

# Finite Type Arithmetic

Computable Existence Analysed by Modified Realisability and  
Functional Interpretation



Klaus Frovin Jørgensen

Master's Thesis, March 19, 2001

Supervisors: Ulrich Kohlenbach and Stig Andur Pedersen

Departments of Mathematics and Philosophy

University of Roskilde

## Abstract in English and Danish

Motivated by David Hilbert's program and philosophy of mathematics we give in the context of natural deduction an introduction to the Dialectica interpretation and compare the interpretation with modified realisability. We show how the interpretations represent two structurally different methods for unwinding computable information from proofs which may use certain prima facie non-constructive (ideal) elements of mathematics. Consequently, the two interpretations also represent different views on what is to be regarded as constructive relative to arithmetic. The differences show up in the interpretations of extensionality, Markov's principle and restricted forms of independence-of-premise. We show that it is computationally a subtle issue to combine these ideal elements and prove that Markov's principle is computationally incompatible with independence-of-premise for negated purely universal formulas.

In the context of extracting computational content from proofs in typed classical arithmetic we also compare in an extensional context (i) the method provided by negative translation + Dialectica interpretation with (ii) the method provided by negative translation +  $A$ -translation + modified realisability. None of these methods can be applied fully to  $E\text{-PA}^\omega$ , since  $E\text{-HA}^\omega$  is not closed under Markov's rule, whereas the method based on the Dialectica interpretation can be used if only weak extensionality is required.

Finally, we present a new variant of the Dialectica interpretation in order to obtain (the well-known) existence property, disjunction property and other closure results for typed intuitionistic arithmetic and extensions hereof. Hence it is shown that functional interpretation can be used also for this purpose.



Vi giver med udgangspunkt i Hilberts program og hans matematikfilosofi en introduktion til Dialectica fortolkningen inden for naturlig deduktion. Denne fortolkning sammenlignes med modificeret realiserbarhed, og vi viser, hvordan de to fortolkninger repræsenterer to strukturelt forskellige metoder til at uddrage beregnbar information fra ikke-konstruktive beviser. Således repræsenterer de to fortolkninger forskellige syn på, hvad der er konstruktivt. Forskellighederne viser sig især i forbindelse med ekstensionalitet, Markovs princip og princippet om uafhængighed af præmis. Vi viser, at det er et subtilt emne at kombinere disse ideal-elementer, og at Markov's princip og princippet om uafhængighed af negeret universelle præmisses i særdeleshed er uforlignelige med hensyn til beregnbarhed.

Med hensyn til beregnbart indhold af klassiske beviser inden for typet aritmetik sammenligner vi i en ekstensionel kontekst (i) metoden givet ved negativ oversættelse + Dialectica fortolkning med (ii) metoden givet ved negativ oversættelse +  $A$ -oversættelse + modificeret realiserbarhed. Ingen af disse metoder kan benyttes i fuld udstrækning inden for  $E\text{-PA}^\omega$ , da  $E\text{-HA}^\omega$  ikke er lukket under Markovs regel. Metoden baseret på Dialectica fortolkningen kan derimod anvendes, hvis det kun kræves, at teorien har svag ekstensionalitet.

Til sidst introduceres en ny variant af Dialectica fortolkningen. Denne benyttes til at vise (de velkendte) eksistensegenskab, disjunktionsegenskab og andre egenskaber med hensyn til lukkethed for typet intuitionistisk aritmetik inklusive en ekstension. Således vises det, at funktional fortolkning også kan benyttes til dette formål.

## Preface

This master's thesis is written for obtainment of a master's degree in both mathematics and philosophy. Hence, the present text treats topics from both fields, as they are connected in logic. The specific purpose is to give a systematic analysis of certain *prima facie* non-constructive elements of mathematics by using different interpretations and translations provided by proof theory. My hope is that the reader will acknowledge this to have significance both within mathematics and philosophy.

My gratitude goes to my supervisors: Ulrich Kohlenbach in mathematics and Stig Andur Pedersen in philosophy. Ulrich Kohlenbach has been most generous with his time and expertise, and all results presented here are obtained in collaboration with him. Stig Andur Pedersen introduced me several years ago to logic and proof theory and to a very fruitful, I think, view on mathematics. They have both in a very friendly and constructive way supported and guided me.

I also want to thank Ulrich Berger for giving me the idea, that one should give a *Dialectica* interpretation within natural deduction. I have gained a lot in pursuing this. With respect to the final proof reading I want to thank Vincent F. Hendricks for his thorough treatment; also Dan Temple adjusted my language. Finally, thanks also to Jesper H. Andersen and Ivar Rummelhoff for useful comments.



# Contents

<b>1</b>	<b>Setting and Object of Survey</b>	<b>1</b>
1.1	Hilbert’s program and view on mathematics . . . . .	3
1.2	Failure of Hilbert’s program—failure of separation . . . . .	5
1.3	The proof interpretation by Heyting . . . . .	7
1.4	Gödel’s view on a constructive foundation . . . . .	9
1.5	Motivation for and focus of this thesis . . . . .	11
<b>2</b>	<b>Introduction to Gödel’s Dialectica Interpretation</b>	<b>15</b>
2.1	A Hilbert system for weakly extensional Heyting arithmetic in all finite types . . . . .	15
2.2	Axioms and rules of $\text{WE-HA}_H^\omega$ . . . . .	17
2.3	$\text{WE-T}_H$ as the quantifier free subsystem of $\text{WE-HA}_H^\omega$ . . . . .	21
2.4	Definition and analysis of Dialectica translation . . . . .	22
2.5	Interpretation theorem for $\text{WE-HA}_H^\omega$ . . . . .	27
<b>3</b>	<b>Interpretation Theorems within Natural Deduction</b>	<b>29</b>
3.1	Formulation of $\text{WE-HA}_{\text{ND}}^\omega$ and $\text{WE-T}_{\text{ND}}$ . . . . .	29
3.2	Discussion of the deduction theorem . . . . .	33
3.3	Translation of derivations under assumptions . . . . .	36
3.4	Interpretation theorem for $\text{WE-HA}_{\text{ND}}^\omega$ . . . . .	37
3.5	Relation between the Dialectica interpretation and the Diller-Nahm interpretation . . . . .	48
3.6	Intuitionistic linear logic and the Dialectica interpretation . . . . .	49
3.7	Interpretation theorem for $\text{WE-HA}^\omega + \text{MP}^\omega + \text{IP}_{\forall}^\omega + \text{AC}$ . . . . .	51
3.8	The non-constructive theory $\text{WE-HA}^\omega + \text{IP}_{\neg\forall}^\omega + \text{MP}^\omega$ . . . . .	53
<b>4</b>	<b>Classical Arithmetic and Kuroda’s Negative Translation</b>	<b>55</b>
4.1	Formulation of $\text{WE-PA}^\omega$ . . . . .	55
4.2	From classical to intuitionistic logic: Kuroda’s negative translation . . . . .	56
4.3	Interpretation theorem for $\text{WE-PA}^\omega + \text{QF-AC}$ and consistency of the theory . . . . .	59
4.4	Extraction theorem and conservativeness . . . . .	60
4.5	The philosophical significance of interpretation theorems . . . . .	62
4.6	$\text{PA}$ as a subsystem of $\text{WE-PA}^\omega$ . . . . .	65
4.7	The no-counterexample interpretation of Peano arithmetic . . . . .	65
<b>5</b>	<b>Modified Realisability, A-translation and Applications</b>	<b>69</b>
5.1	Definition of $\text{E-HA}^\omega$ and modified realisability . . . . .	70
5.2	Realisability interpretation of $\text{E-HA}^\omega + \text{AC} + \text{IP}_{\text{ef}}^\omega$ . . . . .	72
5.3	Contraction and negatively occurring universal quantifiers . . . . .	74
5.4	Markov’s principle intuitionistically unprovable . . . . .	75
5.5	$\text{E-HA}^\omega \pm \text{IP}_{\text{ef}}^\omega \pm \text{AC}$ not closed under Markov’s rule . . . . .	75

5.6	The Friedman-Dragalin $A$ -translation for formulas of HA . . . . .	77
5.7	Realisability with truth: Closure properties . . . . .	80
<b>6</b>	<b>Closure under Rules by Functional Interpretations</b>	<b>84</b>
6.1	There is no ‘Dialectica-with-truth’ interpretation of WE-HA <sup>ω</sup> . . . . .	84
6.2	Definition and soundness of $Q$ -translation . . . . .	86
6.3	Closure properties of intuitionistic arithmetic showed by $Q$ -translation . . . . .	98
6.4	$Q$ -interpretation is not closed under deductions . . . . .	100
6.5	Closure properties of WE-HA <sup>ω</sup> + IP <sub>∇</sub> <sup>ω</sup> + MP <sup>ω</sup> + AC + Γ by Dialectica . . . . .	100
6.6	Constructiveness of WE-HA <sup>ω</sup> + IP <sub>∇</sub> <sup>ω</sup> + MP <sup>ω</sup> + AC + Γ . . . . .	101
<b>7</b>	<b>Conclusions</b>	<b>103</b>
7.1	Interpreting mathematics by methods from proof theory . . . . .	104
7.2	Evaluation of modified realisability and Dialectica interpretation . . . . .	106
7.3	Two strong constructive theories . . . . .	107
7.4	Dialectica as a tool in proof mining and reductive proof theory . . . . .	107
7.5	Different interpretations—different validations: Mathematics and language . . . . .	109
	<b>Bibliography</b>	<b>111</b>
	<b>Index</b>	<b>119</b>

## Setting and Object of Survey

This thesis concerns proof theory and the significance of certain proof theoretical investigations. Proof theory is the part of mathematical logic which studies the concepts of mathematical proofs and mathematical provability. Proofs are an indispensable part of mathematics and therefore proof theory is also a study of the foundations of mathematics. Let us elaborate on the connection between proof theory and philosophy of mathematics.

*First*, proof theory studies proofs as formal objects whereas, generally, mathematical proofs are informal and to a certain degree imprecise. There is, however, in practice a close affinity between formal and informal proofs. Parts of proofs used in mathematical arguments can be more or less formal. But in case there is a disagreement on a proof one can try to make the problematic part of the proof more formal in order to be precise on the matter. One can, in other words, formalise in different degrees and the complete formal proof can be viewed as some kind of idealisation of the mathematical proof. It is in this respect proof theory studies mathematical proofs.

*Second*, from results obtained in the 1930s in mathematical logic it follows that there cannot be formal theories which represent *all* of mathematics: For any given formal theory containing a minimum of arithmetic there will always be mathematical theorems which are independent of that theory. There are, in other words, theorems of the theory which cannot be proved nor disproved. It is, nevertheless, possible to represent big bodies of mathematics in formal frameworks and to study internal relationships. A part of mathematical logic called reverse mathematics is an impressive example of this; see e.g. (Simpson, 1999). Therefore, relative to specific mathematical theories proof theoretical investigations can have important philosophical consequences. We will call such investigations *local*.

But of course, proof theory does not give the full picture of a philosophy of mathematics. General epistemological, ontological, sociological and historical issues – which are treated in philosophy, philosophy of science and history of science – play an important role in the overall philosophical analysis of mathematics. In this thesis, however, we will only touch the epistemological and ontological questions in so far as is necessary in order to derive philosophical consequences of the proof theoretical investigations.

We will apply different methods from proof theory in order to study constructivity of proofs and principles. Now, it is in itself a difficult question to point out what constitutes a constructive proof. Below we will line up certain properties which a constructive system should meet. But as it turns out, there is no unambiguous characterisation of constructivism. Anyhow, a desirable property of a constructive theory is certainly the following: If we have proved an existential statement  $\exists xA(x)$  then we can in fact exhibit an object  $t$  such that  $A$  holds for  $t$ .<sup>1</sup> Similarly, if we have proved a disjunction  $A \vee B$ , with no free variables then we can tell which of the two actually holds. We will now give a simple example of a non-constructive proof, due to H. Friedman.

---

<sup>1</sup>For an interesting and informative survey of the historical roots of these matters (which at least go back to Whitehead and Russell's *Principia Mathematica* from 1910–13) see Mancuso (n.d.). In connection with this paper it should, however, be noted that it is somehow difficult to follow Mancuso's separation of direct and indirect non-constructive proofs, since they both by the end of the day rest on *tertium non datur*.

**Theorem.**  $e - \pi$  is irrational or  $e + \pi$  is irrational.

**Proof.** Assume they both are rational. Since the sum of two rationals is a rational we have that  $2e$  is rational. Contradiction.  $\dashv$

It is questionable how much information this proof contains. Surely, the proof does not tell us which of the two disjuncts is irrational. A little surprising due to the simplicity of the problem, no method or proof on this problem has until now shown us *which* of the two actually is irrational. The above proof only tells us that both cannot be rational, but it does not exhibit a number of the two such that this number is the one—clearly it would have been more satisfying if we knew that.

There are many examples from all over mathematics where existence statements or disjunction statements are proved but the proofs do not provide us with any instances or any algorithms which make it possible for us to see *why* the theorems are true—all we see is that it cannot be the case that their negations are true. This is certainly constructively unsatisfying.<sup>2</sup> Indirect proofs are in general not constructive. But in lots of cases they are or seem to be the only possible; as for instance in the above proof.

At the entrance to the 20th century mathematics had its so-called foundational crisis, where fundamental questions appeared during the development of new mathematical theories. An example of such a theory is set theory as developed by Georg Cantor (1845–1918), who introduced a beautiful and useful theory of transfinite arithmetic—arithmetic on infinite numbers. Another example is the development of a general theory of topology, mainly by Felix Hausdorff (1868–1942), where concepts from analysis such as continuity were generalised to quite abstract notions. Many of these new theories and concepts involved infinite totalities and, moreover, inconsistencies and difficult open foundational problems were found in the very first formulations of set theory. The paradox of Bertrand Russell (1872–1970) showed that not all properties can be used for defining sets and Cantor posed the continuum hypothesis (CH) which asserts that there is no set with cardinality strictly greater than the set of natural numbers but also strictly less than the set of real numbers. If one views the CH as a definite mathematical problem then it has remained open—it has even been shown that it cannot be decided within the ZFC formalisation of set theory.

Meanwhile, also proof techniques were questioned. In 1888 David Hilbert (1862 – 1943) proved his famous basis theorem in invariant theory by non-constructive methods but experts on the field – e.g. P. Gordan – called Hilbert’s proof “theology”.<sup>3</sup> L.E.J. Brouwer (1881 – 1966), the founder of intuitionism, banned in his thesis from 1907 indirect proofs and non-constructive methods in general. Also Hermann Weyl (1885 – 1955), a student of Hilbert, became unsatisfied with set theoretical foundations and developed in (Weyl, 1918) his kind of constructive mathematics, which later became known under the name “predicative mathematics”. The so-called “Grundlagenstreit” was just about to start. Many important mathematicians such as Hilbert, Brouwer and Weyl took part in it, and – as we shall see – so did

---

<sup>2</sup>It is also epistemologically problematic to assume that any mathematical statement has a truth value—also in cases where it completely impossible for us to get to *know* what that value is. We will not, however, deal with that problem here, but see for instance (Jørgensen & Pedersen, 2000). Also: Dummett (1977) argues that truth must be something which is *accessible* to us.

<sup>3</sup>More on this in (Rowe, 2000).

Kurt Gödel (1906 – 1978) and Arend Heyting (1898 – 1980). Now and then the discussions were rather emotional, but many important notions, insights and interpretations arose from this intense period. One of them was proof theory.

### 1.1 Hilbert's program and view on mathematics

Hilbert had a long mathematical career and was contributing in nearly all areas of mathematics such as algebra, geometry, number theory, mathematical physics, mathematical logic, etc. He had worked with the new mathematical theories and seen into the uncountable infinite that Cantor's set theory was about. He had seen and used the effectiveness and elegance of non-constructive methods. Thus, "Aus dem Paradies, daß Cantor uns geschaffen, soll uns niemand vertreiben können", as Hilbert (1926, 170) puts it in his famous metaphor. Therefore he developed his program.

The program was at heart Kantian.<sup>4</sup> It was formulated fully – by Hilbert and his co-workers, especially Paul Bernays (1888–1977) – and put forth in the 1920s (Hilbert, 1922, 1926). The philosophical position behind the program is the following. Mathematics can be split into two parts:

1. The finitary (or contentual) part of mathematics.
2. The ideal part of mathematics.

The *finitary part of mathematics* was meant to be that part of mathematics of which there could be no doubts: Finitary reasoning about the natural numbers, i.e. no unrestricted quantifiers, and simple reasoning on finite graphs and geometrical figures. Today there is a general agreement on, that what Hilbert took to be the finitary part of (informal) mathematics was, at least, what can be coded and justified by primitive recursive arithmetic (PRA), see (Tait, n.d.) for a discussion of this. *Ideal mathematics* was taken to be the highly abstract elements of mathematics of which the ontological status were not immediate. Examples of such ideal parts of mathematics could be (i) completed infinities, such as the set of natural numbers  $\omega$ , needed to develop a theory of the transfinite; (ii) the expansion of the real numbers by the complex number  $i = \sqrt{-1}$ , which enables us to prove the fundamental theorem of algebra; (iii) the 'completion' of Euclidean geometry by an infinite line consisting of points at infinity thus obtaining projective geometry, and so forth. As such, the ideal elements could not, according to Hilbert, be perceived by the senses. They had more the role of completing the finitary and regulating it—in the same way as in Kant's theory of knowledge where ideas of reason regulate knowledge.<sup>5</sup> The ideal elements were supposed to be abstract elements introduced in the development of mathematics in order to simplify, generalise and complete already existing mathematics. But in such a process new mathematics would also arise and this was how Hilbert saw the expansion and progression of mathematics. These

<sup>4</sup>Details on this is provided in Andersen (2000). Andersen also discusses the conflicts between Kant's view on mathematics and Hilbert's view on mathematics—in particular the role of the infinite. A theme we will not touch here, but see also Majer (1993), Detlefsen (1995) and Posy (1995).

<sup>5</sup>Note, that Kant (1781/87) indeed had two kinds of 'ideas': (i) The constitutive use of transcendental ideas causing paralogisms and antinomies of pure reason (Kant, 1781/87, A338-A567); and (ii) the regulative use of general ideas of pure reason, as used in science (Kant, 1781/87, A642-A704).

ideal elements of the mathematical method and universe were, of course, of indispensable value. However, Hilbert was at the same time aware of the fact that this progressiveness of the mathematical method was transcending the secured finitary parts of mathematics and it was therefore in need of some kind of justification.

This justification would consist in showing, mathematically, that the ideal part of mathematics could not prove new purely finitary statements, i.e. could not prove finitary statements which were not provable already in the finitary part of mathematics. This is where Hilbert connected the axiomatic approach, the idea of a ‘proof theory’, with his general view on mathematics as just described. In modern technical terms the goal of the program is described by the following.

Let  $S$  be some formal system representing mathematics—both the ideal and the finitary part. A formula in the language of  $S$  is a finite object and it can therefore be coded effectively by a natural number; proofs in  $S$  can likewise be coded. Thus,  $\text{Proof}_S(x, y)$  is a predicate obtaining between two natural numbers  $x$  and  $y$  expressing that  $x$  encodes a proof in  $S$  of some formula having code  $y$ . As is standard, let  $\ulcorner A \urcorner$  be the code of  $A$ . (The argument below is in the given form a little vague—the details rely on the specific properties of the encoding; see (Smorynski, 1977, sect. 2–4).) In technical terms the essence of Hilbert’s program was that for any finitary statement  $R(x)$  with  $x$  as free variable the reflection principle

$$\text{Proof}_S(u, \ulcorner R(\dot{x}) \urcorner) \rightarrow R(x) \quad (1.1)$$

should be provable by finitary means (where  $\dot{x}$  refers to the  $x$ -th numeral). However, it would be sufficient to establish consistency of  $S$  in a finitary way. For suppose one has a proof in  $S$  of some finitary statement  $R(x)$  containing only  $x$  as free variable, hence

$$\text{Proof}_S(u, \ulcorner R(\dot{x}) \urcorner) \quad (1.2)$$

would be finitarily provable. However, if  $R(x)$  were not true for all  $x$  then for some  $c$ ,  $\neg R(c)$  would be provable within  $S$ . In fact we would have, due to  $\Sigma_1$  completeness,

$$\neg R(x) \rightarrow \text{Proof}_S(v_x, \ulcorner \neg R(\dot{x}) \urcorner), \quad (1.3)$$

where  $v_x$  depends on which value  $x$  takes. If, on the other hand, we could prove consistency of  $S$  by finitary means we would have

$$\neg(\text{Proof}_S(u, \ulcorner R(\dot{x}) \urcorner) \wedge \text{Proof}_S(v, \ulcorner \neg R(\dot{x}) \urcorner)). \quad (1.4)$$

Now, (1.2), (1.3) together with (1.4) implies that  $\neg\neg R(x)$  has a finitary proof, and since  $R(x)$  is a finitary statement this implies  $R(x)$ .

The argument is modern version of Hilbert’s argument as found, for instance, in (Hilbert, 1927, 78):

Aber auch wer sich mit der Widerspruchsfreiheit nicht begnügt und noch weitergehende Gewissenskrupel hat, muß die Bedeutung des Beweises der Widerspruchsfreiheit anerkennen, nämlich als einer allgemeinen Methode aus Beweisen für allgemeine Sätze vom Charakter etwa des Fermatschen Satzes, die mit Hilfe der  $\varepsilon$ -Funktion geführt sind, finite Beweise zu gewinnen.<sup>6</sup>

<sup>6</sup>The  $\varepsilon$  operator was a technical invention within proof theory, and Hilbert saw it as representing a highly ideal aspect of mathematics.

