THE JOURNAL OF SYMBOLIC LOGIC
Volume 70, Number 2, June 2005

INTERPRETING DESCRIPTIONS IN INTENSIONAL TYPE THEORY

JESPER CARLSTRÖM

Abstract. Natural deduction systems with indefinite and definite descriptions (ε-terms and ι-terms) are presented, and interpreted in Martin-Löf's intensional type theory. The interpretations are formalizations of ideas which are implicit in the literature of constructive mathematics: if we have proved that an element with a certain property exists, we speak of 'the element such that the property holds' and refer by that phrase to the element constructed in the existence proof. In particular, we deviate from the practice of interpreting descriptions by contextual definitions.

§1. Introduction. There are two kinds of descriptions to be considered in this paper: $\varepsilon x A(x)$ is *an* element such that $A(x)$ and $\iota x A(x)$ is *the* element such that $A(x)$, thus requiring uniqueness. Formally, $\varepsilon x A(x)$ is an individual if $\exists x A(x)$ is true, while $\iota x A(x)$ is an individual only if $\forall x \forall y (A(x) \,\&\, A(y) \supset x = y)$ is true as well. This difference is the reason for us to distinguish between $\varepsilon x A(x)$ and $\iota x A(x)$. We say that they are *indefinite* and *definite* descriptions, respectively. It is not quite correct to expect that this difference should always be reflected in English by the articles *a/an* and *the*. We say things like 'I met a man. The man was tall', and refer by 'the man' to the man we met, even if there are more than one man in the world [17]. We use a definite article because a man is determined by context, but he is not determined by uniqueness. Because the uniqueness property is what will formally distinguish ε-terms and ι-terms in this paper, this occurrence of 'the man' is better represented by $\varepsilon x A(x)$ than by $\iota x A(x)$ in our formalism.

In English one may express the same thing by saying instead 'I met a man, who was tall' or 'I met a man. He was tall'. We will make no attempts to analyze such grammatical variants, but view them as synonymous with 'I met a man. The man was tall'. Thus the representation in natural language of $\varepsilon x A(x)$ can be 'he', 'who', 'it' etc., as well as the typical one: 'an element such that $A(x)$'. It is sometimes even more natural to think of it as 'one of the elements such that $A(x)$', as we will see.

Our main goal is to understand *definite* descriptions intensionally. That is, we are not satisfied with a proof that one can live without definite descriptions by eliminating them, we actually want to carry out an interpretation of them which explains what they *mean*. Among constructivists, it is often believed that this can be done and the aim of this paper is not only to show that this belief is indeed correct, but also to investigate *how* such an interpretation really works, intensionally.

Therefore, we introduce a ι-calculus and give a formal interpretation of it in Martin-Löf's intensional type theory. To simplify our task, we proceed in two

Received February 23, 2004; revised November 12, 2004.

0022-4812/05/7002-0010/$3.70

steps. First, we introduce an ε-calculus and interpret it in type theory. Then we introduce the ι-calculus and show that proofs in this calculus can be translated into proofs in the ε-calculus, if ι-terms are interpreted as ε-terms:

$$\iota \longrightarrow \varepsilon \longrightarrow \text{MLTT}\,.$$

Interestingly, our interpretation of the ε-calculus relies quite heavily on type-theoretical choice and the strong disjunction elimination which is present in Martin-Löf's type theory. It is an open problem whether a similar interpretation could be carried out in weaker type theories, like logic-enriched type theory [2].

Being more general than the ι-calculus, but still interpretable in type theory, the ε-calculus would be very useful, was it not the case that it suffers from some unfamiliar restrictions: modus ponens and existence introduction are not valid in general. We will explain how this fact is dealt with, and why it is not a problem in the ι-calculus.

§**2. The structure of the paper.** We relate our approach to previous work on descriptions in Section 3 and discuss some properties of our systems. In Section 4, we introduce the ε-calculus, which is extended to include equality in Section 5. We then show, in Section 6, how this calculus is interpreted in Martin-Löf's type theory, and give some examples in Section 7. We introduce the ι-calculus in Section 8 and explain how it is translated into the ε-calculus. We end by some comments about how restricted quantifiers are interpreted (Sect. 9) and how the system can be used to treat partial functions (Sect. 10).

§**3. Background.** Russell [18, 19] proposed a contextual definition of descriptions. It interprets 'the father of Charles II. was executed' as

> It is not always false of x that x begat Charles II. and that x was executed and that 'if y begat Charles II., y is identical with x' is always true of y. [18, p. 482]

The theory makes every proposition of the form $P(\text{the present King of France})$ false, a fact that Russell considers 'a great advantage in the present theory' [p. 482]. In particular, propositions of the form $x = x$ are false according to Russell if 'the present King of France' is substituted for x, and so its negation $\sim(x = x)$ is instead true.

In mathematics, however, it is very unusual to use descriptive phrases that do not refer, even though descriptions as such are very common. At least, it seems to be common practice to require a *hypothetical* reference, that is, that the description refers under some condition. This is the case, for example, when multiplicative inverses are defined by descriptions, as is done by Mines, Richman and Ruitenburg, among others:

> If a and b are elements of a monoid, and $ab = 1$, then we say that a is a *left inverse* of b and b is a *right inverse* of a. If b has a left inverse a and a right inverse c, then $a = a(bc) = (ab)c = c$; in this case we say that a is the *inverse* of b and write $a = b^{-1}$. If b has an inverse we say that b is a *unit*, or that b is *invertible*. [13, p. 36]

According to this passage, the expression 'b^{-1}' is used only when b has an inverse, and it then refers to this inverse. One is, it seems, not even allowed to *use* the expression unless b has an inverse. This is indeed the attitude taken by most mathematicians, in contrast to the attitude advocated by Russell, who seems to claim (when we have translated his examples to more mathematical ones) that we need to say things like '0^{-1} does not exist' and that we have to consider an expression like '$\sim(0^{-1} = 2)$' to be a true proposition. It is rather mathematical practice to consider such expressions meaningless—as malformed propositions. More precisely, a term is used only when it refers to some individual. In order for '$\sim(0^{-1} = 2)$' to be accepted as a legitimate proposition, '0^{-1}' must refer, which it does not.[1]

Our point of departure is Frege's idea [5], that the *reference* of a description may depend on a *presupposition*. In mathematics, the necessary presuppositions are often made explicit beforehand, while it is common in everyday life not to spell them out. Frege distinguishes the *sense* of a term from its *reference*. For instance, he claims that 0^{-1} has a sense (the real number which is the multiplicative inverse of 0), but no reference (because there is no such real number).[2] According to Frege, every assertion presupposes that all terms used refer: 'Wenn man etwas behauptet, so ist immer die Voraussetzung selbstverständlich, daß die gebrauchten einfachen oder zusammengesetzen Eigennamen eine Bedeutung haben' [5, p. 40]. He argues that 'Kepler starb im Elend' presupposes that 'Kepler' refers, but that it cannot be said to be *contained* in the proposition that 'Kepler' refers, because then the negation would be 'Kepler starb nicht im Elend, oder der Name "Kepler" ist bedeutunglos'. Rather, Frege argues, the reference of 'Kepler' is presupposed also in the negation 'Kepler starb nicht im Elend'. It seems that most mathematicians tend to think in this way: the proposition $x^{-1} + y^{-1} = 1$ presupposes that x and y are invertible, but it is usually not said that the proposition *implies* that they are.

This view has some consequences for Russell's examples. All talk about 'the present King of France' presupposes the existence of a present King of France, not necessarily in our ordinary world, but in an imagined context of the utterances. In this context, it is true that 'the present King of France is the present King of France', contrary to Russell's proposal. In the same way, if '$\sim(0^{-1} = 2)$' is to be accepted as a legitimate proposition, we have to presuppose that 0 has an inverse.

In [19, pp. 167–180], Russell considers examples like 'I met a man' and 'I met a unicorn' and argues that the latter one is as meaningful as the first one, even if we assume that unicorns do not exist, because it is perfectly clear what the speaker is trying to communicate. Hence, Russell argues, it must be admitted that descriptive phrases may be meaningful even if they do not refer to anything. It is easy to agree with Russell in this case because 'I met a man' is naturally perceived as 'I met something, which was a man' and 'I met a unicorn' as 'I met something, which was a unicorn'. There is no logical difference between the examples and there is no need to talk about reference, because 'a man' and 'a unicorn' are used as *properties*

[1] We suppose here that we are discussing a non-trivial ring like the real numbers. There is of course the trivial case when 0 has in fact an inverse. Curiously, $\sim(0^{-1} = 2)$ is then *false* rather than true.

[2] Frege's claim is actually not about this specific example, but about a similar empty description: 'Der Ausdruck "die am wenigsten konvergente Reihe" hat einen Sinn; aber man beweist, daß er keine Bedeutung hat, da man zu jeder konvergenten Reihe eine weniger konvergente, aber immer noch konvergente finden kann' [5, p. 28].

rather than as individual terms. However, there are situations where something like indefinite descriptions are involved, and which are not as easily resolved. Consider for instance the expression 'I met one of the unicorns'. This would be a very natural thing to say for a figure in a tale where unicorns are perfectly normal creatures, but in an environment where it is not in general admitted that unicorns exist, we would hardly express ourselves in this manner. Hence it seems that while 'I met a unicorn' does not presuppose anything, 'I met one of the unicorns' does in fact presuppose the existence of unicorns. Moreover, 'one of the unicorns' is more naturally perceived as an individual term than as a property. As an individual term, it needs a reference.

Let us analyze our examples according to a presupposition account. Write '$P(x)$' for 'I met x' and '$Q(x)$' for 'x is a unicorn'. We can then write 'I met one of the unicorns' as '$P(\varepsilon x Q(x))$'. This expression will be admitted as a proposition (under some assumptions) in our system only if $\exists x Q(x)$ can be proved (under the same assumptions). Hence, the inference rule

$$\frac{P(\varepsilon x Q(x))}{\exists x Q(x)}$$

will be *admissible* in the sense that whenever the premise can be derived from certain assumptions, so can the conclusion. But we cannot derive $P(\varepsilon x Q(x)) \supset \exists x Q(x)$, unless we can derive $\exists x Q(x)$.

The natural formalization of 'I met a unicorn' is, as has been argued above, the same as the formalization of 'I met something, which was a unicorn', or 'I met something, and the thing was a unicorn': $\exists x P(x)\ \&\ Q(\varepsilon x P(x))$. It can be proved that this is a proposition in the system we are about to study, and that it implies $\exists x Q(x)$. Furthermore, the interpretation we will propose will interpret $\varepsilon x Q(x)$ as referring to the individual claimed to be a unicorn, in correspondence with how we understand the *claimer* when he speaks about 'the unicorn'. We return to this example in Section 7.

