CrossMark

The Justification of Identity Elimination in Martin-Löf's Type Theory

Ansten Klev[1]

Abstract On the basis of Martin-Löf's meaning explanations for his type theory a detailed justification is offered of the rule of identity elimination. Brief discussions are thereafter offered of how the univalence axiom fares with respect to these meaning explanations and of some recent work on identity in type theory by Ladyman and Presnell.

Keywords Justification of logical laws · Type theory · Identity

1 Introduction

Let us say that a formal language L serves as a foundation for a mathematical theory T if (i) the concepts of T are definable in L—that is, for each concept of T there is a concept in L that 'translates' it; and (ii) any theorem of T (after its translation) is demonstrable in L. For L to serve as such a foundation it is clear that it must contain certain concepts on the basis of which the concepts of T may be defined and support certain modes of reasoning on the basis of which the theorems of T may be demonstrated. Moreover, if we are to recognize L as a foundation for T, it is clear that we must understand these concepts and appreciate that these modes of reasoning are valid. This latter requirement will be met, it seems, if the language L has been given an intended interpretation: if it has been laid down what statements made in L mean and how the symbols of L contribute to this meaning.

✉ Ansten Klev
anstenklev@gmail.com

1   Department of Logic, Institute of Philosophy, Czech
    Academy of Sciences, Jilská 1, 110 00 Praha 1,
    Czech Republic

On the basis of the meaning explanations provided we should, namely, come to an understanding of the concepts of L and come to see that the forms of reasoning supported by L are valid.

On pain of circularity the meaning explanations for L cannot rely on concepts from T. Hence, if T is 'all of mathematics', then the meaning explanations for L cannot themselves rely on a mathematical theory, in particular not on model theory. Thus, axiomatic set theory, which is often taken to be a candidate for a foundation of all of classical mathematics, has been explained in terms of the so-called iterative conception of set (Zermelo 1930; Boolos 1971; Shoenfield 1977). I shall be concerned with Martin-Löf's constructive type theory, which aims to be a foundation for constructive mathematics. In particular, I wish to consider in detail how, on the basis of the meaning explanations offered by Martin-Löf for this language, a certain mode of reasoning supported by it known as the identity elimination rule can be justified.

I shall approach the justification of elimination rules in a manner that differs somewhat from that of Martin-Löf (1984) and Nordström et al. (1990). In order to justify an elimination rule it is necessary to make it evident, on the assumption that the premisses of the rule are known, that a certain object a evaluates to a canonical element of some **set** A (the notation and terminology here will be explained presently). In the cited works an informal description of this evaluation is taken to justify both the elimination rule and the corresponding equality rule. I shall instead take the equality rule to be definitional in nature and to provide the main ingredient of the justification of the elimination rule. More generally, I shall distinguish rules that are definitional, or meaning-giving, in nature from rules, such as the elimination rules, that are postulational in nature and that require justification.

There are at least two reasons why the rule of identity elimination deserves attention. In the paper (Martin-Löf 1982), where Martin-Löf first gave meaning explanations for his type theory, the justification of the rules of the system is left to the reader. In the booklet (Martin-Löf 1984), where both meaning explanations and justifications of the rules are given, another rule of identity elimination is employed, much stronger than the one found in (Martin-Löf 1975) and that interests us. The textbook (Nordström et al. 1990) gives meaning explanations and treats of the relevant rule of identity elimination, but provides only a very cursory justification of it (ibid. p. 58). In many cases it will be obvious to anyone who has seen the justification of, for instance, Σ-elimination how the justification of an elimination rule should go. Owing to certain peculiarities of identity-**set**s, a justification of identity elimination cannot, however, be obtained by a straightforward adaptation of the justification of, say, Σ-elimination. That a thorough justification of identity elimination is thus not found in any of the well-known introductory works on type theory (not in any publication, as far as I am aware) and involves features not found in the justification of other elimination rules is one reason to devote attention to it.

Another reason is that, in a recent paper, Ladyman and Presnell (2015) have claimed that identity elimination has so far not been satisfactorily justified from 'pre-mathematical' principles. More precisely, Ladyman and Presnell claim that the rule called path induction in homotopy type theory (The Univalent Foundations Program 2013, ch. 1.12.1) has so far not been thus justified. Path induction is formally identical to identity elimination—if the two rules differ, it is in the meaning explanations assumed by the homotopy interpretation of type theory on the one hand and the intended interpretation on the other. Ladyman and Presnell emphasize, however, that the kind of justification they seek will not be found in the homotopy interpretation (p. 401); and the justification they themselves offer is free of concepts from homotopy theory (pp. 401–405). What Ladyman and Presnell think about the mode of justification of the rules of type theory employed by Martin-Löf (1984) and Nordström et al. (1990) is not easy to say, since they do not discuss it. But in order to forestall the impression readers of Ladyman and Presnell's paper may get that this mode of justification cannot satisfactorily be applied to identity elimination, it seems worthwhile showing in detail that the contrary is the case. Some discussion of Ladyman and Presnell's own suggested justification will be given at the end of this paper (Sect. 6). After having presented what I shall call the intended, or standard, interpretation of type theory (Sects. 2, 3) and shown how identity elimination (Sect. 4) may be justified on the basis of it, I shall also briefly discuss the univalence axiom of homotopy type theory in the light of this interpretation (Sect. 5).

2 Preliminaries

2.1 Syntax and Notation

Statements made in Martin-Löf's type theory are called 'judgements'. In particular, the premisses and conclusion of an inference made in the system are judgements. A judgement may be categorical or hypothetical. There are two forms of categorical judgement:

$$a : \mathscr{C}$$
$$a = b : \mathscr{C}$$

The first may be read 'a is an object of category $\mathscr{C}$' and the second 'a and b are equal objects of category $\mathscr{C}$'. (The term 'category' is here being used in the traditional, philosophical sense and not in the sense of category theory.) A judgement of the latter form presupposes the judgements $a : \mathscr{C}$ and $b : \mathscr{C}$ for its meaningfulness. In the basic formulation of the theory the categories are

(1) **set**;
(2) A for any **set** A.

Extensions of the basic formulation may be obtained by admitting as categories higher types or, for instance, a category of contexts, as in the so-called substitution calculus. We shall here restrict ourselves to the basic formulation, thereby following Martin-Löf (1975, 1982, 1984) and Nordström et al. (1990). In this formulation the forms of categorical judgement are thus

$$A : \textbf{set}$$
$$A = B : \textbf{set}$$

and for any **set** A,

$$a : A$$
$$a = b : A$$

The form of a hypothetical judgement is

$$x_1 : A_1, \ldots, x_n : A_n \vdash J$$

where J has the form of a categorical judgement.

We write boldface **set** so as to make it clear that a set in the sense of constructive type theory is meant. A **set** in this sense is not at all a set in the sense of axiomatic set theory. The notion in axiomatic set theory that corresponds most closely to a **set** in type theory is not the notion of a set, but rather that of a universe of sets. Namely, a **set** is a domain, or type, or sort, of individuals, just as a universe of, say, ZFC-sets is (cf. Zermelo 1930). A **set** should thus be thought of

as a type of individuals. It is not uncommon in the literature to call the category **set** 'type'. Indeed, this is the terminology employed in Martin-Löf (1975, 1982). With higher types present, it is, however, necessary to distinguish terminologically between types of individuals and types in general.[1] We follow Martin-Löf (1984) in calling the former 'sets'.[2]

2.2 General Remarks on the Intended Interpretation

The intended, or standard, semantics of Martin-Löf's type theory may be said to be determined by three components:

- The explanation of the forms of categorical judgement.
- The explanation of the forms of hypothetical judgement.
- The rules of the system.

The explanation of a form of judgement provides the assertion-conditions for judgements of the form in question. It lays down what one must know in order to have the right to make such a judgement. These explanations in effect also determine what is understood by the four basic notions of being a **set**; being equal **set**s; being an element of a **set** A; and being equal elements of a **set** A.

The rules of the system are of two kinds. Firstly, there are general rules governing judgements of the forms $A = B$: **set** and $a = b$: A, as well as rules of substitution and assumption. These rules can be seen to be justified by the explanations of the forms of judgement.

Secondly, there are rules associated with the primitive constants of the theory. These are of four kinds: formation rule, introduction rule, elimination rule, and equality rule. Formation-, introduction-, and elimination rules come in pairs: one rule has a : $\mathscr{C}$ as conclusion and the other has $a = b$: $\mathscr{C}$ as conclusion, where in formation rules $\mathscr{C}$ is **set**, and in introduction- and elimination rules, $\mathscr{C}$ is A for some **set** A. I will call the rule whose conclusion is a : $\mathscr{C}$ the primary rule, and the rule whose conclusion is $a = b$: $\mathscr{C}$ the secondary rule, of the pair in question. Secondary rules are often left implicit in presentations of type theory. But in a lower-order presentation their presence must be assumed, since without them one cannot justify in general that if $a = b$: $\mathscr{C}$, then a and b are intersubstitutable '*salva identitate*'.