On page 78–79 Hilbert then goes on to explain – by use of Fermat's last theorem as an example – how this can be accomplished. In that argument Hilbert also uses  $\Sigma_1$  completeness, although he due to his time does not recognize it as something special. Note, this also throws light on the late Hilbert's conception of the connection between consistency and existence.

Consequently, Hilbert focused on the program of proving consistency of mathematics in order to provide a finitary foundation for ideal mathematics and thereby justify the use of ideal elements. Bernays sums up the advantage and goal of such a program:<sup>7</sup>

Unter diesem Gesichtspunkt werden wir versuchen, ob es nicht möglich ist, jene transzendenten Annahmen in einer solchen Weise zu begründen, daß nur *primitive anschauliche Erkenntnisse zur Anwendung kommen*. (Bernays, 1922, 11)

Moreover, Bernays stresses the importance that the central problem becomes a *mathematical* problem:

Gerade darin liegt der große Vorzug des Hilbertschen Verfahrens, daß die Probleme und Schwierigkeiten, welche sich in der Grundlegung der Mathematik bieten, aus dem Bereich des Erkenntnistheoretisch-philosophischen in das Gebiet des eigentlich Mathematischen übergeführt werden. (Bernays, 1922, 19)

It is important to note that underlying the program is the idea of separating the foundational studies of mathematics from philosophy and epistemology. Of course, the characterisation of finitism is motivated by a philosophical analysis—but when that is done *it is up to the mathematicians to take care of their own foundation*, “gerade wie ja auch der Astronom die Bewegung seines Standortes berücksichtigen, der Physiker sich um die Theorie seines Apparates kümmern muß und der Philosoph die Vernunft selbst kritisiert” as Hilbert (1918, 155) expresses it.

## 1.2 Failure of Hilbert's program—failure of separation

Gödel, though not a student of Hilbert, was attracted to Hilbert's foundational questions and problems. Around 1930 he proved, within less than a year, not only the completeness of first order predicate logic,<sup>8</sup> but he also showed – when pursuing the program – that Hilbert's original program was unattainable. Gödel (1931) showed that for any consistent theory T containing just a minimum of arithmetic and given by a recursively enumerable set of axioms – and certainly this was included in Hilbert's finitism – two things are the case:<sup>9</sup>

---

<sup>7</sup>Bernays uses the words *anschauliche Erkenntnisse* which belong to the Kantian theory of knowledge. These are often translated into English by *intuitive knowledge*. *Intuitive* here refers to *intuition* which is the translation of *Anschauung*. Thus, *intuitive knowledge* is knowledge based on intuition (*Anschauung*) and, consequently, it does not necessarily mean *immediate*. It rather means that when such knowledge is obtained it is qua its obtainment objective. The justification of the ideal elements that Bernays refers to would thus be absolute.

<sup>8</sup>A problem formulated by Hilbert & Ackermann (1928).

<sup>9</sup>Gödel actually assumed the slightly stronger  $\omega$ -consistency, but J.B. Rosser later replaced this by consistency. Furthermore, theorem 2 was stated by Gödel (1931) without proof. The first published proof of the theorem is in (Bernays & Hilbert, 1939).

**Theorem 1.** T is syntactically incomplete in the sense that there exists a sentence  $A$  in the language of T such that T does not prove  $A$  nor  $\neg A$ .

**Theorem 2.** One cannot in T prove the consistency of T.

The theorems are a disaster for Hilbert's program. First of all, they show that there is no absolute proof of the consistency of all mathematics: The reduction of ideal elements used in proofs of finitary statements cannot be obtained by a finitary consistency proof, since such a proof does not exist.

Secondly, it was more than an implicit assumption of Hilbert and his co-workers that the formal systems were representing mathematics completely. As late as 1930 they conjectured that Peano arithmetic was deductively closed, in the sense that adjoining any sentence not provable in Peano arithmetic would make it inconsistent, see (Bernays, 1976, 59). However, this conjecture is certainly refuted by theorem 1, and this has severe consequences for the program. Due to theorem 1 any consistent formalisation  $S$  of any mathematical theory will be incomplete in the sense, that there will be a sentence  $A$  true in the standard model but  $A$  is undecided by  $S$ .<sup>10</sup> Moreover,  $S + A$  will also be incomplete, and such an iterated process of adding undecidable formulas will go on forever. Therefore, there is no obvious way to choose an  $S$  and such a choice becomes a *central* question.

All this shows that the *global* reduction – to use a phrase of Solomon Feferman (2000) – Hilbert had in mind is impossible. But maybe a more *local* reduction is possible?

However, after theorem 2 there is no obvious choice of a constructive part of mathematics to which some of ideal mathematics can be reduced to. But maybe Hilbert's finitism was only a first approximation of constructivism. In any case we see that the separation of foundational studies from philosophy and epistemology, which Hilbert aimed at, is lost. Questions regarding the nature of mathematical concepts and how we come to have knowledge about them are certainly not eliminated as Hilbert hoped. And what constructivism is, or how a satisfactory interpretation of constructivism can be strong enough to interpret – in some way or another – parts of ideal mathematics are very open ended questions after Gödel (1931).

We will try to follow such questions in this thesis. We will single out certain principles of classical mathematics which certainly look non-constructive and therefore belong to ideal mathematics. On the other hand we will also try to give a characterisation of constructivism. But as it turns out, there are different interpretations of mathematics—and different interpretations validate different classical principles.

After theorem 2 it is not obvious what a possible constructive foundation looks like and what we should understand under the term 'constructive'. How can we interpret mathematics and the mathematical language? There are, as we shall see, different layers of constructivism and different understandings of what is to be regarded as a constructive proof. Intuitionism offers one such account.

---

<sup>10</sup>Such sentences need not to be artificial and without mathematical meaning. Paris & Harrington (1977) gave an example of a finite version Ramsey's theorem which is independent of Peano arithmetic. Later Kirby & Paris (1982) gave a simpler independence result concerning so-called Goodstein sequences. Recently H. Friedman has in a series of papers (which can be downloaded at <http://www.math.ohiostate.edu/~friedman>) provided theorems from finite graph theory, which are even unprovable in predicative mathematics, e.g. the Graph Minor Theorem.

### 1.3 The proof interpretation by Heyting

Brouwer, certainly, had a conception of mathematics that was different from Hilbert's. He thought that intuitionistic reasoning should be *the* reasoning in mathematics. Although Brouwer was not in favour of Hilbert's formal approach, Heyting (1930,a, 1934) – a student of Brouwer – gave a formalisation of intuitionistic logic. The basic element motivating Heyting's formalisation was, indeed, Brouwer's idea of 'constructions'. Heyting asked the question 'how should we understand the logical symbols of mathematics'? Below we will see Heyting's answer, i.e. his interpretation. On the basis of this interpretation Heyting singled out the axioms of *Principia Mathematica* which were sound under the interpretation and this led him to the formalisation of intuitionistic logic. Heyting was to some extent anticipated by Kolmogorov (1925), and the interpretation is therefore referred to as the "BHK interpretation".

The interpretation is very general and is based on the notions of informal proofs and constructions. Actually, a proof here should be understood as a *construction* which informally verifies a statement.

The BHK interpretation takes the meaning of prime formulas for granted. Prime formulas in this setting are formulas with finite meaning: We can determine the truth of any (closed) prime formula e.g.  $13^2 = 169$ . For compound  $A$ , BHK then explains " $p$  proves  $A$ ", which we abbreviate by  $p : A$ , in terms of provability of the components. The clauses defining the interpretation are:

- ( $\perp$ )  $\perp$  denotes contradiction and there is no proof of contradiction.
- ( $\wedge$ )  $p : A \wedge B$  iff  $p$  is a pair  $(p_0, p_1)$  such that  $p_0 : A$  and  $p_1 : B$ .
- ( $\vee$ )  $p : A \vee B$  iff  $p$  is a pair  $(p_0, p_1)$ ,  $p_0 \in \{0, 1\}$  and  $p_1 : A$  if  $p_0 = 0$  and  $p_1 : B$  if  $p_0 = 1$ .
- ( $\rightarrow$ )  $p : A \rightarrow B$  iff  $p$  is a construction taking any  $q$  such that  $q : A$  into  $p(q)$  such that  $p(q) : B$ .
- ( $\neg$ )  $p : \neg A$  iff  $p$  is a construction taking any  $q$  where  $q : A$  into  $p(q)$  such that  $p(q) : \perp$ .
- ( $\forall$ )  $p : \forall x A(x)$  iff  $p$  is a construction taking any  $t$  from the intended domain into  $p(t)$  such that  $p(t) : A(t)$ .
- ( $\exists$ )  $p : \exists x A(x)$  iff  $p$  is a pair  $(p_0, p_1)$ , where  $p_0$  is an object of the domain and  $p_1 : A(p_0)$ .

The interpretation tells us what constitutes a constructive proof. For instance, a proof of an existential statement is constructive if it actually provides an instance together with a proof of the desired property of that instance. Likewise it tells us that a constructive proof of a disjunction is—and under this interpretation our proof of the irrationality of  $e - \pi$  or  $e + \pi$  fails (again) to be constructive. Consequently, the interpretation says under which conditions we (constructively) can assert a formula  $A$ .

As mentioned above, Heyting tested the axioms and rules of *Principia Mathematica* under this interpretation. Most of these passed the test, e.g. modus ponens:

$$\frac{A \quad A \rightarrow B}{B}$$

Inductively one would have  $q : A$  and  $p : A \rightarrow B$ , and since  $p$  converts any proof of  $A$  into a proof of  $B$ ,  $p(q)$  will be a proof of  $B$ . The axiom  $\perp \rightarrow A$ , i.e. any formula  $A$  follows from a contradiction, is trivial under BHK since there will never be a proof of  $\perp$  (guaranteed by the first clause); therefore anything well-formed is a proof of  $\perp \rightarrow A$ . One can think of it in terms of a game. Person<sub>1</sub> has to give a proof of  $\perp$ . For any such proof person<sub>2</sub> has to give a proof of  $A$ . But this is very easy for person<sub>2</sub> since he will never have to do anything. Person<sub>2</sub> therefore has a winning strategy for all (games)  $A$ . However, *tertium non datur*:

$$A \vee \neg A$$

is not sound under BHK—since one would have to provide a universal method for obtaining either a proof of  $A$  or a proof of  $\neg A$  for any  $A$ . But this method would then decide any hitherto undecided statements such as, say, Goldbach’s conjecture.

On the other hand it *seems* plausible that a system sound under BHK would have

- *Existence property*: From a proof of a closed formula  $\exists x A(x)$  one can extract a witness  $t$  such that  $A(t)$ , and
- *Disjunction property*: If one obtains a proof of  $A \vee B$ , for  $A, B$  closed then one can tell which of the formulas is true.

But that intuitionistic logic in fact has these properties is not clear from BHK. We will show this later on by a variant of Gödel’s Dialectica interpretation.

Now, the BHK interpretation has some problematic aspects. There is an inherent vagueness in what is meant by (informal) “proof” and “construction”. Furthermore, the property of being a proof of a statement is impredicative and in general not decidable. We say that a definition is impredicative if it refers to a totality which involves the object that is being defined.<sup>11</sup> The impredicativity of BHK shows up, for instance, in the case of implication.  $p$  is a proof of  $A \rightarrow B$  if for *any* proof  $q$  of  $A$ ,  $p(q)$  is a proof of  $B$ . Moreover, it is not in general decidable whether  $p$  is actually able to do this, since there is no finite procedure that generates all possible proofs of  $A$ .<sup>12</sup> This led Kreisel (1962a) to formulate additional clauses, namely:  $p : A \rightarrow B$  iff  $p$  is a pair  $(v, \tilde{p})$  such that for all  $q$  if  $q : A$  then  $\tilde{p}(q) : B$  and  $v$  is a verification that  $\tilde{p}$  actually does this. Likewise Kreisel added additional clauses for the interpretation of  $\forall$  and  $\neg$ . N. Goodmann (1970) developed a rather complicated theory of constructions based on Kreisel’s additional clauses, but as it turned out they were needless for the result Goodman had in mind, namely showing that intuitionistic arithmetic plus countable choice is conservative over intuitionistic arithmetic itself. It generally seems that second clauses are problematic as they only raise the complexity. We see, however, that under the BHK interpretation one will not in general be able to recognise a proof when one sees it.<sup>13</sup>

<sup>11</sup>For instance, the least upper bound axiom used in classical analysis requires an impredicative definition.

<sup>12</sup>G. Kreisel (1987, 397) has formulated the problem in the following way: “Until the mid fifties I found the subject [intuitionistic logic] distasteful because . . . iterated implications made my head spin.

<sup>13</sup>Recently S. Artëmov has given a BHK-semantics for intuitionistic propositional logic, where one interprets “ $p : A$ ” as “ $p$  is a proof of  $A$  in a formal system”, e.g. PA; hence whether  $p$  is such a proof becomes decidable. The logic Artëmov has developed is thus a logic of operations on proofs and Artëmov shows that the propositional part of intuitionistic logic is complete with respect to this semantics. See Artëmov (2001).

### 1.3.1 Negative translation and consistency of classical logic relative to intuitionistic

Later in the 30s Gödel (1933) and, independently, Gerhard Gentzen (1933) (1909–1945) (who was a student of Hilbert and Bernays) discovered a very important relation between classical logic/arithmetic and intuitionistic logic/arithmetic.<sup>14</sup> Using a so-called negative translation they proved that with respect to consistency intuitionistic logic and arithmetic is no better than classical logic and arithmetic. By a negative translation they embedded classical logic in intuitionistic logic in such a way that provability is preserved. Let  $A'$  denote the translation of  $A$ . It was proved that if  $A$  is classically provable then  $A'$  is intuitionistically provable, thus showing that if classical logic is inconsistent then so is intuitionistic logic.<sup>15</sup> This result together with (Gödel, 1931) showed another interesting thing for the Hilbert school: Apparently finitism is much more restrictive than the foundation advocated by Brouwer, since intuitionism is sufficient for consistency of classical arithmetic while finitism is not. Gödel and Gentzen were probably the first to point out this fact, and it had the following impact on a possible continuation of a generalised Hilbert program, as noted in 1967 by Bernays (who in 1930 had thought (1976, 60) that finitism and intuitionism were coextensive):

It thus became apparent that the “finite Standpunkt” is not the only alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of a restriction to finitist methods of reasoning it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inferences. (Bernays, 1967, 502)

The result by Gödel and Gentzen was, in other words, just stressing the point that there were probably different kinds of constructive foundations. In fact this point will be among the conclusions of this thesis.

## 1.4 Gödel's view on a constructive foundation

With respect to Gödel's view in the 30s and beginning of the 40s on proof theory and constructivism there are three important documents (Gödel, 1933a, 1938, 1941). These documents are scripts Gödel made for lectures, but they were not published until 1995. It seems quite clear from these lectures, that Gödel in those years was closer to the ideas and goals of the Hilbert school than has been generally assumed. The three papers just mentioned witness that Gödel viewed his *Dialectica* interpretation – developed at the end of the 30s – as a contribution to the ongoing discussions on the foundations of mathematics, more specific:

1. Gödel generalised Hilbert's finitism to a constructive theory  $\Sigma$ , which was strong enough for a partial realisation of Hilbert's program.
2. The proposal of  $\Sigma$  as a constructive theory can also be viewed as Gödel's replacement of imprecise notions of intuitionism.

<sup>14</sup>Again, Kolmogorov (1925) anticipated this.

<sup>15</sup>Chapter 4 of this thesis is partly devoted to these matters.

Seen from a strictly constructive point of view, basically two notions of classical mathematics are problematic according to Gödel (1933a): (i) The non-constructive notion of existence (based on *tertium non datur*) and (ii) the use of impredicative definitions. These parts are problematic, Gödel says, because of a necessary Platonist presupposition “which cannot satisfy any critical mind and which does not even produce the conviction that they [the classical axioms] are consistent” (Gödel, 1933a, 19). A Platonistic a priori justification of the ideal parts of mathematics (‘ideal’ in the sense of Hilbert) is not enough. Earlier in the year 1933 Gödel discovered the negative translation and he became dissatisfied with intuitionism as a foundation, due to the problems mentioned above. He was of the opinion (1933a, 22) that

the domain of this intuitionistic mathematics is by no means so uniquely determined as it may seem at first sight. For it is certainly true that there are different notions of constructivity and, accordingly, different layers of intuitionistic or constructive mathematics. As we ascend in the series of these layers, we are drawing nearer to ordinary non-constructive mathematics, and at the same time the methods of proof and construction which we admit are becoming less satisfactory and less convincing.

The most fundamental of the constructive theories could very well be Hilbert’s finitism. But for a constructive foundation for parts of ideal mathematics we need more. A candidate is Hilbert’s finitism + quantifier free transfinite induction along constructive ordinals up to  $\epsilon_0$ . This is the approach of Gentzen (1936) who showed consistency of Peano arithmetic relative to this basis. Gödel, in his *Vortrag bei Zilsel* (1938, 12), emphasizes that with respect to the epistemological side “one will not deny a high degree of intuitiveness to the inference by induction on  $\epsilon_0$  thus defined”.<sup>16</sup> In the lecture Gödel is quite positive towards Gentzen’s approach, however, he has another idea which later became the *Dialectica* interpretation.

Returning to the different layers of constructive theories we have, still further up the hierarchy, intuitionism. Thus, we see the impact of Gödel’s incompleteness theorems: There is apparently no clear cut characterisation of a constructive foundation.

In all three lectures Gödel (1933a, 1938, 1941) discusses which criteria a strictly constructive system should meet and these criteria vary only a little in the three lectures. The clearest account is in (Gödel, 1941, 5–6). It boils down to:

- (a) The primitive functions and relations must be calculable and decidable, respectively.
- (b) Existential quantifiers function only as abbreviations of actual constructed objects and propositional operators cannot be applied to universal statements.

In order to meet criteria (b) universal statements  $\forall x A_{\text{qf}}(x)$  (where  $A_{\text{qf}}$  is quantifier free) can not be negated – only in the sense that one has obtained a counterexample  $\exists x \neg A_{\text{qf}}(x)$ , i.e.  $\neg A_{\text{qf}}(t)$  for some  $t$ . These considerations lead to a class of constructively meaningful statements in  $\exists\forall$ -form. Gödel formulates a finite type theory of primitive recursive functionals  $\Sigma$  which meets these criteria.<sup>17</sup> In (Gödel, 1941) the description of  $\Sigma$  is reasonably detailed

<sup>16</sup>We will hardly discuss this approach of Gentzen any further, but see e.g. Andersen et al. (1996).

<sup>17</sup> $\Sigma$  essentially became Gödel’s system T in the published paper (Gödel, 1958). More details on the theory and the constructiveness hereof in forthcoming chapters.

and likewise how classical arithmetic – via negative translation – can be interpreted in  $\Sigma$ . The interpretation of intuitionistic number theory in  $\Sigma$  is essentially the Dialectica interpretation published in 1958, (Gödel, 1958).

#### 1.4.1 Benefits and applications of Gödel's interpretation

It is clear that Gödel understood his interpretation as a contribution to the foundational discussion. But it is more difficult to say precisely in which sense. Gödel mentions (1941, 26–29) four applications of the interpretation. Let HA denote intuitionistic arithmetic, then:

1. There is a number theoretical formula  $A(x)$  such that for  $C \equiv \neg\forall x(A(x) \vee \neg A(x))$  we have  $\text{HA} + C$  is consistent (given HA is).
2. If HA proves  $\exists xA(x)$  then  $\Sigma$  will prove the *translated* formula  $\exists xA^D(x)$ . And since existential quantifiers in  $\Sigma$  are only abbreviations one has that  $\Sigma$  proves  $A^D(t)$  for some term  $t$  of  $\Sigma$ .
3. Negative translation together with the new interpretation proves consistency of classical arithmetic relative to  $\Sigma$ .
4. The following rule holds: If classical arithmetic proves the closed formula  $\exists xA_{\text{qf}}(x)$  then we can find a number  $n$  such that  $\Sigma$  proves  $A_{\text{qf}}(n)$ , for  $A_{\text{qf}}$  quantifier free.

At the very end of the lecture (1941, 30) Gödel says that “[i]t is perhaps not altogether hopeless to try to generalize these consistency proofs to analysis by means of functions of still higher (i.e. transfinite) type.” In fact Spector (1962) generalised Gödel's interpretation to analysis by adding a generalised version of bar induction (induction along well-founded trees), and later on Friedrich (1985) provided a generalisation to the transfinite types. Gödel's interpretation can, on the other hand, also be used to show that a mathematically rich subsystem of analysis named  $\text{WKL}_0$  can be reduced to PRA. This follows from a result of U. Kohlenbach (1992).

It is probably a combination of all these applications and possible generalisations that were the motives for Gödel. According to A.S. Troelstra (1990, 219) (i) S.C. Kleene reports that the above 1 was the principal goal; (ii) Kreisel says that Gödel wanted to establish that intuitionistic proofs of existential theorems provide explicit realisations (in the sense of 2). Whereas the publication of the interpretation (Gödel, 1958) is most clearly devoted to 3. The motives of Gödel were probably a combination.