The attitude we will take towards descriptions is thus that they are allowed only when they refer to individuals (possibly under assumptions). Formal systems for definite descriptions based on this idea were introduced by Stenlund [23, 24], who also argued philosophically for this view in a much more elaborate way than we have done above. We will introduce systems that are similar to Stenlund's intuitionistic one, but we treat both ε-terms and $\imath$-terms and choose formulations which are suitible for interpretations in Martin-Löf's intensional type theory.

Martin-Löf [12, p. 45] noted that, in his type theory, there is a natural way of making sense of indefinite descriptions by observing that the rules

$$\frac{\exists x A(x)}{\varepsilon x A(x) : I} \qquad \frac{\exists x A(x)}{A(\varepsilon x A(x))}$$

can be viewed as special cases of the rules

$$\frac{p : \exists x A(x)}{\pi_\ell(p) : I} \qquad \frac{p : \exists x A(x)}{\pi_r(p) : A(\pi_\ell(p))}$$

if we make the definition $\varepsilon x A(x) \stackrel{\text{def}}{=} \pi_\ell(p)$, where $\pi_\ell(p)$ is the left projection of the existence proof p. This is the idea also behind the interpretation we are about to consider here. Unfortunately, it is not as easy as it might seem at a first sight. The reason is that in replacing '$\pi_\ell(p)$' by '$\varepsilon x A(x)$', we remove the p, which can be crucial. For instance, consider the following derivation, which looks quite harmless:

$$\dfrac{P(a) \vee P(b) \qquad \dfrac{\dfrac{[P(a)]}{\exists x P(x)}}{P(\varepsilon x P(x))} \qquad \dfrac{\dfrac{[P(b)]}{\exists x P(x)}}{P(\varepsilon x P(x))}}{P(\varepsilon x P(x)).}$$

When we replace all ε-terms by the left projections of the corresponding existence proofs, we get the following (after reduction):

$$\dfrac{P(a) \vee P(b) \qquad \dfrac{\dfrac{[P(a)]}{\exists x P(x)}}{P(a)} \qquad \dfrac{\dfrac{[P(b)]}{\exists x P(x)}}{P(b)}}{?.}$$

So we don't get the same proposition twice above the final line, as we want to. Because of the intensionality, the problem remains also when we have a unique element satisfying P: consider for instance natural numbers and take $a \stackrel{\text{def}}{=} n + 0$ and $b \stackrel{\text{def}}{=} 0 + n$. Then $P(n + 0)$ and $P(0 + n)$ are equivalent, but they are not definitionally equal propositions in intensional type theory. Fortunately, it turns out that this problem can be solved, as will be shown in Section 6.

We note some differences as compared with other first order systems enriched by ε-terms. Maehara [9, 10, 11] and Shirai [22] introduce ε-terms and prove that their calculi are conservative extensions of predicate logic. We want more, we want to interpret ε-terms as individuals, which cannot be done constructively for Maehara's and Shirai's systems. The reason is that ε-terms are allowed in the systems without any presuppositions (but $A(\varepsilon x A(x))$ is required to be true only if $\exists x A(x)$ is true), and this solution makes it nonconstructive to view the ε-terms as individuals. To see this, consider the real numbers and the term $\varepsilon y(xy = 1)$,[3] which we denote by x^{-1}. Because this term depends on x it would be interpreted as a function of x, and because there is no presuppositions on the introduction of x^{-1}, the function would be total. We use the notation x^{-1} also for this function and note that it would satisfy $\forall x(x \neq 0 \supset xx^{-1} = 1)$, where $\neq$ is Heyting's apartness relation [7, 13]. But the existence of such a total function implies LPO.[4] To see this, let $\{a_n\}$ be a binary sequence and let $a = \sum 2^{-n} a_n$. We have either $aa^{-1} \neq 0$ or $aa^{-1} \neq 1$. In the first case, $a \neq 0$, so some of the a_n must be 1. In the second case, we must have $a = 0$ (because if $a \neq 0$ it follows from $\forall x(x \neq 0 \supset xx^{-1} = 1)$ that $aa^{-1} = 1$, which contradicts $aa^{-1} \neq 1$), so no a_n can be 1.

Also Leivant [8] and Mints [14] allow ε-terms without presuppositions, but their ε-terms can be constructively viewed as individuals because of restrictions on their use. These restrictions make $\forall x(x \neq 0 \supset xx^{-1} = 1)$ underivable from

[3] In [11] ε-terms are not allowed to contain free variables, but x may be viewed as a constant symbol in this example.

[4] Bishop's 'limited principle of omniscience' LPO says that in every infinite binary sequence, either a 1 occurs, or no 1 occurs. [3]

$\forall x(x \neq 0 \supset \exists y(xy = 1))$. In Leivant's system, this is because the assumption $x \neq 0$ may not be discharged when x^{-1} occurs in the conclusion. In the system of Mints, $\varepsilon x A(x)$ may occur in the end of a derivation only if the universal closure of $\exists x A(x)$ is assumed, so no result about x^{-1} can be derived unless the false proposition $\forall x \exists y(xy = 1)$ is assumed. We shall see that $\forall x(x \neq 0 \supset xx^{-1} = 1)$ is derivable in the system under consideration in this paper, which is constructively acceptable because we allow ε-terms only under presuppositions.

Let us compare the explanations why the following derivation is erroneous:

$$\cfrac{\cfrac{\cfrac{[\exists x A(x)]^1}{A(\varepsilon x A(x))}}{\exists x A(x) \supset A(\varepsilon x A(x))}\,1}{\exists y(\exists x A(x) \supset A(y)).}$$

Leivant forbids the second step (discharging the assumption), because the ε-term 'depends' on $\exists x A(x)$. Mints agrees that the second step is the flaw provided x is the only free variable in $A(x)$, otherwise his system forbids already the first step. The system studied in this paper blames the third step, because the existence introduction requires that $\varepsilon x A(x)$ refers, which presupposes that $\exists x A(x)$ is true.

In some sense, our system looks a bit like Scott's E-logic [21], in which one has an 'existence predicate' E saying that a term refers.[5] The difference is that instead of an existence *predicate* we use a *judgement*, as did Stenlund [23, 24]. This seems necessary for an interpretation in type theory. In E-logic, definite descriptions can be introduced in the system and interpreted by elimination [21, sect. 6]. However, as we have argued, this is not the kind of result we are interested in, because it says only that descriptions are in some sense harmless (provided E-logic is, which is far from clear from an intensional point of view). Rather, we would like to explain the intended meaning of descriptions.

We should finally mention a system by Abadi, Gonthier and Werner [1], who take the approach of *extending* type theory with ε-terms with a special operational semantics, giving a non-conservative extension of intuitionistic predicate logic. Our aim is instead to keep the type theory as it is, and show that a first order proof involving descriptions can be interpreted in it. Moreover, our ε-terms are individuals, while theirs are *types*. So, in spite of the connections with type theory and computational meaning, our work has little in common.

§4. Indefinite descriptions. We will set up a natural deduction system for first order logic with indefinite descriptions, very much like the systems for definite descriptions introduced by Stenlund [23, 24]. Characteristic for all these systems is the idea that terms containing descriptions may refer to individuals only under some assumptions, and that they are allowed in the system only when they refer (possibly depending on currently open assumptions). Hence we have to incorporate in the system a possibility to judge that a term t refers to an individual. Such a judgement will read '$t : I$'. Further, we allow as propositions only formulas in

[5]Shirai's system *LDJ* [22] is very similar, with a predicate D instead. It is older than Scott's E-logic but less known.

which all terms refer, so we need also the possibility to judge that a formula A is admitted as a proposition, which we do by saying 'A : prop'.[6] We have also a third form of judgement: the usual one, that a proposition is true (under the open assumptions). As is common in first order logic, we write simply 'A' instead of 'A true'. Because of these three different kinds of judgements, our system is more complicated than natural deduction for ordinary first order logic. It looks in fact already like a piece of type theory.

The rules for forming individuals and propositions are as follows.

$$\frac{t_1 : I \quad \cdots \quad t_n : I}{f(t_1, \ldots, t_n) : I}\, fF \qquad \frac{t_1 : I \quad \cdots \quad t_n : I}{P(t_1, \ldots, t_n) : \mathsf{prop}}\, PF \qquad \frac{\exists x A(x)}{\varepsilon x A(x) : I}\, \varepsilon F$$

$$\frac{A : \mathsf{prop} \quad \begin{matrix}[A] \\ \vdots \\ B : \mathsf{prop}\end{matrix}}{A \,\&\, B : \mathsf{prop}}\, \&F \qquad \frac{A : \mathsf{prop} \quad B : \mathsf{prop}}{A \vee B : \mathsf{prop}}\, \vee F \qquad \frac{}{\bot : \mathsf{prop}}\, \bot F$$

$$\frac{A : \mathsf{prop} \quad \begin{matrix}[A] \\ \vdots \\ B : \mathsf{prop}\end{matrix}}{A \supset B : \mathsf{prop}}\, \supset F \qquad \frac{\begin{matrix}[x : I] \\ \vdots \\ A(x) : \mathsf{prop}\end{matrix}}{\forall x A(x) : \mathsf{prop}}\, \forall F \qquad \frac{\begin{matrix}[x : I] \\ \vdots \\ A(x) : \mathsf{prop}\end{matrix}}{\exists x A(x) : \mathsf{prop}}\, \exists F$$

The rules fF and PF are schemata: every primitive n-ary function f and every primitive n-ary predicate P needs such a rule. In particular, we admit the case $n = 0$, giving us constants. (In this case we have zero premises.)

Two kinds of assumptions are allowed. It may be assumed that a variable ranges over the individuals, $x : I$, or that a proposition A is true. In the latter case, we need to know that A is indeed a proposition, so the rule for making such assumptions read

$$\begin{matrix} A : \mathsf{prop} \\ A. \end{matrix}$$

The absence of an inference bar is chosen to stress that this is not an inference of any kind, but an assumption, where the upper part is what is presupposed in the assumption. When such an assumption is discharged, it is replaced by $[A]$, so discharging assumptions involves erasing their derivations of proposition-hood. Instead, these are often moved to the inference bar of the discharging rule. This happens in $\&F$ and $\supset F$ above, but also in the introduction rule for implication below. In some rules, however, the derivations mentioned are not kept, but simply erased. It is thus wise not to write down the derivations of proposition-hood until the rest of the tree is finished and it is clear which assumptions are still in need of derivations of proposition-hood in the end.

[6] Stenlund used the dichotomy 'formula expression' vs. 'formula'.