Introduction- and equality rules, as well as the secondary elimination rules, are definitional, or meaning-giving, in nature. Introduction rules determine what are the so-called canonical elements, and what are equal canonical elements, of a canonical **set** introduced by a formation rule (the notion of canonicity will be explained in more detail below). Equality rules in effect determine how an element of a **set** introduced by means of an elimination rule is to be evaluated to canonical form. A secondary elimination rule determines that the identity of the element a : A in the conclusion of the corresponding primary elimination rule does not depend on the 'mode of presentation' of the b's occurring in the premisses b : $\mathscr{C}$ of this rule. Being definitional, these rules require no justification—which is not to say that any rule one might think of can serve as, for instance, an introduction rule. An example of a purported introduction rule that fails its task as a definition will be given below.

In regarding equality rules as definitional our approach to the standard semantics differs somewhat from that found in Martin-Löf (1984) and Nordström et al. (1990). There, a primary elimination rule and its corresponding equality rule are taken to be justified on the basis of an informal explanation of how an element introduced by the application of an elimination rule is to be evaluated to canonical form. Such an informal explanation can, however, be extracted from the equality rule together with the explanation of the forms of categorical judgement. Since we already need to assume that certain rules are meaning-giving in nature, we might as well also make that assumption regarding the equality rules. Doing so reduces the reliance on informal explanations and, it seems to me, also leads to a cleaner division of labour among the rules.

Formation rules and primary elimination rules are postulational in nature and do require justification. Formation rules postulate that such and such is a canonical **set** and that such and such are equal canonical **set**s; but to have the right to assert that A is a canonical **set** I must know what are the canonical elements of A as well as what are equal canonical elements of A; and to have the right to assert that A and B are equal canonical **set**s I must know that any canonical element of A is a canonical element of B, and vice versa, and that equal canonical elements of A are equal canonical elements of B, and vice versa. A primary elimination rule postulates that such and such, a, is an element of a **set**, and to have the right to assert this I must know that a evaluates to a canonical element.

What all of this means in practice should become clearer in the course of this paper. But one may see already now that, in the context of the explanations of the basic notions of the theory, the justification of a formation rule is provided by the associated introduction rules, and the justification of a primary elimination rule by the associated equality rule(s).

---

[1] The higher-order type structure was developed by Martin-Löf in the early 1980's. A presentation can be found in (Nordström et al. 1990, chs. 19–20) and (Nordström et al. 2000).

[2] The already complicated terminological situation is not helped by the fact that in homotopy type theory, 'set' and 'proposition' are given novel meanings; see definitions 3.1.1 and 3.3.1 in (The Univalent Foundations Program, 2013).

2.3 The Notion of Evaluation

In the explanation of the forms of categorical judgement the notion of evaluation plays an important role. In particular, the relation

a evaluates to b

is assumed as a primitive notion, not itself explained, but used in the explanation of the forms of judgement. The following examples suggest what one should understand by evaluation.

$$(3 + 2)! \times 4 \text{ evaluates to } 480$$
$$\mu x.(Prime(x) \wedge x > (3 + 2)! \times 4) \text{ evaluates to } 487$$

Thus, ordinary arithmetical computation as well as computation in an extended language of arithmetic are instances of evaluation. Another instance of evaluation is the passage from the *definiendum* to the *definiens* of an explicit definition. Hence, if we have defined $\neg A$ as $A \supset \bot$, then

$\neg A$ evaluates to $A \supset \bot$

Also reduction steps in the sense of Prawitz (1965) should be thought of as instances of evaluation.

$$\cfrac{\cfrac{\begin{matrix}\mathcal{D}_1 & \mathcal{D}_2 \\ A & B\end{matrix}}{A \wedge B}}{A} \text{ evaluates to } \cfrac{\mathcal{D}_1}{A}$$

It follows from the explanations of the forms of categorical judgement to be given below that if a evaluates to b, then $a = b : \mathscr{C}$. That evaluation thus preserves equality is reasonable in light of the above examples. The number $(3 + 2)! \times 4$ evaluates to, and is equal to, the number 480. If the proposition $\neg A$ is explicitly defined as $A \supset \bot$, then it is clear that $\neg A$ is the same proposition as $A \supset \bot$. In general, if an a of some category $\mathscr{C}$ is explicitly defined as a', then it is clear that a is the same $\mathscr{C}$ as a'. That a natural deduction derivation is equal to one to which it reduces is a well-known conjecture regarding the identity of proofs due to Martin-Löf (cf. e.g. Prawitz 1971, p. 257–261).[3]

There is a clear sense in which $5! \times 4$ is not the answer wanted when a schoolboy is asked to compute $(3 + 2)! \times 4$. The answer wanted is 480, which we may therefore think of as the 'end point' of the evaluation of $(3 + 2)! \times 4$. We introduce the term 'canonical object' for an object that within type theory serves as an end point of evaluation.[4] A

canonical object is therefore an object that evaluates to itself. What are regarded as the end points of evaluation is, to some extent at least, a matter of convention. In type theory it is stipulated that a **set** A is canonical if and only if $A :$ **set** is the conclusion of a primary formation rule. Whenever a primary formation rule is laid down it must be specified, for any **set** A such that $A :$ **set** is the conclusion of an application of this rule, what a canonical element of A is. In most cases this specification follows directly from the associated primary introduction rule(s), so that a will be a canonical element of a **set** A if and only if $a : A$ is the conclusion of an application of an associated introduction rule. Instead of 'canonical element' we sometimes say 'element of canonical form'.

Evaluation as implemented in constructive type theory operates from without and not from within (cf. Martin-Löf 1982, p. 160).[5] That is to say, whether an object is canonical depends only on the form of its outermost operator. It is therefore necessary also to specify what equal canonical objects of a category are, since we must allow for canonical objects' being equal although they are syntactically different. It is stipulated that A and B are equal canonical **set**s if $A = B :$ **set** is the conclusion of a secondary formation rule. Whenever a primary formation rule is laid down it must be specified, for any **set** A such that $A :$ **set** is the conclusion of an application of this rule, what equal canonical elements of A are. In most cases this specification follows directly from the associated secondary introduction rule(s), so that a and b will be the same canonical element of a **set** A if and only if $a = b : A$ is the conclusion of an application of an associated secondary introduction rule.

In the explanation of the forms of judgement we thus take ourselves to be dealing with objects of which it makes sense to say that they evaluate to canonical objects. More precisely, we are assuming that if $a : \mathscr{C}$ is a valid judgement of the theory, then it makes sense to speak of the evaluation of a to canonical form. That any such a is indeed an *object* follows from the fact that through the explanation of the forms of judgement $a : \mathscr{C}$ and $a = b : \mathscr{C}$, the category $\mathscr{C}$ is provided with criteria of application and identity. A category is therefore what philosophers have called a sortal concept, and most would agree that whatever falls under a sortal concept deserves to be called an object.

2.4 Explanation of the Forms of Categorical Judgement

In explaining the four forms of categorical judgement we shall follow (Martin-Löf 1982). The perhaps more well-known explanations of (Martin-Löf 1984) may be regarded as a

---

[3] On the origin of the conjecture, see also Prawitz (2015, p. 45).

[4] Perhaps the first to introduce a special term for what are here called canonical objects was Husserl, who in the second part of his *Philosophie der Arithmetik* (Husserl 1891, pp. 295–297), speaks of the reduction of *problematische Zahlen* to *Normalzahlen*.

[5] The system of Martin-Löf (1975) differs in this respect, since evaluation works from within.

special case of these. Namely, we may regard the latter explanations as obtained from the former if we assume that all **set**s are canonical. The latter explanations therefore do not suffice when one or more so-called universes are assumed or when something as basic as explicit definitions of **set**s is allowed.

The judgement

A : **set**

means that A evaluates to a canonical **set**. We are free to postulate that C is a canonical **set**—that is, to lay down a formation rule whose conclusion is C : **set**—provided we specify what a canonical element of C is and what equal canonical elements of C are. The latter specification must be done so that the relation of equality between canonical objects in C is rendered reflexive, symmetric, and transitive.

The judgement

$A = B$: **set**

means that A and B evaluate to equal canonical **set**s. We are free to postulate that the canonical **set**s C and D are equal— that is, to lay down a formation rule whose conclusion is $C = D$: **set**—provided we can justify that any canonical element of C is also a canonical element of D, and vice versa; and that equal canonical elements of C are equal canonical elements of D, and vice versa.

Assume A : **set**. Then A evaluates to some canonical set C. The judgement

a : A

means that a evaluates to a canonical element of C. The judgement

$a = b$: A

means that a and b evaluate to equal canonical elements of C.

