Natural Deduction

The "feeling" of natural deduction is very different from that of Hilbert-style proof systems. It is closer to Gentzen-style systems, but there are several differences too. (Natural deduction was also first formulated by Gentzen. So, it isn't entirely proper to reserve the name "Gentzen" only for the sequent calculi of Handout 3.) In particular, in natural deduction there are *no* axiom schemes and *only* inference rules. To compensate for the lack of axioms, natural deduction allows the introduction of wff's as hypotheses at any stage of a derivation. Moreover, while a Gentzen system has only *introduction* rules, natural deduction uses both *introduction* and *elimination* rules.

The presentation in this handout is based on [1]. We can restrict the rules of the system to the logical connectives of a functionally complete set, say $\{\rightarrow, \perp\}$, leaving the other connectives $\{\vee, \wedge, \neg, \leftrightarrow\}$ to be defined in terms of the first two. (The symbol $\perp$ stands for *false.*) However, we prefer to include rules for all the connectives, with the understanding that we can always revert to a system restricted to rules for only $\{\rightarrow, \perp\}$ to simplify an argument (typically an induction on the length of derivations). The inclusion of rules for all the logical connectives not only makes the system more user-friendly, but gives more situations to illustrate the dual mechanism of introducing and cancelling hypotheses. Moreover, when we restrict the system for intuitionism, it will not be possible to define all the connectives in terms of only $\{\rightarrow, \perp\}$.

We start with the simpler case of propositional logic, and later add the necessary rules for first-order logic. For propositional logic, there are introduction and elimination rules for each of the 5 connectives: $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\neg$, but not for $\perp$.

1

Propositional logic

φ, ψ and σ range over the set of all wff's. By including rules for all of the 6 connectives, we get a total of 15 (one of them, →E, can be recognized as Modus Ponens).

1. One introduction rule, called ∧I, and two elimination rules, both called ∧E, for "∧":

$$\frac{\varphi \qquad \psi}{\varphi \wedge \psi} \ \wedge\text{I} \qquad\qquad \frac{\varphi \wedge \psi}{\varphi} \ \wedge\text{E} \qquad\qquad \frac{\varphi \wedge \psi}{\psi} \ \wedge\text{E}$$

2. Two introduction rules, both called ∨I, and one elimination rule, called ∨E, for "∨":

$$\frac{\varphi}{\varphi \vee \psi} \ \vee\text{I} \qquad\qquad \frac{\psi}{\varphi \vee \psi} \ \vee\text{I} \qquad\qquad \frac{\varphi \vee \psi \qquad \overset{[\varphi]}{\underset{\sigma}{\vdots}} \qquad \overset{[\psi]}{\underset{\sigma}{\vdots}}}{\sigma} \ \vee\text{E}$$

3. One introduction rule, called →I, and one elimination rule, called →E, for "→":

$$\frac{\overset{[\varphi]}{\underset{\psi}{\vdots}}}{\varphi \to \psi} \ \to\text{I} \qquad\qquad \frac{\varphi \qquad \varphi \to \psi}{\psi} \ \to\text{E}$$

4. One introduction rule, called ↔I, and two elimination rules, both called ↔E, for "↔":

$$\frac{\overset{[\varphi]}{\underset{\psi}{\vdots}} \qquad \overset{[\psi]}{\underset{\varphi}{\vdots}}}{\varphi \leftrightarrow \psi} \ \leftrightarrow\text{I} \qquad \frac{\varphi \qquad \varphi \leftrightarrow \psi}{\psi} \ \leftrightarrow\text{E} \qquad \frac{\psi \qquad \varphi \leftrightarrow \psi}{\varphi} \ \leftrightarrow\text{E}$$

5. One introduction rule, called ¬I, and one elimination rule, called ¬E, for "¬":

$$\frac{\overset{[\varphi]}{\underset{\bot}{\vdots}}}{\neg\varphi} \ \neg\text{I} \qquad\qquad \frac{\varphi \qquad \neg\varphi}{\bot} \ \neg\text{E}$$

6. Two rules for "$\bot$", the first called $\bot$ and the second `RAA`:

$$\frac{\bot}{\varphi} \quad \bot \qquad\qquad\qquad \begin{array}{c}[\neg\varphi]\\ \vdots \\ \dfrac{\bot}{\varphi}\end{array} \quad \texttt{RAA}$$

Note that the rules for "$\bot$" are the only rules not exhibiting a symmetry between "introduction" and "elimination".