We have the following rules for introducing and eliminating logical operations.[7]

$$\frac{A \quad B}{A \,\&\, B}\,\&I \qquad \frac{A \,\&\, B}{A}\,\&E\ell \quad \frac{A \,\&\, B}{B}\,\&Er \qquad \frac{A : \mathsf{prop} \quad \bot}{A}\,\bot E$$

$$\frac{A \quad B : \mathsf{prop}}{A \vee B}\,\vee I\ell \quad \frac{A : \mathsf{prop} \quad B}{A \vee B}\,\vee Ir \qquad \frac{A \vee B \quad \begin{matrix}[A]\\ \vdots \\ C\end{matrix} \quad \begin{matrix}[B]\\ \vdots \\ C\end{matrix}}{C}\,\vee E$$

$$\frac{A : \mathsf{prop} \quad \begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}\,\supset I \quad \frac{A \quad A \supset B}{B}\,\supset E \qquad \frac{\begin{matrix}[x : I]\\ \vdots \\ A(x)\end{matrix}}{\forall x A(x)}\,\forall I \quad \frac{\forall x A(x) \quad t : I}{A(t)}\,\forall E$$

$$\frac{\begin{matrix}[x : I]\\ \vdots \\ A(x) : \mathsf{prop}\end{matrix} \quad t : I \quad A(t)}{\exists x A(x)}\,\exists I \qquad \frac{\exists x A(x) \quad \begin{matrix}[x : I][A(x)]\\ \vdots \\ B\end{matrix}}{B}\,\exists E \qquad \frac{\exists x A(x)}{A(\varepsilon x A(x))}\,\varepsilon$$

In fact, the rule $\exists E$ is redundant because of the rule ε, but we keep it for convenience.

The usual variable restrictions apply. To be precise: a rule which discharges an assumption of the form $x : I$ is allowed only if x does not occur free in any non-discharged assumption in the same sub-derivation, nor in the conclusion of the rule. The notation $A(t)$ is intended to presuppose that t is free for x in $A(x)$. The definition of 'free for' has to be slightly more complicated than in ordinary first order logic, because variables can be bound also in terms. Thus t is free for x in $\varepsilon y A(x, y)$ if x and y are the same variables (because substitution of t for x in $\varepsilon x A(x, x)$ has no effect), or y does not occur free in t and t is free for x in $A(x, y)$.

We introduce also some unusual restrictions on $\supset E$ (*modus ponens*) and $\exists I$. We will soon specify these restrictions, but first, let us consider an example of a derivation in the system, namely of a familiar property of fields:

$$\forall x(x \neq 0 \supset xx^{-1} = 1)\,.$$

To this end, we introduce some abbreviations:

$$x^{-1} \stackrel{\text{def}}{=} \varepsilon y(xy = 1)$$
$$U(x) \stackrel{\text{def}}{=} \exists y(xy = 1)$$
$$\text{FIELD} \stackrel{\text{def}}{=} \forall x(x \neq 0 \supset U(x))\,,$$

which let us state two useful special cases of the rules εF and ε:

$$\frac{U(x)}{x^{-1} : I}\,\varepsilon F \qquad \frac{U(x)}{xx^{-1} = 1.}\,\varepsilon$$

[7] One could consider having more premises in these rules as is explained in an appendix.

We can now make the following formal derivation in the system:

$$
\begin{array}{c}
\vdots \\
\mathrm{FIELD} : \mathsf{prop}
\end{array}
$$

$$
\cfrac{\cfrac{\cfrac{[x : I]^2 \qquad \cfrac{}{0 : I}}{x \neq 0 : \mathsf{prop}} \qquad \cfrac{\cfrac{[x \neq 0]^1 \qquad \cfrac{\mathrm{FIELD} \qquad [x : I]^2}{x \neq 0 \supset U(x)}}{U(x)}}{xx^{-1} = 1}}{x \neq 0 \supset xx^{-1} = 1}\;{\scriptstyle 1}}{\forall x(x \neq 0 \supset xx^{-1} = 1).}\;{\scriptstyle 2}
$$

This illustrates how the system can be used to define and reason about functions that are only partially defined. We devote Section 10 to some comments about such applications.

Let us now turn to the restrictions on $\supset E$ and $\exists I$. They are motivated by the fact that we have similar restrictions in natural language, restrictions which seem to be unavoidable when references of descriptions depend on assumptions.

If the author of a novel lets the detective conclude ‘if the man was in the room, the murder didn’t take place there’, and later on ‘the man was in the room’, it would be correct to conclude ‘the murder didn’t take place there’ only if it is obvious that the term ‘the man’ refers to the same individual both times. So we need a restriction like the following one:

> In $\supset E$, ε-terms occurring in A refer to the same individual in both occurrences of A.

Unfortunately, this restriction is semantical, which is not satisfying in a formal system, which should allow correctness of proofs to be automatically and syntactically checked. Moreover, it is not very precise unless we define carefully what ‘refer to the same individual’ means. Finally, it is in general not decidable. We therefore replace this condition by a syntactical and decidable one which is met only when the semantical one is too.

One such syntactical restriction would be not to allow ε-terms at all in A. This would be sufficient, but far too restrictive. Our solution uses the fact that the *derivation* of $t : I$ determines the individual to which t refers. Hence, we may check that two occurrences of t refer to the same individual by checking that they have equal derivations of $t : I$.

The restriction we will choose is to require that both occurrences of A in $\supset E$ are proved to be propositions in the same ways, after reduction. Consider the situation

$$
\begin{array}{cc}
\mathscr{D}_1 & \mathscr{D}_2
\end{array}
$$
$$
\cfrac{A \qquad A \supset B}{B}\;{\scriptstyle \supset E}
$$

where $\mathscr{D}_1$ and $\mathscr{D}_2$ are derivations, possibly involving hypotheses (we assume that they include also the end formulas A and $A \supset B$, respectively; but we display these for enhanced clarity). We will define derivations $\mathscr{D}_1^*$ and $\mathscr{D}_2^*$ of $A : \mathsf{prop}$ and

$A \supset B$: prop, respectively, in a moment. By inspection of the formation rules, we see that $\mathscr{D}_2^*$ must look as follows:

$$\dfrac{\begin{matrix} \\ \vdots \\ A : \mathsf{prop} \end{matrix} \quad \begin{matrix} [A] \\ \vdots \\ B : \mathsf{prop} \end{matrix}}{A \supset B : \mathsf{prop}}$$

and so we can pick out the derivation of A : prop. We reduce it and compare the result syntactically with the reduced form of $\mathscr{D}_1^*$, and require them to be equal. This is our condition for accepting a use of $\supset E$.

For the existence introduction rule, we have the following situation:

$$\dfrac{\begin{matrix} [x : I] \\ \mathscr{D}_3 \\ A(x) : \mathsf{prop} \end{matrix} \quad \begin{matrix} \mathscr{D}_4 \\ t : I \end{matrix} \quad \begin{matrix} \mathscr{D}_5 \\ A(t) \end{matrix}}{\exists x A(x).}\ \exists I$$

We define the derivation $\mathscr{D}_5^*$ of $A(t)$: prop and compare it with the one obtained when t is substituted for x in the derivation we had of $A(x)$: prop,

$$\begin{matrix} \mathscr{D}_4 \\ t : I \\ \mathscr{D}_3[t/x] \\ A(t) : \mathsf{prop}. \end{matrix}$$

We also require that t be free for x in $\mathscr{D}_3$ so that the substitution works as intended.

The restriction on $\exists I$ is equally needed because an unrestricted use of it would let us derive the unrestricted version of $\supset E$ from the restricted one, assuming the domain I is inhabited:

$$\dfrac{\dfrac{\begin{matrix} \vdots \\ A : \mathsf{prop} \end{matrix} \quad t : I \quad A}{\exists x A}\ \exists I \qquad \dfrac{[A] \quad A \supset B}{B}\ \supset E}{B.}\ \exists E$$

Let us turn to the definition of $\mathscr{D}^*$. It is given in the proof of the following theorem, which is a sharpening of a theorem by Stenlund [23, Theorem 3.2.6].

THEOREM 1. *If there is a derivation $\mathscr{D}$ of A from some assumptions, there is a derivation $\mathscr{D}^*$ of A* : prop *from the* same *assumptions.*

Let us first isolate a lemma, to be used when $\mathscr{D}$ includes a use of $\vee E$. It will also play a crucial role in the interpretation.

LEMMA 2. *If we have a derivation of C* : prop *from an assumption A (and possibly other assumptions Γ) as well as a derivation of C* : prop *from the assumption B (and Γ), we can construct a derivation of C* : prop *from the assumption $A \vee B$ (and Γ).*

PROOF OF THE LEMMA. Say the derivations are $\mathscr{D}_1$ and $\mathscr{D}_2$, respectively. By inspection of the formation-rules, we see that $\mathscr{D}_1$ and $\mathscr{D}_2$ must *end* in the same ways: the only possible differences occur *before* some application of εF. If there is no such application, the two proofs must be identical, and so in particular, they do not use the assumptions A and B unless they are found in Γ. Hence we are done in that case. Now, assume there is at least one application of εF. There has, then, to be some application of εF below which $\mathscr{D}_1$ and $\mathscr{D}_2$ are identical. So they look as follows:

$$
\begin{array}{c} \vdots \\ A : \mathsf{prop} \\ A \\ \vdots \\ \dfrac{\exists x D(x)}{\varepsilon x D(x) : I}\,\varepsilon F \\ \vdots \\ C : \mathsf{prop} \end{array}
\qquad
\begin{array}{c} \vdots \\ B : \mathsf{prop} \\ B \\ \vdots \\ \dfrac{\exists x D(x)}{\varepsilon x D(x) : I}\,\varepsilon F \\ \vdots \\ C : \mathsf{prop} \end{array}
$$

with the lower dotted parts identical. We now use $\vee E$ and get

$$
\begin{array}{c}
\dfrac{\begin{array}{c} \dfrac{\begin{array}{cc} \vdots & \vdots \\ A : \mathsf{prop} & B : \mathsf{prop} \end{array}}{A \vee B : \mathsf{prop}}\,\vee F \\ A \vee B \end{array} \qquad \begin{array}{c} [A] \\ \vdots \\ \exists x D(x) \end{array} \quad \begin{array}{c} [B] \\ \vdots \\ \exists x D(x) \end{array}}{\dfrac{\exists x D(x)}{\varepsilon x D(x) : I}\,\varepsilon F}\,\vee E \\
\vdots \\
C : \mathsf{prop}.
\end{array}
$$

This procedure is repeated if necessary. The result is a derivation with the required properties. ⊣

PROOF OF THE THEOREM. We define $\mathscr{D}^*$ by recursion.

If $\mathscr{D}$ is an assumption which is *discharged* somewhere below in the tree, then A occurs also as a subformula of a premise in the discharging rule (see the derivation rules in which assumptions are discharged), and $\mathscr{D}^*$ is defined with respect to this occurrence.

If $\mathscr{D}$ ends by an *open* assumption

$$
\begin{array}{c} A : \mathsf{prop} \\ A \end{array}
$$

we get $\mathscr{D}^*$ by simply erasing the last step in $\mathscr{D}$.