2.5 Example

We postulate the following rules of N-formation.

(N-form)   N : **set**   N = N : **set**

Thus, we postulate that N is a canonical **set** and that N is the same canonical **set** as N. In the case of a constant, such as N, the secondary formation rule is, in general, trivial. To be justified in postulating N : **set**, we must specify what the canonical elements of N are and what equal canonical elements of N are. That is done by stipulating the following N -introduction rules.

(N-intro) $0 : N$ $0 = 0 : N$ $\dfrac{n : N}{s(n) : N}$ $\dfrac{n = m : N}{s(n) = s(m) : N}$

By virtue of these rules 0 is a canonical element of N, as is $s(n)$ provided n is a N (n itself does not have to be canonical). Moreover, 0 is the same canonical element of N as 0, and $s(n)$ is the same canonical element of N as $s(m)$ provided

$n = m$: N (n and m do not have to be canonical elements of N). Since 0 is a constant, the secondary introduction rule $0 = 0$: N is trivial.

Having specified what the canonical elements of N are we must justify the secondary N-formation rule. But in this case, that is entirely trivial; namely, it is entirely trivial that any canonical element of N is a canonical element of N and that equal canonical elements of N are equal canonical elements of N.

Since N is a canonical **set**, the judgement

n : N

means that n evaluates to a canonical element of N. For instance, by means of the rule of N-elimination, which will not be stated here, one can define the addition function, +. Making the definitions $1 = s(0)$: N and $2 = s(1)$: N, one can then see that $2 + 2$ is an element of N, since it evaluates to $s(2 + 1)$, which is of canonical form, since we can show $2 + 1$: N. The judgement

$m = n$: N

means that m and n evaluate to equal canonical elements of N. Thus, making the definition $3 = s(2)$: N, we have $2 + 2 = 3 + 1$: N, since the left hand side evaluates to $s(2 + 1)$ and the right hand side to $s(3 + 0)$, and we can show $2 + 1 = 3 + 0$: N.

2.6 Principles Justified by the Explanations

It may be instructive to see that the explanations given justify the following important rule.

$$\dfrac{a : A \qquad A = B : \textbf{set}}{a : B}$$

The premisses presuppose A : **set** and B : **set**. Hence, A evaluates to a canonical **set** C and B to a canonical **set** D. By the first premiss, viz. $a : A$, a therefore evaluates to some canonical element $c : C$. By the second premiss, C and D are equal canonical **set**s. By the explanation of what it is for canonical **set**s to be equal, we may infer that c is a canonical element of D. Hence a evaluates to a canonical element of D, which is a canonical **set** to which B evaluates; whence $a : B$. The rule

$$\dfrac{a = b : A \qquad A = B : \textbf{set}}{a = b : B}$$

can be justified in a similar manner.

The explanations likewise justify rules to the effect that the relations of equality between **set**s and equality between the elements of a **set** A are reflexive, symmetric, and transitive. In what follows I shall rely on all of these rules without explicit mention.

It is stipulated that a canonical object evaluates to itself. Hence, if the **set** A evaluates to the canonical **set** C, then $A = C$: **set** according to the explanation of this form of judgement. Likewise, if $a : A$ evaluates to $c : C$, then $a = c : C$, and therefore also $a = c : A$. Thus, in general, if a evaluates to b in some category $\mathscr{C}$, then $a = b : \mathscr{C}$. This fact is important for the justification of elimination rules.

2.7 Propositions

In constructive type theory the category **prop** of propositions is identified with the category **set**. If A is a **set**, we may therefore also regard it as a **prop**, and vice versa. When regarding A as a **prop**, the judgement $a : A$ may be read as saying that a is a proof of A. The explanation of the forms of categorical judgement given above carries over to explanations of the forms of judgement A : **prop**, $A = B$: **prop**, $a : A$, and $a = b : A$, where A : **prop**. In particular, when stipulating that A is a canonical **prop** we must specify what the canonical proofs, and what equal canonical proofs, of A are. The proposition A is true if a proof of A exists.

2.8 Explanation of the Forms of Hypothetical Judgement

Assume A : **set**. The hypothetical judgement

$x : A \vdash B$: **set**

means that $B[a/x]$: **set** whenever $a : A$ and that $B[a/x] = B[a'/x]$: **set** whenever $a = a' : A$. We here use square brackets to indicate substitution. Thus, $B[a/x]$ is the result of substituting a for x in B. The judgement $x : A \vdash B$: **set** may therefore be thought of as saying that B is a family of **set**s over A, or a **set**-valued function over A. Owing to the condition that $B[a/x] = B[a'/x]$: **set** whenever $a = a' : A$, the family is extensional in the sense that it does not depend on the 'mode of presentation' of the index a. The judgement $x : A \vdash B$: **prop** may be thought of as saying that B is a unary propositional function over A.

The hypothetical judgement

$x : A \vdash B = B'$: **set**

means that $B[a/x] = B'[a/x]$: **set** whenever $a : A$.

Assume $x : A \vdash B$: **set**. The hypothetical judgement

$x : A \vdash b : B$

means that $b[a/x] : B[a/x]$ whenever $a : A$ and that $b[a/x] = b[a'/x] : B[a/x]$ whenever $a = a' : A$. The judgement $x : A \vdash b : B$ may therefore be thought of as saying that b is a function whose domain is A and such that $b[a/x]$ is of category $B[a/x]$; thus the category of $b[a/x]$ may depend on the argument a.

The hypothetical judgement

$x : A \vdash b = b' : B$

means that $b[a/x] = b'[a/x] : B[a/x]$ whenever $a : A$.

Note that we here in effect reduce the explanation of a hypothetical judgement $x : A \vdash J[x]$ to the corresponding categorical judgement $J[a]$. In particular, the judgement $x : A \vdash B$: **set** is taken to mean that $B[a/x]$ evaluates to a canonical **set** whenever $a : A$ and that $B[a/x]$ and $B[a'/x]$ evaluate to equal canonical **set**s whenever $a = a' : A$. The judgement $x : A \vdash b : B$ is taken to mean that $b[a/x]$ evaluates to a canonical element of the canonical **set** to which $B[a/x]$ evaluates and that $b[a/x]$ and $b[a'/x]$ evaluate to equal canonical elements of the canonical **set** to which $B[a/x]$ evaluates.

The above explanations are generalized by induction to hypothetical judgements with an arbitrary number of assumptions. We give one example. Assume

$x_1 : A_1, \ldots, x_{n-1} : A_{n-1} \vdash A_n$: **set**

Let $\bar{a}$ be a sequence $a_1, \ldots, a_n$ such that

$a_1 : A_1, \ldots, a_n : A_n[a_1/x_1, \ldots, a_{n-1}/x_{n-1}]$

We call $x_1 : A_1, \ldots, x_n : A_n$ a context and $\bar{a}$ an environment for this context. Environments $\bar{a}$ and $\bar{a}'$ are said to be equal if

$a_1 = a_1' : A_1, \ldots, a_n = a_n' : A_n[a_1/x_1, \ldots, a_{n-1}/x_{n-1}]$

Note that $\bar{a}$ is a sequence of closed terms. Let $\bar{a}/\bar{x}$ stand for

$a_1/x_1, \ldots, a_n/x_n$

Then

$x_1 : A_1, \ldots, x_n : A_n \vdash B$: **set**

means that

$B[\bar{a}/\bar{x}]$: **set**

whenever $\bar{a}$ is an environment for $x_1 : A_1, \ldots, x_n : A_n$, and that

$B[\bar{a}/\bar{x}] = B[\bar{a}'/\bar{x}]$: **set**

whenever $\bar{a}$ and $\bar{a}'$ are equal environments for $x_1 : A_1, \ldots, x_n : A_n$.

Note that a **set** A_k occurring within a context $x_1 : A_1, \ldots, x_n : A_n$ may be a family of **set**s over $A_1, \ldots, A_{k-1}$.

From these explanations the rules of substitution and assumption may be justified. By rules of substitution we mean the rules

$$\frac{x : A \vdash B : \mathbf{set} \qquad a : A}{B[a/x] : \mathbf{set}} \qquad \frac{x : A \vdash B = B' : \mathbf{set} \qquad a = a' : A}{B[a/x] = B'[a'/x] : \mathbf{set}}$$

$$\frac{x : A \vdash b : B \qquad a : A}{b[a/x] : B[a/x]} \qquad \frac{x : A \vdash b = b' : B \qquad a = a' : A}{b[a/x] = b'[a'/x] : B[a/x]}$$

and their generalizations to hypothetical judgements with longer contexts. The rule of assumption is

$$\frac{A : \mathbf{set}}{x : A \vdash x : A}$$

Thus, the assumption that x is an element of a **set** A is treated as a tautological hypothetical judgement. We are allowed to make the assumption $x : A \vdash x : A$ only if we have established $A : \textbf{set}$. If higher types are present, one can also make assumptions of the form $A : \textbf{set}$.

