Example illustrating the proof of the soundness theorem

Suppose that we have a "toy" proof system with only the following three rules:

Repetition: If $A_1, \dots, A_n$ are formulas and $B \in \{A_1, \dots, A_n\}$, then

$$\begin{array}{|l} A_1 \\ \vdots \\ A_n \\ \hline B \end{array}$$

is a proof of B from the assumptions $A_1, \dots, A_n$.

$\exists$ Introduction If

$$\Pi = \begin{array}{|l} A_1 \\ \vdots \\ A_n \\ \hline \vdots \\ B[t/x] \end{array}$$

is a proof of $B[t/x]$ from the assumptions $A_1, \dots, A_n$, t is a term such that x does not occur in t and no variable in t occurs in B, where $B[t/x]$ denotes the formula obtained from B by replacing every <u>free</u> occurence of x in B by t.

then

$$\begin{array}{|l} A_1 \\ \vdots \\ A_n \\ \hline \vdots \\ B[t/x] \\ \exists x\, B \end{array} \qquad \text{(lines from } \vdots \text{ to } B[t/x] \text{ as in } \Pi)$$

is a proof of $\exists x\, B$ from the assumptions $A_1, \dots, A_n$.

$\exists$ Elimination

If

$$\Pi_1 = \begin{array}{|l} A_1 \\ \vdots \\ A_n \\ \hline \vdots \\ \exists x\, B \end{array}$$

is a proof of $\exists x\, B$ from the assumptions $A_1, \dots, A_n$,

and

$$\Pi_2 = \begin{array}{|l} A_1 \\ \vdots \\ A_n \\ B[y/x] \\ \hline \vdots \\ S \end{array}$$

is a proof of S from the premisses $A_1, \dots, A_n$, $B[y/x]$ where $B[y/x]$ is obtained from B by replacing every free occurence of x by y, and y does not occur in any of $A_1, \dots, A_n, B, S$, then

$$
\begin{array}{|l}
A_1 \\
\vdots \\
A_n \\
\hline
\left.\begin{array}{l} \vdots \\ \exists x\, B \end{array}\right\} \text{ as in } \Pi_1 \\
\begin{array}{|l} B[y/x] \\ \hline \left.\begin{array}{l} \vdots \\ S \end{array}\right\} \text{ as in } \Pi_2 \end{array} \\
S
\end{array}
$$

is a proof of S from the assumptions $A_1, \ldots, A_n$.

Now we want to prove the soundness of this proof system. That is, we like to prove that,

if $A_1, \ldots, A_n \vdash B$

then $A_1, \ldots, A_n \models B$.

In order that '$A_1, \ldots, A_n \models B$' makes sense also for formulas with free variables (which we have to consider in the proof) we use the following definition.

For any formulas (possibly with free variables) $A_1, \ldots, A_n, B$, we define

$$A_1, \ldots, A_n \models B \quad \text{to mean that}$$

for <u>every</u> structure $\mathcal{M}$ (in the language considered) and <u>every</u> valuation w of the variables in $\mathrm{dom}(\mathcal{M})$ (see below)

(0) $\mathcal{M} \models_w A_1 \wedge \ldots A_n \Longrightarrow \mathcal{M} \models_w B$.

A <u>valuation</u> in $\mathrm{dom}(\mathcal{M})$ is a function from the set of variables to $\mathrm{dom}(\mathcal{M})$, so a valuation w assigns a "value" in $\mathrm{dom}(\mathcal{M})$ to every variable.

See my notes about Tarski's truth definition for the definition of '$\models_w$' (and of t^w for terms t).

We want to prove the implication

(1) $A_1, \ldots, A_n \vdash B \Longrightarrow A_1, \ldots, A_n \models B$.

This is done by induction on the length of a (shortest) proof of B from $A_1, \dots, A_n$. By <u>length</u> we mean the number of applications of rules used in the proof.

If the length is 1 then the rule used must be repetition. We leave it as an exercise to show that if $B \in \{A_1, \dots, A_n\}$ then the implication (0) holds for every $\mathcal{M}$ and every valuation in $\mathrm{dom}(\mathcal{M})$ (and therefore also (1) holds).

For the induction step, suppose that (1) holds whenever there is a proof of B from $A_1, \dots, A_n$ of length k.

We have two cases to consider, corresponding to the rules $\exists$ Introduction and $\exists$ Elimination.