If $\mathscr{D}$ ends with $\&I$, it looks as follows:

$$
\begin{array}{cc} \mathscr{D}_1 & \mathscr{D}_2 \end{array}
$$
$$
\dfrac{A \quad B}{A \,\&\, B.}\,\&I
$$

We let $\mathscr{D}^*$ be the derivation

$$\dfrac{\begin{matrix}\mathscr{D}_1^* & \mathscr{D}_2^* \\ A : \mathsf{prop} & B : \mathsf{prop}\end{matrix}}{A \mathbin{\&} B : \mathsf{prop}.}\,\&I$$

If $\mathscr{D}$ ends with $\&E\ell$, it is of the form

$$\dfrac{\begin{matrix}\mathscr{D}_1 \\ A \mathbin{\&} B\end{matrix}}{A.}\,\&E\ell$$

But $\mathscr{D}_1^*$ must be of the form

$$\dfrac{\begin{matrix} & [A] \\ \vdots & \vdots \\ A : \mathsf{prop} & B : \mathsf{prop}\end{matrix}}{A \mathbin{\&} B : \mathsf{prop}.}\,\&F$$

and we let $\mathscr{D}^*$ be the piece ending with $A : \mathsf{prop}$.

If $\mathscr{D}$ ends with $\&Er$, it is of the form

$$\dfrac{\begin{matrix}\mathscr{D}_1 \\ A \mathbin{\&} B\end{matrix}}{B.}\,\&Er$$

But $\mathscr{D}_1^*$ must be as in the previous case and we let $\mathscr{D}^*$ be the tree

$$\begin{matrix}\dfrac{\begin{matrix}\mathscr{D}_1 \\ A \mathbin{\&} B\end{matrix}}{A}\,\&E\ell \\ \vdots \\ B : \mathsf{prop}.\end{matrix}$$

If $\mathscr{D}$ ends with $\perp E$, we let $\mathscr{D}^*$ be the left subderivation above the final inference line. Notice that our form of the rule $\perp E$ is essential; the theorem is not true with Stenlund's rule, in which the premise $A : \mathsf{prop}$ is left out. In Stenlund's proof sketch of the corresponding theorem [24, p. 206], this rule is forgotten. Stenlund has agreed that the new form of the rule is probably the right one (private communication, Jan. 30, 2003).

Suppose now that $\mathscr{D}$ ends with $\vee E$,

$$\dfrac{\begin{matrix} & [A] & [B] \\ \vdots & \vdots & \vdots \\ A \vee B & C & C\end{matrix}}{C.}\,\vee E$$

Let us call the sub-derivations $\mathscr{D}_1$, $\mathscr{D}_2$, $\mathscr{D}_3$, respectively. By recursion, we have derivations $\mathscr{D}_2^*$ and $\mathscr{D}_3^*$ of $C : \mathsf{prop}$ from the assumptions A and B, respectively. According to the lemma, we get a derivation of $C : \mathsf{prop}$ from the assumption $A \vee B$. But we have also a derivation $\mathscr{D}_1$ of $A \vee B$. This derivation, followed by the one proving $C : \mathsf{prop}$ from $A \vee B$ is taken to be $\mathscr{D}^*$.

If $\mathscr{D}$ ends with $\exists E$, it has this form:

$$\dfrac{\begin{matrix} \\ \vdots \\ \exists x A(x)\end{matrix} \qquad \begin{matrix}[x : I][A(x)] \\ \vdots \\ B\end{matrix}}{B.}\ \exists E$$

By recursion, we have a derivation of B : prop using the assumptions $x : I$ and $A(x)$, which we now have to get rid of. First notice that the derivation cannot make use of such assumptions unless εF is used. If it is not, we can take $\mathscr{D}^*$ to be the derivation we have. Otherwise, the derivation has the following form:

$$\begin{matrix} x : I \quad A(x) \\ \vdots \\ \dfrac{\exists y D(y)}{\varepsilon y D(y) : I}\ \varepsilon F \\ \vdots \\ B : \mathsf{prop} \end{matrix}$$

where the last dots represent a part where only formation rules are used. We now transform this tree using $\exists E$:

$$\begin{matrix} \dfrac{\begin{matrix} \\ \vdots \\ \exists x A(x)\end{matrix} \qquad \begin{matrix}[x : I][A(x)] \\ \vdots \\ \exists y D(y)\end{matrix}}{\dfrac{\exists y D(y)}{\varepsilon y D(y) : I}\ \varepsilon F}\ \exists E \\ \vdots \\ B : \mathsf{prop} \end{matrix}$$

(the variable restrictions are met: since x does not occur free in B, it cannot occur free in $\exists y D(y)$ either). We repeat this procedure if necessary and the result is $\mathscr{D}^*$.

If $\mathscr{D}$ ends with the ε-rule, it has this form:

$$\begin{matrix} \vdots \\ \dfrac{\exists x A(x)}{A(\varepsilon x A(x)).}\ \varepsilon \end{matrix}$$

By recursion, we have a derivation of $\exists x A(x)$: prop, hence a derivation of $A(x)$: prop from the assumption $x : I$. By substituting $\varepsilon x A(x)$ for x in this one, we get $\mathscr{D}^*$:

$$\begin{matrix} \vdots \\ \dfrac{\exists x A(x)}{\varepsilon x A(x) : I}\ \varepsilon F \\ \vdots \\ A(\varepsilon x A(x)) : \mathsf{prop}. \end{matrix}$$

The other cases are similar. ⊣

Derivations are said to be equal if they can be converted using reductions:

DEFINITION 3. The relation $\approx$ is the smallest equivalence relation between derivations such that

- each redex is $\approx$-equal to its contractum (we consider the usual reductions, as in Prawitz [16, II. § 2., pp. 35–38][8]),
- if derivations are composed by $\approx$-equal subderivations, then they are themselves $\approx$-equal.

The relation $\approx$ is decidable, since one can normalize and compare the normal forms, which are $\approx$-equal if and only if they are syntactically equal. We are now ready to state precisely what restrictions we put on $\supset E$ and $\exists I$.

Restriction on $\supset E$. The inference

$$\dfrac{\begin{matrix}\mathscr{D}_1 & & \mathscr{D}_2 \\ A & & A \supset B\end{matrix}}{B}\supset E$$

is allowed if $\mathscr{D}_2^*$, which is of the form

$$\dfrac{\begin{matrix} & & [A] \\ \mathscr{D}_3 & & \mathscr{D}_4 \\ A : \mathsf{prop} & & B : \mathsf{prop}\end{matrix}}{A \supset B : \mathsf{prop},}\supset F$$

satisfies $\mathscr{D}_3 \approx \mathscr{D}_1^*$.

Restriction on $\exists I$. The inference

$$\dfrac{\begin{matrix}[x : I] & & & \\ \mathscr{D}_1 & & \mathscr{D}_2 & & \mathscr{D}_3 \\ A(x) : \mathsf{prop} & & t : I & & A(t)\end{matrix}}{\exists x A(x)}\exists I$$

[8]Per Martin-Löf suggested to me the reductions

$$\dfrac{\dfrac{\begin{matrix}[x : I] & & \\ \vdots & \vdots & \vdots \\ A(x) : \mathsf{prop} & t : I & A(t)\end{matrix}}{\exists x A(x)}\exists I}{\varepsilon x A(x) : I}\varepsilon F \quad \rightsquigarrow \quad \begin{matrix}\vdots \\ t : I\end{matrix} \qquad\qquad \dfrac{\dfrac{\begin{matrix}[x : I] & & \\ \vdots & \vdots & \vdots \\ A(x) : \mathsf{prop} & t : I & A(t)\end{matrix}}{\exists x A(x)}\exists I}{A(\varepsilon x A(x))}\varepsilon \quad \rightsquigarrow \quad \begin{matrix}\vdots \\ A(t).\end{matrix}$$

Indeed, these reductions will be justified by the translation we are about to define in the next section, but we omit them anyway. There are several reasons for this. One is that they make the system much more complicated because they introduce the phenomenon that equal proofs may end with syntactically different propositions. Another is that they make $A(t) \supset (\varepsilon x A(x) = t)$ derivable if $A(t) : \mathsf{prop}$ and $t : I$ are derivable. In particular, $\varepsilon x \sim\perp = t$ becomes derivable if $t : I$ is derivable and $=$ is a reflexive relation. Therefore, $\forall x \forall y (x = y)$ would be derivable if we had unrestricted modus ponens. But there are natural non-trivial models of the ε-calculus which model also unrestricted modus ponens. Consider for instance a finite ordered set with decidable predicates, so that excluded middle is valid. Let $\varepsilon x A(x)$ be interpreted as the *first* individual satisfying A in case there is such an element, otherwise as the last element in the set. This interpretation models our system with unrestricted modus ponens, so it does not model the reductions under consideration. We prefer the system to be sound with respect to models of this kind.

is allowed if t is free for x in $\mathscr{D}_1$ and the following derivation is $\approx$-equal to D_3^*.

$$\begin{array}{c} \mathscr{D}_2 \\ t : I \\ \mathscr{D}_1[t/x] \\ A(t) : \mathsf{prop} \end{array}$$

It is thus a trivial but not in general convenient task to decide if an instance of $\supset E$ or $\exists I$ meets the requirement. It seems that human intuition is quite good at guessing right (we know what we refer to by descriptions), so that a human informal derivation often includes only acceptable instances, but it would be desirable to let a computer check the derivations. We will return to these issues in Section 6. Notice however that the conditions are always satisfied when there are no ε-terms at all involved, because then there is, for each formula, at most one way of deriving that it is a proposition.

In Section 8 we will show that if we use *definite* descriptions only, no restrictions will be necessary, except for the usual variable restrictions. In fact, we give a process which repairs all illegal applications of $\supset E$ and $\exists I$. This process works also partially for indefinite descriptions, but not in general.

Observe that $\vee E$ (only!) is similar to $\supset E$ and $\exists I$ in that the same proposition occurs twice among the premises. Surprisingly, no restriction is needed in this case, as the translation given in Section 6 will show.

§**5. Equality.** We have omitted rules for equality, because ε-terms make it possible to define functions that do not preserve equality ('non-extensional' functions). Hence equality is not very natural in the system. However, equality can be introduced as any binary predicate (write '$t = s$' for $P(t, s)$). Reasoning with equality is then performed by using an axiom[9] $\forall x(x = x)$ for reflexivity, and for each primitive predicate $P(x_1, \dots, x_n)$ (with $n \geq 1$), including the binary primitive predicate $=$, an axiom

(P ext)

$$\forall x_1 \cdots \forall y_n \big((x_1 = y_1 \,\&\, \cdots \,\&\, x_n = y_n) \supset (P(x_1, \dots, x_2) \supset P(y_1, \dots, y_2))\big)\,,$$

and, finally, for each n-ary primitive function (with $n \geq 1$) an axiom

(f ext)

$$\forall x_1 \cdots \forall y_n \big((x_1 = y_1 \,\&\, \cdots \,\&\, x_n = y_n) \supset (f(x_1, \dots, x_2) = f(y_1, \dots, y_2))\big)\,.$$

The rules for symmetry and transitivity can then be derived, but the replacement rule will *not* in general be justified by the interpretation in type theory (but see Section 8).

