Proof systems

Robert J. Carroll

Last revised: 5 May 2026

---

1 Motivation

The propositional-logic handout left an open promise. Logical consequence was defined semantically: $\Gamma \models \varphi$ holds when every valuation making all of Γ true also makes φ true. That definition is conceptually clean, but it is not a procedure. Given a stack of premises and a putative conclusion, "check every valuation" is finite for propositional logic but exponential in the number of atoms, and as soon as we move to first-order logic in the next handout the analogous semantic check becomes uncomputable. What we want alongside the semantics is a *proof system*: a finite, mechanical apparatus of inference rules that lets us *derive* the conclusion from the premises step by step. The handout works through one such system in detail (natural deduction), gestures at two others (sequent calculus and Hilbert-style), and then proves the two-way bridge that ties the syntactic and the semantic together: *soundness* (every derivable consequence is a semantic consequence) and *completeness* (every semantic consequence is derivable).

The bridge is what makes the formalism trustworthy. Soundness says the derivation-rules are sane: you cannot derive falsehoods from truths. Completeness says they are powerful enough: every claim that is semantically valid can in principle be derived by these rules. Together they justify treating $\models$ and $\vdash$ as interchangeable for propositional logic, which is what working theorists actually do — they argue informally about validity using whichever notion is easier in the moment, with the soundness–completeness theorem in the background warranting the move.

We work entirely in propositional logic in this handout. The same machinery extends to first-order logic, where the analogous theorem (Gödel's completeness theorem of 1930) is more substantive but follows the same pattern. The first-order case is in the next handout.

2 Natural deduction

When you and a colleague argue informally about whether some chain of reasoning is valid, you typically work step by step: introduce an assumption, conjoin two premises, draw a conclusion under a hypothetical, eventually discharge the hypothetical and report the result. Natural deduction is the formalism that captures exactly this informal style. It was introduced by Gerhard Gentzen in 1935 with the explicit goal of looking like the way mathematicians actually reason. The rules come in pairs — an *introduction* rule for each connective, saying how a formula with that connective is built up, and an *elimination* rule, saying what can be inferred from such a formula — and a derivation is a sequence of formulas, each justified either as a premise, an assumption, or by application of a rule to earlier lines.

The rules are summarized below in tree style, grouped by connective: premises stacked above a horizontal line, conclusion below, label to the right.

Conjunction.

$$\frac{\varphi \qquad \psi}{\varphi \wedge \psi}\,\wedge\mathrm{I} \qquad\qquad \frac{\varphi \wedge \psi}{\varphi}\,\wedge\mathrm{E} \qquad\qquad \frac{\varphi \wedge \psi}{\psi}\,\wedge\mathrm{E}$$

Disjunction.

$$\frac{\varphi}{\varphi \vee \psi}\,\vee\mathrm{I} \qquad\qquad \frac{\psi}{\varphi \vee \psi}\,\vee\mathrm{I} \qquad\qquad \frac{\varphi \vee \psi \qquad \begin{array}{c}[\varphi]^1\\ \vdots\\ \chi\end{array} \qquad \begin{array}{c}[\psi]^2\\ \vdots\\ \chi\end{array}}{\chi}\,\vee\mathrm{E}^{1,2}$$

Implication.

$$\frac{\begin{array}{c}[\varphi]^1\\ \vdots\\ \psi\end{array}}{\varphi \to \psi}\,\to\mathrm{I}^{1} \qquad\qquad \frac{\varphi \qquad \varphi \to \psi}{\psi}\,\to\mathrm{E}\ (\mathrm{MP})$$

Negation and absurdity.

$$\frac{\begin{array}{c}[\varphi]^1\\ \vdots\\ \bot\end{array}}{\neg\varphi}\,\neg\mathrm{I}^{1} \qquad\qquad \frac{\varphi \qquad \neg\varphi}{\bot}\,\neg\mathrm{E} \qquad\qquad \frac{\bot}{\varphi}\,\bot\mathrm{E}$$

Reductio ad absurdum.

$$\frac{\begin{array}{c}[\neg\varphi]^1\\ \vdots\\ \bot\end{array}}{\varphi}\,\mathrm{RAA}^{1}$$

The bracket-and-superscript notation $[\varphi]^1$ marks an assumption introduced for the purpose of a subproof; the matching superscript on the rule label indicates which assumption is being *discharged* when the rule fires. The vertical dots between the bracketed assumption and the conclusion stand for "a subproof goes here" — zero or more intermediate steps connecting the top of the column to the bottom.