### 1.5 Motivation for and focus of this thesis

The view on mathematics underlying Hilbert's program is interesting. It has a high explanatory power: First, it explains clearly why a great body of mathematics *is* robust and unchanged over time. This is simply so because it is finitary or is founded immediately on such grounds. Those parts of mathematics which are finitarily justified<sup>18</sup> are intuitively true and therefore

<sup>18</sup>Notice that we use the word “justified”. By this we mean that what is justified by, say, primitive recursive reasoning is finitarily justified and therefore intuitively true. But note also, that if we allow (free) variables in PRA

objective.<sup>19</sup> Second, it also explains why some objects are introduced in mathematics, then abandoned, but later on – maybe – re-introduced; as infinitesimals for instance. Such objects are simply ideal elements, and the introduction of ideal elements is in general indeed a non-trivial matter. It is a difficult process to define and thereby introduce new (ideal) objects and concepts which are meaningful and useful. Sometimes they are, but sometimes they are not and will perhaps need to be modified.<sup>20</sup> Thus, the philosophy of ideal elements leaves room for the possibility of errors made by mathematicians doing their science. However, fluctuation will only happen on the very edge of the whole science, so to speak. On the other hand, such a philosophy also leaves room for creativity and progressiveness.<sup>21</sup> It gives an account of how and why mathematical theories grow and evolve over time.

The motivation underlying this thesis is the just described generalized Hilbertian view on mathematics. But contrary to Hilbert we will not focus on consistency. Primarily so because mathematicians and logicians have been working for a long time with mathematics and formal theories representing mathematics. Today, only a few sceptics (if any at all) in the mathematical and logical enterprise doubt the consistency of, say, ZFC. We will mainly focus on other themes from the ‘Grundlagenstreit’, namely on the constructivity of proofs and, generally, on the permissibility<sup>22</sup> and justification of ideal methodology. In this respect we will be in line with Kreisel and his program of “unwinding” proofs:

*To determine the constructive (recursive) content or the constructive equivalent of the non-constructive concepts and theorems used in mathematics, particularly arithmetic and analysis. (Kreisel, 1958, 155).*

By not focusing on consistency as Hilbert did we intend to ascribe more specific constructive meaning to non-constructive proofs. The activity of applying this to specific proofs could be named *proof mining* – a name suggested by Dana Scott to Ulrich Kohlenbach – and it has important mathematical applications as for example Kohlenbach has demonstrated clearly, see e.g. Kohlenbach (1992, 1993,a, n.d.). But our survey is also related to reductive proof theory – associated with Solomon Feferman – and there are certainly philosophical consequences too.

The modern mathematician interested in constructive mathematics, or interested in the constructive content of proofs, has many different elements of (classical) mathematics which then we can formulate in the language of PRA questions with no immediate answers. The Goldbach conjecture, for instance, can be represented in PRA as a formula, with one free variable. But this does not rule out the possibility that an answer to Goldbach’s conjecture could be intuitive in the sense of *anschaulich* (in the sense of Kant), although not immediate.

<sup>19</sup>In Kantian terms we could say that they are necessary conditions prior to any knowledge. This guarantees objectivity. On the relation between (axiomatic scientific) knowledge and finitism Bernays (1928, 145) says “Für die Hilbertsche Grundlegung ist aber kennzeichnend, daß hier der finite Standpunkt in Zusammenhang gebracht wird mit der axiomatischen Begründung der theoretischen Wissenschaften. Dadurch stellen sich die Voraussetzungen der finiten Einstellung zugleich als Bedingungen dar für die Möglichkeit theoretischer Naturerkenntnis, ganz im Sinne der Kantischen Problemstellung.” See also (Andersen, 2000, 34–39).

<sup>20</sup>There are many examples of this in the history of mathematics. As mentioned is infinitesimals one example. Another, from at the end of the 19th century, is Frege’s idea of the extension of a general property; a third is Church and Curry’s proposal around 1930 of  $\lambda$ -calculus as a foundation for mathematics.

<sup>21</sup>Similar to what Cantor and Hilbert maintained, namely that the essence of mathematics lies in its *freedom of abstraction*.

<sup>22</sup>The German word *Zulässigkeit* is probably better here.

he must judge between. These elements and principles are very fruitful in the development of mathematics. But which of them can be given a constructive foundation? And what are the mathematical and logical relations between the principles?

We will primarily investigate – locally in the framework of typed arithmetic – the following elements of ideal mathematics:<sup>23</sup>

1. Extensionality: Two functions are equal if they are equal on all arguments.
2. Markov’s principle: If  $A(x)$  is decidable for any  $x$  and if  $\neg\neg\exists xA(x)$  holds then  $\exists xA(x)$ , i.e.

$$\forall x(A(x) \vee \neg A(x)) \wedge \neg\neg\exists xA(x) \rightarrow \exists xA(x).$$

3. Axiom of choice: If for any  $x$  there exists a  $y$  such that  $A(x, y)$  holds, then there exists a total (choice) function  $f$  which takes any  $x$  and gives  $f(x)$  such that  $A(x, f(x))$  holds.
4. Independence-of-premise (for certain classes of formulas). If  $A \rightarrow \exists yB(y)$  holds and  $y$  is not a free variable of  $A$  then  $\exists y(A \rightarrow B(y))$  holds, i.e.

$$(A \rightarrow \exists yB(y)) \rightarrow \exists y(A \rightarrow B(y)).$$

One of our main purposes is to give a systematic analysis of 1–4 in order to provide coherent and consistent systems that partly contain these *prima facie* non-constructive principles. Systems, that nevertheless can be seen as constructively meaningful.

*Generally* speaking, the notions and principles in 1–4 are constructively problematic: Some of them are not justified by the BHK interpretation, and though AC is justified under this interpretation then certainly it is a non-constructive principle over classical logic.<sup>24</sup> As such they are often rejected by intuitionists of today. We will see, however, that if one takes the constructive slogan “existence = computability” seriously then a closer analysis of the apparently non-constructive notions has a lot to say about which methods can be applied where. But we will also see that different interpretations validate different principles. Nevertheless, individually the interpretations give coherent and meaningful views on constructivity and as such we will therefore *not* end up with one true and unique formulation of constructivity. Despite of this we will try to give a comparison of the different interpretations.

Our general agenda, however, will be that:

*If we insist on constructivity with respect to existence proofs, how much of classical logic can then be justified and given a constructive interpretation in the context of typed arithmetic?*

We will thereby investigate the thesis that in order to obtain constructive results it is not necessary to restrict the methodology to the intuitionistic one. This follows views expressed by Feferman and Kreisel amongst others.

<sup>23</sup>Other examples of important principles of mathematics which we will only discuss briefly in this thesis are Church’s thesis (every algorithm is recursive), König’s lemma (an infinite tree branching only finitely many times has an infinite branch) and bar induction.

<sup>24</sup>Recall Zermelo’s classical result from set theory that AC implies that any set can be well-ordered.

We will mainly be concerned with extensions of intuitionistic systems which enjoy fully constructive interpretations in the sense that existence proofs, for instance, allow for explicit realisations. If, however, one is interested in other kinds of constructive information, say effective bounds, many other principles of logic and mathematics than those mentioned above can be used; this approach has been developed by Kohlenbach (1998). We will only touch these perspectives on classical methodology in last chapter. Firstly we will, however, introduce and investigate the proof theoretical tools to be used throughout this thesis. These are mainly the Dialectica interpretation and modified realisability. The following chapters will be rather technical but conclusive results will show up, which form a kind of critique of the BHK interpretation—if this is taken to be a global interpretation.

## Introduction to Gödel's Dialectica Interpretation

In 1958 Kurt Gödel finally published his interpretation (Gödel, 1958). In the paper he interpreted Heyting arithmetic in a quantifier free type theory with primitive recursion in all finite types. This type theory is called Gödel's system T and the interpretation became known as Gödel's 'Dialectica' interpretation – named after the journal in which it was published. As mentioned in the forgoing chapter, Gödel (1933) had earlier interpreted Peano arithmetic in Heyting arithmetic via a negative translation, so by the Dialectica interpretation Gödel proved, among other things, that Peano arithmetic is consistent relative to T. Besides the consistency proof Gödel provided via the soundness theorem a method for extracting terms (programs) which realise theorems of arithmetic.

The interpretation given by Gödel extends easily to typed Heyting arithmetic. This extension of Gödel's result was first carried out in details by Troelstra (1973). For the rest of this thesis HA is the theory of Heyting arithmetic,  $HA^\omega$  is Heyting arithmetic generalized to all finite types,  $HA_H^\omega$  is  $HA^\omega$  formulated in a Hilbert style calculus and  $HA_{ND}^\omega$  is  $HA^\omega$  formulated in a natural deduction calculus.<sup>1</sup> As an introduction to the subject we will state the soundness theorem for  $HA_H^\omega$  (with weak extensionality) and sketch the proof. Thereafter the theorem will be proved for  $HA_{ND}^\omega$  (also with weak extensionality) in detail. First, however, the formal theories.

### 2.1 A Hilbert system for weakly extensional Heyting arithmetic in all finite types

The typed theory of Heyting arithmetic will be formulated with weak extensionality. This theory will therefore have the name WE- $HA^\omega$ . Since WE- $HA^\omega$  is typed it has a type structure,  $\mathcal{T}$ .

1.  $0 \in \mathcal{T}$ ,
2. If  $\sigma \in \mathcal{T}$  and  $\tau \in \mathcal{T}$  then  $\sigma \rightarrow \tau \in \mathcal{T}$ .

Intuitively each type represents a class of objects. Type 0 is thought of as representing the natural numbers and  $\sigma \rightarrow \tau$  is the type of functions from objects of type  $\sigma$  to objects of type  $\tau$ .<sup>2</sup>

That  $A$  is an object of type  $\sigma$  can be written in different ways:  $A^\sigma$  or  $A : \sigma$ . If it is clear from the context what type  $A$  has then it is usually written without indication of type, i.e.  $A$ . Parentheses for types are associated to the right, e.g.,  $\sigma \rightarrow \tau \rightarrow \tau$  is shorthand for  $\sigma \rightarrow (\tau \rightarrow \tau)$ , and quite often we will omit ' $\rightarrow$ ' and just write  $\sigma\tau\tau$ .

To each type  $\sigma$  we can assign a natural number  $\text{lev}(\sigma)$  as its *type level* by:

<sup>1</sup>The general terminology regarding the different names of the theories (with or without extensionality/intensionality) is also due to Troelstra (1973). Furthermore, a precise introduction to the Dialectica interpretation is found in (Troelstra, 1973, 230–249), whereas a modern and comprehensive survey of the subject is found in Avigad & Feferman (1998).

<sup>2</sup>Thus, the type structure can be thought of as a hierarchy given by specifying sets  $\mathcal{T}_\sigma$  for each type  $\sigma$  and an application mapping  $f : \mathcal{T}_{\sigma \rightarrow \tau} \times \mathcal{T}_\sigma \rightarrow \mathcal{T}_\tau$  for any  $\sigma$  and  $\tau$ .

1.  $\text{lev}(0) = 0$ ,
2.  $\text{lev}(\sigma \rightarrow \tau) = \max\{\text{lev}(\sigma) + 1, \text{lev}(\tau)\}$ .

The type structure  $\mathcal{T}$  is said to be of finite type since any type  $\sigma \in \mathcal{T}$  is assigned a finite level.

### 2.1.1 Language, terms, formulas and notation

The language  $\mathcal{L}(\text{WE-HA}^\omega)$  of  $\text{WE-HA}^\omega$  has for every type  $\sigma$  denumerably many variables  $x^\sigma, y^\sigma, z^\sigma, \dots$ . The constant symbols are  $0^0, S^{00}$ , and in all types there are symbols for the projector  $\Pi_{\sigma, \tau} : \sigma\tau\sigma$ , for the combinator  $\Sigma_{\rho, \tau, \sigma} : (\rho\tau\sigma)(\rho\tau)\rho\sigma$ , and for recursion,  $R_\sigma : \sigma(\sigma 0\sigma)0\sigma$ . The logical constants are  $\wedge, \vee, \rightarrow, \forall, \exists$ . Equality is a primitive concept only for type 0, so  $=_0$  belongs to the language, but equality in higher types will be defined in terms of equality in type 0; (in case of equality the type is indicated by *subscript*).

In order to be able to make a distinction between equality in the object language and identity in the meta-language we have  $\equiv$  as meta-symbol. This symbol is used to express syntactical identities or syntactical abbreviations. For definitions we use  $:\equiv$  which could be read as “is defined to be identical to”. However, when no misunderstanding can occur we may also use  $=$  as a meta-symbol. The normal mathematical symbols as  $\Rightarrow, \in$  and so forth are meta-symbols with their standard mathematical meaning.

The terms and formulas of  $\text{WE-HA}^\omega$  can now be defined.

**Definition 2.1.1.** The terms are generated according to:

- (i) Constant symbols  $c^\sigma$  and variables  $x^\sigma$  are terms.
- (ii) If  $t^{\sigma\tau}$  and  $s^\sigma$  are terms, then  $(ts)^\tau$  is a term.

◁

**Definition 2.1.2.** The formulas are generated from:

- (i) Prime formulas which have the form  $s^0 =_0 t^0$ .
- (ii) If  $A$  and  $B$  are formulas, then  $A \diamond B$  is a formula,  $\diamond \in \{\vee, \wedge, \rightarrow\}$ .
- (iii) If  $A$  is a formula and  $x^\sigma$  a variable, then  $Qx^\sigma A$  is a formula,  $Q \in \{\forall, \exists\}$ .

◁

### Notation

Define  $\perp :=_0 =_0 S0$ . Then, since our language contains arithmetical symbols, we can define negation:  $\neg A :=_0 A \rightarrow \perp$ . Also equivalence:  $A \leftrightarrow B :=_0 (A \rightarrow B) \wedge (B \rightarrow A)$ , ( $\wedge$  binds stronger than  $\rightarrow$ ). By vector notation  $\mathbf{x}$  we mean a finite string of variables  $x_1, \dots, x_n$ . Then  $Q\mathbf{x}$  abbreviates  $Qx_1 \dots Qx_n$  for  $Q \in \{\forall, \exists\}$ . The vector notation also applies generally to terms: if  $\mathbf{t} \equiv t_1, \dots, t_n$  and  $\mathbf{s} \equiv s_1, \dots, s_m$  then  $\mathbf{t}\mathbf{s}$  abbreviates  $t_1 s_1 \dots s_m, t_2 s_1 \dots s_m, \dots, t_n s_1 \dots s_m$ . In general  $t_1 t_2 \dots t_n$  is short for  $(\dots ((t_1 t_2) t_3) \dots t_n)$ . In order to refer to the possibly empty set of free variables  $x_1, \dots, x_n$  in a term  $t$  we write  $t[x_1, \dots, x_n]$ . Then  $t[t_1, \dots, t_n]$  denotes the result

of simultaneously substituting  $t_1, \dots, t_n$  for  $x_1, \dots, x_n$ , respectively. This notation also applies to formulas with parentheses instead of the squared brackets, i.e.  $A(b)$  denotes the result of substituting  $b$  for  $x$  in  $A$ .

Since we only have equality in type 0 as a primitive, higher type equations are abbreviations of lower type equations. If  $\sigma$  is not 0 and  $\sigma \equiv \sigma_1 \dots \sigma_n 0$  then  $s^\sigma =_\sigma t^\sigma$  is short for

$$\forall x_1^{\sigma_1} \dots x_n^{\sigma_n} (s^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n} =_0 t^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n}),$$

where the  $x_i$ 's are variables not occurring in  $s^\sigma, t^\sigma$ . We have in other words an extensional notion of equality.

## 2.2 Axioms and rules of WE-HA<sub>H</sub><sup>0</sup>

The logic is intuitionistic and either formulated within natural deduction which gives us WE-HA<sub>ND</sub><sup>0</sup> or formulated in a Hilbert style, WE-HA<sub>H</sub><sup>0</sup>. Dialectica interpretations of Hilbert systems are in many respects considerably less complicated than interpretations of systems formulated within natural deduction. Furthermore, the former are the well-known versions and as an introduction to the interpretation we therefore present WE-HA<sup>0</sup> in Hilbert style and sketch the proof of the interpretation theorem.

The following formulation of the underlying logic is due to Spector (1962). The axiom schemes and rules are:

$$\begin{array}{lll} A \rightarrow A, & & \text{(axioms of implication).} \\ A \rightarrow A \vee B, & B \rightarrow A \vee B, & \text{(axioms of weakening).} \\ A \wedge B \rightarrow A, & A \wedge B \rightarrow B, & \text{(axioms of weakening).} \\ & \perp \rightarrow A, & \text{(ex falso quodlibet).} \\ \frac{A \quad A \rightarrow B}{B} \text{MP} & \frac{A \rightarrow B \quad B \rightarrow C}{A \rightarrow C} \text{Syl} & \frac{A \rightarrow B \quad A \rightarrow C}{A \rightarrow B \wedge C} \text{Con} \\ \frac{A \wedge B \rightarrow C}{A \rightarrow (B \rightarrow C)} \text{Expo} & \frac{A \rightarrow (B \rightarrow C)}{A \wedge B \rightarrow C} \text{Impo} & \frac{A \rightarrow C \quad B \rightarrow C}{A \vee B \rightarrow C} \text{Dis} \end{array}$$

Quantifiers are introduced in the following way:

$$\begin{array}{ll} \frac{B \rightarrow A(b^\sigma)}{B \rightarrow \forall x^\sigma A(x^\sigma)} \text{Q1} & \forall x^\sigma A(x^\sigma) \rightarrow A(t^\sigma), \text{ (Q2)} \\ A(t^\sigma) \rightarrow \exists x^\sigma A(x^\sigma), \text{ (Q3)} & \frac{A(b^\sigma) \rightarrow B}{\exists x^\sigma A(x^\sigma) \rightarrow B} \text{Q4,} \end{array}$$

Here  $b^\sigma$  is eigenvariable, which means that regarding the rules Q1 and Q4  $b^\sigma$  is not allowed to occur free in  $B$ .

### Equality and Extensionality

Since equality in higher types is defined in terms of lower types we only have equality axioms for type 0. These include (the universal closure of):

$$x =_0 x, \quad x =_0 y \rightarrow y =_0 x, \quad x =_0 y \wedge y =_0 z \rightarrow x =_0 z, \quad \mathbf{x} =_0 \mathbf{y} \rightarrow f\mathbf{x} =_0 f\mathbf{y},$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are of the same length,  $\mathbf{x} =_0 \mathbf{y}$  is an abbreviation of  $x_1 =_0 y_1 \wedge \dots \wedge x_n =_0 y_n$  and  $f$  is any  $n$ -ary function symbol of type level 1.

What is to follow is a short discussion of equality and extensionality as a motivation for the quantifier free rule of weak extensionality, QF-ER.

By the abbreviation of equality in higher types we have an extensional notion of equality, e.g. we consider two one-argument number theoretic functions<sup>3</sup>  $f^1$  and  $g^1$  to be equal if they are equal with respect to all arguments:  $f =_1 g \equiv \forall x^0 (fx =_0 gx)$ . But this only says that *equality* is extensional. Another matter is whether or not the functionals behave extensionally. We say a functional is extensional if the results of any application of the functional to two equal arguments are equal, more precisely if  $\sigma \equiv \sigma_1 \dots \sigma_n 0$ :

$$z^\sigma \text{ is extensional} \quad := \quad \forall x_1^{\sigma_1}, y_1^{\sigma_1}, \dots, x_n^{\sigma_n}, y_n^{\sigma_n} \left( \bigwedge_{i=1}^n x_i =_{\sigma_i} y_i \rightarrow z\mathbf{x} =_0 z\mathbf{y} \right). \quad (2.1)$$

Say  $E_\sigma(z^\sigma)$  holds iff  $z^\sigma$  is extensional. Now, if our system under consideration had full extensionality this would mean that all functionals in all types would behave extensionally, i.e.  $\forall z^\sigma E_\sigma(z)$ . This requirement would give us the system E-HA<sup>ω</sup>. However, Howard (1973) has shown that E-HA<sup>ω</sup> does not have a Dialectica interpretation in itself or in any subsystem of itself. For, there is no functional of E-HA<sup>ω</sup> that satisfies the Dialectica translation of the extensionality axiom of type 2:  $\forall z^2 E_2(z)$ . We will come back to the non-interpretability of E-HA<sup>ω</sup> later. Since we want to define a version of typed intuitionistic arithmetic which has a Dialectica interpretation we will have to weaken (the schema of) full extensionality. Therefore WE-HA<sub>H</sub><sup>ω</sup> has the following *rule* of weak extensionality:

$$\frac{A_{\text{qf}} \rightarrow s^\sigma =_\sigma t^\sigma}{A_{\text{qf}} \rightarrow r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau} \text{ QF-ER}$$

where  $A_{\text{qf}}$  is a quantifier free formula, and  $s^\sigma, t^\sigma$  and  $r[x^\sigma]^\tau$  are terms.

In order to propose the connection between QF-ER and full extensionality let the following be mentioned briefly.

