Natural deduction for Foundations

The introduction and elimination rules for the connective $\longrightarrow$ in natural deduction, as introduced by Billy, are

$$\dfrac{\begin{array}{c}[A]^{i}\\ \vdots \\ B\end{array}}{A \longrightarrow B}\,(\to I)^{i} \qquad\qquad \dfrac{\begin{array}{c}\vdots\\ A \longrightarrow B\end{array} \quad \begin{array}{c}\vdots\\ A\end{array}}{B}\,(\to E)$$

The introduction rule says that a deduction of $A \longrightarrow B$ is a process (perhaps a verbal argument instantiated in time) that begins with a deduction of A and ends with a deduction of B. Note that the deduction above the line <u>need not</u> contain A as a hypothesis (because from B we deduce $A \longrightarrow B$ for any A) and it may contain <u>more than one</u> copy of A as a hypothesis (some subset of which may be discharged "simultaneously", see (5) below). These possibilities reflect, respectively, weakening and contraction in sequent calculus.

<u>Example</u>

(1) undischarged hypotheses: A $\qquad A \qquad$ } not a deduction of A

(2) undischarged hypotheses: ϕ $\qquad \dfrac{[A]^1}{A \to A}\,(\to I)^1 \qquad$ } deduction of $A \longrightarrow A$

(3)
$$\dfrac{\dfrac{[A]^1}{A \longrightarrow A}\,(\to I)^1 \qquad \begin{array}{c}\pi\\ \vdots\\ A\end{array}}{A}\,(\to E)$$

This deduction exhibits what is called a "detour" since the occurrence of modus ponens here is redundant and the overall deduction "seems" to be just π.

(4)

$$\dfrac{\dfrac{\dfrac{[A \to A]^2 \qquad [A]^1}{A}\,(\to E)}{A \to A}\,(\to I)^1}{(A \to A) \to (A \to A)}\,(\to I)^2 \tag{2.1}$$

Note that we could have obtained a deduction of $A \to ((A \to A) \to A)$ by performing the introductions in a different order.

(5) In fact there are infinitely many "different" deductions of $(A \to A) \to (A \to A)$. Here is another one:

$$\dfrac{\dfrac{\dfrac{[A \to A]^2 \qquad \dfrac{[A \to A]^2 \qquad [A]^1}{A}\,(\to E)}{A}\,(\to E)}{A \to A}\,(\to I)^1}{(A \to A) \to (A \to A)}\,(\to I)^2 \tag{2.2}$$

This deduction corresponds under Curry-Howard to the Church numeral $\underline{2}$, while (4) is $\underline{1}$. In what sense are these "different" deductions? Is logic concerned with just <u>provability</u> (whether or not this set is empty) or is it also about the <u>structure of this set</u> (whatever that may be).

Let us take the latter point of view seriously for a moment.

The main thing to notice about $\underline{2}$ is that two copies of $A \to A$ are discharged in the final step. Clearly we are not concerned with the fact that the indices $1, 2$ are used to label $(\to I)$, all that matters is the <u>pattern of coincidence</u> of these symbols (i.e. $1 \neq 2$ and 2 appears twice). So we are prepared to believe that the deduction below "is" also $\underline{2}$

$$\dfrac{\dfrac{\dfrac{[A \to A]^f \qquad \dfrac{[A \to A]^f \qquad [A]^x}{A}\,(\to E)}{A}\,(\to E)}{A \to A}\,(\to I)^x}{(A \to A) \to (A \to A)}\,(\to I)^f \tag{3.1}$$

But what are the rules on these symbols $1, 2, x, f, \ldots$ <u>exactly</u>? Here's a scary example

$$\dfrac{\dfrac{\dfrac{\dfrac{[A \to A]^f \qquad \dfrac{[A \to A]^f \qquad [A]^x}{A}\,(\to E)}{A}\,(\to I)^x}{A \to A}\,(\to I)^f}{(A \to A) \to (A \to A)}\,(\to I)^f}{(A \to A) \to ((A \to A) \to (A \to A))} \tag{3.2}$$

← But I mean only one of them!

← and this means the other one, but which?

Am I allowed to re-use the name f? If so, the transcript of my deduction probably contains some information like "that one", or is it "pick one"? Does it even matter? It seems reasonable to say these are different deductions, and so the additional information in the transcript should somehow be embedded in the deduction itself. We could forbid the reuse of names, as long as we continue to insist that these names are introduced "at the moment" of the $(\to I)$ rule, as in

(3.3)

$$\dfrac{\dfrac{A \to A \qquad \dfrac{A \to A \qquad [A]^x}{A}\,(\to E)}{A}\,(\to E)}{A \to A}\,(\to I)^x$$

$t = 0$

$$\dfrac{\dfrac{\dfrac{A \to A \qquad \dfrac{[A \to A]^f \qquad [A]^x}{A}\,(\to E)}{A}\,(\to E)}{A \to A}\,(\to I)^x}{(A \to A) \to (A \to A)}\,(\to I)^f$$

$t = 1$

$$\dfrac{\dfrac{\dfrac{\dfrac{[A \to A]^g \qquad \dfrac{[A \to A]^f \qquad [A]^x}{A}\,(\to E)}{A}\,(\to E)}{A \to A}\,(\to I)^x}{(A \to A) \to (A \to A)}\,(\to I)^f}{(A \to A) \to ((A \to A) \to (A \to A))}\,(\to I)^g$$

$t = 2$

How do we know what undischarged hypotheses are <u>available</u> at any given stage of the deduction? It is intuitively clear: this availability flows down the tree, uninterrupted by $(\to E)$ rules and $(\to I)^x$ removes some specified subset of occurrences of a particular formula. Two deductions are <u>the same</u> if they discharge the same hypotheses at the same places.

<u>Theorem</u> (Curry-Howard) Simply-typed lambda terms which are closed (up to α-equivalence) correspond to natural deductions (up to the above equivalence rel^N of "sameness").

This is a strong indication that the <u>structure of the set of deductions is meaningful</u>. As we have said, (2.1), (2.2) correspond to Church numerals $\underline{1}$, $\underline{2}$ and (3.3) is a binary integer ($\underline{01}$ or $\underline{10}$ depending on conventions).

<u>Remark</u> (1) Introducing names "at the moment of $(\to I)$" is closely related to the policy "contract as late as possible" in the Mints normal form, and "weaken as late as possible".

(2) If you want to extend CH to open terms, which should correspond to deductions with undischarged hypotheses, then you'll be forced to say which undischarged hypotheses are "the same" (that is, you'll have to allow for the ability to "tie the hands" of whoever continues the deduction, in the sense that they have to discharge certain hypotheses as a packet) and this means that the policy of "I'll decide what things are called when I use $(\to)$" from p. 3 has to be abandoned, and some kind of additional book-keeping structure for names has to be added to deductions.

Arguably, the fact that this policy does not survive the generalisation from closed to open terms means it was not sound to begin with.