[9]An *axiom* is here formally the same as an assumption. There is, however, a difference as regards to their interpretations (sect. 6): axioms are interpreted as *proved propositions*, while assumptions are interpreted as assumptions also in type theory. Therefore, axioms can always be used at no cost, unlike assumptions.

§6. **The translation into type theory.** We now turn to the translation of the ε-calculus into intensional type theory, as presented in [15], where also the translation of first order logic is explained. We concentrate here on the things that have to be changed rather than defining the translation from the beginning.

First of all, it is necessary to understand that the translation of a proposition A will be determined, not by its syntactical form, but by the *derivation* of A : prop. For example, consider the derivations

$$\dfrac{\dfrac{\dfrac{\begin{matrix}[x:I]\\ \vdots\\ P(x):\mathsf{prop}\end{matrix}\quad \begin{matrix}\vdots\\ a:I\end{matrix}\quad \begin{matrix}\vdots\\ P(a)\end{matrix}}{\exists x P(x)}}{\varepsilon x P(x):I}}{P(\varepsilon x P(x)):\mathsf{prop}} \qquad \dfrac{\dfrac{\dfrac{\begin{matrix}[x:I]\\ \vdots\\ P(x):\mathsf{prop}\end{matrix}\quad \begin{matrix}\vdots\\ b:I\end{matrix}\quad \begin{matrix}\vdots\\ P(b)\end{matrix}}{\exists x P(x)}}{\varepsilon x P(x):I}}{P(\varepsilon x P(x)):\mathsf{prop}.}$$

The end formula '$P(\varepsilon x P(x))$' will in the left case be translated into a proposition which is definitionally equal to the translation of $P(a)$, while in the right case it will be translated into a proposition which is definitionally equal to the translation of $P(b)$. However, we will have the following facts.

PROPOSITION 4. *If $\mathscr{D}_1$ and $\mathscr{D}_2$ are derivations of A : prop and $\mathscr{D}_1 \approx \mathscr{D}_2$, then both occurrences of A are translated into definitionally equal propositions.*

PROOF. Each derivation will be translated into a derivation in type theory, and each redex will correspond to a redex in type theory. ⊣

PROPOSITION 5. *If $\mathscr{D}_1$ and $\mathscr{D}_2$ are derivations of A and $\mathscr{D}_1 \approx \mathscr{D}_2$, then both occurrences of A are translated into definitionally equal propositions.*

PROOF. Both derivations will be translated into derivations in type theory of $p : A_1$ and $q : A_2$, respectively. Since p and q must be definitionally equal, also A_1 and A_2 must be definitionally equal, by the monomorphic property of type theory [15]. ⊣

PROPOSITION 6. *If $\mathscr{D}_1$ is a derivation of A, then it is translated into a derivation of a judgement of the form $p : A_1$ (in some context), where A_1 is such that the derivation $\mathscr{D}^*$ is translated into a derivation of A_1 : prop (in the same context).*

PROOF. This follows immediately from the definition of $\mathscr{D}^*$ and the following definition of the translation. ⊣

We now give the translation. Fix a set I, which will act as the domain of discourse, i.e., as the interpretation of the symbol 'I'. For each n-ary function symbol f there has to correspond an n-ary function on I. We denote it by the same symbol, i.e., we write $f : (I, \dots, I)I$. Correspondingly, for each n-ary predicate symbol P there has to correspond an n-ary propositional function $P : (I, \dots, I)\mathsf{prop}$. These requirements in themselves justify the rules fF and PF.

The equality is supposed to be interpreted as an equivalence relation $=_I$ on I. The axioms for equality are interpreted in the obvious ways. For instance, the reflexivity axiom $\forall x(x = x)$ is interpreted as $\lambda(\text{refl}) : (\forall x : I)(x =_I x)$. This is the reason why we call them *axioms*, rather than *assumptions*: they are interpreted as *proved* propositions in contrast to assumptions, which must in general be interpreted as assumptions also in type theory; i.e., an assumption A is interpreted as A true, or rather, as $p : A$, where p is a fresh variable.

The rules εF and ε are interpreted as Martin-Löf [12, p. 45] proposed:

$$\frac{p : \exists x A(x)}{\pi_\ell(p) : I} \qquad \frac{p : \exists x A(x)}{\pi_r(p) : A(\pi_\ell(p)).}$$

The rule $\&F$ is interpreted as the general rule of Σ-formation

$$\frac{A : \mathsf{prop} \quad B(x) : \mathsf{prop}\ (x : A)}{\Sigma(A, B) : \mathsf{prop},}$$

rather than the specialized one where $B(x)$ is not allowed to depend on x. Likewise for the rule $\supset F$, which is interpreted as the general rule of Π-formation:

$$\frac{A : \mathsf{prop} \quad B(x) : \mathsf{prop}\ (x : A)}{\Pi(A, B) : \mathsf{prop}.}$$

The other formation rules are interpreted as usual [15].

Among the introduction rules, the only one needing a new idea is $\exists I$. The reason is that this rule has two occurrences of A and t among the premises, and we cannot be sure that they have been interpreted in the same ways. We may therefore be faced with a situation where we would need a rule like this one:

$$\frac{A_1(x) : \mathsf{prop}\ (x : I) \quad t_1 : I \quad A_2(t_2) : \mathsf{prop} \quad p : A_2(t_2)}{(\exists x : I)A_1(x) \text{ true},}$$

which is *not* a valid rule in type theory. However, the restriction put on $\exists I$ in Section 4 gives us $A_2(t_2) = A_1(t_1) : \mathsf{prop}$, so that we get

$$\frac{A_1(x) : \mathsf{prop}\ (x : I) \quad t_1 : I \quad \dfrac{p : A_2(t_2) \quad A_2(t_2) = A_1(t_1) : \mathsf{prop}}{p : A_1(t_1)}}{(t_1, p) : (\exists x : I)A_1(x),}$$

which *is* a valid derivation in type theory.

Among the elimination rules, there are two that need some care: $\supset E$ and $\vee E$. Let us begin with the former one. After having interpreted the premises, we might need a rule like this one:

$$\frac{p : A_1 \quad q : \Pi(A_2, B)}{B(p) \text{ true},}$$

which we do not have. Again, the restrictions put on $\supset E$ in Section 4 gives us $A_1 = A_2 : \mathsf{prop}$ and we may use the following derivation, which is indeed valid:

$$\frac{\dfrac{p : A_1 \quad A_1 = A_2 : \mathsf{prop}}{p : A_2} \qquad q : \Pi(A_2, B)}{\mathsf{app}(q, p) : B(p).}$$

Finally, we should interpret $\vee E$. Assume we are interpreting a derivation $\mathscr{D}$ which ends with $\vee E$. We have the following situation, after having interpreted the premises:

$$\frac{c : A \vee B \quad \begin{array}{c}[x : A]\\ d(x) : C_1\end{array} \quad \begin{array}{c}[y : B]\\ e(y) : C_2\end{array}}{?.}$$

It might well be, as an example on page 492 showed, that C_1 and C_2 are different propositions, so it is not obvious what should be put as conclusion. Looking at the corresponding rule in type theory we see that one premise is lacking:

$$\frac{c : A \vee B \quad \begin{array}{c}[z : A \vee B]\\ C(z) : \mathsf{prop}\end{array} \quad \begin{array}{c}[x : A]\\ d(x) : C(\mathsf{inl}(x))\end{array} \quad \begin{array}{c}[y : B]\\ e(y) : C(\mathsf{inr}(y))\end{array}}{\mathsf{when}(c, d, e) : C(c).}$$

Hence we have to come up with a propositional function C with $C(\mathsf{inl}(x)) = C_1$ and $C(\mathsf{inr}(y)) = C_2$, and such that $C(c)$ is the proposition we get from the interpretation of the derivation $\mathscr{D}^*$ of $C : \mathsf{prop}$, defined in the proof of Theorem 1.

By this theorem, we have derivations of $C : \mathsf{prop}$ from the assumptions A and B, respectively. By Lemma 2, this gives us a derivation of $C : \mathsf{prop}$ from the assumption $A \vee B$. Interpreting this derivation, we get a propositional function $C(z) : \mathsf{prop}\ (z : A \vee B)$. Now, the derivation $\mathscr{D}^*$ was defined by substituting the derivation of $A \vee B$ for the assumption (see the proof of Theorem 1). We were assuming that the derivation of $A \vee B$ had already been interpreted, yielding $c : A \vee B$. Thus, $\mathscr{D}^*$ corresponds precisely to $C(c)$, as we required.

Further, $C(\mathsf{inl}(x))$ corresponds to replacing the assumption $A \vee B$ by the derivation

$$\frac{A \quad B : \mathsf{prop}}{A \vee B.}\vee I\ell$$

in the derivation of $C : \mathsf{prop}$. Reducing the resulting derivation, we get the original derivation from A to $C : \mathsf{prop}$ back. Hence, by Proposition 4, $C(\mathsf{inl}(x)) = C_1$. An analogous argument shows that $C(\mathsf{inr}(y)) = C_2$.

It probably helps to consider an example: the 'problematic' derivation mentioned before (p. 492) is interpreted as follows:

$$\frac{c : P(a) \vee P(b) \quad \begin{array}{c}[y : P(a) \vee P(b)]\\ \vdots\\ P(\pi_\ell(\mathsf{when}(y, (z)(a, z), (z)(b, z)))) : \mathsf{prop}\end{array} \quad \dfrac{\dfrac{\begin{array}{c}[x : I]\\ P(x) : \mathsf{prop}\end{array} \ a : I\ [z : P(a)]}{(a, z) : (\exists x : I)P(x)}}{z : P(a)} \quad \dfrac{\dfrac{\begin{array}{c}[x : I]\\ P(x) : \mathsf{prop}\end{array} \ b : I\ [z : P(b)]}{(a, z) : (\exists x : I)P(x)}}{z : P(b)}}{\mathsf{when}(c, (z)z, (z)z) : P(\pi_\ell(\mathsf{when}(c, (z)(a, z), (z)(b, z)))).}$$

§7. Unicorns. It was promised in Section 3 that we return to unicorns after having defined the translation. Recall that '$P(x)$' means 'I met x' and '$Q(x)$' means 'x is a unicorn'. Hence 'I met one of the unicorns' can be formalized as '$P(\varepsilon x Q(x))$'. In order for this to be a proposition, we must have $\varepsilon x Q(x) : I$, hence we must assume, or prove, $\exists x Q(x)$. So if we can prove that $P(\varepsilon x Q(x))$ is a proposition, we must have assumptions enough for proving also $\exists x Q(x)$.