3 The Justification of Elimination Rules

3.1 General Remarks

Introduction-, equality-, and secondary elimination rules are, as noted, definitional, or meaning-giving, in nature. Doubts about their correctness can therefore only concern their felicity as meaning-giving rules. In particular, doubts concerning an introduction rule can only concern its felicity in determining what are the canonical elements of a purported **set** A. Martin-Löf (1982, p. 166) notes that

> there are certain limits to what verbal explanations can do when it comes to justifying axioms and rules of inference. In the end, everybody must understand for himself.

It may, for instance, not be obvious how one could bring about an understanding of the rules of N-introduction in someone who insists that he fails to grasp by their means what is a canonical element of N. By contrast, it is possible to argue that certain purported introduction rules fail to be meaning-giving. Suppose we postulate a formation rule,

$$(\Lambda\text{-form}) \qquad \frac{A : \textbf{set}}{\Lambda(A) : \textbf{set}}$$

and provide the following introduction rule.[6]

$$(\Lambda\text{-intro}) \qquad \frac{x : \Lambda(A) \vdash a : A}{\text{L}(x.a) : \Lambda(A)}$$

Here $x.a$ indicates that x has become bound in a. Thus L is a variable-binding operation. This introduction rule fails to be meaning-giving because the postulated canonical **set** $\Lambda(A)$ occurs negatively in its premiss.[7] According to the explanation of the forms of hypothetical judgement, the premiss $x : \Lambda(A) \vdash a : A$ means that

$a[b/x] : A$ whenever $b : \Lambda(A)$

Hence, in order to understand this premiss I must know what is an arbitrary element of $\Lambda(A)$. But in order to know what is an arbitrary element of $\Lambda(A)$, I must know what is a canonical element of $\Lambda(A)$, and this I do not know before having grasped the Λ-introduction rule. Understanding the premiss of this rule thus presupposes knowledge that the rule is meant to provide. The rule therefore fails to specify what is a canonical element of $\Lambda(A)$, that is, it fails to be meaning-giving.

Formation- and primary elimination rules are not meaning-giving, but postulational. They must therefore be justified in the manner in which rules of inference in general are to be justified: the conclusion of the rule must be made evident on the assumption that we know the premisses.

It has been stipulated that a **set** A is canonical if and only if $A : \textbf{set}$ is the conclusion of an application of a primary formation rule and that A and B are equal canonical **set**s if and only if $A = B : \textbf{set}$ is the conclusion of an application of a secondary formation rule. To have the right to postulate that A is a canonical **set** I must specify what the canonical elements of A are and what equal canonical elements of A are. In most cases, this specification is given directly through the associated introduction rules. Assuming that the introduction rules for A are felicitous as meaning-giving rules, they therefore directly justify the formation rule whose conclusion is $A : \textbf{set}$. To have the right to postulate that A and B are equal canonical **set**s I must make it evident that any canonical element of A is a canonical element of B, and vice versa, and that equal canonical elements of A are equal canonical elements of B, and vice versa. This justification rests on the specification of what a canonical element is, and what equal canonical elements are, of A and B, together with the explanation of the forms of judgement $A = B : \textbf{set}$ and $a = b : A$.

The rules of the system are applicable under assumptions. For instance, from the assumption $x : \text{N} \vdash x : \text{N}$, I can apply N-introduction to infer $x : \text{N} \vdash \text{s}(x) : \text{N}$. The notion of canonicity may be extended to such cases. Thus, $\text{s}(x)$ may be regarded a canonical element of N in the context $x : \text{N}$.[8] Likewise, the application of a formation rule in a context Γ may be regarded as yielding a canonical **set** in Γ.

Consider a primary elimination rule whose major premiss has the form $a : A$ and whose conclusion has the form $b : B$. To justify the rule we must make the conclusion $b : B$ evident on the assumption that we know the premisses of the rule. We distinguish the cases where the major premiss, $a : A$, is categorical and where it is hypothetical. If the

---

[6] This introduction rule together with the naturally associated elimination- and equality rules give rise to Curry's Paradox (Curry 1942).

[7] That an inductively defined predicate cannot occur negatively in the premiss of its own introduction rule is in effect required by the schemes for such definitions given by Martin-Löf (1971, pp. 182–183). The requirement is also made by Dybjer (1994) for his general form of introduction rules in type theory.

[8] This extension of the notion of canonicity can be used to argue that, for instance, $A \land \neg A$ and $\bot$ are not equal **prop**s. In the context $x : A, y : \neg A$, the term $\langle x, y \rangle$ is a canonical element of $A \land \neg A$, but not of $\bot$.

major premiss is categorical, then—as we shall see in detail below—we can use the equality rule associated with A to argue that b evaluates to a canonical element of C, where C is the canonical **set** to which B evaluates.

Assume next that the major premiss is a hypothetical judgement

$$x_1 : A_1, \ldots, x_n : A_n \vdash a : A$$

The conclusion then has the form

$$x_1 : A_1, \ldots, x_n : A_n, \Gamma \vdash b : B$$

where Γ is a, possibly empty, context stemming from the minor premiss(es). To make this conclusion evident it must be argued that for any environment $\bar{a}$ for $x_1 : A_1, \ldots, x_n : A_n$, we have

$$\Gamma[\bar{a}/\bar{x}] \vdash b[\bar{a}/\bar{x}] : B[\bar{a}/\bar{x}]$$

and that for any equal environments $\bar{a}$ and $\bar{a}'$ we have

$$\Gamma[\bar{a}/\bar{x}] \vdash b[\bar{a}/\bar{x}] = b[\bar{a}'/\bar{x}] : B[\bar{a}/\bar{x}]$$

But, for any such environment $\bar{a}$, the major premiss becomes the categorical judgement $a[\bar{a}/\bar{x}] : A[\bar{a}/\bar{x}]$, and the justification for this case has already been given.

The second part will be seen to follow from the secondary elimination rule.

3.2 Illustration: ∧-elimination

We postulate the following rules of ∧-formation.

$$(\wedge\text{-form}) \qquad \frac{A : \mathbf{prop} \qquad B : \mathbf{prop}}{A \wedge B : \mathbf{prop}} \qquad \frac{A = A' : \mathbf{prop} \qquad B = B' : \mathbf{prop}}{A \wedge B = A' \wedge B' : \mathbf{prop}}$$

To justify this postulation we must, firstly, specify what the canonical elements of $A \wedge B$ are and what equal canonical elements of $A \wedge B$ are. Secondly, we must argue, on the assumption that $A = A' : \mathbf{prop}$ and $B = B' : \mathbf{prop}$, that any canonical element of $A \wedge B$ is a canonical element of $A' \wedge B'$, and vice versa; and that equal canonical elements of $A \wedge B$ are equal canonical elements of $A' \wedge B'$, and vice versa.

What the canonical elements, and what equal canonical elements, of $A \wedge B$ are is specified by the rules of ∧-introduction.

$$(\wedge\text{-intro}) \qquad \frac{a : A \qquad b : B}{\langle a, b \rangle : A \wedge B} \qquad \frac{a = a' : A \qquad b = b' : B}{\langle a, b \rangle = \langle a', b' \rangle : A \wedge B}$$

Thus $\langle a, b \rangle$ is a canonical element of $A \wedge B$ provided $a : A$ and $b : B$; and $\langle a, b \rangle$ and $\langle a', b' \rangle$ are equal canonical elements of $A \wedge B$ provided $a = a' : A$ and $b = b' : B$. This justifies the primary formation rule.

Towards justifying the secondary formation rule, assume the premisses $A = A' : \mathbf{prop}$ and $B = B' : \mathbf{prop}$. Any canonical element of $A \wedge B$ has the form $\langle a, b \rangle$ with $a : A$ and $b : B$. Then $a : A'$ and $b : B'$ by the premisses;

therefore $\langle a, b \rangle : A' \wedge B'$ by the primary ∧-introduction rule. Any equal canonical elements of $A \wedge B$ have the form $\langle a, b \rangle$ and $\langle a', b' \rangle$, with $a = a' : A$ and $b = b' : B$. Then $a = a' : A'$ and $b = b' : B'$ by the premisses; therefore $\langle a, b \rangle = \langle a', b' \rangle : A' \wedge B'$ by the secondary ∧-introduction rule.