The *derived* rule

$$\frac{s^\sigma =_\sigma t^\sigma}{r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau} \text{ EXT-R}$$

is a trivial consequence of QF-ER (assume we have derived  $s^\sigma =_\sigma t^\sigma$ ; this formula can be weakened to  $0 = 0 \rightarrow s^\sigma =_\sigma t^\sigma$  then apply QF-ER to obtain  $0 = 0 \rightarrow r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau$  and then

<sup>3</sup>Note that type 00 is also called type 1.

apply MP on this and  $0 = 0$ ).<sup>4</sup>

Now, say  $z$  is of type  $\rho$ , where  $\rho$  is  $\rho_1 \dots \rho_n 0$  and assume we have derived the premise of the right hand side of (2.1), namely

$$\text{for all } i \leq n (y_i =_{\rho_i} \tilde{y}_i).$$

We are then able to conclude  $zy =_0 z\tilde{y}$  with the use of EXT-R by taking  $r[x^\sigma]^\tau$  to be  $z^\rho x_1^{\rho_1} \dots x_n^{\rho_n}$ . Thus, the schema of full extensionality is replaced by a (weaker) rule in order to obtain WE-HA<sub>H</sub><sup>0</sup>. We will see later that because of this the deduction theorem does *not* hold for the theory with weak extensionality.

### Arithmetic

The arithmetical axioms for the successor symbol  $S^{00}$  are

$$Sx =_0 0 \rightarrow \perp, \quad Sx =_0 Sy \rightarrow x =_0 y,$$

and the axiom schema of complete induction is:

$$(IA) \quad A(0^0) \wedge \forall x^0 (A(x) \rightarrow A(Sx)) \rightarrow \forall x A(x).$$

In order to do primitive recursive arithmetic inside WE-HA<sub>H</sub><sup>0</sup> we have the following axioms for projector  $\Pi_{\sigma,\tau}$  and combinator  $\Sigma_{\rho,\tau,\sigma}$ :

$$\begin{aligned} \Pi_{\sigma,\tau} x^\sigma y^\tau &=_{\sigma} x^\sigma, & \Pi_{\sigma,\tau} &: \sigma\tau\sigma, \\ \Sigma_{\rho,\tau,\sigma} x y z &=_{\sigma} xz(yz), & x : \rho\tau\sigma, y : \rho\tau, z : \rho, \Sigma_{\rho,\tau,\sigma} &: (\rho\tau\sigma)(\rho\tau)\rho\sigma, \end{aligned}$$

and for  $R_\sigma$ :

$$\left. \begin{aligned} R_\sigma x y 0 &=_{\sigma} x \\ R_\sigma x y S z &=_{\sigma} y(R_\sigma x y z) \end{aligned} \right\} x : \sigma, y : \sigma 0\sigma, z : 0, R_\sigma : \sigma(\sigma 0\sigma)0\sigma.$$

One can think of the equations as reduction rules: The terms on the left ‘reduces’ to the terms on the right, thus defining rules for calculations.

By using projector  $\Pi_{\sigma,\tau}$  and combinator  $\Sigma_{\rho,\tau,\sigma}$  we now introduce the  $\lambda$ -operator as a *defined* notion. This is done by induction on the complexity of the term  $t[x^\sigma]^\tau$ , where  $\Pi$  and  $\Sigma$  are of suitable types:

$$\begin{aligned} \lambda x.x &::= \Sigma\Pi\Pi \\ \lambda x.t &::= \Pi t, \text{ if } x \notin \text{FV}(t) \\ \lambda x.(ts) &::= \Sigma(\lambda x.t)(\lambda x.s), \text{ if } x \in \text{FV}(ts) \end{aligned}$$

From this it follows that

<sup>4</sup>Bezem (1988) has shown that EXT-R over WE-HA<sub>H</sub><sup>0</sup> is deductively equivalent to QF-ER. He has in other words shown that when we have the replacement axiom for type 0 (i.e.  $x =_0 y \rightarrow fx =_0 fy$ ) then if we have  $A_{\text{qf}} \rightarrow s^\sigma =_{\sigma} t^\sigma$  then we are able to conclude  $A_{\text{qf}} \rightarrow r[s^\sigma]^\tau =_{\tau} r[t^\sigma]^\tau$  by use of the rule EXT-R.

- (i)  $\lambda x^\sigma . t^\tau$  is of type  $\sigma\tau$ ,
- (ii)  $FV(\lambda x^\sigma . t[x^\sigma]^\tau) = FV(t[x^\sigma]^\tau) \setminus \{x^\sigma\}$ ,
- (iii) The following equality is derivable in WE-HA<sub>II</sub><sup>0</sup>:

$$(\lambda x^\sigma . t[x^\sigma])s^\sigma =_\tau t[s^\sigma]^\tau.$$

When equations are seen as reduction rules the last equation is called a  $\beta$ -contraction.

That the defined  $\lambda$ -operator in fact has these three properties is proved by induction on the complexity of the term  $t$ . Here we just sketch that the equality given by  $\beta$ -contraction is provable by using  $\Pi$  and  $\Sigma$ . The proof is by induction of the complexity of  $t$ .

- 1a.  $t \equiv y \neq x$ .  $(\lambda x . y)s = (\Pi y)s = y$ .
- 1b.  $t \equiv x$ .  $(\lambda x . x)s = (\Sigma \Pi \Pi)s = \Pi s(\Pi s) = s$ .
- 2.  $t \equiv s_1 s_2$ . Assume  $x \in FV(s_1 s_2)$ . From the definition of  $\lambda$  it follows that

$$(\lambda x . (s_1 s_2))s = \Sigma(\lambda x . s_1)(\lambda x . s_2)s = (\lambda x . s_1[x])s((\lambda x . s_2[x])s) \stackrel{\text{IH}}{=} s_1[s]s_2[s],$$

what equals  $(s_1 s_2)[s]$ .

### Arithmetical examples

Let us see how some very basic arithmetic works with the  $\lambda$ -operator and the recursion operator. First of all we want to show that absolute difference  $|x - y|$  is definable. For this we need to define a plus operator  $+$  and a cutoff operator  $\div$ . The sum of two numbers  $x^0$  and  $y^0$  is defined as follows:

$$x + y := R_0 x(\lambda w . u . S w)y.$$

On the basis of this definition it is immediate that  $x + 0 = x$  and  $x + S y = S(x + y)$ . The next thing we need is the predecessor of a number:

$$\text{prd}^1 := \lambda x . R_0 0(\lambda w . u . u)x.$$

From this definition it is also immediate that  $\text{prd}(0) = 0$  and that  $\text{prd}(S z) = z$ . Finally we define cutoff:

$$x \div y := R_0 x(\lambda w . u . \text{prd}(w))y$$

Intuitively the operator works (or computes) like this: It takes  $y$  cuts 1 of  $y$  and writes  $\text{prd}$  on a list. When  $y$  is turned into 0 it takes  $x$  and applies the list of  $\text{prd}$  to  $x$ . Thus, it applies  $\text{prd}$  to  $x$  exactly  $y$  times and we see that if  $y$  is greater than  $x$  then  $x \div y = 0$  else  $\div$  tells us how much  $x$  is bigger than  $y$ .

With  $\div$  it is easy to define formally  $x$  is greater than  $y$ , namely:  $x > y := x \div y \neq 0$ , and  $x \geq y := (x > y) \vee (x = y)$ . Furthermore,  $x < y := y > x$ .

Also using  $\dot{-}$  we can define absolute difference:

$$|x - y| := (x \dot{-} y) + (y \dot{-} x).$$

Now we will see that ‘definition by cases’ is also definable on the basis of the recursion operator. Define:

$$\text{Cond} := \lambda x^\sigma, y^\sigma, z^0. \text{R}_{\sigma x}(\lambda v^\sigma, w^0. y)z$$

Then we read  $\text{Cond}t_1t_2z^0$  as “if  $z =_0 0$  then  $t_1$  else  $t_2$ .” For the sake of readability we will usually write  $\text{Cond}(x, y, z)$  instead of  $\text{Cond}xyz$ . With  $\text{Cond}$  and  $\dot{-}$  we are able to define operators that give us the maximum and minimum of two numbers:

$$\begin{aligned} \max & := \lambda x^0, y^0. \text{Cond}(y, x, x \dot{-} y) \\ \min & := \lambda x^0, y^0. \text{Cond}(x, y, x \dot{-} y) \end{aligned}$$

As in the case of  $\text{Cond}$  we will write  $\max(m, n)$  and  $\min(m, n)$  instead of  $\max mn$  and  $\min mn$ , respectively.

### 2.3 WE-T<sub>H</sub> as the quantifier free subsystem of WE-HA<sub>H</sub><sup>0</sup>

The weakly extensional version of Gödel’s system T, called WE-T, is essentially the quantifier free subsystem of WE-HA<sup>0</sup>. Again this system can be formulated either in natural deduction or in Hilbert style. The interpretation of WE-HA<sub>H</sub><sup>0</sup> in WE-T is a little easier when we formulate WE-T in Hilbert style.

WE-T<sub>H</sub> arises from WE-HA<sub>H</sub><sup>0</sup> when we drop the quantifiers of WE-HA<sub>H</sub><sup>0</sup>. So, the type structure underlying WE-T<sub>H</sub> is the same as the structure underlying WE-HA<sub>H</sub><sup>0</sup>; the language  $\mathcal{L}(\text{WE-T}_H)$  of WE-T<sub>H</sub> is same as  $\mathcal{L}(\text{WE-HA}_H^0)$ , just without quantifiers; terms and formulas are constructed in the same way as for WE-HA<sub>H</sub><sup>0</sup> but without quantifiers. WE-T<sub>H</sub> has the same axioms and rules as WE-HA<sub>H</sub><sup>0</sup> except for: (i) the rules and axioms regarding the quantifiers, i.e. Q1, Q2, Q3 and Q4 as given on page 17; (ii) The axiom of complete induction (IA) is replaced by the quantifier free rule of induction:

$$\frac{A(0) \quad A(x^0) \rightarrow A(Sx^0)}{A(x^0)} \text{QF-IR}$$

Furthermore, (iii) the absence of Q1 and Q2 force us to introduce a substitution rule:

$$\frac{A(x^\sigma)}{A(t^\sigma)} \text{Sub}$$

A remark should be made pertaining to the rule of weak extensionality (QF-ER) and higher type equations in WE-T<sub>H</sub>. QF-ER remains essentially the same but since quantifiers are not at our disposal we have to view higher type equations as abbreviations of lower type equations containing fresh variables that *in the presence* of quantifiers could be universally quantified: If  $\sigma$  is not 0 and  $\sigma = \sigma_1 \dots \sigma_n 0$  then  $s^\sigma =_\sigma t^\sigma$  is an abbreviation of

$s^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n} =_0 t^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n}$ , where the  $x_i$ 's are variables not occurring in  $s^\sigma$ ,  $t^\sigma$  and which are not free anywhere in formulas implying  $s^\sigma =_\sigma t^\sigma$ . Moreover, we have to require that  $s^\sigma =_\sigma t^\sigma$  occurs positively, in order to guarantee that *if* we were allowed to quantify, the quantifiers would be universal quantifiers. In particular all this means that the new variables are not allowed to occur in the hypothesis  $A_{\text{qf}}$  of QF-ER.

### Characteristic terms

The prime formulas of WE- $T_H$  are decidable. This is proved by induction on the complexity of the terms, see (Troelstra & van Dalen, 1988, Vol. 1). From this it follows that all formulas of WE- $T_H$  are decidable, and we have classical truth reasoning in WE- $T_H$ . This allows for the elimination of logical connectives in favour of a set of defined constants called propositional functions. If we define  $\text{sg} := \lambda x^0. R_0 0(\lambda \mu, v. S0)x$  and  $\overline{\text{sg}} := \lambda x^0. R_0(S0)(\lambda \mu, v. 0)x$  then we have

$$\begin{aligned} \text{sg}(0) &= 0 & \text{sg}(Sx) &= S0, \\ \overline{\text{sg}}(0) &= S0 & \overline{\text{sg}}(Sx) &= 0. \end{aligned}$$

Now we define the propositional functions as:

$$\begin{aligned} \text{con} &:= \max, \\ \text{dis} &:= \min, \\ \text{imp} &:= \lambda x, y. \min(\overline{\text{sg}}(x), y). \end{aligned}$$

Then we have primitive recursive functions  $\text{con}$ ,  $\text{dis}$  and  $\text{imp}$  of type 000 such that

$$\begin{aligned} \text{con}(Sx, y) &= \text{con}(y, Sx) \neq 0, & \text{con}(0, 0) &= 0, \\ \text{dis}(0, x) &= \text{dis}(x, 0) = 0, & \text{dis}(Sx, Sy) &\neq 0, \\ \text{imp}(Sx, y) &= \text{imp}(z, 0) = 0, & \text{imp}(0, Sx) &\neq 0. \end{aligned}$$

One then constructs inductively for any formula  $A$  in the language of WE- $T_H$  with  $\text{FV}(A) = \{\mathbf{x}\}$  a closed term  $t_A$  (type of  $t_A$  determined by  $\mathbf{x}$ ) such that

$$\text{WE-}T_H \vdash t_A \mathbf{x} =_0 0 \leftrightarrow A(\mathbf{x}). \quad (2.2)$$

For  $t_{s[\mathbf{x}] = r[\mathbf{y}]}$  we take absolute difference  $\lambda \mathbf{x}, \mathbf{y}. |s[\mathbf{x}] - r[\mathbf{y}]|$ . Assume  $\text{FV}(A) = \{\mathbf{x}\}$  and  $\text{FV}(B) = \{\mathbf{y}\}$  then  $t_{A \wedge B} := \lambda \mathbf{x}, \mathbf{y}. \text{con}(t'_A, t'_B)$ ;  $t_{A \vee B} := \lambda \mathbf{x}, \mathbf{y}. \text{dis}(t'_A, t'_B)$  and  $t_{A \rightarrow B} := \lambda \mathbf{x}, \mathbf{y}. \text{imp}(t'_A, t'_B)$ , where  $t'_A$  and  $t'_B$  are identical to  $t_A$  and  $t_B$ , respectively, just without the lambda abstraction.

That the equivalence stated in (2.2) is derivable in WE- $T_H$  is proved by induction on the complexity of  $A$ . See (Troelstra & van Dalen, 1988, Vol. 1, 120–125) for some of the basic primitive recursive arithmetic which is needed.

### 2.4 Definition and analysis of Dialectica translation

To each formula  $A(\mathbf{a})$  of  $\mathcal{L}(\text{WE-HA}^\omega)$ , where  $\text{FV}(A) = \{\mathbf{a}\}$ , we now inductively associate its Dialectica translation  $(A(\mathbf{a}))^D$ :

$$(A(\mathbf{a}))^D \equiv \exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}, \mathbf{a}) \quad (2.3)$$

where  $A_D$  is a formula of  $\mathcal{L}(\text{WE-T})$  and thereby quantifier free.  $\mathbf{x}$  and  $\mathbf{y}$  are sequences of fresh variables. We note that  $\text{FV}(A) = \text{FV}(A^D)$ , and that  $\mathbf{x}$  and  $\mathbf{y}$  can be empty sequences. Very intuitively: think of (2.3) in the following way: If  $A$  is provable then according to the translation  $A^D$  there are  $\mathbf{x}$  making  $A_D$  ‘true’ for all possible instances of  $\mathbf{y}$ .

Now,  $(\cdot)^D$  and  $(\cdot)_D$  are defined simultaneously, where  $(A(\mathbf{a}))^D \equiv \exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}, \mathbf{a})$  and  $(B(\mathbf{b}))^D \equiv \exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v}, \mathbf{b})$ :

**Definition 2.4.1.** (Dialectica translation).

$$\begin{aligned}
(P^D) \quad (A(\mathbf{a}))^D &::= A_D(\mathbf{a}) := A(\mathbf{a}), \text{ if } A(\mathbf{a}) \text{ is prime,} \\
(\wedge^D) \quad (A(\mathbf{a}) \wedge B(\mathbf{b}))^D &::= \exists \mathbf{x}, \mathbf{u} \forall \mathbf{y}, \mathbf{v} (A_D(\mathbf{x}, \mathbf{y}, \mathbf{a}) \wedge B_D(\mathbf{u}, \mathbf{v}, \mathbf{b})), \\
(\vee^D) \quad (A(\mathbf{a}) \vee B(\mathbf{b}))^D &::= \exists z^0, \mathbf{x}, \mathbf{u} \forall \mathbf{y}, \mathbf{v} ((z = 0 \rightarrow A_D(\mathbf{x}, \mathbf{y}, \mathbf{a})) \wedge \\
&\quad (z \neq 0 \rightarrow B_D(\mathbf{u}, \mathbf{v}, \mathbf{b}))), \\
(\exists^D) \quad (\exists z^\sigma A(z, \mathbf{a}))^D &::= \exists z^\sigma, \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}, z, \mathbf{a}), \\
(\forall^D) \quad (\forall z^\sigma A(z, \mathbf{a}))^D &::= \exists \mathbf{X} \forall \mathbf{y}, z^\sigma A_D(\mathbf{X}z, \mathbf{y}, z, \mathbf{a}), \\
(\rightarrow^D) \quad (A(\mathbf{a}) \rightarrow B(\mathbf{b}))^D &::= \exists \mathbf{U}, \mathbf{Y} \forall \mathbf{x}, \mathbf{v} (A_D(\mathbf{x}, \mathbf{Y}\mathbf{x}\mathbf{v}, \mathbf{a}) \rightarrow B_D(\mathbf{U}\mathbf{x}, \mathbf{v}, \mathbf{b})).
\end{aligned}$$

◁

### 2.4.1 Analysis of translation

Let us discuss this definition in order to make some sense of it. One of the questions in this connection is what makes it possible to prove the equivalence  $A \leftrightarrow A^D$ ? The definitions  $(P^D)$ ,  $(\wedge^D)$ ,  $(\vee^D)$  and  $(\exists^D)$  are straight-forward. For instance, if we have  $\exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \wedge \exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})$  and the variables are fresh then, of course, we are able to get the quantifiers to the front of the whole formula by intuitionistic logic. What is special about  $(\vee^D)$  is, that it ‘eliminates’  $\vee$  from the formula. On the other hand it could be said that the translation just spells out the intuitionistic meaning of disjunction. Actually one can inside arithmetic dispense with disjunction, since  $A \vee B$  can be defined as  $\exists z((z = 0 \rightarrow A) \wedge (z \neq 0 \rightarrow B))$ , and then prove the axioms for  $\vee$  to be derivable from the other axioms on the basis of this definition. We find it more natural, however, to include  $\vee$  as primitive. But we see from this remark that  $A^D \vee B^D \leftrightarrow (A \vee B)^D$  is intuitionistically provable within the framework of arithmetic.<sup>5</sup>

The intuitive motivation for  $(\forall^D)$  is that if a functional should make  $(\forall z A(z))^D$  true then this functional will, possibly among other arguments, take  $z$  as an argument. Now, assume we have  $\forall z \exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}, z)$ . The constructive reading, i.e. the BHK interpretation, of this is that there are functionals  $\mathbf{X}$  which take any given  $z$  as an argument and produce a sequence  $\mathbf{x}$  making  $A_D$  true for every  $\mathbf{y}$ . The ‘formal motivation’ makes use of the axiom of choice. In the context of typed arithmetic the axiom of choice is given by the schema

$$\text{AC} ::= \bigcup_{\sigma, \tau \in \mathcal{T}} \{\text{AC}^{\sigma, \tau}\}$$

<sup>5</sup>What also motivates this definition is that one can prove about intuitionistic logic, that for sentences  $A$  and  $B$ , if we can prove  $A \vee B$  then we can prove either  $A$  or  $B$  and we can tell which one. This can be proved for intuitionistic predicate logic by normalization on proofs, e.g. cut elimination for sequent systems. For intuitionistic arithmetic this can, as we will see later on, be proved by realisability or by  $Q$ -interpretation, which we will introduce as a variant of the Dialectica interpretation.

where<sup>6</sup>

$$\text{AC}^{\sigma, \tau} : \quad \forall x^\sigma \exists y^\tau A(x, y) \rightarrow \exists Y^{\sigma\tau} \forall x^\sigma A(x, Yx).$$

In connection with the D-translation this gives

$$\forall z \exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}, z) \rightarrow \exists \mathbf{X} \forall z \forall \mathbf{y} A_D(\mathbf{X}z, \mathbf{y}, z).$$

The motivation and analysis of  $(\rightarrow^D)$  is more delicate. To make an implication  $A \rightarrow B$  true requires basically two things: (i) If  $A$  is true we have to make  $B$  true and (ii) if  $B$  is false we have to make  $A$  false. These requirements should the functionals  $\mathbf{U}$  and  $\mathbf{Y}$  satisfy for all possible  $\mathbf{x}$  and  $\mathbf{v}$ . The following motivation is given by Gödel (1941, 196-7) and can be seen in the light of (i) and (ii). Suppose we have an expression of the form  $\exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})$ . The meaning of this is that if there exist objects  $\mathbf{x}$  satisfying a certain condition then there exist objects  $\mathbf{u}$  satisfying a certain other condition. Reading this in a constructive setting means that there exist procedures or functionals  $\mathbf{U}$  taking any  $\mathbf{x}$  witnessing the existential quantifiers in the antecedent into  $\mathbf{U}\mathbf{x}$  witnessing the existential quantifiers in the conclusion, i.e.

$$\exists \mathbf{U} \forall \mathbf{x} (\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \forall \mathbf{v} B_D(\mathbf{U}\mathbf{x}, \mathbf{v})). \quad (2.4)$$

But we have not yet arrived at the Dialectica normal form; how to interpret  $\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \forall \mathbf{v} B_D(\mathbf{U}\mathbf{x}, \mathbf{v})$ . Well, what would in general  $\forall x C(x) \rightarrow \forall y D(y)$  mean. A quite natural (computational) interpretation hereof would be: For any counterexample of  $D$  we can construct a counterexample of  $C$ , i.e.  $\exists X \forall y (\neg D(y) \rightarrow \neg C(Xy))$ . Under this interpretation (2.4) transforms into

$$\exists \mathbf{U} \forall \mathbf{x} \exists \mathbf{Y} \forall \mathbf{v} (\neg B_D(\mathbf{U}\mathbf{x}, \mathbf{v}) \rightarrow \neg A_D(\mathbf{x}, \mathbf{Y}\mathbf{v})).$$

Now we have an implication between  $A_D$  and  $B_D$  and we take the contraposition of this and obtain  $\neg \neg A_D(\mathbf{x}, \mathbf{Y}\mathbf{v}) \rightarrow \neg \neg B_D(\mathbf{U}\mathbf{x}, \mathbf{v})$ . Since we have stability for quantifier free formulas:  $\neg \neg C_{\text{qf}} \leftrightarrow C_{\text{qf}}$ , we can discharge the double negations and we arrive at:

$$\exists \mathbf{U} \forall \mathbf{x} \exists \mathbf{Y} \forall \mathbf{v} (A_D(\mathbf{x}, \mathbf{Y}\mathbf{v}) \rightarrow B_D(\mathbf{U}\mathbf{x}, \mathbf{v})).$$

The constructive reading of  $\forall \mathbf{x} \exists \mathbf{Y}$  gives us the Dialectica normal form.