More interesting is 'I met a unicorn', perceived as 'I met something, and the thing was a unicorn'. Formally, this is $\exists x P(x)\ \&\ Q(\varepsilon x P(x))$. It can be proved to be a proposition in the following way:

$$\dfrac{\dfrac{\dfrac{[x:I]}{P(x):\mathsf{prop}}\,PF}{\exists xP(x):\mathsf{prop}}\,\exists F \qquad \dfrac{\dfrac{[\exists xP(x)]}{\varepsilon xP(x):I}\,\varepsilon F}{Q(\varepsilon xP(x)):\mathsf{prop}}\,QF}{\exists xP(x)\;\&\;Q(\varepsilon xP(x)):\mathsf{prop}.}\,\&F$$

This, as may be checked by the reader, makes $\exists xP(x)\;\&\;Q(\varepsilon xP(x))$ interpreted in type theory as

$$(\exists p : (\exists x : I)P(x))Q(\pi_\ell(p))\,.$$

We may derive from this proposition that there are unicorns (notice that the restriction on $\exists I$ is met!):

$$\dfrac{\dfrac{[x:I]}{Q(x):\mathsf{prop}}\,QF \qquad \dfrac{\dfrac{\begin{array}{c}\vdots\\ \exists xP(x)\;\&\;Q(\varepsilon xP(x)):\mathsf{prop}\\ \exists xP(x)\;\&\;Q(\varepsilon xP(x))\end{array}}{\exists xP(x)}\,\&E}{\varepsilon xP(x):I}\,\varepsilon F \qquad \dfrac{\begin{array}{c}\vdots\\ \exists xP(x)\;\&\;Q(\varepsilon xP(x)):\mathsf{prop}\\ \exists xP(x)\;\&\;Q(\varepsilon xP(x))\end{array}}{Q(\varepsilon xP(x))}\,\&E}{\exists xQ(x)}\,\exists I$$

Hence, $\varepsilon xQ(x)$ is interpreted as

$$\pi_\ell(\pi_\ell(\pi_\ell(q)), \pi_r(q)) : I \ (q : (\exists p : (\exists x : I)P(x))Q(\pi_\ell(p)))\,,$$

which reduces to

$$\pi_\ell(\pi_\ell(q)) : I \ (q : (\exists p : (\exists x : I)P(x))Q(\pi_\ell(p)))\,.$$

In particular, if $a : I$, $b : P(a)$, and $c : Q(a)$, we may take $q \overset{\text{def}}{=} ((a, b), c)$ and so $\varepsilon xQ(x)$ is interpreted as something which is definitionally equal to a. This is to say, if the claim 'I met an individual which was a unicorn' is in fact true, then $\varepsilon xQ(x)$ refers to that individual.

§8. Definite descriptions. We now extend our system with definite descriptions, obtaining a system very similar to Stenlund's [23, 24], except for some minor changes, and the fact that we can have ε-terms simultaneous in the system.

When no ε-terms are involved, but we have definite descriptions only, some new inference rules can be justified, for instance unrestricted modus ponens, as we will show. One could therefore introduce these rules in the system if ε-terms were abandoned. This would, however, make it necessary to extend also the translation, which is much more complicated than one would expect. We will therefore instead *derive* the new rules. In other words, we show that a $\imath$-calculus *with* the new rules formally added, can be interpreted in the ε-calculus. The translation from the $\imath$-calculus to the ε-calculus then works by removing all applications of the new rules, replacing them by derivations of these rules, and replacing also every $\imath$ by an ε.

Formally, we add the following two rules only:

$$\dfrac{\exists xA(x) \qquad \forall x\forall y(A(x)\;\&\;A(y) \supset x = y)}{\imath xA(x) : I}\,\imath F$$

$$\frac{\exists x A(x) \qquad \forall x \forall y (A(x) \,\&\, A(y) \supset x = y)}{A(\iota x A(x)).}\,\iota$$

Equality is treated and interpreted as in Section 5. In particular, all primitive relation symbols and function symbols are supposed to be interpreted as equality-preserving relations and functions. Hence we assume that we have the axioms for reflexivity and extensionality (P ext) (f ext) at hand, so that we need no special inference rules for equality. When we use 'Γ' to denote an arbitrary set of assumption formulas, we will always assume that it contains the necessary axioms for equality, because we can do so at no cost, keeping in mind that we know how to interpret axioms.

The definition of $\mathscr{D}^*$ is extended in the obvious way: a derivation that ends with the ι-rule is replaced by the same derivation but ending with the ιF-rule, followed by the derivation from $\iota x A(x) : I$ to $A(\iota x A(x)) : \mathsf{prop}$ (for details, see the case of the ε-rule in the definition of $\mathscr{D}^*$, Theorem 1).

Terms of the form $\iota x A(x)$ are interpreted exactly as $\varepsilon x A(x)$. In particular, the interpretations of the rules ι and ιF do not make use of the premises about uniqueness. These premises are there solely because they allow us to prove the following meta-mathematical result.

THEOREM 7. *The following rules are derivable*:

1. *Unrestricted modus ponens*

$$\frac{A \qquad A \supset B}{B}$$

when A does not contain any ε-term.

2. *Unrestricted*[10] *existence introduction*

$$\frac{\begin{array}{c}[x : I] \\ \vdots \\ A(x) : \mathsf{prop}\end{array} \qquad t : I \quad A(t)}{\exists x A(x)}$$

when $A(t)$ does not contain any ε-term.

3. *The replacement rule*

$$\frac{A(t) \quad t = s}{A(s)}$$

when $A(t)$ does not contain any ε-term.

We need a number of lemmas.

LEMMA 8. *Assume A does not contain any ε-term. If $\mathscr{D}_1, \mathscr{D}_2$ are derivations $\Gamma \vdash A : \mathsf{prop}$, then there is a derivation $\mathscr{D} : \Gamma \vdash A \supset A$ such that $\mathscr{D}^*$ is*

$$\frac{\begin{array}{c}\mathscr{D}_1 \\ A : \mathsf{prop}\end{array} \quad \begin{array}{c}\mathscr{D}_2 \\ A : \mathsf{prop}\end{array}}{A \supset A : \mathsf{prop}.}\,\supset F$$

[10]By *unrestricted*, we mean without restrictions on how $A(x) : \mathsf{prop}$ and $t : I$ are derived. It is still, of course, necessary to have variable restrictions.

PROOF. First note that if A does not contain any ι-term, the lemma is trivial (since, by assumption, A does not contain ε-terms either). We need to take care of the case when A contains some ι-term(s). We will do that by making a simultaneous induction over terms and formulas.

Define the following measure:

$$\begin{aligned}
\mu(x) &= 0 && \text{for variables } x\\
\mu(f(t_1,\dots,t_n)) &= \max \mu(t_i) + 1\\
\mu(\iota x A(x)) &= \mu(A(x)) + 1\\
\mu(P(t_1,\dots,t_n)) &= \max \mu(t_i) + 1 && \text{`$=$' included}\\
\mu(A \mathbin{\&} B) &= \max(\mu(A), \mu(B)) + 1\\
\mu(A \vee B) &= \max(\mu(A), \mu(B)) + 1\\
\mu(A \supset B) &= \max(\mu(A), \mu(B)) + 1\\
\mu(\bot) &= 1\\
\mu(\forall x A(x)) &= \mu(A(x)) + 1\\
\mu(\exists x A(x)) &= \mu(A(x)) + 1
\end{aligned}$$

When $\mu(A) = 1$, the lemma is obvious, because there is no ι-term in A. Suppose the lemma is proved for $\mu(A) < N$ (with $N \geq 2$), we shall prove it when $\mu(A) = N$. There are a number of cases.

A is of the form $P(t_1,\dots,t_n)$, with $n \geq 1$. Use (P ext) to reduce the problem to proving that for given derivations $\mathscr{D}_3$, $\mathscr{D}_4$ of $t : I$ from Γ, there is a derivation $\mathscr{D}$ of $t = t$ from Γ such that $\mathscr{D}^*$ is

$$\dfrac{\begin{matrix}\mathscr{D}_3 & \mathscr{D}_4\\ t : I & t : I\end{matrix}}{t = t : \mathsf{prop.}}\,{=}F$$

If t is a variable, or a constant, it is trivial. If t is $f(t_1,\dots,t_n)$, with $n \geq 1$, we reduce the problem to deriving $t_i = t_i$ in an appropriate way. Because

$$\mu(t_i = t_i) = \mu(t_i) + 1 \leq \mu(t) < N\,,$$

it follows by the induction hypothesis that $t_i = t_i \supset t_i = t_i$ can be derived in an appropriate way, so that the following derivation solves the problem:

$$\dfrac{\dfrac{\forall x(x = x) \quad \begin{matrix}\vdots\\ t_i : I\end{matrix}}{t_i = t_i}\,\forall E \qquad \begin{matrix}\vdots\\ t_i = t_i \supset t_i = t_i\end{matrix}}{t_i = t_i.}\,{\supset}E$$

It remains the case when t is of the form $\iota x B(x)$ (the form $\varepsilon x B(x)$ is excluded by assumption). Then $\mathscr{D}_3$ and $\mathscr{D}_4$ end by the rule ιF. Let $\mathscr{D}_3'$ and $\mathscr{D}_4'$ be the derivations of $B(t)$ obtained by replacing the final rule ιF by the rule ι in $\mathscr{D}_3$ and $\mathscr{D}_4$, respectively. Since there is a derivation of $\forall x \forall y(B(x) \mathbin{\&} B(y) \supset x = y)$ in $\mathscr{D}_3$ (and $\mathscr{D}_4$), it suffices to derive $B(t) \mathbin{\&} B(t)$ in a way which makes it possible to apply

$\supset E$. This can be done as follows (the careful reader will notice that this derivation is not reduced, simply because the argument will be easier).

$$\dfrac{\dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_3'}{B(t)} \quad \dfrac{\dfrac{\genfrac{}{}{0pt}{}{[x:I]}{\genfrac{}{}{0pt}{}{\vdots}{B(x)\supset B(x)}}}{\forall x(B(x)\supset B(x))}{\scriptstyle\forall I} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_3}{t:I}}{B(t)\supset B(t)}{\scriptstyle\forall E}}{B(t)}{\scriptstyle\supset E} \quad \dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_4'}{B(t)} \quad \dfrac{\dfrac{\genfrac{}{}{0pt}{}{[x:I]}{\genfrac{}{}{0pt}{}{\vdots}{B(x)\supset B(x)}}}{\forall x(B(x)\supset B(x))}{\scriptstyle\forall I} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_4}{t:I}}{B(t)\supset B(t)}{\scriptstyle\forall E}}{B(t)}{\scriptstyle\supset E}}{B(t)\,\&\,B(t).}{\scriptstyle\&I}$$

It is straightforward to check that both applications of $\supset E$ are allowed if the dotted parts are filled in appropriately, which is possible to do according to the induction hypothesis, because $\mu(B(x)) < \mu(t) < N$.