We postulate the following rules of ∧-elimination.

$$(\wedge\text{-elim}) \qquad \frac{c : A \wedge B}{\mathrm{l}(c) : A} \qquad \frac{c : A \wedge B}{\mathrm{r}(c) : B}$$

$$\frac{c = c' : A \wedge B}{\mathrm{l}(c) = \mathrm{l}(c') : A} \qquad \frac{c = c' : A \wedge B}{\mathrm{r}(c) = \mathrm{r}(c') : B}$$

The secondary rules, on the second line here, are definitional; but the primary rules require justification. Namely, it must be made evident, on the assumption that we know the premiss $c : A \wedge B$, that $\mathrm{l}(c)$ and $\mathrm{r}(c)$ evaluate to canonical form. The most important ingredient of this justification is provided by the ∧-equality rules, which tell us how to evaluate $\mathrm{l}(\langle a, b \rangle)$ and $\mathrm{r}(\langle a, b \rangle)$.

$$(\wedge\text{-eq}) \qquad \frac{a : A \qquad b : B}{\mathrm{l}(\langle a, b \rangle) = a : A} \qquad \frac{a : A \qquad b : B}{\mathrm{r}(\langle a, b \rangle) = b : B}$$

We may now justify primary ∧-elimination as follows. We first assume that the premiss $c : A \wedge B$ holds categorically, hence that c is a closed term. By this premiss, c evaluates to some $\langle a, b \rangle : A \wedge B$, where $a : A$ and $b : B$. Hence we get an equality $c = \langle a, b \rangle : A \wedge B$. By the secondary ∧-elimination rules and ∧-equality we thence get $\mathrm{l}(c) = \mathrm{l}(\langle a, b \rangle) = a : A$ and $\mathrm{r}(c) = \mathrm{r}(\langle a, b \rangle) = b : B$. Let D be the canonical **set** to which A evaluates and let E be the canonical **set** to which B evaluates. Since $a : A$, a evaluates to a canonical element d of D; since $\mathrm{l}(c) = a : A$, $\mathrm{l}(c)$ evaluates to a canonical element d', equal to d, of D. Likewise, since $b : B$, b evaluates to a canonical element e of E; since $\mathrm{r}(c) = b : B$, $\mathrm{r}(c)$ evaluates to a canonical element e', equal to e, of E.

When the premiss is hypothetical, the first ∧-elimination rule has the form

$$\frac{x_1 : A_1, \ldots, x_n : A_n \vdash c : A \wedge B}{x_1 : A_1, \ldots, x_n : A_n \vdash \mathrm{l}(c) : A}$$

To make the conclusion of this rule evident we must, firstly, argue that for any environment $\bar{a}$ for $x_1 : A_1, \ldots, x_n : A_n$,

$$\mathrm{l}(c[\bar{a}/\bar{x}]) : A[\bar{a}/\bar{x}]$$

Here $c[\bar{a}/\bar{x}]$ is a closed term, so the justification just given shows that $\mathrm{l}(c[\bar{a}/\bar{x}])$ evaluates to canonical form. Secondly, we must argue that for any equal environments $\bar{a}$ and $\bar{a}'$,

$$\mathrm{l}(c[\bar{a}/\bar{x}]) = \mathrm{l}(c[\bar{a}'/\bar{x}]) : A[\bar{a}/\bar{x}]$$

But for any such equal environments the premiss $x_1 : A_1, \ldots, x_n : A_n \vdash c : A \wedge B$ yields

$c[\bar{a}/\bar{x}] = c[\bar{a}'/\bar{x}] : (A \wedge B)[\bar{a}/\bar{x}]$

The secondary $\wedge$-elimination rule for l therefore yields the required equality.

4 The Justification of Id-elimination

The elimination rules for a **set** A give us means to define functions from A into a **set** or a family of **set**s. They do not give us means to define functions from A into **set** itself; that is, they do not give us means to define families of **set**s, or propositional functions, over A. Thus, although we may define addition by means of N-elimination, we cannot thus define the relation $\leq$ on N. To provide for the possibility of defining **set**-valued functions over a **set** we postulate a formation rule whose premisses include judgements of the form $a : A$. More specifically, we postulate the following rules of Id-formation.

(Id-form)
$$\frac{A : \mathbf{set} \qquad a : A \qquad b : A}{\mathrm{Id}(A, a, b) : \mathbf{prop}}$$

$$\frac{A = A' : \mathbf{set} \qquad a = a' : A \qquad b = b' : A}{\mathrm{Id}(A, a, b) = \mathrm{Id}(A', a', b') : \mathbf{prop}}$$

Provided we have justified the first of these rules, viz. the primary Id-formation rule, then whenever $A : \mathbf{set}$, we may form a binary propositional function over A,

$x : A, y : A \vdash \mathrm{Id}(A, x, y) : \mathbf{prop}$

From this, further relations may be defined by means of the same or other operators available in the language.

Since the relation of identity—unlike, say, the relation of less than—is defined over any individual domain, and since $\mathrm{Id}(A, a, b)$ is to be defined whenever $A : \mathbf{set}$, $a : A$, and $b : A$, we intend $\mathrm{Id}(A, a, b)$ to be a proposition to the effect that a and b are identical. Martin-Löf (1971), in a paper concerned with natural deduction, and not type theory, provides a general scheme for natural deduction introduction- and elimination rules, as well as reduction procedures, for inductively defined predicates. As examples of such predicates captured by this scheme of rules Martin-Löf gives the 0-ary predicate of absurdity, the unary predicate of being a natural number, and the binary relation of identity. Identity, which Martin-Löf designates by E, is the predicate that has the introduction rule

(E-intro) $\qquad\qquad\qquad Exx$

with no premisses and where x is a variable. Martin-Löf's scheme determines that E should have the following elimination rule.

(E-elim)
$$\frac{Etu \qquad C[z, z]}{C[t, u]}$$

Here C is any formula of the language, and $C[z, z]$ is the result of substituting the variable z for, say, x and y in C. The terms t and u may be any terms of the language. The rule can be read as saying that whenever C is a reflexive relation, and Etu holds, then $C[t, u]$ holds as well; thus, E is the least reflexive relation. The reduction rule for E becomes

$$\frac{Ett \qquad \begin{array}{c}\mathcal{D}\\C[z, z]\end{array}}{C[t, t]} \quad \Rightarrow \quad \begin{array}{c}\mathcal{D}[t]\\C[t, t]\end{array}$$

Here $\mathcal{D}[t]$ is the result of substituting t for z everywhere in $\mathcal{D}$, where variables in t may have to be renamed to avoid unintended binding.

The introduction-, elimination-, and equality rules for Id follow the pattern of the corresponding rules for E. The rules of Id-introduction are as follows.

(Id-intro)
$$\frac{A : \mathbf{set} \qquad a : A}{\mathrm{refl}(A, a) : \mathrm{Id}(A, a, a)}$$

$$\frac{A = A' : \mathbf{set} \qquad a = a' : A}{\mathrm{refl}(A, a) = \mathrm{refl}(A', a') : \mathrm{Id}(A, a, a)}$$

Thus, provided $a : A$, we stipulate that $\mathrm{refl}(A, a)$ is a proof of $\mathrm{Id}(A, a, a)$.

We cannot use these introduction rules directly in specifying what a canonical element of $\mathrm{Id}(A, a, b)$ is and what equal canonical elements of $\mathrm{Id}(A, a, b)$ are. We cannot do so, since the $\mathscr{C}$ of the conclusion of the introduction rules has the form $\mathrm{Id}(A, a, a)$ and not the form $\mathrm{Id}(A, a, b)$. If $a = b : A$, then the judgements $\mathrm{refl}(A, a) : \mathrm{Id}(A, a, b)$ and $\mathrm{refl}(A, b) : \mathrm{Id}(A, a, b)$ follow from Id-introduction and the rules of substitution. But since evaluation in type theory operates from without, we cannot expect that the evaluation of an arbitrary $p : \mathrm{Id}(A, a, b)$ yields either $\mathrm{refl}(A, a)$ or $\mathrm{refl}(A, b)$. We stipulate that

a canonical element of $\mathrm{Id}(A, a, b)$ is any $\mathrm{refl}(A', a')$ such that

$A' = A : \mathbf{set}$
$a' = a : A$
$a' = b : A$

This stipulation makes sense, since $A' = A : \mathbf{set}$, $a' : A'$ and $a, b : A$ entail that a', a, b are all elements of A. The stipulation, moreover, agrees with the primary Id-introduction rule in the sense that it deems $\mathrm{refl}(A, a)$ to be a canonical element of $\mathrm{Id}(A, a, a)$.

We stipulate that

$\mathrm{refl}(A', a')$ and $\mathrm{refl}(A'', a'')$ are equal canonical elements of $\mathrm{Id}(A, a, b)$ provided

$A = A' = A'' : \mathbf{set}$
$a = a' = a'' : A$
$b = a' = a'' : A$

This stipulation agrees with the secondary Id-introduction rule in the sense that it deems $\mathrm{refl}(A, a)$ and $\mathrm{refl}(A', a')$ to be equal canonical elements of $\mathrm{Id}(A, a, a)$ provided $A = A' :$ **set** and $a = a' : A$. By means of these two stipulations the primary Id-formation rule has been justified.

