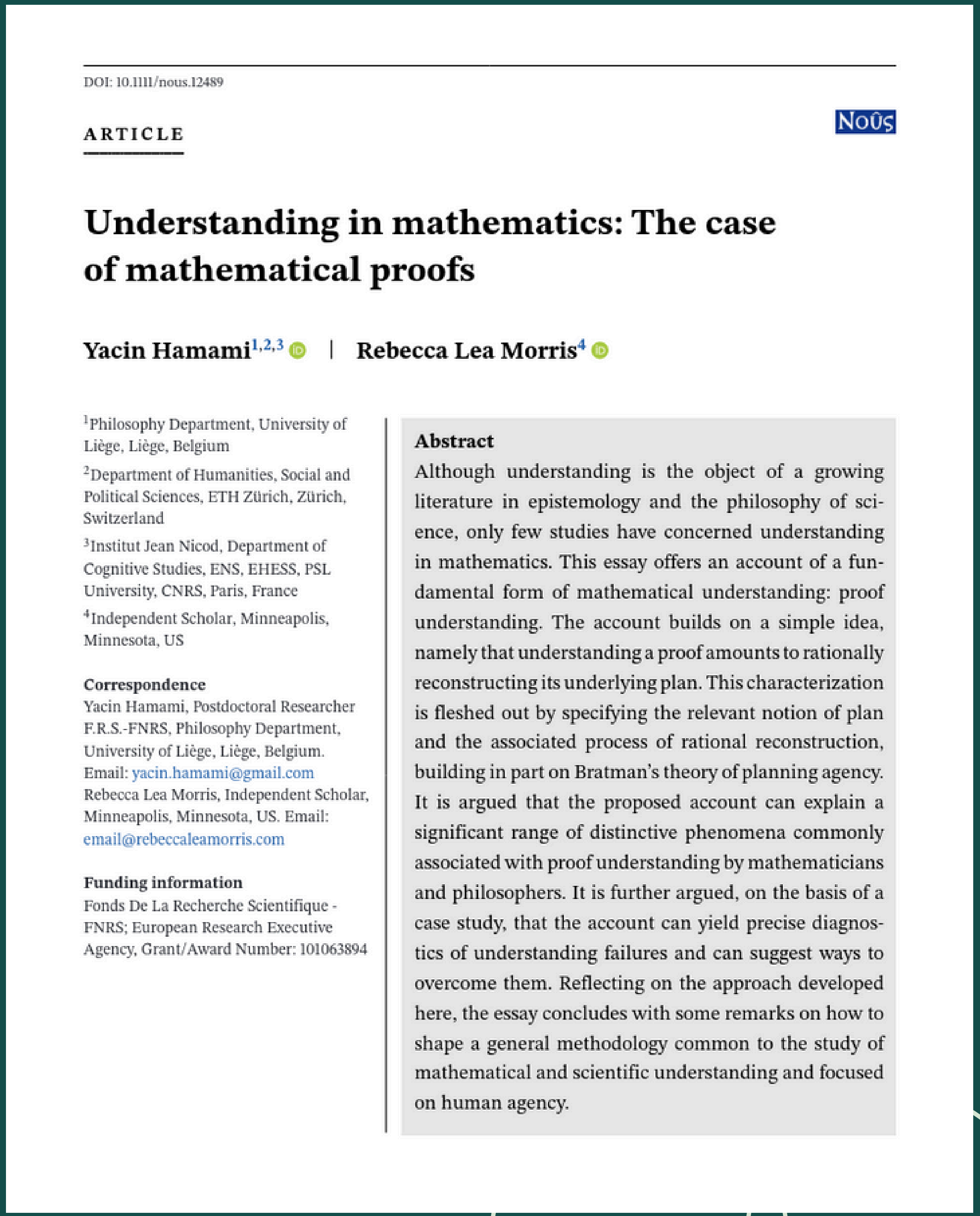
The Proof as Activities

1. **Proof**: the written sequence of inferences
2. **Proof Activity**: the epistemic performance described by the text
3. **Proof Plan**: the hierarchical structure of intentions underlying the activity

Theory of Planning

- Following Bratman, H&M sees proof as action of agents, deriven by hierarchical intentions ('to prove,' 'to show,' 'to infer')

Plan → Proof Activity → Proof!


DOI: 10.1111/nous.12489

ARTICLE

Noûs

Understanding in mathematics: The case of mathematical proofs

Yacin Hamami[1,2,3] | Rebecca Lea Morris[4]

[1]Philosophy Department, University of Liège, Liège, Belgium
[2]Department of Humanities, Social and Political Sciences, ETH Zürich, Zürich, Switzerland
[3]Institut Jean Nicod, Department of Cognitive Studies, ENS, EHESS, PSL University, CNRS, Paris, France
[4]Independent Scholar, Minneapolis, Minnesota, US

Correspondence
Yacin Hamami, Postdoctoral Researcher F.R.S.-FNRS, Philosophy Department, University of Liège, Liège, Belgium. Email: yacin.hamami@gmail.com
Rebecca Lea Morris, Independent Scholar, Minneapolis, Minnesota, US. Email: email@rebeccaleamorris.com

Funding information
Fonds De La Recherche Scientifique - FNRS; European Research Executive Agency, Grant/Award Number: 101063894

Abstract
Although understanding is the object of a growing literature in epistemology and the philosophy of science, only few studies have concerned understanding in mathematics. This essay offers an account of a fundamental form of mathematical understanding: proof understanding. The account builds on a simple idea, namely that understanding a proof amounts to rationally reconstructing its underlying plan. This characterization is fleshed out by specifying the relevant notion of plan and the associated process of rational reconstruction, building in part on Bratman's theory of planning agency. It is argued that the proposed account can explain a significant range of distinctive phenomena commonly associated with proof understanding by mathematicians and philosophers. It is further argued, on the basis of a case study, that the account can yield precise diagnostics of understanding failures and can suggest ways to overcome them. Reflecting on the approach developed here, the essay concludes with some remarks on how to shape a general methodology common to the study of mathematical and scientific understanding and focused on human agency.

Plan-Based Account

Understanding in mathematics: The case of mathematical proofs

Proof as Ordered Tree

- A proof plan is an ordered tree:

1. **Root**: prove theorem
2. **Branches**: subgoals
3. **Leaves**: concrete inferential steps.

Understanding

- (Rationally) reconstructing the tree!