On the other hand a complete *formal* examination of the translation (as spelled out by Spector (1962)) is given by the equivalences

$$\begin{aligned} (\exists \mathbf{x} \forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})) & \stackrel{(i)}{\leftrightarrow} \\ \forall \mathbf{x} (\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})) & \stackrel{(ii)}{\leftrightarrow} \\ \forall \mathbf{x} \exists \mathbf{u} (\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})) & \stackrel{(iii)}{\leftrightarrow} \\ \forall \mathbf{x} \exists \mathbf{u} \forall \mathbf{v} (\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow B_D(\mathbf{u}, \mathbf{v})) & \stackrel{(iv)}{\leftrightarrow} \\ \forall \mathbf{x} \exists \mathbf{u} \forall \mathbf{v} \exists \mathbf{y} (A_D(\mathbf{x}, \mathbf{y}) \rightarrow B_D(\mathbf{u}, \mathbf{v})) & \stackrel{(v)}{\leftrightarrow} \\ \exists \mathbf{U}, \mathbf{Y} \forall \mathbf{x}, \mathbf{v} (A_D(\mathbf{x}, \mathbf{Y}\mathbf{x}\mathbf{v}) \rightarrow B_D(\mathbf{U}\mathbf{x}, \mathbf{v})) & \end{aligned}$$

<sup>6</sup>Note that the BHK interpretation of AC is trivial: The interpretation of  $\forall x^\sigma \exists y^\tau A(x, y)$  is that there is a procedure taking any  $x$  and produces  $y$  such that  $A(x, y)$ . But this is essentially the same as the interpretation of the conclusion of AC.

These equivalences are all justified from a classical point of view, whereas only (i) and (iii) are intuitionistically valid. The last equivalence (v) is true in virtue of (AC) and is, thus, constructively justified. In this way we see that the decisive factor in proving the equivalence between the top and the bottom are (ii) and (iv).

### Independence-of-premise

Let ‘independence-of-premise’ for type  $\sigma$  be the following principle:

$$\text{IP}^\sigma : (A \rightarrow \exists x^\sigma B(x)) \rightarrow \exists x^\sigma (A \rightarrow B(x)),$$

where  $x^\sigma$  is not free in  $A$ . Let  $\text{IP}^\omega$  be the union of all instances of  $\text{IP}^\sigma$  for all types  $\sigma$ . Now we restrict  $\text{IP}^\omega$  to the case where the premise  $A$  is purely universal: call the restricted principle  $\text{IP}_{\forall}^\omega$ . It then follows that (ii) is justified on the basis of  $\text{IP}_{\forall}^\omega$ . In general, the constructiveness of  $\text{IP}^\omega$  is questionable. The BHK interpretation of  $A \rightarrow \exists x B(x)$  would be that given any proof of  $A$  we can from this proof construct a witness  $c$  for the existential quantifier and a proof of  $B(c)$ . Thus, our construction of the pair consisting of  $c$  and the proof of  $B(c)$  will in general depend on the proof of  $A$ . But the principle ‘independence-of-premise’ says that we can find such a  $c$  independent of the proof of  $A$ . Thus,  $\text{IP}^\omega$  is not sound under the BHK interpretation. We will, nevertheless, later on see that the Dialectica interpretation verifies  $\text{IP}_{\forall}^\omega$  and that this therefore is a constructive principle. At the same time we will see that – in connection with the Dialectica interpretation – this is the best possible: Let  $\text{IP}_{\neg\forall}^\omega$  denote the principle where the premise is negated purely universal; at the end of the next chapter (page 53) we will see that this principle is not D-interpretable.

### Markov’s principle

Equivalence (iv) is validated by the so-called Markov’s principle. The general logical form of this principle is

$$\forall x(A(x) \vee \neg A(x)) \wedge \neg\neg\exists x A(x) \rightarrow \exists x A(x).$$

In our setup the decidable formulas are the quantifier free formulas. Furthermore the theory is typed. Therefore Markov’s principle,  $\text{MP}^\omega$ , in the context of typed Heyting arithmetic is the union of all instances of the following schema for all  $\sigma$ :

$$\text{MP}^\sigma : \neg\neg\exists x^\sigma A_{\text{qf}}(x) \rightarrow \exists x^\sigma A_{\text{qf}}(x),$$

where  $A_{\text{qf}}$  is quantifier free. In our justification of (iv)

$$\forall \mathbf{x} \exists \mathbf{u} \forall \mathbf{v} (\forall \mathbf{y} A_D(\mathbf{x}, \mathbf{y}) \rightarrow B_D(\mathbf{u}, \mathbf{v}))$$

is given. By classical logic one could derive  $\forall \mathbf{x} \exists \mathbf{u} \forall \mathbf{v} \exists \mathbf{y} (A_D(\mathbf{x}, \mathbf{y}) \rightarrow B_D(\mathbf{u}, \mathbf{v}))$  and therefore intuitionistic logic gives

$$\forall \mathbf{x} \exists \mathbf{u} \forall \mathbf{v} \neg\neg \exists \mathbf{y} (A_D(\mathbf{x}, \mathbf{y}) \rightarrow B_D(\mathbf{u}, \mathbf{v})),$$

and then with  $MP^0$  we get the desired conclusion. In general, Markov's principle is not accepted by constructivists. But assuming decidability of quantifier free formulas and that the variable  $x$  is ranging over natural numbers the reading is: given  $\neg\neg\exists x A_{qf}$  holds (i.e. it is contradictory that for all  $n$   $A(n)$  does not hold)<sup>7</sup>, then in fact there exists an  $n$  such that  $A(n)$ . The argument *for* Markov's principle of type 0 is simply: Assume we have a decidable procedure for testing whether  $A(n)$  holds or not and assume furthermore that it cannot be the case that for every natural number  $n$   $\neg A(n)$  then we can find a natural number  $m$  by testing whether or not  $A$  holds for 0, for 1 and so forth; sooner or later we shall find a number  $m$  such that  $A(m)$  holds. In terms of Turing machines the argument is: If it is impossible that the Turing machine will compute forever, then the clear algorithm for obtaining an output is by continuing the computation until the machine halts.

In this form Markov's principle is accepted by some constructivists, e.g. the so-called Russian school of constructivity.

### Arbitrarily large types needed

The analysis of  $(\rightarrow^D)$  displays another interesting feature. The translation of an implication raises the type of the functionals needed, since either  $\mathbf{U}$  or  $\mathbf{Y}$  (or both sequences) are of higher type than any functional occurring in  $\exists x \forall y A_D(\mathbf{x}, \mathbf{y})$  and  $\exists \mathbf{u} \forall \mathbf{v} B_D(\mathbf{u}, \mathbf{v})$ . Now, the use of  $\rightarrow$  can be iterated as often as we want to the effect that we need functionals of arbitrarily high type in order to express the functionals needed for the interpretation of any implication.

In the discussion of the translation of implication there is another important remark to make. There are four different ways of getting the quantifiers to the front i.e. to get the formula on  $\exists \forall$  normal form (see (Troelstra, 1973, 232–233)). The Dialectica makes use of one of them – the one we have just seen. As we will see on the next couple of pages the Dialectica translation gives rise to an interpretation in WE-T. But the other three translations of implication ask for non-recursive realisations already for  $A \rightarrow A$  when  $A$  is  $\exists x \forall y \neg T z x y$ , and  $T$  is Kleene's  $T$ -predicate.<sup>8</sup> Therefore the specific translation given by  $\rightarrow^D$  is the only possible of these four. For further details in this respect see (Troelstra, 1973, 232–233) and (Kohlenbach, 1998a, 39–41).

### Foundational discussion of translation

The conclusion we consequently arrive at is:

$$WE-HA^\omega + MP^\omega + IP_{\forall}^\omega + AC \vdash A \leftrightarrow A^D. \quad (2.5)$$

We see that the faithfulness of the translation cannot be validated alone by intuitionistic principles. *But this is not the point either.* The point is *not* that the translation should preserve the intuitionistic meaning of the formulas, but that we can come up with a translation such that the soundness theorem below is provable. The soundness theorem gives a reduction of

<sup>7</sup>Note that  $\neg\neg\exists x A(x) \leftrightarrow \neg\forall x \neg A(x)$  is intuitionistically provable.

<sup>8</sup>Kleene's  $T$  is a primitive recursive predicate and  $Txyz$  expresses that Turing machine with Gödel number  $x$  applied to input  $y$  terminates with a computation whose Gödel number is  $z$ , see for instance (Troelstra & van Dalen, 1988, Vol. 1).

typed arithmetic using quantifiers to the quantifier free WE-T. The value of this reduction is independent of the analysis of or motivation for the translation. The three ‘extra’ principles of (2.5) are validated by the interpretation and therefore we can in fact use MP<sup>ω</sup>, IP<sub>∀</sub><sup>ω</sup> and AC on top of intuitionistic logic and *still* extract constructive content. This will be discussed later, but we can already say it shows in the context of WE-HA<sup>ω</sup> that these partly classical principles do carry computational content. This support the general view on the methodology of mathematics that Solomon Feferman advocates: “I do not see the necessity, insisted upon by Brouwer and his followers, to restrict to constructive reasoning in order to obtain constructive results . . . ” (Feferman, 1998, ix). This is most definitely also in line with Hilbert’s view on the methodology of mathematics. Hilbert wanted with his program to show that the use of ideal elements in proofs of finitary statements could in principle be eliminated. Thus, the use of ideal reasoning could safely be used in order to obtain elegant proofs of finitary statements.

### 2.5 Interpretation theorem for WE-HA<sub>H</sub><sup>ω</sup>

We will now present the interpretation theorem for WE-HA<sub>H</sub><sup>ω</sup> and give some examples from the proof. The theorem in this form is stated and proved in all details by Troelstra (1973, 234–237).

**Theorem 2.5.1.** (Soundness of *D*-translation).

$$\text{If } \text{WE-HA}_H^\omega \vdash A(\mathbf{a}) \text{ then } \text{WE-T}_H \vdash A_D(\mathbf{T}\mathbf{a}, \mathbf{y}, \mathbf{a}),$$

for a certain sequence of closed terms  $\mathbf{T}$  which can be extracted from a derivation of  $A(\mathbf{a})$  in WE-HA<sub>H</sub><sup>ω</sup>.

**Proof.** The proof is by induction on the length of the WE-HA<sub>H</sub><sup>ω</sup>-derivation. Here we present three cases from the proof that display the idea.

**Case 1.** Axiom  $A(\mathbf{a}) \rightarrow A(\mathbf{a})$ . This translates to

$$\exists \mathbf{X}, \mathbf{Y} \forall \mathbf{x}, \mathbf{y} (A_D(\mathbf{x}, \mathbf{Y}\mathbf{x}\mathbf{y}, \mathbf{a}) \rightarrow A_D(\mathbf{X}\mathbf{x}, \mathbf{y}, \mathbf{a})).$$

From this we see that with  $\mathbf{T}_1 := \lambda \mathbf{a}, \mathbf{x}. \mathbf{x}$  and  $\mathbf{T}_2 := \lambda \mathbf{a}, \mathbf{x}, \mathbf{y}. \mathbf{y}$  we have

$$\text{WE-T}_H \vdash A_D(\mathbf{x}, \mathbf{T}_2 \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{a}) \rightarrow A_D(\mathbf{T}_1 \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}).$$

**Case 2.** MP. Assume as induction hypothesis

- (i) WE-T<sub>H</sub> ⊢ A<sub>D</sub>(T<sub>1</sub>a, y, a),
- (ii) WE-T<sub>H</sub> ⊢ A<sub>D</sub>(x, T<sub>2</sub>cxv, a) → B<sub>D</sub>(T<sub>3</sub>cx, v, b),

for given T<sub>1</sub>, T<sub>2</sub>, and T<sub>3</sub>;  $\mathbf{c}$  is written for  $a_1, \dots, a_n, b_1, \dots, b_m$ . Find T<sub>4</sub> such that WE-T<sub>H</sub> ⊢ B<sub>D</sub>(T<sub>4</sub>b, v, b). Set  $\mathbf{x}$  in (ii) to T<sub>1</sub>a and let  $\mathbf{y}$  in (i) be T<sub>2</sub>c(T<sub>1</sub>a)v. Then use MP (in WE-T<sub>H</sub>) to obtain

$$\text{WE-T} \vdash B_D(\mathbf{T}_3 \mathbf{c}(\mathbf{T}_1 \mathbf{a}), \mathbf{v}, \mathbf{b}).$$

Let  $\sigma = \sigma_1 \dots \sigma_n 0$  and define  $\circ^\sigma := \lambda x^{\sigma_1} \dots x^{\sigma_n}. 0^0$ . Now substitute all free variables in  $\mathbf{T}_3 \mathbf{c}(\mathbf{T}_1 \mathbf{a})$  that do not occur in  $\mathbf{b}$  by  $\circ$  of the corresponding type; name the result  $\mathbf{T}$ . If we put  $\mathbf{T}_4 := \lambda \mathbf{b}. \mathbf{T}$  we have the required result.

**Case 3.** We will see that ‘definition by cases’ corresponds to a kind of contraction. We will see this by verifying rule Con.

$$\frac{A \rightarrow B \quad A \rightarrow C}{A \rightarrow B \wedge C} \text{Con}$$

For notational simplicity we omit free variables. Then the induction hypothesis is:

$$\begin{aligned} \text{WE-T}_H \vdash A_D(\mathbf{x}, \mathbf{T}_1 \mathbf{xv}) &\rightarrow B_D(\mathbf{T}_2 \mathbf{x}, \mathbf{v}) \text{ and} \\ \text{WE-T}_H \vdash A_D(\mathbf{x}, \mathbf{T}_3 \mathbf{xq}) &\rightarrow C_D(\mathbf{T}_4 \mathbf{x}, \mathbf{q}), \end{aligned}$$

for given sequences of terms  $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$  and  $\mathbf{T}_4$ . When we translate  $A \rightarrow B \wedge C$  we get:

$$\exists \mathbf{U}, \mathbf{P}, \mathbf{Y} \forall \mathbf{x}, \mathbf{v}, \mathbf{q} (A_D(\mathbf{x}, \mathbf{Y} \mathbf{xv} \mathbf{q}) \rightarrow B_D(\mathbf{U} \mathbf{x}, \mathbf{v}) \wedge C_D(\mathbf{P} \mathbf{x}, \mathbf{q})).$$

From this we see that we have to provide sequences of terms  $\mathbf{S}_1, \mathbf{S}_2$  and  $\mathbf{S}_3$  such that:

$$\text{WE-T}_H \vdash A_D(\mathbf{x}, \mathbf{S}_1 \mathbf{xv} \mathbf{q}) \rightarrow B_D(\mathbf{S}_2 \mathbf{x}, \mathbf{v}) \wedge C_D(\mathbf{S}_3 \mathbf{x}, \mathbf{q}).$$

This is done by taking  $\mathbf{S}_2 := \lambda \mathbf{x}. \mathbf{T}_2 \mathbf{x}$  and  $\mathbf{S}_3 := \lambda \mathbf{x}. \mathbf{T}_4 \mathbf{x}$ . For  $\mathbf{S}_1$  we need ‘definition by cases’ and characteristic terms for formulas of WE-T. Take

$$\mathbf{S}_1 := \begin{cases} \lambda \mathbf{x}, \mathbf{v}, \mathbf{q}. \mathbf{T}_3 \mathbf{xq}, & \text{if } t_{B_D}(\mathbf{T}_2 \mathbf{x}) \mathbf{v} =_0 0, \\ \lambda \mathbf{x}, \mathbf{v}, \mathbf{q}. \mathbf{T}_1 \mathbf{xv}, & \text{else,} \end{cases}$$

where  $t_{B_D}$  is the characteristic term for  $B_D$ .  $\mathbf{S}_1$  is definable in WE-T since as we have seen Cond is definable in terms of the recursion operator.

*The rest of the verification* of the logical axioms follows along the same line. All the arithmetical axioms, except induction, are immediately interpreted by their counterparts in WE-T<sub>H</sub>. Induction is interpreted by the recursion operator. We will see this when we verify the interpretation for the natural deduction system in the next chapter.  $\dashv$

### Example

As seen above, contraction as represented by the Con rule is crucial for the Dialectica interpretation: Characteristic terms and decidability of prime formulas are needed. Due to the importance of this phenomenon we will give an example of contraction. In some formulations of the Hilbert style calculus  $A \rightarrow A \wedge A$  is an axiom. In those formulations this axiom represents contraction and, as in the case of the Con rule, the realising terms are defined by definition by cases. From the translation of  $A(\mathbf{a}) \rightarrow A(\mathbf{a}) \wedge A(\mathbf{a})$  we see that we have to provide sequences of terms  $\mathbf{T}_1, \mathbf{T}_2$  and  $\mathbf{T}_3$  such that

$$\text{WE-T}_H \vdash A_D(\mathbf{x}, \mathbf{T}_1 \mathbf{a} \mathbf{x} \mathbf{y}_1 \mathbf{y}_2, \mathbf{a}) \rightarrow (A_D(\mathbf{T}_2 \mathbf{a} \mathbf{x}, \mathbf{y}_1, \mathbf{a}) \wedge A_D(\mathbf{T}_3 \mathbf{a} \mathbf{x}, \mathbf{y}_2, \mathbf{a})). \quad (2.6)$$

Take  $\mathbf{T}_1 := \lambda \mathbf{a}, \mathbf{x}, \mathbf{y}_1, \mathbf{y}_2. \text{Cond}(\mathbf{y}_1, \mathbf{y}_2, t_{A_D} \mathbf{x} \mathbf{y}_2 \mathbf{a})$  and  $\mathbf{T}_2 := \mathbf{T}_3 := \lambda \mathbf{a}, \mathbf{x}. \mathbf{x}$ . This yields (2.6).

We will now go directly to natural deduction and thus postpone the mathematical and philosophical consequences of the interpretation.

## Interpretation Theorems within Natural Deduction

After having seen the interpretation of  $\text{WE-HA}^\omega$  formulated in Hilbert style we turn to natural deduction. We will do this for two reasons: Apparently the Dialectica interpretation has never been presented within natural deduction. This is of course a motivation in itself, but for practical reasons it would be interesting for the future to have such an interpretation, since natural deduction provides a flexibility which is not present within Hilbert style. Another reason is that as the interpretation unfolds new insights show up. These connect, in particular, with a “contraction lemma” which is necessary for our natural deduction interpretation. But first we will introduce the system and discuss some of the central features regarding natural deduction.

The discovery of natural deduction as an elegant and natural way of formalising predicate logic, intuitionistic as well as classic, is often ascribed to Gentzen (1935), but apparently Gentzen was anticipated by S. Jáskowski who developed this formalism for classical logic in a kind of linear style, see (Troelstra & Schwichtenberg, 1996, 47). It is, nevertheless, Gentzen who introduced natural deduction in the form we know it today and him who made the first deep investigations of the formalism. One of his main reasons for choosing natural deduction (and giving it its name) was that it comes very close to informal reasoning. This feature makes it easy to do derivations inside natural deduction. The essential difference between doing derivations in Hilbert style contra natural deduction is that in the latter one does derivations under assumptions. This feature has, of course, crucial consequences when we want to do the Dialectica interpretation.

A derivation in natural deduction takes the form of a rooted tree. The formula appearing at the root node at the bottom of the tree is the proven formula and the formulas at the top of the branches, i.e. the leaves, are the assumptions of the derivation. Some rules discharge some of the assumptions of the derivation; when we discharge an assumption we indicate this by writing one of the letters  $u, v, w$  to the left of the formula and simultaneous writing the label to the right of the name of the rule. Actually a whole class of formulas of the same form occurring at different places in the tree can be discharged, namely the formula occurrences with the same label. An *assumption class* consisting of formulas of the same form, say  $A$ , is written in squared brackets  $[A]$ . A derivation in natural deduction starts typically by assuming one or more formulas, and then – in an ongoing proces – a tree is generated by the rules of inference. Some rules used in the process then discharge some of the assumptions.