Towards justifying the secondary Id-formation rule, assume the premisses $A = A' :$ **set**, $a = a' : A$ and $b = b' : A$. Assume that $\mathrm{refl}(A'', a'')$ is a canonical element of $\mathrm{Id}(A, a, b)$. Then $A'' = A :$ **set**, $a'' = a : A$, and $a'' = b : A$, whence by the premisses, $A'' = A' :$ **set**, $a'' = a' : A$, and $a'' = b'$, so $\mathrm{refl}(A'', a'')$ is also a canonical element of $\mathrm{Id}(A', a', b')$. Assume next that $\mathrm{refl}(A'', a'')$ and $\mathrm{refl}(A''', a''')$ are equal canonical elements of $\mathrm{Id}(A, a, b)$. Then $A = A'' = A''' :$ **set**, $a = a'' = a''' : A$ and $b = a'' = a''' : A$, whence by the premisses, $A' = A'' = A''' :$ **set**, $a' = a'' = a''' : A$ and $b' = a'' = a''' : A$, so $\mathrm{refl}(A'', a'')$ and $\mathrm{refl}(A''', a''')$ are also equal canonical elements of $\mathrm{Id}(A', a', b')$.

A direct adaptation of the above natural deduction E-elimination rule to the syntax of type theory with Id instead of E yields the following rule.

$$(\text{Id-elim*}) \qquad \frac{p : \mathrm{Id}(A, a, b) \qquad z : A \vdash c : C[z, z]}{\mathrm{J}^*(p, z.c) : C[a, b]}$$

The premisses presuppose $A :$ **set**, $a : A$, $b : A$, and $x : A, y : A \vdash C :$ **prop**. From now on we shall use the simplified substitution notation exemplified by $C[a, b]$ in place of $C[a/x, b/y]$. The minor premiss, $z : A \vdash c : C[z, z]$, says that $c[a]$ is a proof of $C[a, a]$ whenever $a : A$. The conclusion says that $\mathrm{J}^*(p, z.c)$ is a proof of $C[a, b]$. The notation $z.c$ indicates that z has become bound in c. The reduction rule for E yields the following equality rule.

$$(\text{Id-eq*}) \qquad \frac{a : A \qquad z : A \vdash c : C[z, z]}{\mathrm{J}^*(\mathrm{refl}(A, a), z.c) = c[a] : C[a, a]}$$

Thus the proof of $C[a, a]$ got by applying Id-elimination* to $\mathrm{refl}(A, a)$ and c such that $z : A \vdash c : C$ is stipulated to be the same proof as $c[a]$.

In the language of type theory elimination rules may, however, be generalized by allowing the **prop** C appearing in the conclusion of such a rule to depend on the **prop** A of its major premiss. In particular, we may let the C occurring in the conclusion of Id-elimination be a ternary propositional function.

$$x : A, y : A, w : \mathrm{Id}(A, x, y) \vdash C : \textbf{prop}$$

The resulting rules are as follows.

$$(\text{Id-elim}) \qquad \frac{p : \mathrm{Id}(A, a, b) \qquad z : A \vdash c : C[z, z, \mathrm{refl}(A, z)]}{\mathrm{J}(p, z.c) : C[a, b, p]}$$

$$\frac{p = q : \mathrm{Id}(A, a, b) \qquad z : A \vdash c = d : C[z, z, \mathrm{refl}(A, z)]}{\mathrm{J}(p, z.c) = \mathrm{J}(q, z.d) : C[a, b, p]}$$

I f $\quad x : A, y : A \vdash C :$ **set** , t h e n a l s o $x : A, y : A, p : \mathrm{Id}(A, x, y) \vdash C :$ **set**, by weakening, hence

Id-elimination is indeed a generalization of Id-elimination*. In many applications of Id-elimination the third argument in C is not needed, but it may be needed if one wishes to prove things about elements of Id-**set**s.

The secondary Id-elimination rule is definitional, but the primary rule requires justification. Namely, assuming that we know the premisses of the rule, it must be made evident that $\mathrm{J}(p, z.c)$ evaluates to a canonical element of $C[a, b, p]$. An important ingredient of this justification is provided by the Id-equality rule, which tells us how to evaluate $\mathrm{J}(\mathrm{refl}(A, a), z.c)$.

$$(\text{Id-eq}) \qquad \frac{a : A \qquad z : A \vdash c : C[z, z, \mathrm{refl}(A, z)]}{\mathrm{J}(\mathrm{refl}(A, a), z.c) = c[a] : C[a, a, \mathrm{refl}(A, a)]}$$

We may now justify Id-elimination. Assume that we know its premisses, viz. $p : \mathrm{Id}(A, a, b)$ and $z : A \vdash c : C[z, z, \mathrm{refl}(A, z)]$. Let us first assume that the premiss $p : \mathrm{Id}(A, a, b)$ holds categorically, in particular that p is a closed term. Then p evaluates to a canonical element of $\mathrm{Id}(A, a, b)$, that is, p evaluates to some $\mathrm{refl}(A', a')$, where $A' = A :$ **set**, $a' = a : A$, and $a' = b : A$. Therefore $p = \mathrm{refl}(A', a') : \mathrm{Id}(A, a, b)$. Furthermore, from the equalities holding for A' and a' we may infer

$$A = A = A' : \textbf{set}$$
$$a = a = a' : A$$
$$b = a = a' : A$$

showing that $\mathrm{refl}(A', a')$ and $\mathrm{refl}(A, a)$ are equal canonical elements of $\mathrm{Id}(A, a, b)$. Thus

$$p = \mathrm{refl}(A', a') = \mathrm{refl}(A, a) : \mathrm{Id}(A, a, b)$$

Secondary Id-elimination and Id-equality then yields

$$\mathrm{J}(p, z.c) = \mathrm{J}(\mathrm{refl}(A', a'), z.c) = \mathrm{J}(\mathrm{refl}(A, a), z.c)$$
$$= c[a] : C[a, a, \mathrm{refl}(A, a)]$$

Note that in order to apply Id-equality here, the first argument of refl must be A, since the minor premiss we have assumed is $z : A \vdash c : C[z, z, \mathrm{refl}(A, z)]$ and not $z : A' \vdash c : C[z, z, \mathrm{refl}(A', z)]$.

Let D be the canonical **set** to which $C[a, a, \mathrm{refl}(A, a)]$ evaluates. By the equalities already established, and the rules of substitution, we have

$$C[a, a, \mathrm{refl}(A, a)] = C[a, b, p] : \textbf{set}$$

Whence, $C[a, b, p]$ evaluates to some canonical **set** D' equal to D. By the premiss $z : A \vdash c : C[z, z, \mathrm{refl}(A, z)]$ we know that $c[a]$ evaluates to a canonical element e of D. From $\mathrm{J}(p, z.c) = c[a] : C[a, a, \mathrm{refl}(A, a)]$, we may infer that $\mathrm{J}(p, z.d)$ evaluates to some canonical element e', equal to e, in D. Since $D = D' :$ **set**, e' is also a canonical element of D'. This justifies $\mathrm{J}(p, z.c) : C[a, b, p]$.

It should be clear that this justification remains valid if we assume that C has more than three arguments, that is, if we assume

$$x : A, y : A, p : \mathrm{Id}(A, x, y), \Gamma \vdash C : \mathbf{set}$$

The additional context Γ may be carried along without causing any additional complications.

Assume next that the major premiss of Id-elimination is hypothetical.

$$x_1 : A_1, \ldots, x_n : A_n \vdash p : \mathrm{Id}(A, a, b)$$

Then we must also assume

$$x_1 : A_1, \ldots, x_n : A_n \vdash A : \mathbf{set}$$
$$x_1 : A_1, \ldots, x_n : A_n \vdash a : A$$
$$x_1 : A_1, \ldots, x_n : A_n \vdash b : A$$

The minor premiss then becomes

$$x_1 : A_1, \ldots, x_n : A_n, z : A \vdash c : C[z, z, \mathrm{refl}(z)]$$

where

$$x_1 : A_1, \ldots, x_n : A_n, x : A, y : A, w : \mathrm{Id}(A, x, y) \vdash C : \mathbf{set}$$

For any environment $\bar{a}$ for $x_1 : A_1, \ldots, x_n : A_n$, we have

$$\mathrm{J}(p, z.c)[\bar{a}/\bar{x}] \equiv \mathrm{J}(p[\bar{a}/\bar{x}], z.c[\bar{a}/\bar{x}])$$
$$C[a, b, p][\bar{a}/\bar{x}] \equiv C[\bar{a}/\bar{x}][a[\bar{a}/\bar{x}], b[\bar{a}/\bar{x}], p[\bar{a}/\bar{x}]]$$
$$\mathrm{Id}(A, a, b)[\bar{a}/\bar{x}] \equiv \mathrm{Id}(A[\bar{a}/\bar{x}], a[\bar{a}/\bar{x}], b[\bar{a}/\bar{x}])$$

where $\equiv$ stands for syntactic identity.