DOI: 10.1111/nous.12489

ARTICLE

Noûs

Understanding in mathematics: The case of mathematical proofs

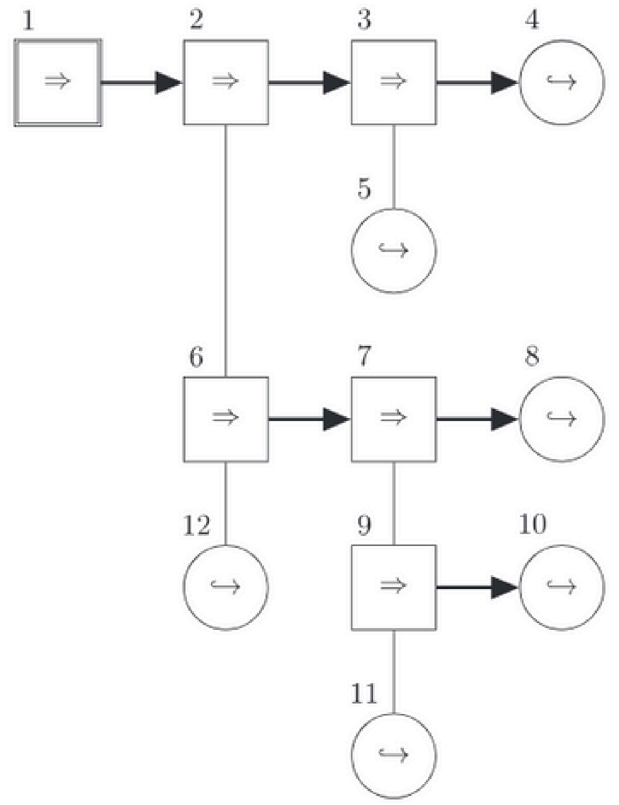


FIGURE 1 A schematic representation of a proof plan.

Squared nodes represent proving intentions of type 'to show' and circled nodes represent proving intentions of type 'to infer'. The root of the tree is the proving intention to show the theorem at hand. In a complete proof plan such as this one, every intention of type 'to show' is decomposed into further intentions organized in a subplan. Executing the plan amounts to traverse this tree in the expected way (which is numbered in the figure) and for each proving intention of type 'to infer' encountered, to carry out the corresponding deductive inference. The execution of the plan leads then to the realization of a proof activity, i.e., a sequence of deductive inferences.

Plan-Based Account

Understanding in mathematics: The case of mathematical proofs

Rationale Reconstruction

- Rational reconstruction consists of two parts:

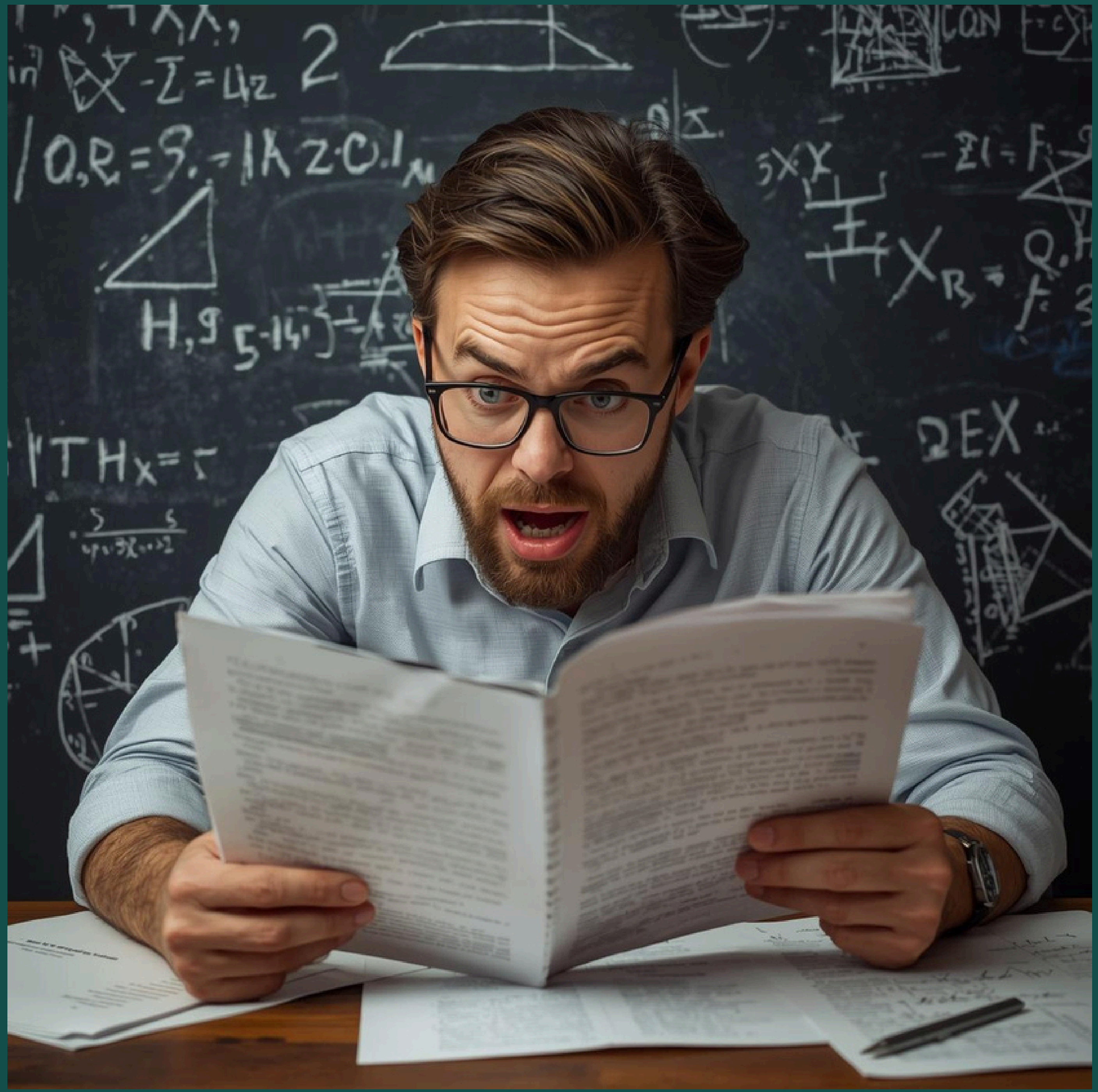
1. **Tracking**: reconstructing the structure of the proof plan – how intentions decompose into subplans. (What was done)
2. **Rationality**: reproducing the author's practical reasoning – why those subplans were chosen. (Why it was done)

Understanding

- Understanding(P) $\Longleftrightarrow$ *Tracking + Rational Reconstruction*

Degrees of Understanding

- Degrees of understanding depend on how much of this plan you can rationally reconstruct.



Post-Hoc Planning

The Problem of Post-Hoc Planning and the Robbins Conjecture

Proof as Plan, but without plan?

