A derivation system, such as natural deduction, is *sound* if it cannot derive things that do not actually follow. Soundness is thus a kind of guaranteed safety property for derivation systems. Depending on which proof theoretic property is in question, we would like to know for instance, that

1. every derivable sentence is valid;

2. if a sentence is derivable from some others, it is also a consequence of them;

3. if a set of sentences is inconsistent, it is unsatisfiable.

These are important properties of a derivation system. If any of them do not hold, the derivation system is deficient—it would derive too much. Consequently, establishing the soundness of a derivation system is of the utmost importance.

**Theorem ntd.1 (Soundness).** If \( \varphi \) is derivable from the undischarged assumptions \( \Gamma \), then \( \Gamma \models \varphi \).

**Proof.** Let \( \delta \) be a derivation of \( \varphi \). We proceed by induction on the number of inferences in \( \delta \).

For the induction basis we show the claim if the number of inferences is 0. In this case, \( \delta \) consists only of a single sentence \( \varphi \), i.e., an assumption. That assumption is undischarged, since assumptions can only be discharged by inferences, and there are no inferences. So, any structure \( \mathcal{M} \) that satisfies all of the undischarged assumptions of the proof also satisfies \( \varphi \).

Now for the inductive step. Suppose that \( \delta \) contains \( n \) inferences. The premise(s) of the lowermost inference are derived using sub-derivations, each of which contains fewer than \( n \) inferences. We assume the induction hypothesis: The premises of the lowermost inference follow from the undischarged assumptions of the sub-derivations ending in those premises. We have to show that the conclusion \( \varphi \) follows from the undischarged assumptions of the entire proof.

We distinguish cases according to the type of the lowermost inference. First, we consider the possible inferences with only one premise.

1. Suppose that the last inference is \( \neg \)Intro: The derivation has the form

\[
\Gamma, [\varphi]^n \\
\vdots \\
\delta_1 \\
\vdots \\
\frac{\vdots}{n \quad \frac{\vdots}{\neg \varphi} \neg \text{ Intro}}
\]
By inductive hypothesis, \( \bot \) follows from the undischarged assumptions \( \Gamma \cup \{ \varphi \} \) of \( \delta_1 \). Consider a structure \( M \). We need to show that, if \( M \models \Gamma \), then \( M \not\models \varphi \). Suppose for reductio that \( M \models \Gamma \), but \( M \not\models \varphi \), i.e., \( M \models \varphi \). This would mean that \( M \models \Gamma \cup \{ \varphi \} \). This is contrary to our inductive hypothesis. So, \( M \models \neg \varphi \).

2. The last inference is \( \land \text{Elim} \): There are two variants: \( \varphi \) or \( \psi \) may be inferred from the premise \( \varphi \land \psi \). Consider the first case. The derivation \( \delta \) looks like this:

\[
\begin{array}{c}
\Gamma \\
\vdots \\
\delta_1 \\
\vdots \\
\varphi \land \psi \\
\varphi \quad \land \text{Elim}
\end{array}
\]

By inductive hypothesis, \( \varphi \land \psi \) follows from the undischarged assumptions \( \Gamma \) of \( \delta_1 \). Consider a structure \( M \). We need to show that, if \( M \models \Gamma \), then \( M \not\models \varphi \land \psi \). Suppose \( M \models \Gamma \); by our inductive hypothesis \( (\Gamma \models \varphi \land \psi) \), we know that \( M \models \varphi \land \psi \). By definition, \( M \models \varphi \land \psi \) iff \( M \models \varphi \) and \( M \models \psi \). (The case where \( \psi \) is inferred from \( \varphi \land \psi \) is handled similarly.)