Several of the rules above allow for a hypothesis φ to be *cancelled* (or *discharged*) which is indicated by enclosing φ between square brackets, i.e. by writing "$[\varphi]$". `RAA` stands for *reductio ad absurdum*, which formalizes the principle of a "proof by contradiction": If we can derive a contradiction, i.e. $\bot$, from $\neg\varphi$ then we can derive φ (without the hypothesis $\neg\varphi$). A few examples will make precise these notions.

Let N_0 denote the above system of rules for propositional logic.

Example 1. A derivation according to the rules of N_0:

$$\frac{\dfrac{[\varphi \wedge \psi]^1}{\psi}\ \wedge\texttt{E} \qquad \dfrac{[\varphi \wedge \psi]^1}{\varphi}\ \wedge\texttt{E}}{\dfrac{\psi \wedge \varphi}{(\varphi \wedge \psi)\ \rightarrow\ (\psi \wedge \varphi)}\ \rightarrow\texttt{I}_1}\ \wedge\texttt{I}$$

Note how we pair off the hypothesis $\varphi \wedge \psi$ with the rule $\rightarrow\texttt{I}$ that cancels it, by attaching the same index "1" to both the cancellation brackets and the corresponding rule, i.e. by writing $[\ ]^1$ and $\rightarrow\texttt{I}_1$. ∎

Example 2. Another derivation in N_0:

$$\frac{\dfrac{\dfrac{[\varphi]^2 \qquad [\varphi \rightarrow \bot]^1}{\bot}\ \rightarrow\texttt{E}}{(\varphi \rightarrow \bot) \rightarrow \bot}\ \rightarrow\texttt{I}_1}{\varphi \rightarrow ((\varphi \rightarrow \bot) \rightarrow \bot)}\ \rightarrow\texttt{I}_2$$

There are two cancellations in this example, for two different hypotheses, thus the two indexes, "1" and "2". ∎

Example 3. A more interesting derivation in N_0:

$$\cfrac{\cfrac{\cfrac{[\varphi \wedge \psi]^1}{\psi} \; \wedge\text{E} \qquad \cfrac{\cfrac{[\varphi \wedge \psi]^1}{\varphi} \; \wedge\text{E} \qquad [\varphi \to (\psi \to \sigma)]^2}{\psi \to \sigma} \; \to\text{E}}{\sigma} \; \to\text{E}}{\cfrac{(\varphi \wedge \psi) \to \sigma}{(\varphi \to (\psi \to \sigma)) \to ((\varphi \wedge \psi) \to \sigma)} \; \to\text{I}_2} \; \to\text{I}_1}$$

There are two cancelled hypotheses in this example, with two occurrences of the first and one occurrence of the second. ∎

It should be clear that any derivation using the rules of natural deduction can be organized in the form of a tree, with one wff attached to every node of the tree. The wff's at the leaf nodes are the *hypotheses* used in the derivation (some or all of them cancelled), and the single wff at the root node is the *conclusion* of the derivation. For the derivation in Example 3 the resulting tree is:
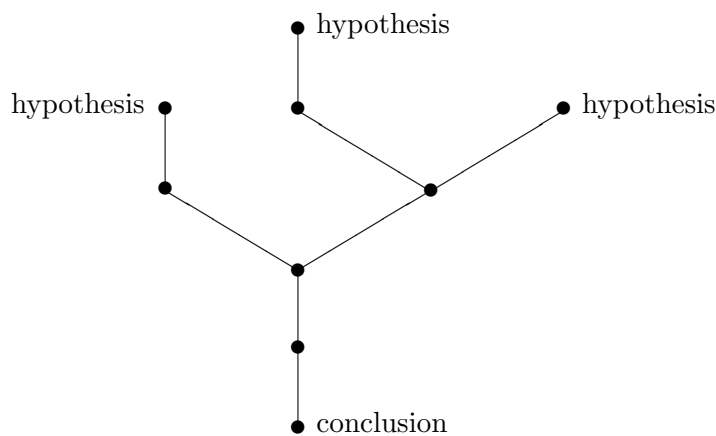
Given a set Γ of wff's and a wff φ, we write $\Gamma \vdash \varphi$ iff there is a derivation of the conclusion φ from uncancelled hypotheses that are all in Γ. Thus, for the derivations shown in Examples 1, 2, and 3, we have:

$$\vdash (\varphi \wedge \psi) \to (\psi \wedge \varphi) , \qquad \vdash \varphi \to ((\varphi \to \bot) \to \bot) , \qquad \vdash (\varphi \to (\psi \to \sigma)) \to ((\varphi \wedge \psi) \to \sigma) ,$$

respectively, and in all three cases we take $\Gamma = \emptyset$. Although not shown in these examples, we do not require that Γ be exactly the set of all uncancelled hypotheses: Γ may contain many (even infinitely many) wff's that do not appear at all in the derivation.