- Hamami and Morris assume that for a proof, there always is a rational proof to recover.
- But what if the proof wasn't planned at all? Sheer luck? or AI-driven?

Computational Proofs

- In computational proofs there's often:

1. No single planner,
2. No rigid, pre-specified plans.

- *Only a discovered traces!*

Post-Hoc Planning

The Problem of Post-Hoc Planning and the Robbins Conjecture

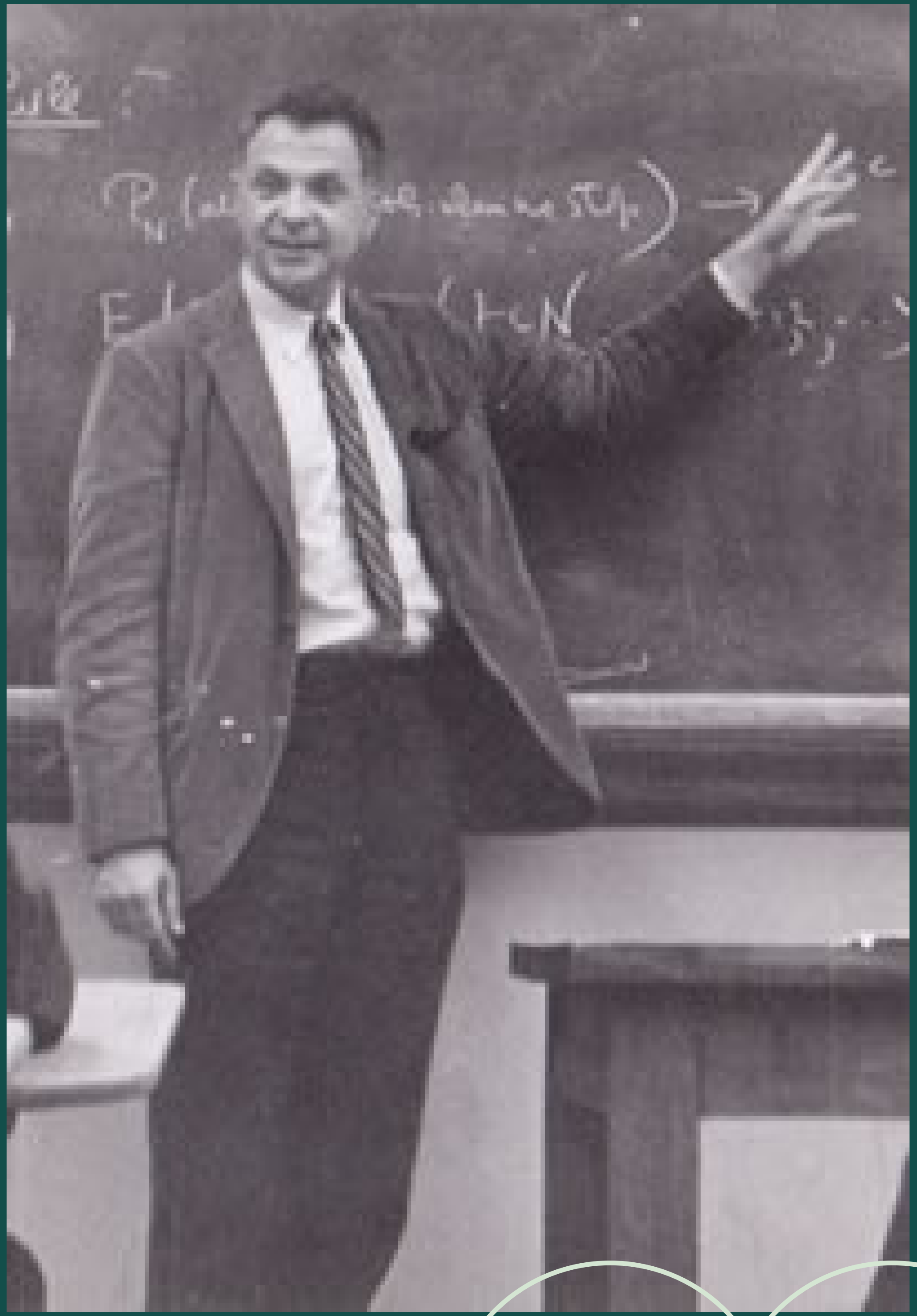
The Robbins Conjecture

- Herbert Robbins conjectured that every Robbins algebra is a Boolean algebra.

A Little Detail

- **Boolean algebra**: A system of logic that deals with two values: true/false. Typically defined by axioms involving three operations: $\vee$, $\wedge$, and $\neg$.
- **Robbins algebra**: An algebra $(A, \vee, \mp)$ is a Robbins algebra if it satisfies the following:

1. Associativity: $x \vee (y \vee z) = (x \vee y) \vee z$
2. Commutativity: $x \vee y = y \vee x$
3. Robbins Identity: $\mp(x \vee y) \vee \mp(x \vee \mp(y \vee z)) = x$