Suppose that

$$
\begin{array}{|l}
A_1 \\
\vdots \\
A_n \\
\hline
\vdots \\
B[t/x] \\
\exists x B
\end{array}
$$

is a proof of $\exists x B$ of length $k+1$, where the last rule is $\exists$ Introduction, By induction we have

(2) $A_1, \dots, A_n \vdash B[t/x] \Rightarrow A_1, \dots, A_n \models B[t/x]$,

so for every $\mathcal{M}$ and every valuation w in $\mathrm{dom}(\mathcal{M})$,

(3) $\mathcal{M} \models_w A_1, \dots, A_n \Longrightarrow \mathcal{M} \models_w B[t/x]$.

Now let $\mathcal{M}$ be any structure an w_0 any valuation in $\mathrm{dom}(\mathcal{M})$. Suppose that

(4) $\mathcal{M} \models_{w_0} A_1, \dots, A_n$.

We must show that

(5) $\mathcal{M} \models_{w_0} \exists x B$.

Since (3) holds for every w it holds, in particular for w_0, so (4) gives

(6) $\mathcal{M} \models_{w_0} B[t/x]$

By Tarski's truth definition, we have

(7) $\mathcal{M} \models_{w_0} \exists x B \Longleftrightarrow \mathcal{M} \models_{w_0[a/x]} B$ for some $a \in \mathrm{dom}(\mathcal{M})$.

(See my notes for definition of $w[a/x]$.)

Since no variable of t occurs in B and x does not occur in t, it follows that $w_0[t^{w_0}/x]$ is a well defined valuation, and from (6) we get

$$\mathcal{M} \models_{w_0[t^{w_0}/x]} B .$$

That the previous line indeed follows from (6) and our assumptions can be proved by induction on the structure of B.

If we let $a = t^{w_0}$ we get

$$\mathcal{M} \models_{w_0[a/x]} B$$

which together with (7) implies (5).

This concludes the proof of the first case in the induction step of the proof.

The second, and last case (for our "toy" proof system) is when the last and $(k+1)$st application of a rule is $\exists$Elimination in which case the proof looks like

$$
\begin{array}{|l}
A_1 \\
\vdots \\
A_n \\
\hline
\vdots \\
\exists x B \\
\begin{array}{|l}
B[y/x] \\
\hline
\vdots \\
S
\end{array} \\
S
\end{array}
$$

where y does not occur in any of $A_1, \ldots, A_n$, B

and
$$
\begin{array}{|l}
A_1 \\
\vdots \\
A_n \\
\hline
\vdots \\
\exists x B
\end{array}
$$
and
$$
\begin{array}{|l}
A_1 \\
\vdots \\
A_n \\
B[y/x] \\
\hline
\vdots \\
S
\end{array}
$$

are shorter proofs of $\exists x B$ and S, respectively.

By the induction hypothesis we have

(8) $A_1, \ldots, A_n \models \exists x B$ and

(9) $A_1, \ldots, A_n, B[y/x] \models S$.

Let $\mathcal{M}$ be any structure and w any valuation in $dom(\mathcal{M})$.

Suppose that

(10) $\mathcal{M} \models_w A_1, \ldots, A_n$.

We need to prove that

(11) $\mathcal{M} \models_w S$.

By (8) and 10 we get

(12) $\mathcal{M} \models_w \exists x B$,

which, by Tarski's truth definition, means that there is $a \in dom(\mathcal{M})$ such that

(13) $\mathcal{M} \models_{w[a/x]} B$,

and since y does not occur in B we get

(14) $\mathcal{M} \models_{w[a/y]} B[y/x]$.

Since y does not occur in any of $A_1, \dots, A_n, S$ we have by Lemma 2.10 in my notes about Tarski's truth definition, that

(15) $\mathcal{M} \models_w A_1, \dots, A_n \iff \mathcal{M} \models_{w[a/y]} A_1, \dots, A_n$.

By the definition of '$\models$' and (9) we get

(16) $\mathcal{M} \models_{w[a/y]} A_1, \dots, A_n, B[y/x] \Rightarrow \mathcal{M} \models_{w[a/y]} S$,

which together with (10), (14) and (15) gives $\mathcal{M} \models_{w[a/y]} S$, which, since y does not occur in S implies

$\mathcal{M} \models_w S$, so (11) is proved and we are done.