We write $\Gamma \vdash \varphi$ to mean that there is a derivation ending in φ whose undischarged assumptions are all in Γ.

A worked derivation, in step-by-step format. *Modus tollens*: $\varphi \to \psi, \neg\psi \vdash \neg\varphi$.

1.	$\varphi \to \psi$	Premise
2.	$\neg\psi$	Premise
3.	$\quad [\varphi]^a$	Assumption
4.	$\quad \psi$	$\to$E (1, 3)
5.	$\quad \bot$	$\neg$E (2, 4)
6.	$\neg\varphi$	$\neg\mathrm{I}^a$ (3–5)

The reasoning, in English: assume φ for the sake of argument (line 3); then by modus ponens with the first premise we get ψ (line 4); but the second premise is $\neg\psi$, so we have a contradiction (line 5); discharge the assumption to conclude $\neg\varphi$ (line 6). This is exactly how one reasons informally; natural deduction just makes the bookkeeping explicit.

Example 1 (A derivation with bicameralism)**.** Take the simplified bicameralism formula $L \leftrightarrow (H \wedge S)$ from the propositional-logic handout, which we expand to its two implication directions: $L \to (H \wedge S)$ and $(H \wedge S) \to L$. From the first direction together with the observation L (the bill became law), we can derive $H \wedge S$ in three lines: 1. $L \to (H \wedge S)$ (Premise); 2. L (Premise); 3. $H \wedge S$ ($\to$E, 1, 2). Conjunction-elimination then extracts H or S as desired. This is the formal version of the (correct) inference: "the bill became law; bicameralism requires that both chambers passed it; therefore both chambers passed it."

The rules above are sufficient for all of classical propositional logic. (A version of the system that includes only the introduction and elimination rules and *not* RAA is intuitionistic propositional logic, in which not every classical tautology is derivable. The differences become substantive in first-order logic and beyond, but at the level of propositional logic RAA is the difference, and the standard mathematical practice is to keep it.)

3 Sequent calculus and Hilbert systems

There is no canonical proof system for propositional logic. Several different formalisms exist, all proven equivalent in derivability ($\Gamma \vdash \varphi$ holds in one if and only if it holds in the others), but each one foregrounds different features and has different strengths in practice. We mention two in passing.

Sequent calculus, also due to Gentzen, takes *sequents* $\Gamma \Rightarrow \Delta$ as its primitive objects — "from the conjunction of Γ one can derive the disjunction of Δ" — with rules that operate on the sequent's left side, right side, or both. Its theoretical importance is the *cut-elimination* theorem: every derivation can be rewritten so that no formula appears on the right of one sequent and the left of the next (no "cut" on a lemma). Cut-elimination is the workhorse of structural proof theory and underwrites several deep results about proof complexity, but for everyday derivations natural deduction is more ergonomic.

Hilbert-style systems (sometimes called *axiomatic systems*) take a small list of axioms together with a single inference rule, modus ponens. The standard set has three axiom schemes: $\varphi \to (\psi \to \varphi)$; $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$; $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$. Hilbert systems are the most parsimonious of the three: minimal apparatus, easy to analyze. They are also the most awkward in which to actually prove anything: derivations of even simple tautologies often run dozens of lines. Their primary modern use is theoretical, including in the original soundness and completeness proofs.