Post-Hoc Planning

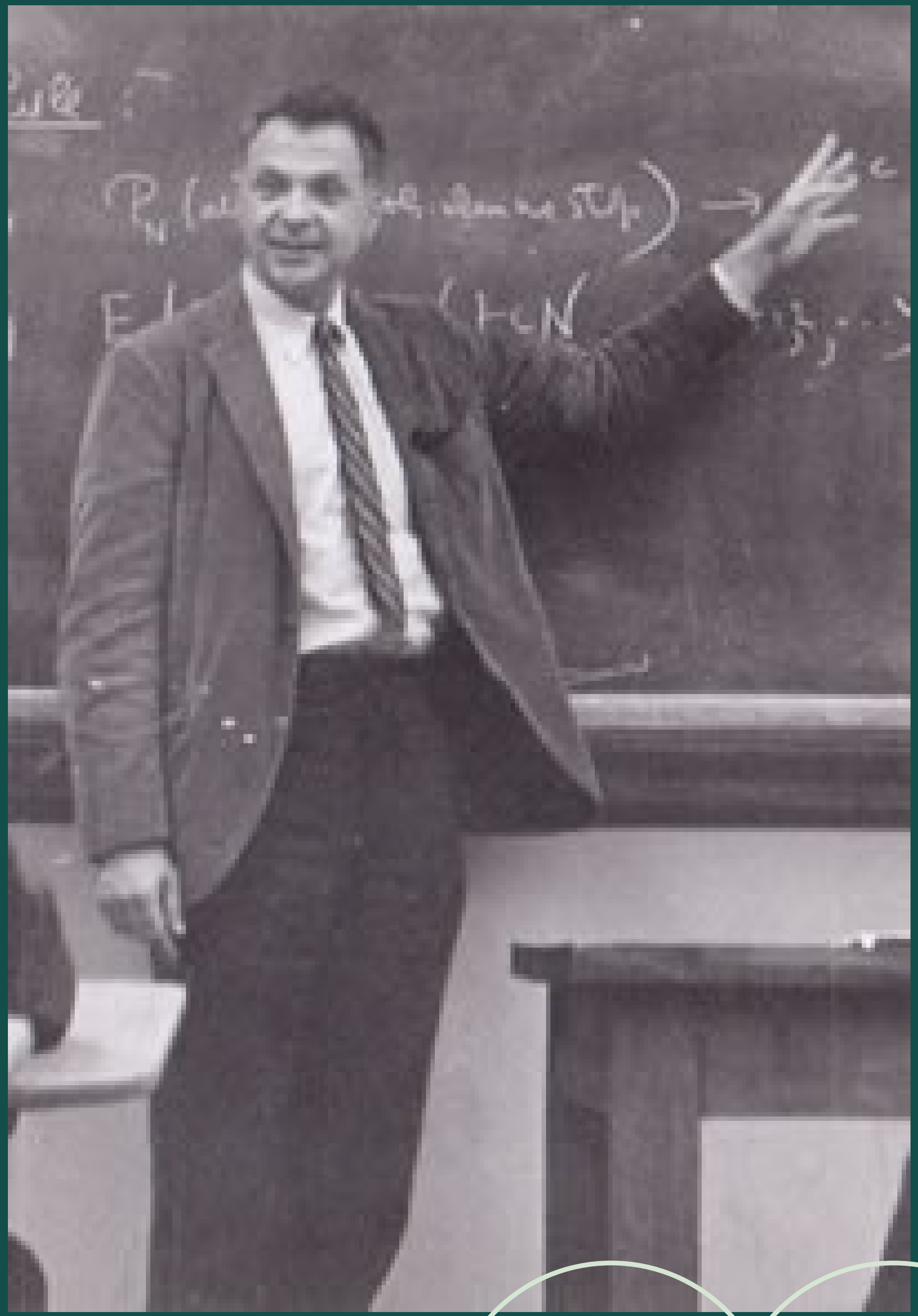
The Problem of Post-Hoc Planning and the Robbins Conjecture

The Robbins Conjecture (Continues)

- To prove the conjecture, we need to know little history.
- But the point is that it was necessary to prove that they satisfied another crucial Boolean property, such as *Idempotence*, or the *existence of a zero element*.

The Difficulty

- Robbins algebra is very simple.
- But the very simplicity made it harder to work with.
- It lacked the structural complexity of the Huntington axiom that might have provided a quicker path to proving the necessary equivalent properties.



Post-Hoc Planning

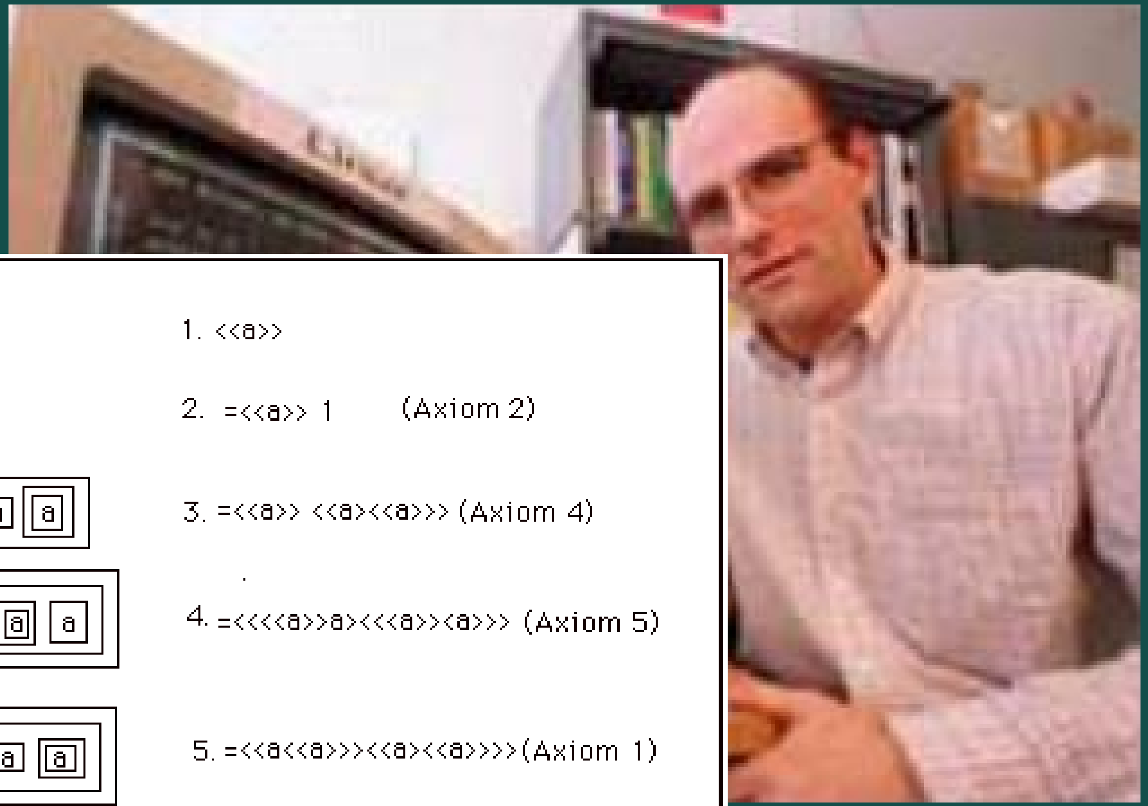
The Problem of Post-Hoc Planning and the Robbins Conjecture

McCune’s Proof of Robbins Conjecture [1997]

- The proof was found by the *automated theorem prover* EQP in 1996.
- It was extremely long, complex, and highly non-intuitive.
- The machine-generated proof was a sequence of low-level logical deductions that were not guided by any human-understandable mathematical "insight" or new theory.

The Difficulty

- EQP ran millions of derivations and found the proof — but no human-level plan preceded it!