3. The last inference is \( \lor \text{Intro} \): There are two variants: \( \varphi \lor \psi \) may be inferred from the premise \( \varphi \) or the premise \( \psi \). Consider the first case. The derivation has the form

\[
\begin{array}{c}
\Gamma \\
\vdots \\
\delta_1 \\
\vdots \\
\varphi \\
\varphi \lor \psi \quad \lor \text{Intro}
\end{array}
\]

By inductive hypothesis, \( \varphi \) follows from the undischarged assumptions \( \Gamma \) of \( \delta_1 \). Consider a structure \( M \). We need to show that, if \( M \models \Gamma \), then \( M \not\models \varphi \lor \psi \). Suppose \( M \models \Gamma \); then \( M \models \varphi \lor \psi \). Suppose \( M \models \varphi \lor \psi \); then \( M \models \varphi \) since \( \Gamma \models \varphi \) (the inductive hypothesis). So it must also be the case that \( M \models \varphi \lor \psi \). (The case where \( \varphi \lor \psi \) is inferred from \( \varphi \land \psi \) is handled similarly.)

4. The last inference is \( \to \text{Intro} \): \( \varphi \to \psi \) is inferred from a subproof with assumption \( \varphi \) and conclusion \( \psi \), i.e.,

\[
\begin{array}{c}
\Gamma, [\varphi]^n \\
\vdots \\
\delta_1 \\
\vdots \\
\varphi \to \psi \\
\varphi \to \psi \quad \to \text{Intro}
\end{array}
\]
By inductive hypothesis, \( \psi \) follows from the undischarged assumptions of \( \delta_1 \), i.e., \( \Gamma \cup \{ \varphi \} \models \psi \). Consider a structure \( \mathcal{M} \). The undischarged assumptions of \( \delta \) are just \( \Gamma \), since \( \varphi \) is discharged at the last inference. So we need to show that \( \Gamma \models \varphi \rightarrow \psi \). For reductio, suppose that for some structure \( \mathcal{M} \), \( \mathcal{M} \models \Gamma \) but \( \mathcal{M} \not\models \varphi \rightarrow \psi \). So, \( \mathcal{M} \models \varphi \) and \( \mathcal{M} \not\models \psi \). But by hypothesis, \( \psi \) is a consequence of \( \Gamma \cup \{ \varphi \} \), i.e., \( \mathcal{M} \models \psi \), which is a contradiction. So, \( \Gamma \models \varphi \rightarrow \psi \).

5. The last inference is \( \bot \)I: Here, \( \delta \) ends in

\[
\begin{array}{c}
\Gamma \\
\vdots \\
\delta_1 \\
\vdots \\
\bot
\end{array}
\]

By induction hypothesis, \( \Gamma \models \bot \). We have to show that \( \Gamma \models \varphi \). Suppose not; then for some \( \mathcal{M} \) we have \( \mathcal{M} \models \Gamma \) and \( \mathcal{M} \not\models \varphi \). But we always have \( \mathcal{M} \not\models \bot \), so this would mean that \( \Gamma \not\models \bot \), contrary to the induction hypothesis.

6. The last inference is \( \bot \)C: Exercise.

7. The last inference is \( \forall \)Intro: Then \( \delta \) has the form

\[
\begin{array}{c}
\Gamma \\
\vdots \\
\delta_1 \\
\vdots \\
\varphi(a)
\end{array}
\]

\( \forall \)Intro

The premise \( \varphi(a) \) is a consequence of the undischarged assumptions \( \Gamma \) by induction hypothesis. Consider some structure, \( \mathcal{M} \), such that \( \mathcal{M} \models \Gamma \). We need to show that \( \mathcal{M} \models \forall x \varphi(x) \). Since \( \forall x \varphi(x) \) is a sentence, this means we have to show that for every variable assignment \( s \), \( \mathcal{M}, s \models \varphi(x) \) (??). Since \( \Gamma \) consists entirely of sentences, \( \mathcal{M}, s \models \psi \) for all \( \psi \in \Gamma \) by ??.

Let \( \mathcal{M}' \) be like \( \mathcal{M} \) except that \( a^{\mathcal{M}'} = s(x) \). Since \( a \) does not occur in \( \Gamma \), \( \mathcal{M}' \models \Gamma \) by ??.