The 5 rules that allow cancellation of hypotheses are: $\vee$E, $\to$I, $\leftrightarrow$I, $\neg$I, and RAA, which are called accordingly *cancellation rules*. These require some care, as they can be used in a more liberal fashion than suggested by the notation. When we use one of these rules:

- *A wff enclosed in [ ], in the statement of the rule, does not mean that this wff has to actually appear as a hypothesis in the derivation.*

Hence, in particular, when we use $\to$I, $\neg$I and RAA (but not $\vee$E and $\leftrightarrow$I) which introduce in their conclusion a wff φ (or $\neg\varphi$) not mentioned among their premises, it is even possible that this φ (or $\neg\varphi$) appears nowhere in the derivation. This is illustrated in the next example.

Example 4.

$$
\cfrac{
  \cfrac{\dfrac{[\varphi \wedge \psi]^1}{\psi}\ \wedge\text{E} \qquad \dfrac{[\varphi \wedge \psi]^1}{\varphi}\ \wedge\text{E}}{\psi \wedge \varphi}\ \wedge\text{I}
  \Big/ (\varphi \wedge \psi)\ \to\ (\psi \wedge \varphi)
}{\sigma\ \to\ (\varphi \wedge \psi)\ \to\ (\psi \wedge \varphi)}
$$

The last use of $\to$I, with index "2", does not cancel any hypothesis occurrence and, moreover, introduces a fresh wff σ into the derivation. ∎

Moreover, when we use one of the cancellation rules:

- *Even if a wff enclosed in [ ] actually appears as a hypothesis in the derivation, not all of its occurrences have to be cancelled.*

This is justified as there is no harm in adding redundant hypotheses in a derivation. For example, we can cancel only one of the two occurrences of $\varphi \wedge \psi$ in Example 4, resulting in:

$$\varphi \wedge \psi\ \vdash\ \sigma\ \to\ ((\varphi \wedge \psi)\ \to\ (\psi \wedge \varphi))$$

or we can cancel neither of the two occurrences, resulting again in:

$$\varphi \wedge \psi\ \vdash\ \sigma\ \to\ ((\varphi \wedge \psi)\ \to\ (\psi \wedge \varphi))$$

or we can cancel the two occurrences separately (by using the rule $\to$I twice), resulting in:

$$\vdash\ \sigma\ \to\ ((\varphi \wedge \psi)\ \to\ ((\varphi \wedge \psi)\ \to\ (\psi \wedge \varphi)))$$

We define one more book-keeping device, before turning to first-order logic. Let $\mathcal{D}$ be a derivation viewed as a tree. Consider a node in $\mathcal{D}$, more precisely the wff φ attached to this node in $\mathcal{D}$, which is obtained by the use of a cancellation rule. For definiteness, let this rule be $\to\mathtt{I}$ and its use instance be $\to\mathtt{I}_n$ (an index n gets attached to it when used in a derivation) so that, in particular, φ must be of the form $\sigma \to \tau$. We use the term *parcel*, or *parcel of hypotheses*, to refer to all occurrences of σ enclosed in $[\ ]^n$. If we need to be more specific, we may say the "parcel with index n" or "parcel n". A parcel consists therefore of finitely many (possibly none) occurrences of the same hypothesis that are discharged together by one of the rules in $\{\vee\mathtt{E}, \to\mathtt{I}, \leftrightarrow\mathtt{I}, \neg\mathtt{I}, \mathtt{RAA}\}$. Our convention, already used in the preceding examples, is to uniquely identify a parcel of hypotheses by an index $n \in \mathbb{N}$.