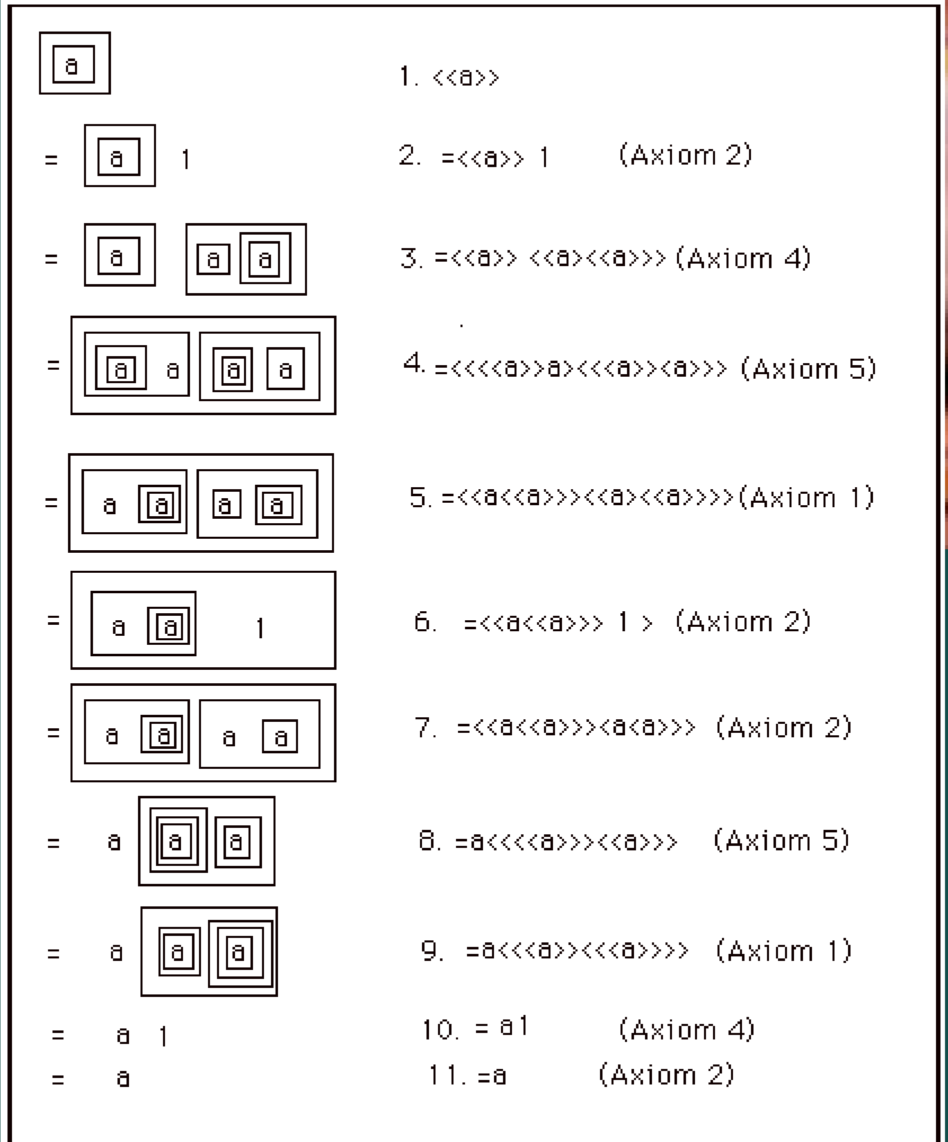


Figure 1. A Boolean Box Derivation

Post-Hoc Planning

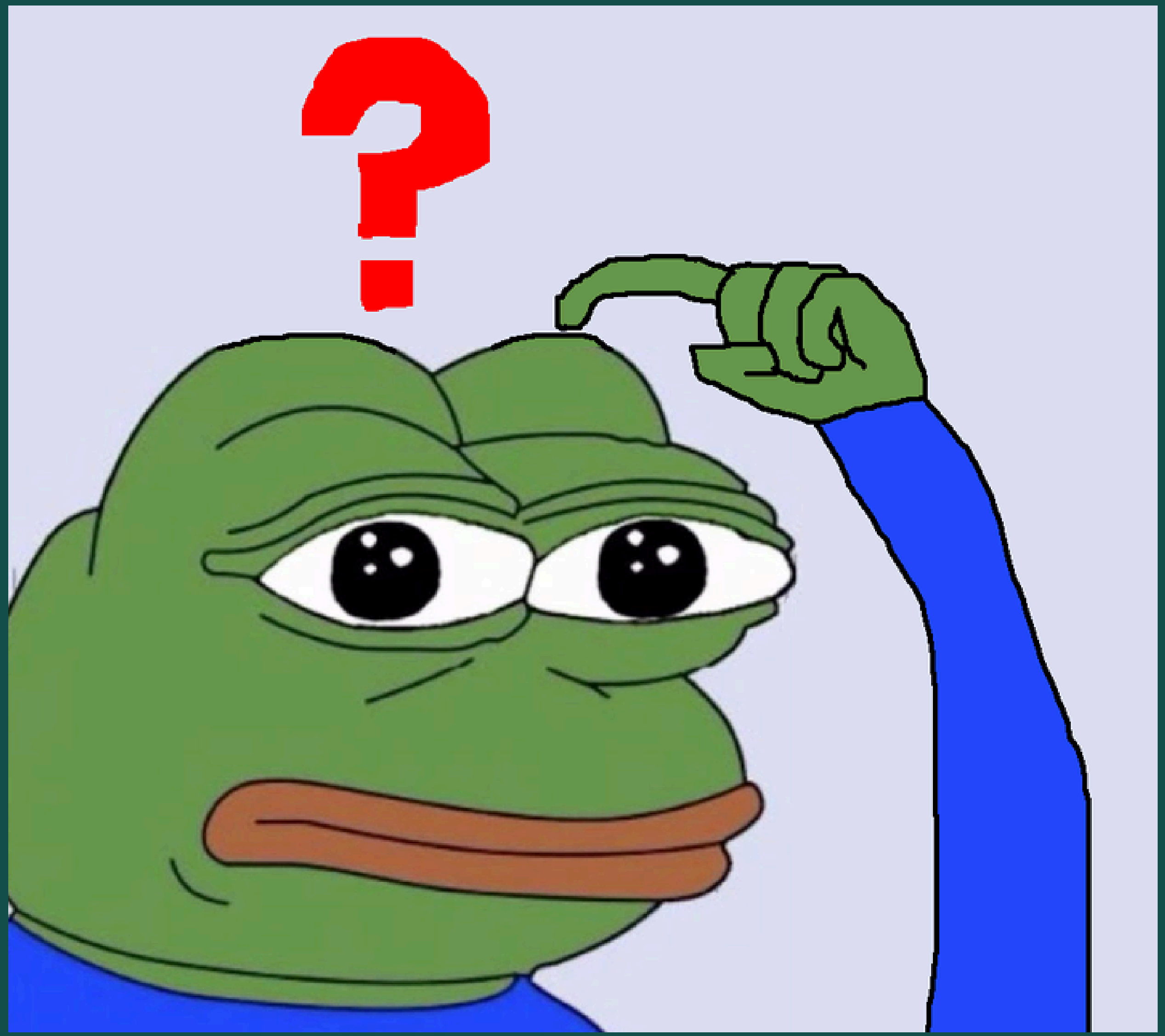
The Problem of Post-Hoc Planning and the Robbins Conjecture

Post-Hoc Reconstruction

- EQP's output: long, unstructured derivation.
- Mathematicians later extracted a conceptual rationale.
- The plan didn't guide the proof — it was reconstructed, and rationalised afterward.