Since \( \mathcal{M} \models \varphi(a) \), \( \mathcal{M}' \models \varphi(a) \). Since \( \varphi(a) \) is a sentence, \( \mathcal{M}', s \models \varphi(a) \) by ??.

\( \mathcal{M}', s \models \varphi(x) \) iff \( \mathcal{M}' \models \varphi(a) \) by ?? (recall that \( \varphi(a) \) is just \( \varphi(x)[a/x] \)). So, \( \mathcal{M}', s \models \varphi(x) \). Since \( a \) does not occur in \( \varphi(x) \), by ??, \( \mathcal{M}, s \models \varphi(x) \). But \( s \) was an arbitrary variable assignment, so \( \mathcal{M} \models \forall x \varphi(x) \).

8. The last inference is \( \exists \)Intro: Exercise.

9. The last inference is \( \forall \)Elim: Exercise.
Now let’s consider the possible inferences with several premises: ∨Elim, ∧Intro, →Elim, and ∃Elim.

1. The last inference is ∧Intro. φ ∧ ψ is inferred from the premises φ and ψ and δ has the form

\[ \begin{align*} & \Gamma_1 \quad \Gamma_2 \\ & \vdots \quad \vdots \\ & \delta_1 \quad \delta_2 \\ & \phi \quad \psi \\ & \hline \end{align*} \]

\[ φ \land ψ \land \text{Intro} \]

By induction hypothesis, φ follows from the undischarged assumptions Γ₁ of δ₁ and ψ follows from the undischarged assumptions Γ₂ of δ₂. The undischarged assumptions of δ are Γ₁ ∪ Γ₂, so we have to show that Γ₁ ∪ Γ₂ ⊨ φ ∧ ψ. Consider a structure M with M ⊨ Γ₁ ∪ Γ₂. Since M ⊨ Γ₁, it must be the case that M ⊨ φ as Γ₁ ⊨ φ, and since M ⊨ Γ₂, M ⊨ ψ since Γ₂ ⊨ ψ. Together, M ⊨ φ ∧ ψ.

2. The last inference is ∨Elim: Exercise.

3. The last inference is →Elim. ψ is inferred from the premises φ → ψ and φ. The derivation δ looks like this:

\[ \begin{align*} & \Gamma_1 \quad \Gamma_2 \\ & \vdots \quad \vdots \\ & \delta_1 \quad \delta_2 \\ & \phi \quad \psi \\ & \hline \phi → ψ \quad \rightarrow \text{Elim} \end{align*} \]

By induction hypothesis, φ → ψ follows from the undischarged assumptions Γ₁ of δ₁ and φ follows from the undischarged assumptions Γ₂ of δ₂. Consider a structure M. We need to show that, if M ⊨ Γ₁ ∪ Γ₂, then M ⊨ ψ. Suppose M ⊨ Γ₁ ∪ Γ₂. Since M ⊨ φ → ψ, M ⊨ φ → ψ. Since Γ₂ ⊨ φ, we have M ⊨ φ. This means that M ⊨ ψ (For if M ⊭ ψ, since M ⊨ φ, we’d have M ⊭ φ → ψ, contradicting M ⊨ φ → ψ).

4. The last inference is ¬Elim: Exercise.

5. The last inference is ∃Elim: Exercise.

\[ \square \]

**Problem ntd.1.** Complete the proof of Theorem ntd.1.

**Corollary ntd.2.** If ⊨ φ, then φ is valid.

**Corollary ntd.3.** If Γ is satisfiable, then it is consistent.
Proof. We prove the contrapositive. Suppose that $\Gamma$ is not consistent. Then $\Gamma \vdash \bot$, i.e., there is a derivation of $\bot$ from undischarged assumptions in $\Gamma$. By Theorem ntd.1, any structure $\mathfrak{M}$ that satisfies $\Gamma$ must satisfy $\bot$. Since $\mathfrak{M} \not\models \bot$ for every structure $\mathfrak{M}$, no $\mathfrak{M}$ can satisfy $\Gamma$, i.e., $\Gamma$ is not satisfiable. \qed

Photo Credits

Bibliography