### 3.1 Formulation of $\text{WE-HA}_{\text{ND}}^\omega$ and $\text{WE-T}_{\text{ND}}$

The language, definition of terms and formulas, notation and the type structure underlying the theory are the same as for  $\text{WE-HA}_{\text{H}}^\omega$ . It is the formulation of the logic and arithmetic that makes the difference.

3.1.1 Rules of WE-HA<sub>ND</sub><sup>0</sup>

The rules of WE-HA<sub>ND</sub><sup>0</sup> operate on *derivations*. In this respect the formalism is on a higher level than Hilbert style, since within the Hilbert style formalism one only operates on formulas. The rules of natural deduction are inductive rules: They are used on given derivations to obtain new derivations – we therefore have above an application of a rule one, two or three derivations pertaining to the specific rule: The rule  $\wedge I$ , for instance, requires two derivations whereas  $\wedge E$  only requires one. The initial derivation is the trivial derivation: an assumption of  $A$  is by definition also a derivation of  $A$  from the only assumption  $A$ .

One convention regarding natural deduction is that when we write down the *rules* we only mention those assumptions that are of interest to the rule we are writing down, but there can in general be any finite number of assumptions to a given derivation. The logical rules are:

$$\begin{array}{c}
\frac{A \quad B}{A \wedge B} \wedge I \\
\\
\frac{A_i}{A_1 \vee A_2} \vee I_i, \\
\\
\frac{[u : A] \quad \vdots \quad B}{A \rightarrow B} \rightarrow I, u \\
\\
\frac{A(b^\sigma)}{\forall x^\sigma A(x^\sigma)} \forall I \\
\\
\frac{A(t^\sigma)}{\exists x^\sigma A(x^\sigma)} \exists I \\
\\
\frac{\perp}{A} \perp I
\end{array}
\qquad
\begin{array}{c}
\frac{A_1 \wedge A_2}{A_i} \wedge E_i, \\
\\
\frac{[u : A] \quad \vdots \quad C \quad [v : B] \quad \vdots \quad C}{C} \vee E, u, v \\
\\
\frac{A \quad A \rightarrow B}{B} \rightarrow E \\
\\
\frac{\forall x^\sigma A(x^\sigma)}{A(t^\sigma)} \forall E \\
\\
\frac{[u : A(b^\sigma)] \quad \vdots \quad C}{\exists x^\sigma A(x^\sigma) \quad C} \exists E, u
\end{array}$$

The rules in the left column, except for  $\perp I$  are naturally called introduction rules and the rules in the right column are called elimination rules. The subscript I in  $\perp I$  denotes that the rule is intuitionistic. The variable  $b^\sigma$  in  $\forall I$  and  $\exists E$  is the eigenvariable. This means that with respect to the application of  $\forall I$ ,  $b^\sigma$  must not occur free in (non-discharged) assumptions of the derivation. With respect to  $\exists E$ ,  $b^\sigma$  must not occur free in assumptions on which  $C$  depend, except for  $A(b^\sigma)$ . Furthermore  $\exists E$  has the restriction that  $b^\sigma$  is not free in  $C$ , i.e.  $b^\sigma \notin FV(C)$ .

A discharged assumption is naturally not an assumption (anymore). We stress that not

necessarily all assumptions of the same form occurring above an application of  $\rightarrow I$ ,  $\vee E$  or  $\exists E$  are being discharged—we discharge only those formula occurrences that are labelled (which can be zero). We are working with what is called CDC (crude discharge convention) if we always discharge every assumption that is dischargeable.

We will try to explain two of the rules. First  $\vee E$ . Why is this an elimination rule? The rule can be seen as corresponding to the informal method of “proof by cases”: Assume we have established  $A$  or  $B$  and that we want to show  $C$ . Then it suffices to show the simpler cases—that  $C$  follows from  $A$  and that  $C$  also follows from  $B$ . On the more formal level we see that if we want to derive  $C$  from  $A \vee B$  we can *eliminate*  $\vee$  and derive  $C$  from  $A$  and from  $B$ , separately.

Second  $\exists E$ . Here  $\exists E$  serves to eliminate the existential quantifier from  $\exists xA(x)$  when we want to derive  $C$  from  $\exists xA(x)$  and possibly other assumptions. The rule corresponds to the standard mathematical proof procedure “there exists an  $x$  such that  $A(x)$ ; now pick such an object  $b$ , then ...”. If our conclusion does not contain that specific object  $b$  and does not with respect to any other assumption depend on  $b$ , then we can conclude that  $C$  indeed follows from  $\exists xA(x)$ .

### Contraction and notation

In the Hilbert style formalism presented earlier contraction is present by the rule:

$$\frac{A \rightarrow B \quad A \rightarrow C}{A \rightarrow B \wedge C} \text{Con}$$

In linear logic this is not allowed, since one has to be ‘economical’ with respect to assumptions. As J.-Y. Girard puts it: It does not follow from the fact that I can buy a pack of cigarettes for a dollar and the fact that I can buy a lighter for another dollar that I can buy a pack of cigarettes *and* a lighter for a dollar. Therefore Con is not a rule of linear logic. Contraction is present in natural deduction in a way similar to Con. When we for instance discharge a class  $[A]$  from our assumptions using  $\rightarrow I$ , we are allowed to discharge all *occurrences* of  $A$  as an assumption above the rule. Having this in mind, implication introduction,  $\rightarrow I$ , says:

$$\Gamma, A, \dots, A \vdash B \Rightarrow \Gamma \vdash A \rightarrow B.$$

One should therefore pay attention to all occurrences of assumptions when discussing the rules of natural deduction. We will especially have to be careful in connection with the Dialectica interpretation, since, as we saw in the foregoing chapter, contraction plays a central role corresponding to ‘definition by cases’.

Another notation for natural deduction is with context. The notation is horizontal compared with the notation displayed above. With context notation one sees at the bottom what the assumptions of the proven formula are, i.e. what the context is. The notation is useful in connection with the so-called Curry-Howard correspondence and is essentially the same notation as Gentzen’s sequent calculus. We will later in this chapter use this notation in connection with linear logic, but here we will use Gentzen’s original notation for natural deduction with formula-trees.

### Equality and arithmetic

The equality rules of  $\text{WE-HA}_{\text{ND}}^0$  are the following rules, where all terms are of type 0 (the first rule is naturally called a zero-premise rule):

$$\frac{}{t =_0 t} \text{E}_1 \quad \frac{t =_0 s}{s =_0 t} \text{E}_2 \quad \frac{t =_0 s \quad s =_0 r}{t =_0 r} \text{E}_3$$

Rule of weak extensionality:

$$\frac{\begin{array}{c} \Gamma_{\text{qf}} \\ \vdots \\ s^\sigma =_\sigma t^\sigma \end{array}}{r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau} \text{QF-ER},$$

where  $\Gamma_{\text{qf}}$  only consists of quantifier free formulas, and  $s^\sigma, t^\sigma$  and  $r[x^\sigma]^\tau$  are terms.

Defining rules for  $S^1$ , where all symbols except  $S^1$  belong to type 0:

$$\frac{Sx = 0}{\perp} \text{S}_1 \quad \frac{Sx = Sy}{x = y} \text{S}_2$$

Note, that the converse of  $\text{S}_2$ , i.e.  $x = y \rightarrow Sx = Sy$  follows from QF-ER.

Rule of induction:

$$\frac{\begin{array}{c} [u : A(b^0)] \\ \vdots \\ A(0^0) \quad A(Sb^0) \end{array}}{A(y^0)} \text{Ind}, u$$

where  $b^0$  is eigenvariable not occurring free in any assumptions on which  $A(Sb^0)$  depends, except  $A(b^0)$  and  $y^0$  is any variable not occurring free in any assumptions.

Defining equations for  $\Pi_{\rho,\tau}$ ,  $\Sigma_{\rho,\tau,\sigma}$  and  $\text{R}_\sigma$  are the same as in the Hilbert system, and with these we again obtain lambda abstraction as a defined notion.

#### 3.1.2 The subsystem $\text{WE-T}_{\text{ND}}$

Again, the weakly extensional version of Gödel's system T is to be considered as the quantifier free subsystem of the theory. Therefore  $\text{WE-T}_{\text{ND}}$  arises when we 'remove all quantifiers' from  $\text{WE-HA}_{\text{ND}}^0$ . Thus, the rules are mainly the same; we omit the quantifier rules but as a result of this we introduce a substitution rule:

$$\frac{A(x^\sigma)}{A(t^\sigma)} \text{Sub},$$

$t^\sigma$  is any term and  $x^\sigma$  not in free in any assumptions. The induction rule remains the same which is also the case regarding equality rules and the axioms for  $\Pi_{\rho,\tau}$ ,  $\Sigma_{\rho,\tau,\sigma}$  and  $\text{R}_\sigma$ . The

rule of weak extensionality has to be slightly modified in order to cope with the definition of higher type equations within WE-T<sub>ND</sub>. If  $\sigma$  is not 0 and  $\sigma = \sigma_1 \dots \sigma_n 0$  then  $s^\sigma =_\sigma t^\sigma$  is within WE-T<sub>ND</sub> short for  $s^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n} =_0 t^\sigma x_1^{\sigma_1} \dots x_n^{\sigma_n}$ , where the  $x_i$ 's are variables not occurring in  $s^\sigma$ ,  $t^\sigma$ , and which are not free in assumptions on which  $s^\sigma =_\sigma t^\sigma$  depends. Therefore, in WE-T<sub>ND</sub> the rule of weak extensionality has the form

$$\frac{\begin{array}{c} \Gamma \\ \vdots \\ s^\sigma =_\sigma t^\sigma \end{array}}{r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau} \text{QF-ER}_T,$$

where none of the  $x_i$ 's which are hidden in  $s^\sigma =_\sigma t^\sigma$  occur in  $\Gamma$ .

### 3.2 Discussion of the deduction theorem

Before turning to the interpretation of WE-HA<sub>ND</sub><sup>ω</sup> into WE-T<sub>ND</sub> let us examine the deduction theorem in the context of WE-HA<sup>ω</sup>—both Hilbert style and natural deduction. We do this since we have to do Dialectica translations of deductions under assumptions, since derivations in natural deduction are done under assumptions. Furthermore, Troelstra makes the following remark when discussing different formulations of HA<sup>ω</sup> (either with intensional equality, I-HA<sup>ω</sup> or with extensional equality E-HA<sup>ω</sup> or WE-HA<sup>ω</sup>):

WE-HA<sup>ω</sup> as an intermediate possibility [between I-HA<sup>ω</sup> and E-HA<sup>ω</sup>] is not very attractive: the deduction theorem does not hold for this the theory. (Troelstra, 1990, 231).

Let us begin with an examination of the point that the deduction theorem does not hold for WE-HA<sup>ω</sup>. But, first of all, we will have to specify which version of the deduction theorem we are talking about.

Let  $S$  be a theory,  $A(\mathbf{x})$  a formula where  $FV(A) = \{\mathbf{x}\}$ . Now, “;” denotes the operation when we assume, for the sake of an argument, a formula, say  $A(\mathbf{x})$ , whereas “+” denotes the operation when we add a formula as an axiom to a theory and no longer regard that formula as an assumption (in the technical sense of an assumption that can be discharged). When we add an axiom it is implicitly understood to be universally closed. Now, using this simple difference between “+” and “;” there are at least two different versions of the deduction theorem:

**Deduction theorem 1.** If  $S, A(\mathbf{x}) \vdash B$  then  $S \vdash A(\mathbf{x}) \rightarrow B$ .

**Deduction theorem 2.** If  $S + A(\mathbf{x}) \vdash B$  then  $S \vdash \forall \mathbf{x} A(\mathbf{x}) \rightarrow B$ .

When we do deductions under open assumptions by the operation “;” we must in the context of Hilbert style make some restrictions with respect to the quantifiers:

$$C_1(\mathbf{c}_1), \dots, C_n(\mathbf{c}_n) \vdash B(\mathbf{b}) \rightarrow A(a) \Rightarrow C_1(\mathbf{c}_1), \dots, C_n(\mathbf{c}_n) \vdash B(\mathbf{b}) \rightarrow \forall x A(x),$$

in case  $a \notin \{\mathbf{b}, \mathbf{c}_1, \dots, \mathbf{c}_n\}$ . Likewise

$$C_1(\mathbf{c}_1), \dots, C_n(\mathbf{c}_n) \vdash A(a) \rightarrow B(\mathbf{b}) \Rightarrow C_1(\mathbf{c}_1), \dots, C_n(\mathbf{c}_n) \vdash \exists x A(x) \rightarrow B(\mathbf{b}),$$

in case  $a \in \{\mathbf{b}, \mathbf{c}_1, \dots, \mathbf{c}_n\}$ .

Following (Kleene, 1952, 98), these restrictions could also be made by displaying variables that cannot be used as eigenvariables by superscript on the turnstile, i.e.  $\vdash^{\mathbf{x}}$  means that none of the variables occurring in  $\mathbf{x}$  can be used as eigenvariables.

The following argument is the standard argument when one shows the failure of the deduction theorem  $\text{WE-HA}^\omega$ , see for instance (Troelstra, 1973, 242). It takes place in the context of Hilbert style.

Let  $f^1$  and  $g^1$  be functions of type 1. Assume  $f =_1 g$ ; this gives us a derivation of  $f =_1 g$ . Now apply QF-ER where  $r[x^\sigma]^\tau$  is  $z^2 x^1$  thereby obtaining  $z^2 f =_0 z^2 g$ .<sup>1</sup> If deduction theorem 1 were true we could derive  $f =_1 g \rightarrow z^2 f =_0 z^2 g$  and introduce universal quantifiers for the free variables  $f, g$  and  $z$ . We would then have derived the extensionality axiom  $\forall z^2 E_2(z)$  in  $\text{WE-HA}_H^\omega$ . In the foregoing chapter we saw that  $\text{WE-HA}_H^\omega$  has a functional interpretation into  $\text{WE-T}_H$ . But Howard (1973) has shown by using the notion of majorizability that no functional of  $\text{WE-T}_H$  can Dialectica interpret  $\forall z^2 E_2(z)$ . Therefore deduction theorem 1 fails for  $\text{WE-HA}_H^\omega$ .

Drawbacks of the argument:

- (i) It does not provide a counterexample to deduction theorem 2.
- (ii) It does not work for our natural deduction formulation,  $\text{WE-HA}_{\text{ND}}^\omega$ .

In the context of natural deduction we have to distinguish strictly between assumptions and axioms, and here we *cannot* take  $f =_1 g$  as an open assumption and then apply QF-ER, since  $f =_1 g$  is not quantifier free. With respect to (i) we will of course not introduce  $f =_1 g$  as axiom, since it would make our theory inconsistent. Thus, we see that the argument simply does not work for  $\text{WE-HA}_{\text{ND}}^\omega$  and it does not produce a counterexample for deduction theorem 2.

A more interesting and conclusive argument showing version 2 of the deduction theorem to fail can be made with the axiom  $\text{Con}_{\text{PA}}$  (Kohlenbach, 2001).  $\text{Con}_{\text{PA}}$  expresses the consistency of Peano arithmetic: Let  $\text{Proof}(x, y)$  be a primitive recursive predicate expressing that  $x$  is (the Gödel number of) a proof in PA of a formula (with the Gödel number)  $y$ . Then  $\text{Con}_{\text{PA}}$  is the closed  $\Pi_1^0$ -formula

$$\forall x \neg \text{Proof}(x, \ulcorner 0 = 1 \urcorner).$$

Let  $t_{\text{PA}}$  of type 1 be the characteristic term for  $\neg \text{Proof}(x, \ulcorner 0 = 1 \urcorner)$ , i.e.  $\text{WE-T}$  proves the equivalence  $t_{\text{PA}} x =_0 0 \leftrightarrow \neg \text{Proof}(x, \ulcorner 0 = 1 \urcorner)$ . By a relatively short argument it is shown in (Kohlenbach, 2001) that deduction theorem 2 fails with respect to the axiom  $\text{Con}_{\text{PA}}$ . For

<sup>1</sup>Note that we can always weaken a derived formula  $A$  to  $B \rightarrow A$ , for any formula  $B$ : from the axiom  $A \wedge B \rightarrow A$  derive by Expo  $A \rightarrow (B \rightarrow A)$  and then apply MP.

$B \equiv (z^2(t_{\text{PA}}) =_0 z^2(\lambda x.0))$ , where  $z^2$  is a free variable of type 2, we have<sup>2</sup>

$$\text{WE-HA}^\omega + \text{Con}_{\text{PA}} \vdash B, \text{ but } \text{WE-HA}^\omega \not\vdash \text{Con}_{\text{PA}} \rightarrow B.$$

Apart from showing the failure of the deduction theorem this is also interesting in itself: If we tell  $\text{WE-HA}^\omega$  that Peano arithmetic is consistent then  $B$  is provable, but  $\text{WE-HA}^\omega$  cannot, on the other hand, prove that  $B$  follows from the consistency of Peano arithmetic.

Now, why is it important that the deduction theorem holds? There are mainly three reasons here. The first is principal. If the deduction theorem holds for a theory  $S$ , then we know that there is formal agreement between our notion of logical deduction in  $S$ ,  $\vdash_S$ , and the formal symbol for implication  $\rightarrow$ . This expresses, in other words, that  $\rightarrow$  *formalises in  $S$*  the concept of deduction in  $S$ . This is of course in itself a desirable property.

The second reason is that the deduction theorem in connection with Herbrand's theorem has an interesting application. A simplified version of Herbrand's theorem for classical predicate logic runs as follows:

**Simplified Herbrand's theorem.** If  $A_{\text{qf}}$  is a quantifier free formula with just  $x$  as free variable and  $\vdash \exists x A_{\text{qf}}(x)$  is true in classical logic then there exists a finite sequence of closed terms  $t_1, \dots, t_n$  such that  $\vdash A_{\text{qf}}(t_1) \vee \dots \vee A_{\text{qf}}(t_n)$ .

In connection with classical theories given by purely universal axioms this allows for a reduction to logic *if* the deduction theorem holds: If we have derived  $\exists x A_{\text{qf}}(x)$  from a set  $\Gamma$  of purely universal axioms  $A_1, \dots, A_n$ , where  $A_{\text{qf}}$  is quantifier free with just  $x$  as free variable, then we can use the deduction theorem and derive without use of axioms (others than the axioms for logic)  $A_1 \rightarrow \dots \rightarrow A_n \rightarrow \exists x A_{\text{qf}}(x)$ . Then we get all the quantifiers to the front where they become existential and we are able to use Herbrand's theorem. We then get a finite sequence of formulas where the formulas are implications of the form  $A'_1 \rightarrow \dots \rightarrow A'_n \rightarrow A_{\text{qf}}(t_i)$  where  $A'_j$  has closed terms replacing the originally quantified variables in  $A_j$  and the index of  $t_i$  is running from 1 to (some)  $m$ . Thus  $A'_j$  follows from  $A_j$  and we can by assuming the original axioms  $\Gamma$  derive  $A_{\text{qf}}(t_1) \vee \dots \vee A_{\text{qf}}(t_m)$ .<sup>3</sup>

The third reason for wishing that the deduction theorem (here version 1) should hold is practical. If we want to derive  $A \rightarrow B$  we can, hypothetically, assume  $A$  then derive  $B$  and use the deduction theorem to derive  $A \rightarrow B$ . Anyone with the slightest experience regarding derivations in Hilbert style calculi knows, that this is of *great* convenience. Because it shows – as also said above – that with respect to assumptions our formal concept of derivation is consistent with our informal concept of proof, which is much more flexible. Kleene formulates this in the following way:

The property of deducibility expressed by the next theorem [the deduction theorem] corresponds to a familiar method in our informal reasoning. To establish

<sup>2</sup>Actually the argument in (Kohlenbach, 2001) shows the deduction theorem to fail also for the classical theory:  $\text{WE-PA}^\omega$ .

<sup>3</sup>A distinguishing feature about intuitionistic logic is the *Explicit Definability*: If we have proved  $\vdash \exists x B(x)$  for a sentence  $\exists x B(x)$  then we can find a closed term  $t$  such that  $\vdash B(t)$ . But this fails for classical logic. However, Herbrand's theorem shows that the next best thing is the case with respect to quantifier free formulas, namely that we can find a *finite* sequence of closed terms  $t_1, \dots, t_m$  such that  $A_{\text{qf}}$  holds for at least one of them.

an implication “if  $A$ , then  $B$ ”, we often assume  $A$  “for the sake of the argument” and undertake to deduce  $B$ . (Kleene, 1967, 39).

Now, if the deduction theorem does not hold for the theory we are working with we feel ‘unsafe’. The formal and informal concepts of proofs do not agree, and it is a lot more troublesome to make formal derivations.

### 3.2.1 Deduction theorem 1 holds for $WE-HA_{ND}^{\omega}$

In natural deduction we distinguish between assumptions and axioms, as the latter are not counted among the assumptions. This means that within natural deduction the deduction theorem is true w.r.t. (open) assumptions simply by definition of  $\rightarrow$ I: If we have derived  $A$  under the assumption  $B$  we can conclude  $B \rightarrow A$ , and discharge  $B$  from our assumptions.

**Deduction theorem 1.** (Restated).

If  $C_1, \dots, C_n, B \vdash A$  in  $WE-HA_{ND}^{\omega}$  then  $C_1, \dots, C_n \vdash B \rightarrow A$  in  $WE-HA_{ND}^{\omega}$ .