Example 5.

$$
\cfrac{[(\varphi \wedge \psi) \vee \sigma]^5 \quad \cfrac{\cfrac{\cfrac{[\varphi \wedge \psi]^1}{\varphi}\ \wedge\mathtt{E}}{\varphi \vee \sigma}\ \vee\mathtt{I} \quad \cfrac{[\sigma]^2}{\varphi \vee \sigma}\ \vee\mathtt{I}}{\varphi \vee \sigma}\ \vee\mathtt{E}_{1,2} \qquad \cfrac{[(\varphi \wedge \psi) \vee \sigma]^5 \quad \cfrac{\cfrac{\cfrac{[\varphi \wedge \psi]^3}{\psi}\ \wedge\mathtt{E}}{\psi \vee \sigma}\ \vee\mathtt{I} \quad \cfrac{[\sigma]^4}{\psi \vee \sigma}\ \vee\mathtt{I}}{\psi \vee \sigma}}{\psi \vee \sigma}\ \vee\mathtt{E}_{3,4}}{\cfrac{\cfrac{(\varphi \vee \sigma) \wedge (\psi \vee \sigma)}{(\varphi \wedge \psi) \vee \sigma\ \to\ (\varphi \vee \sigma) \wedge (\psi \vee \sigma)}\ \to\mathtt{I}_5}{}}\ \wedge\mathtt{I}
$$

In this derivation there are 5 parcels. Each use of $\vee\mathtt{E}$ cancels two distinct parcels (parcels 1 and 2, parcels 3 and 4), whereas the use of $\to\mathtt{I}$ cancels only one parcel (parcel 5). ∎

First-order logic

The system for propositional logic, N_0, is extended by adding rules for the quantifiers and, if an equality symbol $\approx$ is included in the syntax of wff's, by adding rules for equality too. φ and ψ range over the set of wff's, x, y and z over variables, and t over terms. Call N the resulting system.

1. One introduction rule, called $\forall\mathtt{I}$, and one elimination rule, called $\forall\mathtt{E}$, for "$\forall$":

$$\frac{\varphi(x)}{\forall x\ \varphi(x)}\ \forall\mathtt{I} \qquad\qquad \frac{\forall x\ \varphi(x)}{\varphi(t)}\ \forall\mathtt{E}$$

   where in $\forall\mathtt{I}$ the variable x does not occur free in any hypothesis on which $\varphi(x)$ depends, i.e. in any uncancelled hypothesis in the derivation of $\varphi(x)$, and in $\forall\mathtt{E}$ the substitution of t for x is legal, i.e. no free variable in t is captured by a quantifier in φ.

2. One introduction rule, called $\exists\mathtt{I}$, and one elimination rule, called $\exists\mathtt{E}$, for "$\exists$":

$$\frac{\varphi(t)}{\exists x\ \varphi(x)}\ \exists\mathtt{I} \qquad\qquad \frac{\exists x\ \varphi(x) \qquad \overset{\displaystyle[\varphi(x)]}{\underset{\displaystyle\psi}{\vdots}}}{\psi}\ \exists\mathtt{E}$$

   where in $\exists\mathtt{I}$ the substitution of t for x is legal, i.e. no free variable in t is captured by a quantifier in φ, and in $\exists\mathtt{E}$ the variable x does not occur free in ψ nor in any uncancelled hypothesis (other than $\varphi(x)$) on which $\varphi(x)$ depends, i.e. in any uncancelled hypothesis in the subderivation with conclusion ψ.

3. The rules for equality simulate the axioms for equality, used in Hilbert systems (Handout 1) or in Gentzen systems (Handout 3). There is no symmetry here between "introduction" and "elimination" rules, in contrast to rules for the logical connectives and the quantifiers. In $\mathtt{EQ4}$, f is an arbitrary function symbol of arity $n \geqslant 0$, and in $\mathtt{EQ5}$, P is an arbitrary predicate symbol of arity $n \geqslant 0$.

$$\frac{}{x\ \approx\ x}\ \mathtt{EQ1} \qquad \frac{x\ \approx\ y}{y\ \approx\ x}\ \mathtt{EQ2} \qquad \frac{x\ \approx\ y \qquad y\ \approx\ z}{x\ \approx\ z}\ \mathtt{EQ3}$$

$$\frac{x_1\ \approx\ y_1\ \cdots\ x_n\ \approx\ y_n}{f x_1\cdots x_n\ \approx\ f y_1\cdots y_n}\ \mathtt{EQ4} \qquad\qquad \frac{x_1\ \approx\ y_1\ \cdots\ x_n\ \approx\ y_n \quad P x_1\cdots x_n}{P y_1\cdots y_n}\ \mathtt{EQ5}$$