The Difficulty

- Understanding became a *post-hoc rationalisation!*



Post-Hoc Planning

The Problem of Post-Hoc Planning and the Robbins Conjecture

The “As-If” Clause

- Hamami and Morris were well-aware of this post-hoc realisation issue:

> An assumption underlying our characterization is that, throughout the understanding process, the agent proceeds *as if* there is a rational plan underlying the proof P she aims to understand, i.e., *as if* P has been produced by a rational planning agent. In most cases, this is a genuine hypothesis

Cost to Pay

- The added *as if* clause can save the framework. But with a cost:

1. If every proof is as-if planned, the theory becomes *unfalsifiable*.
2. The real problem is the notion of “*proof activity*” itself.

Hierarchical Proof Architecture

Moving beyond *as-if*

Key Ideas

- Hamami & Morris: understanding = reconstructing the plan behind the proof.
- But many modern proofs lack a single antecedent plan.
- We need a model that captures partial intention + procedural realisation.

Hierarchical Proof Architecture

- Proofs as layered systems linking human policies to machine procedures.

Hierarchical Proof Architecture

Moving beyond *as-if*

From Unified Plan to Partial Plan

- Classical proofs: unified plan, deliberately designed by mathematicians from start to finish.
- Computational proofs: only high-level policy is planned.
- Subplans emerge through search and automation.

Hierarchical Proof Architecture

- Proofs as layered systems linking human policies to machine procedures.

Proof of Robbins Conjecture

(Human Policy)
↓
(Strategic Choice)
↓
EQP →(Heuristic/Tactical Process)
↓
(Computer Operation)

Intentional control fades as we descend the hierarchy.

Hierarchical Proof Architecture

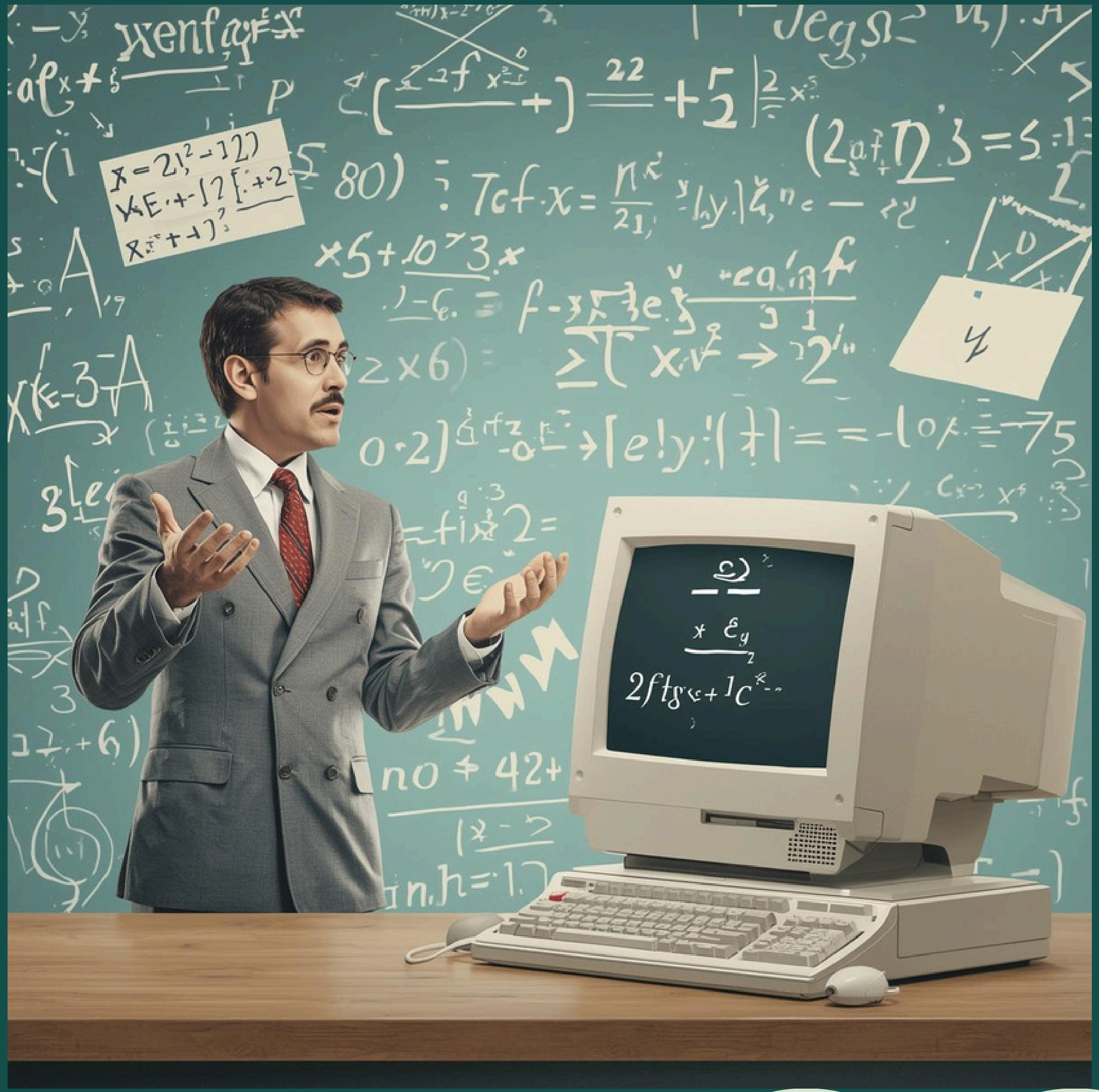
Moving beyond *as-if*

Why HPA?

- Proof activity is hierarchically organised:

1. Upper Layers: intentional policies (L_0–L_1)
2. Lower layers: procedural realisations (L_2–L_3)

- The epistemic unity of the proof comes from how realisations implement the policies that constrain them.