Note that the variable conditions with respect to quantifiers (which were stated earlier in the case of Hilbert systems) are automatically satisfied by definition of derivation in natural deduction.

Summing up, this points to the following: The deduction theorem holds for  $WE-HA_{ND}^{\omega}$  w.r.t. assumptions, because our very formulation of QF-ER blocks for the problematic cases. Furthermore, the deduction theorem holds for axioms as long as these axioms are not leaves of branches where QF-ER is used. We see that the way we usually want to *use* the deduction theorem, namely w.r.t. assumptions, is valid, and this is of practical importance when we do derivations because it is precisely w.r.t. assumptions we want the deduction theorem to hold.

Analysing the deduction theorem in this way we do not find the failure of the full deduction theorem *that* inconvenient as Troelstra does.

We note that this way of using the deduction theorem corresponds to the use of  $\oplus$  in (Kohlenbach, 2001), though the analysis here in the context of natural deduction displays more detail.

## 3.3 Translation of derivations under assumptions

In  $WE-HA_H^{\omega}$  the term extraction starts with the terms given by the verification of the ‘real axioms’. From these terms we build up realising terms according to the algorithm given by the verification of the rules; an example is MP which corresponds to application of terms. But how does the term extraction start in natural deduction? A derivation in natural deduction begins typically from an assumption, say  $A$  which at the same time is a derivation of  $A$ . In a moment we will see, that the verification of this gives the initial realising terms.

Again the difference between natural deduction and Hilbert style is that in the former we are considering derivations under assumptions. This was not the case when we defined the Dialectica translation on page 23 and we must therefore extend the translation in order to include derivations under assumptions. This is actually straightforward. If we have a derivation in  $WE-HA_{ND}^{\omega}$  of  $A$  from the assumptions  $A_1, \dots, A_n$ , i.e.  $A_1, \dots, A_n \vdash A$  then – since

the deduction theorem holds for assumptions – this is equivalent to  $\vdash A_1 \rightarrow \dots \rightarrow A_n \rightarrow A$ , modulus any permutation of the  $n$   $A_i$ 's (parenthesis associated to the right). Therefore, since we have complete agreement between  $\vdash$  and  $\rightarrow$  w.r.t. assumptions we Dialectica translate  $A_1, \dots, A_n \vdash A$  just as we would translate  $A_1 \rightarrow \dots \rightarrow A_n \rightarrow A$ . Thus we see that the translation of  $A_1, \dots, A_n \vdash A$  is that there exist  $\mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}$  such that for all  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}$

$$A_1(\mathbf{x}_1, \mathbf{Y}_1 \mathbf{x}_1 \dots \mathbf{x}_n \mathbf{y}), \dots, A_n(\mathbf{x}_n, \mathbf{Y}_n \mathbf{x}_1 \dots \mathbf{x}_n \mathbf{y}) \vdash A(\mathbf{X} \mathbf{x}_1 \dots \mathbf{x}_n, \mathbf{y}),$$

where  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}, \mathbf{Y}_1, \dots, \mathbf{Y}_n, \mathbf{X}$  are fresh variables.

### 3.4 Interpretation theorem for WE-HA<sub>ND</sub><sup>0</sup>

We know that WE-HA<sub>ND</sub><sup>0</sup> and WE-HA<sub>H</sub><sup>0</sup> are equivalent in the sense that everything which is derivable in the former is also derivable in the latter and vice versa. Therefore the soundness theorem is also true for WE-HA<sub>ND</sub><sup>0</sup>. However, when we want to extract terms we will have to do it by recursion on derivations in WE-HA<sub>H</sub><sup>0</sup>. To get an algorithm for extracting terms from derivations in WE-HA<sub>ND</sub><sup>0</sup> we have to give a independent proof of the soundness theorem.

First we need some lemmas.

**Lemma 3.4.1.** *Let  $A(x^\sigma)$  be quantifier free, then*

$$\frac{\begin{array}{c} \Gamma_{\text{qf}} \\ \vdots \\ t^\sigma = s^\sigma \quad A(t^\sigma) \end{array}}{A(s^\sigma)} \text{QF-ER}',$$

is derivable in WE-T<sub>ND</sub>, where none of the  $x_i$ 's hidden in  $t^\sigma = s^\sigma$  occur free in  $\Gamma_{\text{qf}}$ .

**Proof.** Let  $t_A$  of type  $\sigma_0$  be the characteristic term for  $A$ , i.e.  $\text{WE-T}_{\text{ND}} \vdash t_A x^\sigma =_0 0 \leftrightarrow A(x^\sigma)$ . From this follows

$$\frac{\frac{\frac{\frac{\Gamma_{\text{qf}}}{\vdots} t^\sigma =_\sigma s^\sigma}{t_A t^\sigma =_0 t_A s^\sigma} \text{QF-ER}_T}{t_A s^\sigma =_0 t_A t^\sigma} \text{E}_2}{t_A s^\sigma =_0 0} \frac{\frac{A(t^\sigma) \quad A(t^\sigma) \rightarrow t_A t^\sigma =_0 0}{t_A t^\sigma =_0 0} \rightarrow \text{E}}{\text{E}_3} \rightarrow \text{E}}{A(s^\sigma)} \frac{t_A s^\sigma =_0 0 \rightarrow A(s^\sigma)}{\rightarrow \text{E}}$$

□

### Notation

In the following it will be convenient to use some notation for finite sequences of formulas. Say

$$(C^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x}))_{i=1}^n := C^1(\mathbf{x}_1, \mathbf{T}_1 \mathbf{x}), \dots, C^n(\mathbf{x}_n, \mathbf{T}_n \mathbf{x}).$$

This notation will also be used when all the  $C$ 's are one and the same formula, just with at most that many different free variables and terms; then without indexing  $C$ , i.e.  $(C(\mathbf{x}_i, \mathbf{T}_i \mathbf{x}))_{i=1}^n$ .

The following contraction lemma takes care of the contractions which are present in the rules that discharge assumptions.

**Lemma 3.4.2.** (Contraction lemma). *Let  $A, B$  and  $C^1, \dots, C^n$  be formulas of  $\mathcal{L}(\text{WE-T})$ , and say  $\mathbf{x}$  denotes  $\mathbf{x}_1 \dots \mathbf{x}_n$ . If*

$$(C^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{xz}\check{\mathbf{z}}\mathbf{v}))_{i=1}^n, A(\mathbf{z}, \mathbf{T}' \mathbf{xz}\check{\mathbf{z}}\mathbf{v}), A(\check{\mathbf{z}}, \mathbf{T}'' \mathbf{xz}\check{\mathbf{z}}\mathbf{v}) \vdash B(\mathbf{T} \mathbf{xz}\check{\mathbf{z}}, \mathbf{v})$$

for sequences of closed terms  $\mathbf{T}_1, \dots, \mathbf{T}_n, \mathbf{T}', \mathbf{T}'', \mathbf{T}$  then

$$(C^i(\mathbf{x}_i, \mathbf{S}_i \mathbf{xz}\mathbf{v}))_{i=1}^n \vdash A(\mathbf{z}, \mathbf{S}^* \mathbf{xz}\mathbf{v}) \rightarrow B(\mathbf{S} \mathbf{xz}\mathbf{v})$$

for certain sequences of closed terms  $\mathbf{S}_1, \dots, \mathbf{S}_n, \mathbf{S}^*, \mathbf{S}$ .

**Proof.** From the assumption of the lemma we have by substitution of  $\mathbf{z}$  for  $\check{\mathbf{z}}$ :

$$(C^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{xz}\mathbf{z}\mathbf{v}))_{i=1}^n, A(\mathbf{z}, \mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v}), A(\mathbf{z}, \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) \vdash B(\mathbf{T} \mathbf{xz}\mathbf{z}, \mathbf{v}) \quad (3.1)$$

Let  $t_A$  be the characteristic term for  $A$ . Then define

$$\mathbf{S}^* := \lambda \mathbf{x}, \mathbf{z}, \mathbf{v}. \text{Cond}(\mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v}, \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}, t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v})).$$

From this definition and the general definition of  $\text{Cond}$  (see page 21) we see that in  $\text{WE-T}_{\text{ND}}$  it is true that

$$\begin{aligned} t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) =_0 0 &\rightarrow \mathbf{S}^* = \lambda \mathbf{x}, \mathbf{z}, \mathbf{v}. \mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v}, \\ t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) \neq_0 0 &\rightarrow \mathbf{S}^* = \lambda \mathbf{x}, \mathbf{z}, \mathbf{v}. \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}. \end{aligned} \quad (3.2)$$

We have in  $\text{WE-T}_{\text{ND}}$  that any term of type 0 reduces (computes) to a number. Therefore

$$\text{WE-T}_{\text{ND}} \vdash (t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) =_0 0) \vee (t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) \neq_0 0).$$

If we assume  $t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) = 0$  we get  $\mathbf{S}^* \mathbf{xz}\mathbf{v} = \mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v}$  from (3.2) by use of  $\rightarrow\text{E}$ , QF-ER' and  $\beta$ -contraction, therefore

$$\frac{\mathbf{S}^* \mathbf{xz}\mathbf{v} = \mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v} \quad A(\mathbf{z}, \mathbf{S}^* \mathbf{xz}\mathbf{v})}{A(\mathbf{z}, \mathbf{T}' \mathbf{xz}\mathbf{z}\mathbf{v})} \text{QF-ER}',$$

and at the same time, since  $t_A$  is the characteristic term for  $A$ ,

$$\frac{t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) = 0 \quad t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) = 0 \rightarrow A(\mathbf{z}, \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v})}{A(\mathbf{z}, \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v})} \rightarrow\text{E}.$$

On the other hand, if  $t_A \mathbf{z}(\mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}) \neq 0$  we have by (3.2) that  $\mathbf{S}^* \mathbf{xz}\mathbf{v} = \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v}$  which gives us

$$\frac{\mathbf{S}^* \mathbf{xz}\mathbf{v} = \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v} \quad A(\mathbf{z}, \mathbf{S}^* \mathbf{xz}\mathbf{v})}{A(\mathbf{z}, \mathbf{T}'' \mathbf{xz}\mathbf{z}\mathbf{v})} \text{QF-ER}'. \quad (3.3)$$

Again, since  $t_A$  is the characteristic term for  $A$  and since it is generally valid to take the contraposition, we have

$$\frac{t_A z(\mathbf{T}'' x z z \mathbf{v}) \neq 0 \quad A(z, \mathbf{T}'' x z z \mathbf{v}) \rightarrow t_A z(\mathbf{T}'' x z z \mathbf{v}) = 0}{\neg A(z, \mathbf{T}'' x z z \mathbf{v})} \text{Contraposition.}$$

Still under the assumption that  $t_A z(\mathbf{T}'' x z z \mathbf{v}) \neq 0$ , this together with (3.3) gives us

$$\frac{A(z, \mathbf{T}'' x z z \mathbf{v}) \quad \neg A(z, \mathbf{T}'' x z z \mathbf{v})}{\perp} \rightarrow E$$

$$\frac{\perp}{A(z, \mathbf{T}' x z z \mathbf{v})} \perp I$$

Thus, independently of whether  $t_A z(\mathbf{T}'' x z z \mathbf{v})$  equals 0 we see that from  $A(z, \mathbf{S}^* x z \mathbf{v})$  we can in  $WE-T_{ND}$  derive both  $A(z, \mathbf{T}' x z z \mathbf{v})$  and  $A(z, \mathbf{T}'' x z z \mathbf{v})$ . In the derivation given by (3.1) we can therefore replace the assumptions  $A(z, \mathbf{T}' x z z \mathbf{v})$  and  $A(z, \mathbf{T}'' x z z \mathbf{v})$  by the above derivations from the assumption  $A(z, \mathbf{S}^* x z \mathbf{v})$ .

Define therefore

$$\mathbf{S}_i := \lambda x, z, \mathbf{v}. \mathbf{T}_i x z z \mathbf{v} \quad \text{and} \quad \mathbf{S} := \lambda x, z. \mathbf{T} x z z.$$

We hereby have, together with  $\mathbf{S}^*$ , that

$$(C^i(x_i, \mathbf{S}_i x z \mathbf{v}))_{i=1}^n, A(z, \mathbf{S}^* x z \mathbf{v}), A(z, \mathbf{S}^* x z \mathbf{v}) \vdash B(\mathbf{S} x z, \mathbf{v}).$$

One application of  $\rightarrow I$  gives the conclusion of the lemma.  $\dashv$

For the interpretation of the induction rule we need the following induction lemma.

**Lemma 3.4.3.** (Induction lemma). *If  $\Gamma \vdash C(0^0, \mathbf{v})$  and  $\Gamma \vdash C(x^0, \mathbf{T} x^0 \mathbf{v}) \rightarrow C(\mathbf{S} x^0, \mathbf{v})$  in  $WE-T_{ND}$  where  $x^0$  and  $\mathbf{v}$  are not free in  $\Gamma$ , then  $\Gamma \vdash C(x^0, \mathbf{v})$  in  $WE-T_{ND}$ .*

**Proof.** The proof of the lemma is particularly lengthy. One needs among other things simultaneous recursion and various kinds of arithmetical operations. See (Troelstra, 1973, 51-56) for a detailed proof.  $\dashv$

**Theorem 3.4.4.** (Soundness of  $D$ -translation). *If  $(C^i(c_i))_{i=1}^n \vdash A(\mathbf{a})$  in  $WE-HA_{ND}^{\omega}$ , then*

$$(C_D^i(x_i, \mathbf{T}_i c a x y, c_i))_{i=1}^n \vdash A_D(\mathbf{T} c a x, y, \mathbf{a}),$$

*in  $WE-T_{ND}$  for certain sequences of closed terms  $\mathbf{T}_1, \dots, \mathbf{T}_n, \mathbf{T}$  which can be extracted from a derivation of  $A(\mathbf{a})$  in  $WE-HA_{ND}^{\omega}$ , where  $\mathbf{c} \equiv c_1 \dots c_n$  and  $\mathbf{x} \equiv x_1 \dots x_n$ .*

**Remark 3.4.5.**

1. The special case of the theorem where the deduction has no assumptions appears as:  $WE-HA_{ND}^{\omega} \vdash A(\mathbf{a}) \Rightarrow WE-T_{ND} \vdash A_D(\mathbf{T} \mathbf{a}, y, \mathbf{a})$  for suitable closed terms  $\mathbf{T}$ . Note the flexibility natural deduction provides compared with Hilbert style. There is of course a price to pay. This will be clear from the following proof of the theorem. This points towards an essential difference between the two calculi: Natural deduction gives us flexibility and naturalness, whereas Hilbert style meta-mathematically provides short and elegant proofs.

2. The theorem is formulated with *sequences* of formulas as assumptions. But with respect to the mere existence of programs (terms) Dialectica realising the derived formula we could have written the assumptions as a *set* of formulas. The actual program depends, in other words, on the formula occurrences in a concrete proof, whereas the claim of the existence only depends on the set of assumptions – not on a sequence.

**Proof.** The proof is by induction on the length of the derivation of  $A(\mathbf{a})$  in  $\text{WE-HA}_{\text{ND}}^{\omega}$ . We will in the following proof not indicate explicitly whether the derivations are in  $\text{WE-T}_{\text{ND}}$  or in  $\text{WE-HA}_{\text{ND}}^{\omega}$ , since it should be obvious from the context. Furthermore, we will abbreviate a  $\lambda$ -abstraction  $\lambda x_1, \dots, x_n$  by  $\lambda x_1 \dots x_n$ . Notice, that sometimes we say “functionals” thereby meaning closed terms of  $\text{WE-T}$ .

**Base case.**  $A(\mathbf{a}) \vdash A(\mathbf{a})$ . As previously discussed in this chapter the translation of  $A(\mathbf{a}) \vdash A(\mathbf{a})$  corresponds to  $A(\mathbf{a}) \rightarrow A(\mathbf{a})$ . This was verified on page 27 and we therefore see that if we define  $\mathbf{T}_1 := \lambda \mathbf{a} \mathbf{x} \mathbf{y} . \mathbf{y}$  and  $\mathbf{T}_2 := \lambda \mathbf{a} \mathbf{x} . \mathbf{x}$  we have in  $\text{WE-T}_{\text{ND}}$

$$A_D(\mathbf{x}, \mathbf{T}_1 \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{a}) \vdash A_D(\mathbf{T}_2 \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}).$$

**Induction case.** We are considering derivations from assumptions and we will have to consider all formula occurrences because different terms can realise the same formula at different occurrences in the derivation. We therefore assume some enumeration of all the occurrences of assumptions. This could for instance be done by enumerating the formulas on the top of the branches of the derivation-tree from left to right. Sometimes we will omit free variables, which is done for sake of simplicity of notation.

**Subcase 1.** Last rule of the derivation is  $\wedge\text{I}$ .

$$\frac{A(\mathbf{a}) \quad B(\mathbf{b})}{A(\mathbf{a}) \wedge B(\mathbf{b})} \wedge\text{I}$$

Let  $\Gamma \equiv C^1(\mathbf{c}_1), \dots, C^n(\mathbf{c}_n)$  and  $\Delta \equiv C^{n+1}(\mathbf{c}_{n+1}), \dots, C^m(\mathbf{c}_m)$ , for some enumeration of the assumptions of the derivations of  $A(\mathbf{a})$  and  $B(\mathbf{b})$ , respectively. Since  $\Gamma$  and  $\Delta$  are sequences it can happen that  $C^i(\mathbf{c}_i) \equiv C^j(\mathbf{c}_j)$  for some  $i$  and  $j$ . We assume that  $\text{FV}(C^i) = \{\mathbf{c}_i\}$  and furthermore we write  $\mathbf{c}$  for  $\mathbf{c}_1 \dots \mathbf{c}_n$ ;  $\tilde{\mathbf{c}}$  for  $\mathbf{c}_{n+1} \dots \mathbf{c}_m$ ;  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$  and  $\tilde{\mathbf{x}}$  for  $\mathbf{x}_{n+1} \dots \mathbf{x}_m$ . Then the induction hypothesis (IH) is:

$$\begin{aligned} (C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n &\vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}), \\ (C_D^i(\mathbf{x}_i, \mathbf{T}_i \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}} \mathbf{v}, \mathbf{c}_i))_{i=n+1}^m &\vdash B_D(\mathbf{T}^* \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}}, \mathbf{v}, \mathbf{b}). \end{aligned} \quad (3.4)$$

Write  $\underline{\mathbf{c}}$  for  $\mathbf{c}_1 \dots \mathbf{c}_m$  and  $\underline{\mathbf{x}}$  for  $\mathbf{x}_1 \dots \mathbf{x}_m$ . We have to provide  $\mathbf{S}_i, \mathbf{S}, \mathbf{S}^*$  such that

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \underline{\mathbf{c}} \mathbf{a} \mathbf{b} \underline{\mathbf{x}} \mathbf{y} \mathbf{v}, \mathbf{c}_i))_{i=1}^m \vdash A_D(\mathbf{S} \underline{\mathbf{c}} \mathbf{a} \mathbf{b} \underline{\mathbf{x}}, \mathbf{y}, \mathbf{a}) \wedge B_D(\mathbf{S}^* \underline{\mathbf{c}} \mathbf{a} \mathbf{b} \underline{\mathbf{x}}, \mathbf{v}, \mathbf{b}). \quad (3.5)$$

From IH, (3.4), we derive in  $\text{WE-T}_{\text{ND}}$  by using  $\wedge\text{I}$ :

$$\frac{\begin{array}{c} (C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \\ \vdots \\ A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}) \end{array} \quad \begin{array}{c} (C_D^i(\mathbf{x}_i, \mathbf{T}_i \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}} \mathbf{v}, \mathbf{c}_i))_{i=n+1}^m \\ \vdots \\ B_D(\mathbf{T}^* \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}}, \mathbf{v}, \mathbf{b}) \end{array}}{A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}) \wedge B_D(\mathbf{T}^* \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}}, \mathbf{v}, \mathbf{b})}$$

Now we expand on all terms realising this in order to get terms of the right type, i.e. terms that fit and realise (3.5). By using QF-ER' we see that the following sequences of terms will do that.

$$\begin{aligned} \mathcal{S}_i &::= \begin{cases} \lambda \underline{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y} \mathbf{v}. \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, & \text{if } 1 \leq i \leq n, \\ \lambda \underline{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y} \mathbf{v}. \mathbf{T}_i \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}} \mathbf{v}, & \text{if } n < i \leq m, \end{cases} \\ \mathcal{S} &::= \lambda \underline{c} \mathbf{a} \mathbf{b} \mathbf{x}. \mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x}, \quad \mathcal{S}^* ::= \lambda \underline{c} \mathbf{a} \mathbf{b} \mathbf{x}. \mathbf{T}^* \tilde{\mathbf{c}} \mathbf{b} \tilde{\mathbf{x}}. \end{aligned}$$

**Subcase 2.** The last rule of the derivation is  $\wedge E_1$ :

$$\frac{A(\mathbf{a}) \wedge B(\mathbf{b})}{A(\mathbf{a})} \wedge E_1$$

Assume some enumeration of the occurrences of the assumptions  $\Gamma$ . Say, for this enumeration  $\Gamma$  consists of  $C^1(\mathbf{c}_1), \dots, C^n(\mathbf{c}_n)$ . When we write  $\mathbf{c}$  for  $\mathbf{c}_1 \dots \mathbf{c}_n$  and  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$  the IH is

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y} \mathbf{v}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{y}, \mathbf{a}) \wedge B_D(\mathbf{T}^* \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{v}, \mathbf{b}).$$

We have to provide sequences of closed terms  $\mathcal{S}_i, \mathcal{S}$  such that

$$(C_D^i(\mathbf{x}_i, \mathcal{S}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathcal{S} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}). \quad (3.6)$$

By using  $\wedge E$  in WE-T<sub>ND</sub> on the IH we get

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y} \mathbf{v}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{y}, \mathbf{a})$$

Now replace every  $b_j$  in  $\mathbf{b}$  which do not occur in  $\mathbf{a}$  nor in  $\mathbf{c}_i$  by the zero functional  $\mathbf{o}_j$ ; let the result be  $\mathbf{b}'$ . Replace likewise  $\mathbf{v}$  by a corresponding  $\mathbf{o}'$ . It is now clear that the following terms are closed and that they satisfy (3.6):

$$\begin{aligned} \mathcal{S}_i &::= \lambda \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}. \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{b}' \mathbf{x} \mathbf{y} \mathbf{o}', \quad 1 \leq i \leq n, \\ \mathcal{S} &::= \lambda \mathbf{c} \mathbf{a} \mathbf{x}. \mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b}' \mathbf{x}. \end{aligned}$$

Rule  $\wedge E_2$  is symmetric to  $\wedge E_1$ .