We will not develop sequents or Hilbert systems further. The point is that the proof-system landscape is wider than "natural deduction or nothing," and the soundness–completeness results below hold for any of these formalisms (one proves them once, then transfers across by inter-derivability).[1]

[1] The relationship between the three is closer than the surface presentation suggests. Hilbert systems can be viewed as natural deduction with all the rules pushed into the assumption set as axioms; sequent calculus can be viewed as natural deduction with the assumption-and-discharge bookkeeping flattened out into a left/right structure. The three formalisms are not just *equivalent in derivability* but *translations of each other*, in a precise sense made rigorous

4 Soundness

Soundness is the easy direction of the bridge between $\vdash$ and $\models$, and the conceptually clarifying one. "The proof rules are sane" is the informal claim — you can derive falsehoods only from falsehoods, not from truths. Spelled out:

Theorem 2 (Soundness of natural deduction for propositional logic)**.** *If* $\Gamma \vdash \varphi$*, then* $\Gamma \models \varphi$*.*

Proof sketch. By induction on the structure (equivalently, the length) of a derivation. The base case is when φ is a premise or an assumption that has not been discharged at the end of the derivation: then $\varphi \in \Gamma$ trivially, and any valuation satisfying Γ also satisfies φ. The inductive step is to check that each rule preserves the property "every valuation that makes the (undischarged) assumptions true makes the conclusion true." For $\wedge$I, if v satisfies the assumptions of both subderivations, $v(\varphi) = T$ and $v(\psi) = T$, hence $v(\varphi \wedge \psi) = T$. For $\to$I, if v satisfies the assumptions of the outer derivation, then in any sub-context where φ is additionally assumed, the conclusion ψ is true; hence $v(\varphi \to \psi) = T$ (the only way it could be false is $v(\varphi) = T, v(\psi) = F$, which the subderivation rules out). The remaining rules are similarly direct. RAA is the slightly subtle case: if the subderivation from $\neg\varphi$ ends in $\bot$, then $v(\neg\varphi) = T$ would force a valuation in which $\bot$ holds, which is impossible; so $v(\varphi) = T$. The rule is sound classically; intuitionistically it would need more care. □

5 Completeness

Completeness is the harder direction and the deeper theorem: *every* semantically valid consequence is derivable by the rules. The semantic side is doing infinitely much work — ranging over all valuations — but the syntactic side is finitary — a derivation is a finite tree of finite formulas with finitely many rule applications. That a finite combinatorial procedure can capture an infinitary semantic notion is non-trivial.

Theorem 3 (Completeness of natural deduction for propositional logic)**.** *If* $\Gamma \models \varphi$*, then* $\Gamma \vdash \varphi$*.*

The standard proof uses two ingredients: the notion of a *consistent* set of formulas, and *Lindenbaum's lemma* on extending consistent sets.

A set Γ of formulas is *(syntactically) consistent* if $\Gamma \nvdash \bot$. (Equivalently, by RAA: there is no formula ψ such that $\Gamma \vdash \psi$ and $\Gamma \vdash \neg\psi$.) It is *maximal consistent* if it is consistent and no strict superset is consistent.

Lemma 4 (Lindenbaum)**.** *Every consistent set of formulas* Γ *extends to a maximal consistent set* Γ^* *with* $\Gamma \subseteq \Gamma^*$*.*

Proof sketch. Enumerate all formulas $\varphi_1, \varphi_2, \varphi_3, \ldots$ (the language is countable). Define $\Gamma_0 := \Gamma$, and at stage n, set $\Gamma_n := \Gamma_{n-1} \cup \{\varphi_n\}$ if that union is consistent, and $\Gamma_n := \Gamma_{n-1}$ otherwise. Set

by the Curry–Howard correspondence (which connects natural-deduction proofs to typed lambda-calculus terms) and its extensions to sequent calculus. For working applications, the choice of system is mostly a matter of which rules are most convenient. For automated reasoning, sequents are dominant (cleaner structure for search algorithms); for human reasoning, natural deduction is the native idiom; for theoretical proof analysis, Hilbert systems are still occasionally the right choice. Enderton (2001) gives a careful Hilbert-style treatment; Dalen (2013) works through natural deduction.