It must be argued, firstly, that for any such environment $\bar{a}$,

$$\mathrm{J}(p, z.c)[\bar{a}/\bar{x}] : C[a, b, p][\bar{a}/\bar{x}]$$

But the major premiss is then the categorical judgement

$$p[\bar{a}/\bar{x}] : \mathrm{Id}(A[\bar{a}/\bar{x}], a[\bar{a}/\bar{x}], b[\bar{a}/\bar{x}]),$$

and the justification in this case has already been given.

Secondly, it must be argued that for any equal environments $\bar{a}$ and $\bar{a}'$, we have

$$\mathrm{J}(p, z.c)[\bar{a}/\bar{x}] = \mathrm{J}(p, c.d)[\bar{a}'/\bar{x}] : C[a, b, p][\bar{a}/\bar{x}]$$

For any such environments the major premiss yields

$$p[\bar{a}/\bar{x}] = p[\bar{a}'/\bar{x}] : \mathrm{Id}(A, a, b)[\bar{a}/\bar{x}]$$

The minor premiss yields

$$z : A[\bar{a}/\bar{x}] \vdash c[\bar{a}/\bar{x}] = c[\bar{a}'/\bar{x}] : C[z, z, \mathrm{refl}(A, z)][\bar{a}/\bar{x}]$$

The secondary Id-elimination rule therefore yields the required equality.

It should be clear this justification remains valid if we assume that C has more than $n + 3$ arguments, that is, if we assume

$$x_1 : A_1, \ldots, x_n : A_n, x : A, y : A, w : \mathrm{Id}(A, x, y), \Gamma \vdash C : \mathbf{set}$$

The main aim of this paper has now been reached. The two next sections may be regarded as appendices. They presuppose a bit more familiarity with type theory than what has been presupposed in the foregoing.

5 Remarks on the Univalence Axiom

Homotopy type theory adds to Martin-Löf's type theory the so-called univalence axiom as well as so-called higher inductive types. A thorough discussion of these additions, and of how they may be justified, lies outside the compass of this paper; but a few remarks on how the univalence axiom fares with respect to the standard interpretation of type theory presented above may not be out of place.

It seems that it will be difficult, if not impossible, to justify the univalence axiom on the grounds of this interpretation. There are at least two problems.

The univalence axiom is the postulation, for any $A, B : \mathbf{set}$, of a judgement

$$\mathrm{ua}(A, B) : \mathrm{Univ}(A, B)$$

where $\mathrm{Univ}(A, B)$ is a certain **set** whose definition is not important for current purposes. Since $\mathrm{Univ}(A, B) : \mathbf{set}$, it evaluates to some canonical $C : \mathbf{set}$. The judgement $\mathrm{ua}(A, B) : \mathrm{Univ}(A, B)$ is therefore justified according to the standard interpretation if it can be made evident that $\mathrm{ua}(A, B)$ evaluates to a canonical element of C. A first problem for the univalence axiom from the point of view of the standard interpretation is then that the postulation is made without directions as to how $\mathrm{ua}(A, B)$ is to be evaluated to such a canonical element. In particular, the postulation is not accompanied by anything like equality rules or informal descriptions of the evaluation. Hence we have not been given the information needed to see that the postulation is justified according to the standard interpretation. This is in effect a problem that attends all outright postulations in type theory, for instance also that of the law of the excluded middle in the form

$$\frac{A : \mathbf{prop}}{\mathrm{lem}(A) : A \vee \neg A}$$

Without directions as to how $\mathrm{lem}(A)$ is to be evaluated to a canonical element of $A \vee \neg A$ this postulation remains unjustified on the standard interpretation.

The second problem is more specific to the univalence axiom. A judgement of the form $p : \mathrm{Id}(A, a, b)$ means, according to our explanations above, that p evaluates to some $\mathrm{refl}(A', a')$ where $A' = A : \mathbf{set}$, $a' = a : A$, and $a' = b : A$. It follows, however, from the univalence axiom that for some $A : \mathbf{set}$ and $a, b : A$ there is a $p : \mathrm{Id}(A, a, b)$ that cannot be taken to evaluate to such a canonical element. Let us consider one example.

Let **2** be the **set** of booleans, whose canonical elements are t and f. A universe $\mathcal{U}$ is, intuitively, a **set** whose elements are themselves **set**s, so that we may for instance have $\mathbf{2} : \mathcal{U}$. A canonical element of $\mathcal{U}$ is a canonical **set**. With one universe present one can form the proposition

$$\mathrm{Id}(\mathcal{U}, \mathbf{2}, \mathbf{2})$$

that is, a proposition to the effect that **2** is identical to **2**. Let us make the definition

$$\mathbf{I_2} = \mathrm{Id}(\mathcal{U}, \mathbf{2}, \mathbf{2}) : \mathbf{prop}$$

According to our explanations, a canonical element of this **set** is any $\mathrm{refl}(\mathcal{A}, \mathcal{B})$ such that

$$\mathcal{A} = \mathcal{U} : \mathbf{set}$$
$$\mathcal{B} = \mathbf{2} : \mathcal{U}$$

From the univalence axiom and certain basic facts concerning **2** it follows that there is a function

$$F : \mathbf{2} \to \mathbf{I_2}$$

that is injective and surjective in a suitable sense. We write $\mathrm{ap}(F, a)$ for the application of F to the argument a. For $A : \mathbf{set}$ we introduce a new form of judgement, A true, governed by the rule

$$\frac{a : A}{A \text{ true}}$$

The injectivity of F is the following property: for any $x, y : \mathbf{2}$, if

$$\mathrm{Id}(\mathbf{I_2}, \mathrm{ap}(F, x), \mathrm{ap}(F, y)) \text{ true}$$

then

$$\mathrm{Id}(\mathbf{2}, x, y) \text{ true}$$

The surjectivity of F is the following property: for any $p : \mathbf{I_2}$ there is $x : \mathbf{2}$ such that

$$\mathrm{Id}(\mathbf{I_2}, \mathrm{ap}(F, x), p) \text{ true}$$

Without loss of generality, we may assume that t : **2** is such that

$$\mathrm{Id}(\mathbf{I_2}, \mathrm{ap}(F, \mathsf{t}), \mathrm{refl}(\mathcal{U}, \mathbf{2})) \text{ true}$$

Then $\mathrm{ap}(F, \mathsf{f}) : \mathbf{I_2}$ cannot be evaluated to canonical form. For, if it did, we should have

$$\mathrm{ap}(F, \mathsf{f}) = \mathrm{refl}(\mathcal{U}, \mathbf{2}) : \mathbf{I_2}$$

and therefore also

$$\mathrm{Id}(\mathbf{I_2}, \mathrm{ap}(F, \mathsf{t}), \mathrm{ap}(F, \mathsf{f})) \text{ true}$$

The injectivity of F would then allow us to infer

$$\mathrm{Id}(\mathbf{2}, \mathsf{t}, \mathsf{f}) \text{ true}$$

but it is a theorem of type theory with one universe that

$$\neg\mathrm{Id}(\mathbf{2}, \mathsf{t}, \mathsf{f}) \text{ true}$$

Thus, not only is it not clear how to justify the univalence axiom on the grounds of the standard interpretation—the univalence axiom appears to be in direct conflict with this interpretation. Precisely how one should think of homotopy type theory as a foundation for mathematics, be it classical or constructive, is thus not straightforward. The standard interpretation provides Martin-Löf type theory, and therefore also the mathematics based on it, with solid foundations; but it seems that those foundations are not strong enough to carry the extension of that theory that is homotopy type theory.

6 Other Suggested Justifications

Assume that Π and Σ have been introduced into type theory in the usual way and that moreover Id-formation and Id-introduction have been laid down. Then, for any $A : \mathbf{set}$ and C such that $z : A \vdash C : \mathbf{prop}$, we may define

$$\mathrm{Ind}(A, C) = (\Pi x, y : A)(\mathrm{Id}(A, x, y) \supset (C[x] \supset C[y])) : \mathbf{prop}$$

The proposition $\mathrm{Ind}(A, C)$ says that if $\mathrm{Id}(A, x, y)$ and $C[x]$ are true, then so is $C[y]$; in other words, that C does not discern elements of A that stand in the $\mathrm{Id}(A)$-relation. For any $A : \mathbf{set}$ and $a : A$ we may also define

$$E(A, a) = (\Sigma x : A)\mathrm{Id}(A, a, x) : \mathbf{prop}$$

and

$$G(A, a) = (\Pi z : E(A, a))\, \mathrm{Id}(E(A, a), \langle a, \mathrm{refl}(A, a)\rangle, z) : \mathbf{prop}$$

A canonical element of $E(A, a)$ is a pair $\langle b, p \rangle$ where $b : A$ and $p : \mathrm{Id}(A, a, b)$. An element of $G(A, a)$ is a function f such that for any $b : A$ and $p : \mathrm{Id}(A, a, b)$,

$$\mathrm{ap}(f, \langle b, p \rangle) : \mathrm{Id}(E(A, a), \langle a, \mathrm{refl}(A, a)\rangle, \langle b, p \rangle)$$

It is known that if we postulate

$$\text{(Ind)} \qquad\qquad \mathrm{subst}(A, C) : \mathrm{Ind}(A, C)$$

and

$$\text{(Uniq)} \qquad\qquad \mathrm{uniq}(A, a) : G(A, a)$$

then the rule of Id-elimination may be derived.