A is of the form $B\,\&\,C$. The derivation then looks as follows:

$$\dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_1}{B\,\&\,C : \mathsf{prop}} \quad \dfrac{\dfrac{\dfrac{[B\,\&\,C]}{B} \quad \genfrac{}{}{0pt}{}{\vdots}{B\supset B}}{B} \quad \dfrac{\dfrac{[B\,\&\,C]}{C} \quad \genfrac{}{}{0pt}{}{\vdots}{C\supset C}}{C}}{B\,\&\,C}}{B\,\&\,C\supset B\,\&\,C}$$

where the derivations of $B \supset B$ and $C \supset C$ exist by the induction hypothesis. The derivation of $C \supset C$ may depend on assumptions B, but these can be derived, in appropriate ways, from $B\,\&\,C$ (which is currently open and hence available) and $B \supset B$. It is easy to verify that $\mathscr{D}^*$ has the stated form.

Implication is treated like conjunction. The other cases are even easier and therefore omitted. ⊣

This lemma is enough for proving the first two parts of the theorem.

Proof of Theorem 7, parts 1–2. For part 1, use the transformation

$$\dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_1}{A} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_2}{A\supset B}}{B} \quad \leadsto \quad \dfrac{\dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_1}{A} \quad \genfrac{}{}{0pt}{}{\vdots}{A\supset A}}{A}{\scriptstyle\supset E} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_2}{A\supset B}}{B}{\scriptstyle\supset E}$$

and the lemma to conclude that the dots can be filled in.

For part 2, use the transformation

$$\dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_1}{A(x):\mathsf{prop}} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_2}{t:I} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_3}{A(t)}}{\exists x A(x)} \quad \leadsto \quad \dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_1}{A(x):\mathsf{prop}} \quad \genfrac{}{}{0pt}{}{\mathscr{D}_2}{t:I} \quad \dfrac{\genfrac{}{}{0pt}{}{\mathscr{D}_3}{A(t)} \quad \genfrac{}{}{0pt}{}{\vdots}{A(t)\supset A(t)}}{A(t)}{\scriptstyle\supset E}}{\exists x A(x).}{\scriptstyle\exists I}$$

⊣

In the following, we will therefore freely use modus ponens as a derived rule. In particular, to prove the third part of the theorem, we need only prove that $A(t) \supset A(s)$ is derivable from Γ if $A(t)$ and $t = s$ are derivable from Γ. To do that, we need a couple of more lemmas.

LEMMA 9. *Assume $A(t)$ does not contain any ε-term, nor does s, and that* $\Gamma \vdash A(t) : \mathsf{prop}, A(s) : \mathsf{prop}, t = s$. *Then* $\Gamma \vdash A(t) \supset A(s)$.

PROOF. By induction on μ, as defined in the proof of the previous lemma. We assume that the lemma is true whenever $\mu(A(t) \supset A(s)) < N$, and prove it in the case $\mu(A(t) \supset A(s)) = N$.

If $A(t)$ is atomic, $(P \text{ ext})$ reduces the problem to deriving $t_i(t) = t_i(s)$. If t_i is a variable or a constant, it is trivial. If t_i is of the form $f(\dots)$, $(f \text{ ext})$ reduces the problem to the arguments of f. Finally, if $t_i(t)$ is of the form $\imath x B(t,x)$ we need to derive $\imath x B(t,x) = \imath x B(s,x)$.

Because $\Gamma \vdash \imath x B(s,x) : I$, we have $\Gamma \vdash \forall x \forall y (B(s,x) \mathbin{\&} B(s,y) \supset x = y)$, so it suffices to derive $\Gamma \vdash B(s, \imath x B(t,x)) \mathbin{\&} B(s, \imath x B(s,x))$. The second conjunct is easy to derive: just change the last rule in the derivation $\Gamma \vdash \imath x B(s,x) : I$. In order to derive $\Gamma \vdash B(s, \imath x B(t,x))$, use that $\mu(B(t,x) \supset B(s,x)) < \mu(t_i(t)) + \mu(t_i(s)) < \mu(A(t)) + \mu(A(s)) < \mu(A(t) \supset A(s))$, so the induction hypothesis gives $x : I, \Gamma \vdash B(t,x) \supset B(s,x)$. Now substitute $\imath x B(t,x)$ for x in this derivation.

We now have to treat the cases when $A(t)$ is a composite formula. These cases are similar to each other. We exemplify with conjunction.

$$
\dfrac{\begin{matrix} \mathscr{D}_1 \\ B(s) \mathbin{\&} C(s) : \mathsf{prop} \end{matrix} \qquad \dfrac{\dfrac{\dfrac{[B(s) \mathbin{\&} C(s)]}{B(s)} \qquad \begin{matrix} \vdots \\ B(s) \supset B(t) \end{matrix}}{B(t)} \qquad \dfrac{\dfrac{[B(s) \mathbin{\&} C(s)]}{C(s)} \qquad \begin{matrix} \vdots \\ C(s) \supset C(t) \end{matrix}}{C(t)}}{B(t) \mathbin{\&} C(t)}}{B(s) \mathbin{\&} C(s) \supset B(t) \mathbin{\&} C(t)}
$$

The derivations of $B(s) \supset B(t)$ and $C(s) \supset C(t)$ exist by the induction hypothesis. In the latter case, the derivation may depend on assumptions $B(s)$ and $B(t)$, but these can be derived from $B(s) \mathbin{\&} C(s)$ (which is currently open and hence available) and $B(s) \supset B(t)$. $\dashv$

The next lemma shows that the assumption $\Gamma \vdash A(s) : \mathsf{prop}$ can be removed from the previous lemma.

LEMMA 10. *Assume $A(t)$ does not contain any ε-term, nor does s. If* $\Gamma \vdash A(t) : \mathsf{prop}, t = s$, *then* $\Gamma \vdash A(s) : \mathsf{prop}$.

PROOF. Induction on μ again. We assume this time that the lemma is true for $\mu(A(t)) < N$ and prove it in the case $\mu(A(t)) = N$.

If $A(t)$ is atomic, say $P(t_1(t), \dots, t_n(t))$, then PF reduces the problem to deriving $t_i(s) : I$. If t_i is a variable or constant, this is trivial. If t_i is $f(\dots)$, it reduces, by fF, to the arguments of f. If $t_i(t)$ is $\imath x B(t,x)$, then $\Gamma, x : I \vdash B(t,x) : \mathsf{prop}$ and hence, by the induction hypothesis, $\Gamma, x : I \vdash B(s,x) : \mathsf{prop}$. Hence, by the previous lemma, $\Gamma, x : I \vdash B(t,x) \supset B(s,x)$, so, since $\Gamma \vdash \exists x B(t,x)$, we conclude $\Gamma \vdash \exists x B(s,x)$. A similar argument applies to the uniqueness premise of the $\imath F$-rule. Hence $\Gamma \vdash \imath x B(s,x) : I$.

If $A(t)$ is composite, it is an easy exercise to prove the lemma using the induction hypothesis. The most complicated case is conjunction (or implication, which requires precisely the same argument). If $A(t)$ is $B(t) \,\&\, C(t)$, we have $\Gamma \vdash B(t) : \mathsf{prop}$ and $\Gamma, B(t) \vdash C(t) : \mathsf{prop}$. By induction hypothesis, $\Gamma \vdash B(s) : \mathsf{prop}$ and $\Gamma, B(t) \vdash C(s) : \mathsf{prop}$. Since $\Gamma \vdash B(s) : \mathsf{prop}$, we have $\Gamma \vdash B(s) \supset B(t)$ (the previous lemma), hence $\Gamma, B(s) \vdash C(s) : \mathsf{prop}$. Hence $\Gamma \vdash B(s) \,\&\, C(s) : \mathsf{prop}$. $\dashv$

Proof of Theorem 7, part 3. First notice that it is sufficient to prove the theorem when s is a fresh variable z, because if this has been done, we have

$$\Gamma,\ z : I,\ t = z \vdash A(z)\,,$$

so we can, by replacing all z by s, and each assumption $z : I$ by a derivation of $s : I$ (which exists because $\Gamma \vdash t = s$), and each assumption $t = z$ by a derivation of $t = s$, derive $A(s)$ from Γ. We therefore assume in the following that s is a variable.

Now the previous lemma gives $\Gamma \vdash A(s) : \mathsf{prop}$. Hence, by Lemma 9, $\Gamma \vdash A(t) \supset A(s)$. Using that modus ponens is derivable, we conclude $\Gamma \vdash A(s)$. $\dashv$

§9. Restricted quantifiers. The translation interprets propositions $U(x)$, where x is a free variable, as propositional functions U. Taking the subsets-as-propositional-functions attitude towards subsets in type theory [15, 20, 4], and writing $x \mathrel{\varepsilon} U$ for $U(x)$, we are justified in saying that propositions depending on variables are interpreted as subsets, and restricted quantifiers are interpreted as follows:

$$\begin{aligned} \forall x(U(x) \supset A(x)) &\rightsquigarrow (\forall x : I)(\forall p : x \mathrel{\varepsilon} U)A(x, p) \\ \exists x(U(x) \,\&\, A(x)) &\rightsquigarrow (\exists x : I)(\exists p : x \mathrel{\varepsilon} U)A(x, p)\,, \end{aligned}$$

where $A(x, p)$ may depend on p. Using the notation in [4], we can write

$$\begin{aligned} \forall x(U(x) \supset A(x)) &\rightsquigarrow (\forall x_p \mathrel{\varepsilon} U)A(x_p) \\ \exists x(U(x) \,\&\, A(x)) &\rightsquigarrow (\exists x_p \mathrel{\varepsilon} U)A(x_p)\,. \end{aligned}$$

Thus, restricted quantifiers are indeed interpreted as restricted quantifiers in the sense of [4]. The notions of restricted quantifiers in [15, 20] are however in general too restrictive, because they do not allow $A(x, p)$ to depend on p. For example, the field property, derived on page 496, is interpreted as follows:

$$\forall x\big(U(x) \supset x \cdot x^{-1} = 1\big) \rightsquigarrow (\forall x_p \mathrel{\varepsilon} U)(x \cdot x_p^{-1} = 1)\,,$$

where x_p^{-1} is the left projection of p. In this case, $A(x, p)$ depends on p in a crucial way.