$\Gamma^* := \bigcup_n \Gamma_n$. Three checks: Γ^* is consistent (any inconsistency would arise at some finite stage); Γ^* is maximal (every φ_n was either added or its addition would have been inconsistent); $\Gamma \subseteq \Gamma^*$ (immediate). □

The maximal consistent set has two key features. First, it is *deductively closed*: $\Gamma^* \vdash \psi$ implies $\psi \in \Gamma^*$. (If $\psi \notin \Gamma^*$, maximality fails when we try to add ψ: $\Gamma^* \cup \{\psi\}$ is inconsistent, so $\Gamma^* \vdash \neg\psi$ by RAA, which together with $\Gamma^* \vdash \psi$ gives $\Gamma^* \vdash \bot$, contradicting consistency.) Second, it is *prime* for the connectives: $\neg\psi \in \Gamma^*$ iff $\psi \notin \Gamma^*$; $\psi \wedge \chi \in \Gamma^*$ iff both are; $\psi \vee \chi \in \Gamma^*$ iff at least one is; $\psi \to \chi \in \Gamma^*$ iff $\psi \notin \Gamma^*$ or $\chi \in \Gamma^*$.

These two features let us define a valuation directly from the maximal consistent set:

Lemma 5 (Canonical valuation)**.** *Let Γ^* be maximal consistent. Define $v : \mathcal{P} \to \{T, F\}$ by $v(p) = T$ iff $p \in \Gamma^*$. Then for every formula ψ, $v(\psi) = T$ iff $\psi \in \Gamma^*$.*

Proof sketch. Induction on the structure of ψ. The atomic case is the definition. The inductive cases are the four prime-for-connective bullets above; each says exactly that the truth-functional definition of $v(\psi)$ matches membership in Γ^*. □

The two lemmas together prove completeness:

Proof of Theorem 3. We prove the contrapositive: if $\Gamma \nvdash \varphi$, then $\Gamma \nvDash \varphi$. Suppose $\Gamma \nvdash \varphi$. Then $\Gamma \cup \{\neg\varphi\}$ is consistent (otherwise $\Gamma \vdash \varphi$ by RAA), so by Lindenbaum it extends to a maximal consistent set Γ^* containing $\Gamma \cup \{\neg\varphi\}$. The canonical valuation v from Lemma 5 satisfies every formula in Γ^*, in particular every formula in Γ and the formula $\neg\varphi$. So v is a valuation that satisfies Γ but not φ, witnessing $\Gamma \nvDash \varphi$. □

The structure of the argument is the canonical-model construction: take a syntactic object (a maximal consistent set) and build a semantic object (a valuation, here) from it. The same template extends to first-order logic, where it produces *Henkin's proof* of Gödel's completeness theorem (1949) — the canonical structure built from a maximal consistent set is the model whose existence the completeness theorem asserts. We will see this in the next handout.[2]

6 What's next

The next handout extends all of this to *first-order logic*, where the same soundness–completeness pattern holds (Gödel 1930) but the canonical-model construction is more elaborate and the consequences considerably richer. The first-order completeness theorem is what underwrites the practice of proving theorems by "constructing a model" in (e.g.) algebra, set theory, and abstract economic theory — a syntactically consistent theory has a model, full stop. It also underwrites compactness

[2]The completeness theorem has a remarkable corollary that one cannot prove without it: *compactness*. A set of formulas is *satisfiable* if some valuation makes them all true; the compactness theorem says that an infinite set of formulas is satisfiable if and only if every finite subset is. The forward direction is trivial. The reverse uses completeness: if every finite subset is satisfiable, then no finite subset entails $\bot$, so by completeness no finite subset *derives* $\bot$ either; since derivations are finite, the whole set fails to derive $\bot$, hence is consistent, hence (by Lindenbaum and the canonical valuation) is satisfiable. Compactness is the bridge between finiteness on the syntactic side and a non-finite property on the semantic side, and it has substantial mathematical consequences in first-order logic — for example, it underlies the existence of non-standard models of arithmetic and of nonstandard analysis.

in first-order logic, which is the foundation of the model-theoretic apparatus we will use in the FOL handout.

The proof-systems story has a separate continuation in computability theory and proof complexity. *What can a proof system efficiently prove?* is the entry point to a substantial area of theoretical computer science, with connections to the P versus NP problem (since determining propositional satisfiability is NP-complete, and proving unsatisfiability efficiently in any standard proof system is widely believed to be infeasible). For applied work the takeaway is small but real: "checking whether $\Gamma \models \varphi$" is computationally hard in general, so even when soundness and completeness give us *some* decision procedure, finding a *fast* one for any given application is a separate question.