The justification of Id-elimination suggested by Ladyman and Presnell (2015, p.401ff.) consists in treating (Ind) and (Uniq) as primitive principles of type theory and showing how Id-elimination may be derived on that basis.

It is not clear to me what is gained by this approach. Although the postulation of (Ind) and (Uniq) allows one to derive Id-elimination, these postulations themselves of course have to be justified. According to the standard interpretation it must thus be made evident that $\mathrm{subst}(A, C)$ and $\mathrm{uniq}(A, a)$ evaluate to elements of canonical form. Ladyman and Presnell, however, provide no indication—in the form of equality rules or informal description, say—of how such

an evaluation should proceed. Instead they provide only certain heuristic considerations. Thus, they say of (Ind) that it expresses 'a fundamental part of our pre-mathematical understanding of identity' that 'any formalisation of identity must respect' (p. 404). One can agree with this without regarding it as a justification of (Ind). That a proposition ought to be true is no justification that it is true.

The heuristic considerations offered in support of (Uniq) seems to be that some propositional uniqueness principle ought to hold for Id-**set**s, since such principles hold for the other **set**-forming operators of Martin-Löf type theory.[9] It can, for instance, be shown that for every element z of $A \wedge B$ there are $a : A$ and $b : B$ such that $Id(A \wedge B, z, \langle a, b \rangle)$ is true; in other words, that every element of $A \wedge B$ stands in the $Id(A \wedge B)$-relation to a canonical element. It is, however, difficult to see why this heuristic in the case of Id should lead to the postulation of (Uniq) rather than to the postulation of

$$uip(A, a) : (\forall p : Id(A, a, a)) \, Id(Id(A, a, a), p, refl(A, a))$$

where $A :$ **set** and $a : A$. Let us call the displayed judgement (UIP). Canonical elements of Id-**set**s have the form $refl(A, a)$, and, in particular, when $a : A$, then $refl(A, a)$ is a canonical element of $Id(A, a, a)$. According to (UIP), every element of $Id(A, a, a)$ stands in the $Id(Id(A, a, a))$-relation to $refl(A, a)$. This judgement therefore expresses a uniqueness principle. The postulation of (UIP) is, however, inconsistent with the univalence axiom, as the example of $Id(\mathcal{U}, \mathbf{2}, \mathbf{2})$ above shows.[10]

In another paper, Ladyman and Presnell (2016) present what they call a semantics for type theory (§§ 5.2, 7.2). They take type theory furnished with this semantics to be a theory that may provide what they call an autonomous foundation for mathematics (§§ 2.2, 9). The so-called semantics amounts, however, to saying that a judgement of the form $a : A$ means that a is a specific concept that falls under the general concept A. In addition, some very general remarks on concepts are made, such as: the existence of a concept is independent of whether anything real falls under it; concepts 'may or may not depend for their existence on mental activity' (§ 5.2 point (3)); concepts are intensional in nature; and they have structure and may be composed from other concepts. All of this might be reasonable, but does not suffice as a semantics in any serious sense. It is, for instance, quite impossible to use this semantics to determine whether the various postulations considered above are justified or not. Being told merely that $uniq(A, a) : G(A, a)$ means that the concept $uniq(A, a)$ falls under the concept $G(A, a)$, I am

not in a position to say whether this judgement is justified or not. In order to determine that I would at least need to know more precisely what is understood by specific and general concepts; by something's falling under a concept; and which concepts $uniq(A, a)$ and $G(A, a)$ are.

The approach suggested by Ladyman and Presnell moreover breaks the general pattern in type theory of furnishing **set**-forming operators with introduction- and elimination rules, since the Id-elimination rule is now replaced by two postulates that cannot be regarded as elimination rules. It seems to me a great insight, first developed in Martin-Löf (1971), that propositional identity can be treated by means of such rules. As a constant of the language, the propositional identity symbol is thereby shown to admit of a treatment similar to that of the standard logical constants. Ladyman and Presnell have given no justification for why we should let go of this insight.

Walsh (2017), in a paper presented as a response to Ladyman and Presnell (2015), suggests a very different approach to the topic. He describes how the proof-theoretical notion of harmony can be characterized in category-theoretic terms and how various logical operators, including identity, is captured by this characterization. This is then meant to constitute a justification of the elimination rules of these operators. In Walsh's own words, Id-elimination is in harmony with Id-introduction because 'path induction [i.e. Id-elimination] is determined by the adjoint which defines identity' (p. 319). This result will certainly be of interest to those fluent in the language of category theory, but the sense in which it could be taken to constitute a justification for Id-elimination is not that which has been assumed here. There, the justification proceeds by translation into another mathematical theory, viz., category theory. Here, we have provided informal yet precise explanations of the basic notions and of the forms of statement of type theory that do not rely on any other mathematical theory. The explanations have been rigorous enough both to render meaningful the question of whether a certain postulation is justified and to answer that question positively in the case of Id-elimination.

References

Boolos G (1971) The iterative conception of set. J Philos 68:215–232

---

[9] In the case of Π the higher-order presentation is needed; see Garner (2009, Thm. 4.2').

[10] That (UIP) is not derivable in Martin-Löf type theory with the Id-elimination rule assumed in this paper is a non-trivial meta-mathematical result first proved by Hofmann and Streicher (1998).

Curry HB (1942) The inconsistency of certain formal logics. J Symb Log 7:115–117

Dybjer P (1994) Inductive families. Form Asp Comput 6:440–465

Garner R (2009) On the strength of dependent products in the type theory of Martin-Löf. Ann Pure Appl Log 160:1–12

Hofmann M, Streicher T (1998) The groupoid interpretation of type theory. In: Sambin G, Smith JM (eds) Twenty-five years of constructive type theory. Oxford University Press, Oxford, pp 83–111

Husserl E (1891) Philosophie der Arithmetik. C.E.M. Pfeffer, Halle

Ladyman J, Presnell S (2015) Identity in homotopy type theory, part I: the justification of path induction. Philos Math 23:386–406

Ladyman J, Presnell S (2016) Does homotopy type theory provide a foundation for mathematics? Br J Philos Sci. doi:10.1093/bjps/axw006

Martin-Löf P (1971) Hauptsatz for the intuitionistic theory of iterated inductive definitions. In Fenstad JE (ed) Proceedings of the second Scandinavian logic symposium, pp 179–216. North-Holland, Amsterdam

Martin-Löf P (1975) An intuitionistic theory of types: predicative part. In: Rose HE, Shepherdson JC (eds) Logic colloquium '73. North-Holland, Amsterdam, pp 73–118

Martin-Löf P (1982) Constructive mathematics and computer programming. In: Cohen JL, Łoś J et al (eds) Logic, methodology and philosophy of science VI. North-Holland, Amsterdam, pp 153–175

Martin-Löf P (1984) Intuitionistic type theory. Bibliopolis, Naples

Nordström B, Petersson K, Smith J (1990) Programming in Martin-Löf's type theory. Oxford University Press, Oxford

Nordström B, Petersson K, Smith JM (2000) Martin-Löf's type theory. In: Abramsky S, Gabbay D, Maibaum TSE (eds) Handbook of logic in computer science. Logic and algebraic methods, vol 5. Oxford University Press, Oxford, pp 1–37

Prawitz D (1965) Natural deduction. Almqvist & Wiksell, Stockholm

Prawitz D (1971) Ideas and results in proof theory. In Fenstad JE (ed) Proceedings of the second Scandinavian logic symposium, North-Holland, Amsterdam, pp 235–307

Prawitz D (2015) A short scientific autobiography. In: Wansing H (ed) Dag prawitz on proofs and meaning. Outstanding contributions to logic. Springer, Dordrecht, pp 33–64

Shoenfield J (1977) The axioms of set theory. In: Barwise J (ed) Handbook of mathematical logic. North-Holland, Amsterdam, pp 321–344

The Univalent Foundations Program (2013) Homotopy type theory: Univalent foundations of mathematics. Institute for Advanced Study, Princeton. http://homotopytypetheory.org/book

Walsh P (2017) Categorical harmony and path induction. Rev Symb Log 10:301–321

Zermelo E (1930) Über Grenzzahlen und Mengenbereiche. Fundam Math 16:29–47