Example 6. A derivation in N:

$$\cfrac{\cfrac{\cfrac{\cfrac{[\forall x\ (\varphi(x) \wedge \psi(x))]^1}{\varphi(x) \wedge \psi(x)}\ \forall\mathtt{E}}{\varphi(x)}\ \wedge\mathtt{E}}{\forall x\ \varphi(x)}\ \forall\mathtt{I} \qquad \cfrac{\cfrac{\cfrac{[\forall x\ (\varphi(x) \wedge \psi(x))]^1}{\varphi(x) \wedge \psi(x)}\ \forall\mathtt{E}}{\psi(x)}\ \wedge\mathtt{E}}{\forall x\ \psi(x)}\ \forall\mathtt{I}}{\cfrac{(\forall x\ \varphi(x)) \wedge (\forall x\ \psi(x))}{\forall x\ (\varphi(x) \wedge \psi(x))\ \rightarrow\ (\forall x\ \varphi(x)) \wedge (\forall x\ \psi(x))}\ {\rightarrow}\mathtt{I}_1}\ \wedge\mathtt{I}$$

Example 7. Another derivation in N:

$$\cfrac{[\exists x\ (\varphi(x) \vee \psi(x))]^4 \qquad \cfrac{[\varphi(x) \vee \psi(x)]^3 \qquad \cfrac{\cfrac{[\varphi(x)]^1}{\exists x\ \varphi(x)}\ \exists\mathtt{I}}{(\exists x\ \varphi(x)) \vee (\exists x\ \psi(x))}\ \vee\mathtt{I} \qquad \cfrac{\cfrac{[\psi(x)]^2}{\exists x\ \psi(x)}\ \exists\mathtt{I}}{(\exists x\ \varphi(x)) \vee (\exists x\ \psi(x))}\ \vee\mathtt{I}}{(\exists x\ \varphi(x)) \vee (\exists x\ \psi(x))}\ \vee\mathtt{E}_{1,2}}{\cfrac{(\exists x\ \varphi(x)) \vee (\exists x\ \psi(x))}{\exists x\ (\varphi(x) \vee \psi(x))\ \rightarrow\ (\exists x\ \varphi(x)) \vee (\exists x\ \psi(x))}\ {\rightarrow}\mathtt{I}_4}\ \exists\mathtt{E}_3$$

The symbol "$\vdash_H$" is for derivability relative to one of the Hilbert systems in Handout 1 (which all derive precisely the same set of wff's), and "$\vdash_N$" is for derivability relative to the rules of natural deduction. For comparisons with systems in previous handouts, take $\bot$ as an abbreviation for $\alpha \wedge \neg\alpha$, for some fixed but otherwise arbitrary wff α.

Theorem 1. *For an arbitrary set of wff's Γ and an arbitrary wff φ, $\Gamma \vdash_H \varphi$ if and only if $\Gamma \vdash_N \varphi$.*

Proof: A proof can be found in [4], pp 148-159. A sketch of a proof, with useful comments, is also in [3], pp 26-32. Another proof is to first show: $\Gamma \vdash_N \varphi$ iff $\vdash_G \Gamma|\varphi$ (with the restriction that Γ is finite), and then invoke Theorem 1 of Handout 3. For the equivalence between natural deduction and a Gentzen system (when both are restricted to the intuitionistic case), there are proofs in [4], pp 168-186, and in [2], Ch. 5. ∎

Restrictions for intuitionism

A natural-deduction system for intuitionistic propositional logic (resp. first-order logic) is obtained by omitting just one rule from N_0 (resp. N): RAA.

All the derivations so far, in Examples 1 to 7, are acceptable intuitionistically, because none uses the rule RAA.