§10. Partial functions. A common use of definite descriptions is for defining partial functions. Let $R(x, y)$ be a binary partial functional relation, i.e., $x : I,\ y : I,\ \Gamma \vdash R(x, y) : \mathsf{prop}$ and $\Gamma \vdash \forall x \forall y \forall z(R(x, y) \,\&\, R(x, z) \supset y = z)$. Let '$x \mathrel{\varepsilon} D$' denote the interpretation in type theory of $\exists y R(x, y)$. Then

$$\dfrac{\exists y R(x, y) \qquad \dfrac{\begin{matrix}\vdots\\ \forall x \forall y \forall z(R(x, y) \,\&\, R(x, z) \supset y = z) \qquad x : I\end{matrix}}{\forall y \forall z(R(x, y) \,\&\, R(x, z) \supset y = z)}\,\forall E}{\iota y R(x, y) : I}\,\iota$$

is interpreted as

$$\pi_\ell(p) : I\ (\ldots, x : I,\ p : x \varepsilon D)\,,$$

hence as a partial function from D to $I/{=_I}$ in the sense of [4]. It is not in general a partial function in the sense of [20], because $(\forall p, q : x \varepsilon D)\, \mathsf{Id}(I, \pi_\ell(p), \pi_\ell(q))$ is not in general true.

Suppose, on the other hand, that we have, in type theory, a partial function on I, that is, an extensional domain of definition $D \subseteq I$

$$x \varepsilon D : \mathsf{prop}\ (x : I)$$

and a function, which we in the notation of [4] can write

$$f(x_p) : I\ (x_p \varepsilon D)\,,$$

and which satisfies $(\forall x_p \varepsilon D, q : x \varepsilon D)(f(x_p) =_I f(x_q))$ as was required in [4]. Consider our $\imath$-calculus with a unary primitive predicate symbol $D(x)$, whose interpretation is supposed to be $x \varepsilon D$, and a binary primitive relation symbol $R(x, y)$, whose interpretation is supposed to be $(\exists p : x \varepsilon D)(f(x_p) =_I y)$. Then the following axioms are justified, in the sense that their interpretations can be proved in type theory:

$$\begin{aligned} &\forall x(D(x) \supset \exists y R(x, y)) \\ &\forall x \forall y \forall z(R(x, y)\ \&\ R(x, z) \supset y = z)\,. \end{aligned}$$

Also the required extensionality properties for D and R can be proved. Now, introduce $f(x)$ as a notation for $\imath y R(x, y)$. We can then derive the following:

$$\dfrac{\dfrac{D(x) \qquad \dfrac{\forall x(D(x) \supset \exists y R(x, y)) \quad x : I}{D(x) \supset \exists y R(x, y)}\,\forall E}{\exists y R(x, y)}\,\supset E \qquad \begin{matrix}\vdots \\ \forall y \forall z(R(x, y)\ \&\ R(x, z) \supset y = z)\end{matrix}}{f(x) : I.}\,\imath F$$

If the axiom $\forall x(D(x) \supset \exists y R(x, y))$ is interpreted as

$$\lambda x \lambda p(f(x_p), (p, \mathrm{refl}(f(x_p)))) : (\forall x_p \varepsilon D)(\exists y : I)(\exists q : x \varepsilon D)(f(x_p) =_I y)\,,$$

where refl is the reflexivity proof of $=_I$, then $f(x) : I$, as derived above, will be interpreted as the type-theoretical $f(x_p) : I$, after reduction. Hence the interpretation captures the intended meaning of f, up to definitional equality.

§11. Summary and conclusions. We have given systems for indefinite and definite descriptions and interpreted them in intensional type theory. The system for indefinite descriptions suffers from unusual restrictions on modus ponens and existence introduction, while the unrestricted versions of these rules can be derived for definite descriptions. In fact, we had in the end just *one* system, in which both indefinite and definite descriptions could coexist, the unrestricted versions of the rules being derivable for formulas without ε-terms.

In the translation, we used strong disjunction elimination in an essential way (p. 505), and it is therefore unclear if the interpretation would carry over to type theories without this rule. Moreover, the type-theoretical axiom of choice was implicitly used in the interpretation of descriptions, in the form of left projections of existence proofs (p. 504).

It seems that the interpretation follows closely what constructive mathematicians have in mind when they speak about descriptions. One conclusion is therefore that strong disjunction elimination and the intensional ('type-theoretical') axiom of choice seem to be important for a natural formalization of constructive mathematics with descriptions.

A conclusion from Sections 9–10 is that restricted quantifiers and partial functions in the sense of first order logic with descriptions correspond to restricted quantifiers and partial functions in the sense of [4], rather than in the sense of [15, 20].

Acknowledgements. I would like to thank Per Martin-Löf for many useful comments. His idea to let the primitive functions and predicates be partial would be very interesting to carry out, but this has to be postponed to another paper. Many thanks also to Sören Stenlund for sending me his papers on the subject. A discussion with Erik Palmgren helped me make some decisions about the rules in the ε-calculus. Finally, the referee made many useful comments and provided some references. The prooftrees were typeset by Paul Taylor's package prooftree.sty.

Appendix. In the logical inference rules, we could have, among the premises, some premises that express that all formulas involved are indeed propositions. Conjunction introduction, for instance, would then look as follows:

$$\dfrac{A : \mathsf{prop} \qquad \begin{array}{c}[A]\\ \vdots \\ B : \mathsf{prop}\end{array} \qquad A \qquad B}{A \,\&\, B}$$

and disjunction elimination as follows:

$$\dfrac{A : \mathsf{prop} \qquad B : \mathsf{prop} \qquad \begin{array}{c}[A \vee B]\\ \vdots \\ C : \mathsf{prop}\end{array} \qquad A \vee B \qquad \begin{array}{c}[A]\\ \vdots \\ C\end{array} \qquad \begin{array}{c}[B]\\ \vdots \\ C\end{array}}{C\,.}$$

This was the form my ε-calculus had originally, but Theorem 1 showed that it was conservative to simplify it to the present one.

Although this more verbose ε-calculus is in a sense more natural, it is very heavy to produce derivations in it, and, contrary to what one might first think, the translation into type theory is more difficult. The reason is that every rule except $\perp E$ gets the undesirable property that the same proposition occurs several times among the premises. Since the same propositions could be interpreted in different ways in different places, we would need to introduce restrictions on every rule except $\perp E$ in order to get the interpretation through. Alternatively, we could perform the translation by just throwing all extra premises away and proceed as in this paper, but it is then unclear what would be gained by having them in the first place.

REFERENCES

[1] M. Abadi, G. Gonthier, and B. Werner, *Choice in dynamic linking*, ***Foundations of software science and computation structures, 7th international conference, FOSSACS 2004, Barcelona, Spain, March 29–April 2, 2004, proceedings*** (Igor Walukiewicz, editor), Lecture Notes in Computer Science, vol. 2987, Springer, 2004, pp. 12–26.

[2] P. Aczel and N. Gambino, *Collection principles in dependent type theory*, ***Types for proofs and programs, international workshop, TYPES 2000, Durham, UK, December 8–12, 2000, selected papers*** (P. Callaghan, Z. Luo, J. McKinna, and R. Pollack, editors), Lecture Notes in Computer Science, vol. 2277, Springer, 2002, pp. 1–23.

[3] E. Bishop, ***Foundations of constructive analysis***, McGraw-Hill Book Co., New York, 1967.

[4] J. Carlström, *Subsets, quotients and partial functions in Martin-Löf's type theory*, ***Types for proofs and programs, second international workshop, TYPES 2002, Berg en Dal, The Netherlands, April 24–28, 2002, selected papers*** (H. Geuvers and F. Wiedijk, editors), Lecture Notes in Computer Science, vol. 2646, Springer, 2003, pp. 78–94.

[5] G. Frege, *Über Sinn und Bedeutung*, ***Zeitschrift für Philosophie und philosophische Kritik***, vol. NF 100 (1892), pp. 25–50, English translation in [6].

[6] ———, *Sense and reference*, ***The Philosophical Review***, vol. 57 (1948), no. 3, pp. 209–230, English translation of [5].

[7] A. Heyting, ***Intuitionism: An introduction***, North-Holland, Amsterdam, 1956.

[8] D. Leivant, *Existential instantiation in a system of natural deduction for intuitionistic arithmetics*, Technical Report ZW 13/73, Stichtung Mathematisch Centrum, Amsterdam, 1973.

[9] S. Maehara, *The predicate calculus with ε-symbol*, ***Journal of the Mathematical Society of Japan***, vol. 7 (1955), pp. 323–344.

[10] ———, *Equality axiom on Hilbert's ε-symbol*, ***Journal of the Faculty of Science University of Tokyo, Section 1***, vol. 7 (1957), pp. 419–435.

[11] ———, *A general theory of completeness proofs*, ***Annals of the Japan Association for Philosophy of Science***, vol. 3 (1970), pp. 242–256.

[12] P. Martin-Löf, ***Intuitionistic type theory***, Bibliopolis, Naples, 1984, Notes by Giovanni Sambin.

[13] R. Mines, F. Richman, and W. Ruitenburg, ***A course in constructive algebra***, Springer-Verlag, New York, 1988.

[14] G. E. Mints, *Heyting predicate calculus with epsilon symbol*, ***Journal of Soviet Mathematics***, vol. 8 (1977), pp. 317–323.

[15] B. Nordström, K. Petersson, and J. Smith, ***Programming in Martin-Löf's type theory***, Oxford University Press, 1990, `http://www.cs.chalmers.se/Cs/Research/Logic/book/`.

[16] D. Prawitz, ***Natural deduction: a proof-theoretical study***, Stockholm Studies in Philosophy, Almqvist & Wiksell, 1965.

[17] A. Ranta, ***Type-theoretical grammar***, Oxford University Press, Oxford, 1994.

[18] B. Russell, *On denoting*, ***Mind***, vol. 14 (1905), pp. 479–493.

[19] ———, ***Introduction to mathematical philosophy***, second ed., Allen & Unwin, London, 1920.

[20] G. Sambin and S. Valentini, *Building up a toolbox for Martin-Löf's type theory: subset theory*, ***Twenty-five years of constructive type theory (Venice, 1995)*** (G. Sambin and J. Smith, editors), Oxford Logic Guides, vol. 36, Oxford University Press, 1998, pp. 221–244.

[21] D. Scott, *Identity and existence in intuitionistic logic*, ***Applications of sheaves*** (M. P. Fourman, C. J. Mulvey, and D. S. Scott, editors), Lecture Notes in Mathematics, vol. 753, Springer, Berlin, 1979, pp. 660–696.

[22] K. Shirai, *Intuitionistic predicate calculus with ε-symbol*, ***Annals of the Japan Association for Philosophy of Science***, vol. 4 (1971), pp. 49–67.

[23] S. Stenlund, ***The logic of description and existence***, Filosofiska studier, no. 18, Filosofiska föreningen och Filosofiska institutionen vid Uppsala universitet, Uppsala, 1973.

[24] ———, *Descriptions in intuitionistic logic*, ***Proceedings of the third Scandinavian logic symposium*** (S. Kanger, editor), 1975, pp. 197–212.

DEPARTMENT OF MATHEMATICS
STOCKHOLM UNIVERSITY
STOCKHOLM, SWEDEN
URL: http://www.math.su.se/~jesper