For deeper treatments, see Enderton (2001) (Hilbert-style; clean and rigorous), Dalen (2013) (natural deduction; closer to the modern logician's working idiom), or Troelstra and Schwichtenberg (2000) for proof theory proper.[3]

7 Exercises

Exercise 6. Give a natural-deduction derivation of $\varphi \wedge \psi \vdash \psi \wedge \varphi$.

Exercise 7. Give a natural-deduction derivation of $\vdash \varphi \to (\psi \to \varphi)$. (Notice that no premises are needed: the formula is a tautology.)

Exercise 8. Give a natural-deduction derivation of $\varphi \to \psi, \psi \to \chi \vdash \varphi \to \chi$ (hypothetical syllogism).

Exercise 9. Give a natural-deduction derivation of $\vdash (\varphi \wedge \psi) \to (\varphi \vee \psi)$.

Exercise 10. Show that $\vdash$ is monotonic: if $\Gamma \vdash \varphi$ and $\Gamma \subseteq \Gamma'$, then $\Gamma' \vdash \varphi$. Argue from the definition of derivation, not from soundness and completeness. (The point is that $\vdash$ has a syntactic monotonicity that does not require the semantic detour.)

Exercise 11 (Article V as a derivation)**.** Recall the Article V exercise from the propositional-logic handout: $A \leftrightarrow ((C \vee N) \wedge R)$, where A stands for "the amendment is enacted," C for "two-thirds of each chamber proposes it," N for "two-thirds of state legislatures call a convention proposing it," and R for "three-fourths of state legislatures ratify it." From the rule together with A, give a natural-deduction derivation of R ("ratification happened") and a separate derivation of $C \vee N$ ("one of the two proposal routes was used"). The political content is: enactment licenses both inferences.

Exercise 12. Show that the empty set is consistent: $\emptyset \nvdash \bot$. (*Hint*: use soundness rather than analyzing all possible derivations from no premises.)

[3] A practical aside on the role of formal proof systems in working mathematics. Mathematicians rarely write down formal natural-deduction (or sequent, or Hilbert-style) proofs. They argue in *informal mathematical English* — the metalanguage flagged in the project's preface — with the formal apparatus serving as the warrant in the background that the informal argument *could* in principle be made formal. The justification of the practice is the soundness–completeness theorem itself: the formal system is provably sufficient, so the informal argument is provably reducible (in principle) to a formal derivation. The recent development of *interactive theorem provers* — Coq, Lean, Isabelle/HOL, Agda — is making the "in principle" more practical: substantial chunks of mathematics (the Feit–Thompson theorem, the proof of the Kepler conjecture, large parts of the Liquid Tensor Experiment) have now been fully formalized and machine-checked. The political-economy applications are still nascent but include formal verification of strategic-equilibrium computations and machine-checked proofs of impossibility theorems in social choice. The gap between "everyday math" and "machine-formalized math" is shrinking, but it is real, and the proof-system theory developed in this handout is what underwrites both sides.

Exercise 13 (Compactness from completeness)**.** Show: if every finite subset of a set Γ of propositional formulas is satisfiable, then Γ itself is satisfiable. (Argue by contradiction. If Γ is unsatisfiable, then $\Gamma \models \bot$, so by completeness $\Gamma \vdash \bot$. Now use that derivations are finite.)

Exercise 14. Sketch in your own words the canonical-model argument for the completeness theorem: from $\Gamma \nvdash \varphi$, construct a valuation that satisfies Γ but not φ. Identify which steps depend on the language being countable and which steps would need adjustment if it were uncountable.

Exercise 15 (Soundness as a sanity check)**.** Suppose someone claims to have a derivation of $p \vdash q$, where p and q are distinct atomic propositions. Without examining the derivation, explain why the claim must be wrong.

References

Dalen, Dirk van (2013). *Logic and Structure*. 5th ed. London: Springer.

Enderton, Herbert B. (2001). *A Mathematical Introduction to Logic*. 2nd ed. San Diego: Academic Press.

Troelstra, A. S. and H. Schwichtenberg (2000). *Basic Proof Theory*. 2nd ed. Cambridge: Cambridge University Press.