Example 8. Here is a derivation which is not allowed intuitionistically:

$$\frac{\dfrac{[\neg\varphi]^1 \quad [\neg\varphi \to \bot]^2}{\dfrac{\bot}{\varphi} \ \text{RAA}_1} \to\text{E}}{(\neg\varphi \to \bot) \to \varphi} \to\text{I}_2$$

If we take $\neg\varphi$ as an abbreviation for $\varphi \to \bot$, then we have here:

$$\vdash \ \neg\neg\varphi \ \to \ \varphi$$

which is certainly accepted classically. A subtle point: With the forementioned abbreviation, the derivation in Example 2 shows that

$$\vdash \ \varphi \ \to \ \neg\neg\varphi$$

which *is* acceptable intuitionistically. There is no contradiction here: Intuitionism does not take φ and $\neg\neg\varphi$ as equivalent wff's. ∎

Normalization

A fundamental result has to do with the elimination of superfluous parts in derivations. The motivation is best given by examples.

Example 9.

$$\frac{\dfrac{\dfrac{\sigma \quad \sigma \to \varphi}{\varphi} \to\text{E} \quad [\psi]^1}{\dfrac{\dfrac{\varphi \wedge \psi}{\varphi} \ \wedge\text{E}}{\psi \to \varphi} \to\text{I}_1} \wedge\text{I}}{}$$

The conjunction $\varphi \wedge \psi$ is introduced only to be immediately eliminated. It is clearly more efficient to write instead:

$$\frac{\dfrac{\sigma \quad \sigma \to \varphi}{\varphi} \to\text{E}}{\psi \to \varphi} \to\text{I}_1$$

The consecutive uses of $\wedge\text{I}$ and $\wedge\text{E}$ are now removed. ∎

Example 10. The following is a more interesting derivation:

$$\frac{\dfrac{[\sigma \wedge \varphi]^3}{\varphi} \wedge \mathtt{E} \qquad [\varphi \to \psi]^2}{\dfrac{\psi}{\dfrac{\sigma}{\dfrac{(\varphi \to \psi) \to \sigma}{(\sigma \wedge \varphi) \to ((\varphi \to \psi) \to \sigma)} \to \mathtt{I}_3}} \to \mathtt{I}_2} \to \mathtt{E} \qquad \dfrac{\dfrac{[\sigma \wedge \varphi]^3}{\sigma} \wedge \mathtt{E}}{\psi \to \sigma} \to \mathtt{I}_1}{\to \mathtt{E}}$$

More efficiently, we can write the following derivation:

$$\frac{\dfrac{\dfrac{[\sigma \wedge \varphi]^3}{\sigma} \wedge \mathtt{E}}{(\varphi \to \psi) \to \sigma} \to \mathtt{I}_2}{(\sigma \wedge \varphi) \to ((\varphi \to \psi) \to \sigma)} \to \mathtt{I}_3$$

We have cut out the consecutive uses of $\to\mathtt{I}$ (with index 1) and $\to\mathtt{E}$. ∎

A derivation in which an introduction is never followed by an elimination is called *normal*.

Theorem 2 (Normal Form Theorem). *Every derivation $\mathcal{D}$ is equivalent to a normal derivation $\mathcal{D}'$, i.e. if $\mathcal{D}$ is a derivation of φ from Γ then there is a* normal *derivation $\mathcal{D}'$ of φ from Γ.*

A *reduction step* consists in the removal of a superfluous introduction followed by an elimination. Theorem 3 says that the process of going from $\mathcal{D}$ to $\mathcal{D}'$ in Theorem 2 can be carried out effectively.

Theorem 3 (Normalization Theorem). *For every derivation $\mathcal{D}$ there is a finite sequence of reduction steps that reduces $\mathcal{D}$ into a normal derivation $\mathcal{D}'$ equivalent to $\mathcal{D}$.*

An even stronger result than the preceding two is the *strong normalization theorem*.

Theorem 4 (Strong Normalization Theorem). *Every sequence of reduction steps applied to a derivation $\mathcal{D}$ terminates in a normal derivation $\mathcal{D}'$ equivalent to $\mathcal{D}$.*

References

[1] van Dalen, D., *Logic and Structure*, Third Edition, Springer-Verlag, New York, 1994.

[2] Girard, J.-Y., Lafont, Y., and Taylor, P., *Proofs and Types*, Cambridge University Press, Cambridge, UK, 1989.

[3] Hodges, W., "Elementary Predicate Logic", in *Handbook of Philosophical Logic*, Vol. 1, ed. Gabbay and Guenthner, pp. 1-132, Reidel, 1983.

[4] Sundholm, G., "Systems of Deductions", in *Handbook of Philosophical Logic*, Vol. 1, ed. Gabbay and Guenthner, pp. 133-188, Reidel, 1983.