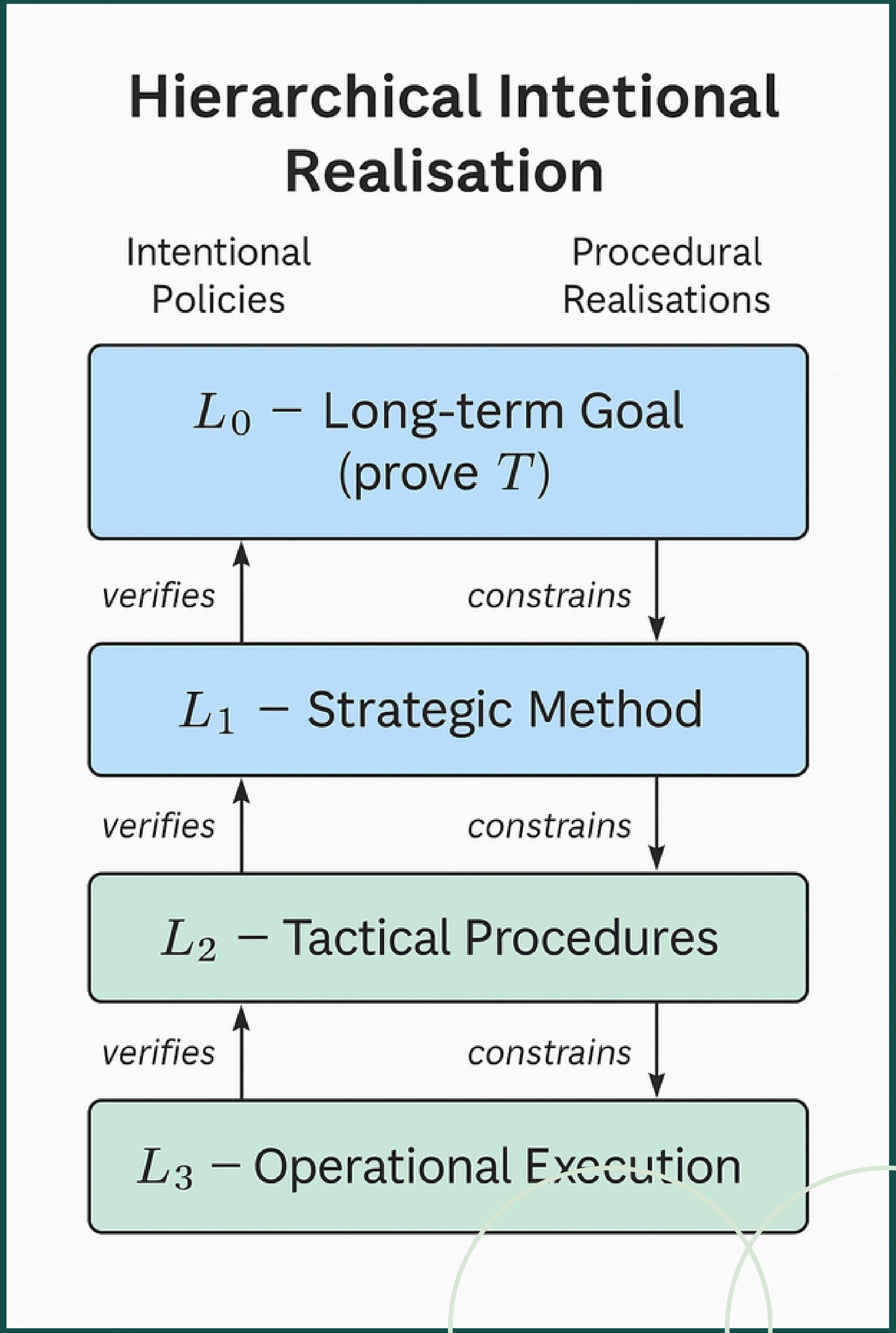


Hierarchical Proof Architecture

Moving beyond *as-if*

Four Layers of Proof Architecture

Layer	Role	Nature
L_0	Long-term goal ("prove T")	intentional policy
L_1	Strategic method (induction, reduction)	intentional policy
L_2	Lemmas, heuristics, representations	quasi-intentional procedure
L_3	Rewriting, substitution, verification	non-intentional procedure



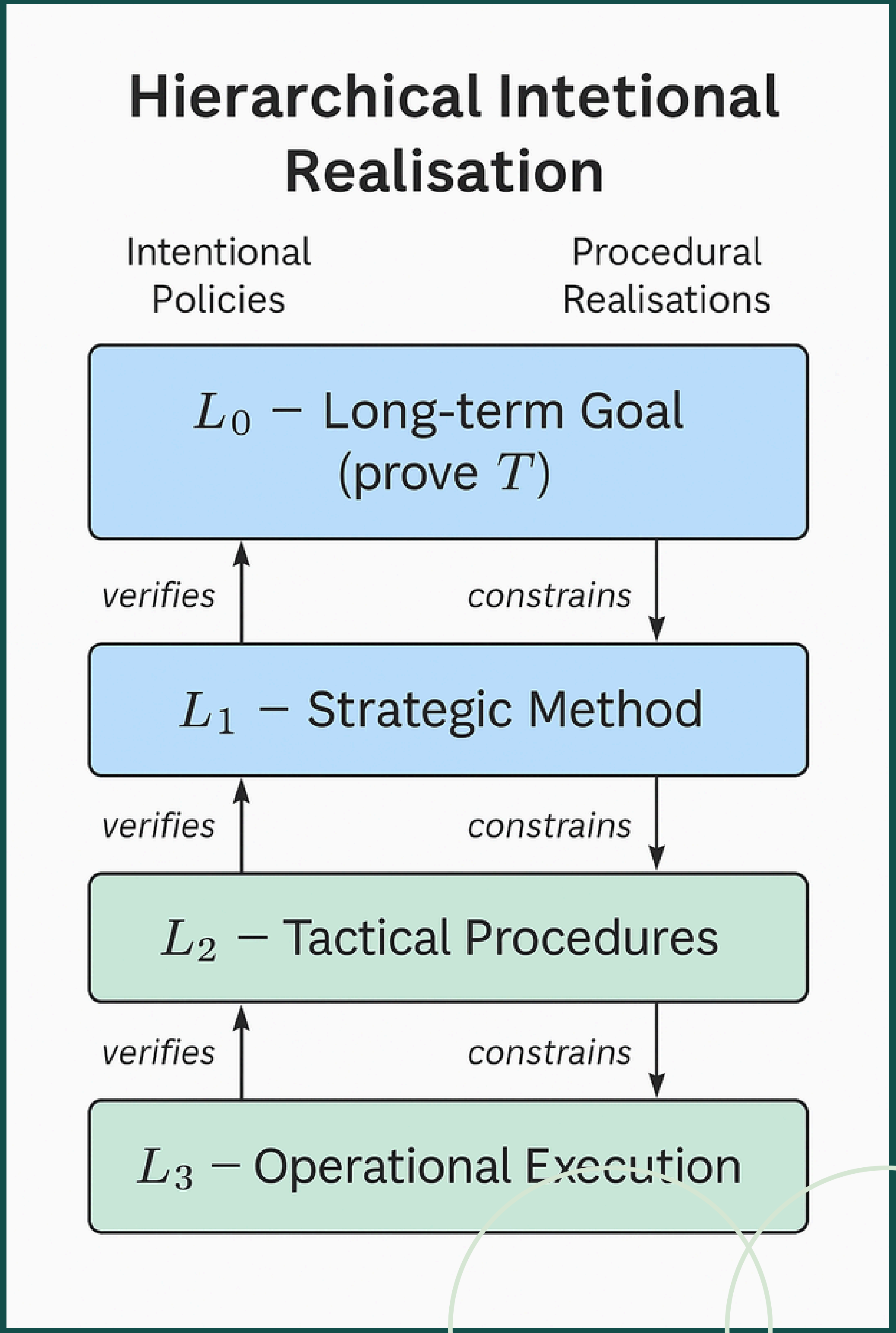
Hierarchical Proof Architecture

Understanding as Hierarchical Reconstruction

Understanding in HPA

$$\text{Understanding}(P) = \text{Reconstruction}(L_0, L_1) + \text{Interpretive Mapping}(L_2) + \text{Justified Trust}(L_3)$$

- Understanding = grasping how policies and procedures interlock.
- Human agents:
  - fully reconstruct L_0–L_1,
  - interpret L_2 (post-hoc mapping),
  - trust L_3 (certified verification).
- Asymmetrical but substantive: rational reconstruction + delegated trust.

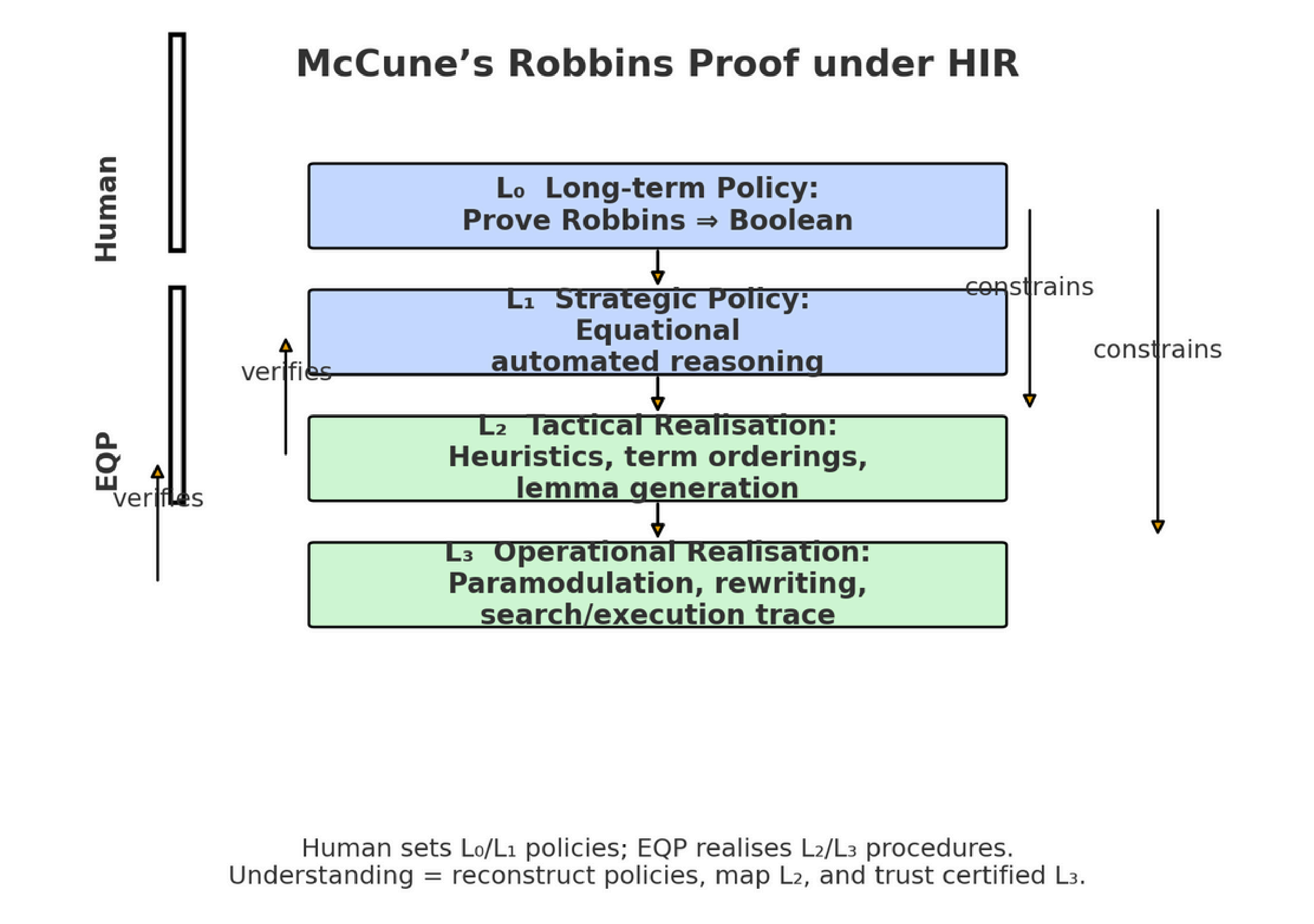


Hierarchical Proof Architecture

Case Study: The Robbins Conjecture

Use of EQP is the proof

- McCune [1997]: EQP automated theorem prover proved it.
- Humans → L_0 (formulate goal), L_1 (select equational reasoning).
- EQP → L_2 (search strategies), L_3 (paramodulation steps).
- Some EQP heuristics approximate meta-strategy → blur into lower L_1.



Hierarchical Proof Architecture

The Epistemic Picture of HPA

The Epistemic Picture

- Proofs are neither wholly rational nor wholly mechanical.
- Understanding = recognising intentional policies and procedural realisations fit.
- Epistemic trust replaces surveyability.
- Understanding is layered, distributed, mediated.

The Upshot

- Bridges normative rationality and actual computational practice.
- Work as diagnostic tools to map the computer's role.
- Clarifies why humans can still understand even when proof is unsurveyable.

Hierarchical Proof Architecture

The Final Takeaway

Proof as Activities

- Proof activities are organised by a hierarchical architecture that links intentional policies with procedural realisers.

Understanding Proof

- Understanding = rational reconstruction of policies + justified reliance on their procedural realisations.

Proofs are no longer merely texts to read — they are architectures to reconstruct.

Thank you for listening!

Email
mail@mincheolseo.com

Website
mincheolseo.com