**Subcase 3.**  $\rightarrow I$ . The last rule of the proof in WE-HA<sub>ND</sub><sup>ω</sup> is

$$\frac{\begin{array}{c} [u : A], \Gamma \\ \vdots \\ B \end{array}}{A \rightarrow B} \rightarrow I, u$$

Assume some enumeration of the formula occurrences in  $\Gamma$ :  $C^1, \dots, C^n$ , and assume also some enumeration of all the occurrences of  $A$  labelled by  $u$ . Say there are  $m - n$  occurrences of  $u : A$ . We omit free variables. Comparing with the Dialectica translation of  $C^1 \rightarrow \dots \rightarrow C^n \rightarrow A \rightarrow \dots \rightarrow A \rightarrow B$  we see that when we write  $\underline{\mathbf{x}}$  for  $\mathbf{x}_1 \dots \mathbf{x}_m$ , the IH is:

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \underline{\mathbf{x}} \mathbf{v}))_{i=1}^n, (A_D(\mathbf{x}_i, \mathbf{T}_i \underline{\mathbf{x}} \mathbf{v}))_{i=n+1}^m \vdash B_D(\mathbf{T} \underline{\mathbf{x}}, \mathbf{v}).$$

We write  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ . Closed terms  $\mathbf{S}_1, \dots, \mathbf{S}_n, \mathbf{S}^*, \mathbf{S}$  are required such that

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \mathbf{x}_{n+1} \mathbf{v}))_{i=1}^n \vdash A_D(\mathbf{x}_{n+1}, \mathbf{S}^* \mathbf{x}_{n+1} \mathbf{v}) \rightarrow B_D(\mathbf{S} \mathbf{x}_{n+1}, \mathbf{v}). \quad (3.7)$$

Write  $\underline{\mathbf{x}}'$  for  $\mathbf{x}_1 \dots \mathbf{x}_{m-1}$ . We use the contraction lemma on the IH and get

$$(C_D^i(\mathbf{x}_i, \mathbf{R}_i \underline{\mathbf{x}}' \mathbf{v}))_{i=1}^n, (A_D(\mathbf{x}_i, \mathbf{R}_i \underline{\mathbf{x}}' \mathbf{v}))_{i=n+1}^{m-2} \vdash A_D(\mathbf{x}_{m-1}, \mathbf{R}_{m-1} \underline{\mathbf{x}}' \mathbf{v}) \rightarrow B_D(\mathbf{R} \underline{\mathbf{x}}', \mathbf{v}).$$

Now take  $A_D(\mathbf{x}_{m-1}, \mathbf{R}_{m-1} \underline{\mathbf{x}}' \mathbf{v})$  as an assumption and derive by use of  $\rightarrow E$

$$(C_D^i(\mathbf{x}_i, \mathbf{R}_i \underline{\mathbf{x}}' \mathbf{v}))_{i=1}^n, (A_D(\mathbf{x}_i, \mathbf{R}_i \underline{\mathbf{x}}' \mathbf{v}))_{i=n+1}^{m-1} \vdash B_D(\mathbf{R} \underline{\mathbf{x}}', \mathbf{v}).$$

We are now in a position where we can use the contraction lemma again. This process is repeated  $m - n - 1$  times in total. At the end we get sequences of realising terms,  $\mathbf{S}_1, \dots, \mathbf{S}_n, \mathbf{S}^*, \mathbf{S}$  for (3.7).

**Subcase 4.**  $\rightarrow E$ .

$$\frac{A \quad A \rightarrow B}{B} \rightarrow E,$$

where  $\Gamma \equiv C^1, \dots, C^n$  are the assumptions of  $A$  and  $\Delta \equiv C^{n+1}, \dots, C^m$  are the assumptions of  $A \rightarrow B$ . We omit free variables—the treatment of free variables is essentially the same as under the Hilbert style verification of MP, see page 27. If we write  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ ;  $\tilde{\mathbf{x}}$  for  $\mathbf{x}_{n+1} \dots \mathbf{x}_m$  and  $\underline{\mathbf{x}}$  for  $\mathbf{x}_1 \dots \mathbf{x}_m$  the IH becomes

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x} \mathbf{y}))_{i=1}^n \vdash A_D(\mathbf{T}^* \mathbf{x}, \mathbf{y}) \quad (3.8)$$

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \tilde{\mathbf{x}} \mathbf{v}))_{i=n+1}^m \vdash A_D(\mathbf{x}', \mathbf{T}' \tilde{\mathbf{x}} \mathbf{v}) \rightarrow B_D(\mathbf{T} \tilde{\mathbf{x}} \mathbf{v}, \mathbf{v}), \quad (3.9)$$

and we have to find sequences of closed terms such that

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \underline{\mathbf{x}} \mathbf{v}))_{i=1}^m \vdash B_D(\mathbf{S} \underline{\mathbf{x}}, \mathbf{v}).$$

In (3.8) we substitute  $\mathbf{T}' \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}) \mathbf{v}$  for  $\mathbf{y}$  to and in (3.9) substitute  $\mathbf{T}^* \mathbf{x}$  for  $\mathbf{x}'$ . Then applying  $\rightarrow E$  gives:

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x} (\mathbf{T}' \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}) \mathbf{v})))_{i=1}^n, (C_D^i(\mathbf{x}_i, \mathbf{T}_i \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}) \mathbf{v}))_{i=n+1}^m \vdash B_D(\mathbf{T} \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}), \mathbf{v}).$$

Defining the following sequences will do what is required:

$$\begin{aligned} \mathbf{S}_i &::= \begin{cases} \lambda \underline{\mathbf{x}} \mathbf{v}. \mathbf{T}_i \mathbf{x} (\mathbf{T}' \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}) \mathbf{v}), & \text{if } 1 \leq i \leq n, \\ \lambda \underline{\mathbf{x}} \mathbf{v}. \mathbf{T}_i \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}) \mathbf{v}, & \text{if } n < i \leq m, \end{cases} \\ \mathbf{S} &::= \lambda \underline{\mathbf{x}}. \mathbf{T} \tilde{\mathbf{x}} (\mathbf{T}^* \mathbf{x}). \end{aligned}$$

**Subcase 5.**  $\vee I_1$ .

$$\frac{A}{A \vee B} \vee I_1$$

Omit free variables and say that the assumptions of the derivation of  $A$  are  $C^1, \dots, C^n$ . Write  $\mathbf{x}$  for  $x_1 \dots x_n$ . Then the IH is  $(C_D^i(x_i, T_i \mathbf{x} \mathbf{y}))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{x}, \mathbf{y})$ . From the Dialectica translation of  $C^1 \rightarrow \dots \rightarrow C^n \rightarrow A \vee B$  we see that we have to find sequences  $\mathcal{S}_i, \mathcal{S}^*, \mathcal{S}'$  such that

$$(C_D^i(x_i, \mathcal{S}_i \mathbf{x} \mathbf{y} \mathbf{v}))_{i=1}^n \vdash (\mathcal{S}^* \mathbf{x} =_0 0 \rightarrow A_D(\mathcal{S} \mathbf{x}, \mathbf{y})) \wedge (\mathcal{S}^* \mathbf{x} \neq_0 0 \rightarrow B_D(\mathcal{S}' \mathbf{x}, \mathbf{v})).$$

Read of the type that the terms in the sequence  $\mathcal{S}'$  should have. Let  $\mathbf{o}$  be the corresponding sequence of zero functionals. Now we derive from IH in WE-T<sub>ND</sub>:

$$\frac{\frac{\frac{\text{IH}}{\vdots} \quad A_D(\mathbf{T} \mathbf{x}, \mathbf{y})}{0 = 0 \rightarrow A_D(\mathbf{T} \mathbf{x}, \mathbf{y})} \rightarrow \text{I} \quad \frac{\frac{0 = 0 \quad u : 0 \neq 0}{\perp} \rightarrow \text{E} \quad \frac{\perp}{B_D(\mathbf{o} \mathbf{x}, \mathbf{v})} \perp \text{I}}{0 \neq 0 \rightarrow B_D(\mathbf{o} \mathbf{x}, \mathbf{v})} \rightarrow \text{I}, u}{(0 = 0 \rightarrow A_D(\mathbf{T} \mathbf{x}, \mathbf{y})) \wedge (0 \neq 0 \rightarrow B_D(\mathbf{o} \mathbf{x}, \mathbf{v}))} \wedge \text{I}}$$

From this derivation we see that the following sequences will do what is required:

$$\begin{aligned} \mathcal{S}_i &::= \lambda \mathbf{x} \mathbf{y} \mathbf{v}. T_i \mathbf{x} \mathbf{y}, & \mathcal{S}^* &::= \lambda \mathbf{x}. 0, \\ \mathcal{S} &::= \lambda \mathbf{x}. \mathbf{T} \mathbf{x}, & \mathcal{S}' &::= \mathbf{o}. \end{aligned}$$

Rule  $\vee \text{I}_2$  is similar. The difference is that we derive  $1 = 0 \rightarrow A_D(\mathbf{o} \mathbf{x}, \mathbf{y})$  and define  $\mathcal{S}^*$  by  $\lambda \mathbf{x}. 1$ .

**Subcase 6.**  $\vee \text{E}$ .

$$\frac{A \vee B \quad \frac{[u : A] \quad \vdots \quad C}{C} \quad \frac{[v : B] \quad \vdots \quad C}{C}}{C} \vee \text{E}, u, v$$

For some enumeration, the assumptions of  $A \vee B$  are  $D^1, \dots, D^n$ . In the second sub-tree the assumptions of  $C$  apart from  $[u : A]$  are  $D^{n+1}, \dots, D^{n'}$ , and in the third sub-tree the assumptions of  $C$  are, apart from  $[v : B]$ ,  $D^{n'+1}, \dots, D^m$ . Write  $\mathbf{x}$  for  $x_1 \dots x_n$ ;  $\tilde{\mathbf{x}}$  for  $x_{n+1} \dots x_{n'}$ ;  $\mathbf{x}'$  for  $x_{n'+1} \dots x_m$  and  $\underline{\mathbf{x}}$  for  $x_1 \dots x_m$ . After the use of the contraction lemma the IH becomes

$$(D_D^i(x_i, T_i \mathbf{x} \mathbf{y} \mathbf{v}))_{i=1}^n \vdash (\bar{T} \mathbf{x} =_0 0 \rightarrow A_D(\mathbf{T} \mathbf{x}, \mathbf{y})) \wedge (\bar{T} \mathbf{x} \neq_0 0 \rightarrow B_D(\tilde{T} \mathbf{x}, \mathbf{v})), \quad (3.10)$$

$$(D_D^i(x_i, T_i \tilde{\mathbf{x}} \mathbf{x}_0 \mathbf{q}))_{i=n+1}^{n'} \vdash A_D(\mathbf{x}_0, T' \tilde{\mathbf{x}} \mathbf{x}_0 \mathbf{q}) \rightarrow C_D(\mathbf{T}_0 \tilde{\mathbf{x}} \mathbf{x}_0, \mathbf{q}), \quad (3.11)$$

$$(D_D^i(x_i, T_i \mathbf{x}' \mathbf{u} \mathbf{q}))_{i=n'+1}^m \vdash B_D(\mathbf{u}, \tilde{T}' \mathbf{x}' \mathbf{u} \mathbf{q}) \rightarrow C_D(\tilde{T}'_0 \mathbf{x}' \mathbf{u}, \mathbf{q}). \quad (3.12)$$

We have to find closed terms  $\mathcal{S}_1, \dots, \mathcal{S}_m, \mathcal{S}$  such that

$$(D_D^i(x_i, \mathcal{S}_i \underline{\mathbf{x}} \mathbf{q}))_{i=1}^m \vdash C_D(\mathcal{S} \underline{\mathbf{x}}, \mathbf{q})$$

In (3.10) we substitute  $T'\tilde{x}(Tx)q$  for  $y$  and  $\tilde{T}'x'(\tilde{T}x)q$  for  $v$ ; in (3.11) we substitute  $Tx$  for  $x_0$  and in (3.12) we substitute  $\tilde{T}x$  for  $u$ . This gives

$$(D_D^i(x_i, T_i x(T' \tilde{x}(Tx)q)(\tilde{T}' x'(\tilde{T}x)q)))_{i=1}^n \vdash \\ (\tilde{T}x = 0 \rightarrow A_D(Tx, T' \tilde{x}(Tx)q)) \wedge (\tilde{T}x \neq 0 \rightarrow B_D(\tilde{T}x, \tilde{T}' x'(\tilde{T}x)q)), \quad (3.13)$$

$$(D_D^i(x_i, T_i \tilde{x}(Tx)q))_{i=n+1}^{n'} \vdash A_D(Tx, T' \tilde{x}(Tx)q) \rightarrow C_D(T_0 \tilde{x}(Tx), q), \quad (3.14)$$

$$(D_D^i(x_i, T_i x'(\tilde{T}x)q))_{i=n'+1}^m \vdash B_D(\tilde{T}x, \tilde{T}' x'(\tilde{T}x)q) \rightarrow C_D(\tilde{T}_0 x'(\tilde{T}x), q). \quad (3.15)$$

We use computability of type 0 terms of WE-T<sub>ND</sub> which gives us WE-T<sub>ND</sub>  $\vdash (\tilde{T}x = 0) \vee (\tilde{T}x \neq 0)$ . Therefore either the first or the second antecedent in (3.13) is derivable and we thereby have  $A_D(Tx, T' \tilde{x}(Tx)q)$  for (3.14) or  $B_D(\tilde{T}x, \tilde{T}' x'(\tilde{T}x)q)$  for (3.15). In both cases we are able to conclude  $C_D$ , though  $C_D$  will contain different terms pertaining to the specific case. Terms satisfying the interpretation are therefore

$$\mathcal{S}_i := \begin{cases} \lambda \underline{x} q. T_i x(T' \tilde{x}(Tx)q)(\tilde{T}' x'(\tilde{T}x)q), & \text{if } 1 \leq i \leq n, \\ \lambda \underline{x} q. T_i \tilde{x}(Tx)q, & \text{if } n < i \leq n', \\ \lambda \underline{x} q. T_i x'(\tilde{T}x)q, & \text{if } n' < i \leq m, \end{cases}$$

$$\mathcal{S} := \lambda \underline{x}. \text{Cond}(T_0 \tilde{x}(Tx), \tilde{T}_0 x'(\tilde{T}x), \tilde{T}x).$$

**Subcase 7.**  $\perp_I$ .

$$\frac{\perp}{A} \perp_I$$

Let the assumptions of  $\perp$  be  $C^1, \dots, C^n$  and write  $\mathbf{x}$  for  $x_1 \dots x_n$ . IH is

$$(C_D^i(x_i, T_i \mathbf{x}))_{i=1}^n \vdash \perp.$$

From the translation of  $C^1 \rightarrow \dots \rightarrow C^n \rightarrow A$  we see that we should find sequences such that

$$(C_D^i(x_i, \mathcal{S}_i \mathbf{x} y))_{i=1}^n \vdash A_D(\mathcal{S} \mathbf{x}, y).$$

From this we read of the types that the terms in  $\mathcal{S}$  should have. Let the sequence  $\mathbf{o}$  of zero functionals correspond hereto. Then we derive in WE-T<sub>ND</sub>

$$\frac{(C_D^i(x_i, T_i \mathbf{x}))_{i=1}^n \vdash \perp}{A_D(\mathbf{o} \mathbf{x}, y)} \perp_I$$

Define  $\mathcal{S}_i := \lambda \mathbf{x} y. T_i \mathbf{x}$  and  $\mathcal{S} := \mathbf{o}$ .

**Subcase 8.**  $\exists I$ .

$$\frac{A(\mathbf{a}, t^\sigma)}{\exists z^\sigma A(\mathbf{a}, z)} \exists I$$

We will include free variables. Say the assumptions are  $C^1(\mathbf{c}_1), \dots, C^n(\mathbf{c}_n)$  and that  $\text{FV}(t^\sigma) = \{\mathbf{b}\}$ . Furthermore we write  $\mathbf{c}$  for  $\mathbf{c}_1 \dots \mathbf{c}_n$  and  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ . IH is

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{y}, \mathbf{a}, t),$$

and we should provide sequences of closed terms such that

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{S} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}, \mathbf{S}' \mathbf{c} \mathbf{a} \mathbf{x}).$$

This is done in the following way: Replace in  $\mathbf{b}$  and in  $t^\sigma$  every  $b_j^\tau$  that does not occur in  $\mathbf{c}, \mathbf{a}$  by a corresponding zero-functional  $\mathbf{o}^\tau$ , in other words: if  $b_j^\tau \notin \{\mathbf{a}, \mathbf{c}\}$  then replace  $b_j^\tau$  by  $\mathbf{o}^\tau$ . Let the result be  $\mathbf{b}'$  and  $t'$ . Given these terms we use the usual  $\eta$ -expansion to get the required closed terms:

$$\mathbf{S}_i := \lambda \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}. \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{b}' \mathbf{x} \mathbf{y}, \quad \mathbf{S} := \lambda \mathbf{c} \mathbf{a} \mathbf{x}. \mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b}' \mathbf{x}, \quad \mathbf{S}' := \lambda \mathbf{c} \mathbf{a} \mathbf{x}. t'.$$

**Subcase 9.**  $\exists E$ .

$$\frac{\begin{array}{c} [u : A(b)] \\ \vdots \\ B \end{array}}{\exists z A(z)} \exists E, u$$

where  $b$  is not free in  $B$  nor in assumptions of  $B$  except  $A(b)$ . Assume again some enumeration of all formula occurrences of assumptions, and say the assumptions of  $\exists z A(z)$  are  $C^1, \dots, C^n$  and that the assumptions of  $B$  apart from  $[u : A(b)]$  are  $C^{n+1}, \dots, C^m$ . Write  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ ;  $\tilde{\mathbf{x}}$  for  $\mathbf{x}_{n+1} \dots \mathbf{x}_m$  and  $\underline{\mathbf{x}}$  for  $\mathbf{x}_1 \dots \mathbf{x}_m$ . Suppose that  $k$  is the number of times where  $u : A(b)$  is occurring as an assumption in the right part of the derivation. After the use of the contraction lemma  $k - 1$  times on the (right part of the) IH it becomes:

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x} \mathbf{y}))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{x}, \mathbf{y}, \bar{\mathbf{T}} \mathbf{x}), \quad (3.16)$$

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{b} \tilde{\mathbf{x}} \mathbf{x}_0 \mathbf{v}))_{i=n+1}^m \vdash A_D(\mathbf{x}_0, \mathbf{T}^* \mathbf{b} \tilde{\mathbf{x}} \mathbf{x}_0 \mathbf{v}, b) \rightarrow B_D(\mathbf{T}' \mathbf{b} \tilde{\mathbf{x}} \mathbf{x}_0, \mathbf{v}). \quad (3.17)$$

We have to find terms  $\mathbf{S}_1, \dots, \mathbf{S}_m, \mathbf{S}$  such that:

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \underline{\mathbf{x}} \mathbf{v}))_{i=1}^m \vdash B_D(\mathbf{S} \underline{\mathbf{x}}, \mathbf{v}). \quad (3.18)$$

In (3.16) we substitute  $\mathbf{T}^*(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}) \mathbf{v}$  for  $\mathbf{y}$  and in (3.17) we substitute  $\mathbf{T} \mathbf{x}$  for  $\mathbf{x}_0$  and  $\bar{\mathbf{T}} \mathbf{x}$  for  $b$ . This gives us in WE-T<sub>ND</sub> the following:

$$\begin{aligned} (C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x}(\mathbf{T}^*(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}) \mathbf{v})))_{i=1}^n &\vdash A_D(\mathbf{T} \mathbf{x}, \mathbf{T}^*(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}) \mathbf{v}, \bar{\mathbf{T}} \mathbf{x}), \\ (C_D^i(\mathbf{x}_i, \mathbf{T}_i(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}) \mathbf{v}))_{i=n+1}^m &\vdash A_D(\mathbf{T} \mathbf{x}, \mathbf{T}^*(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}) \mathbf{v}, \bar{\mathbf{T}} \mathbf{x}) \rightarrow \\ &B_D(\mathbf{T}'(\bar{\mathbf{T}} \mathbf{x}) \tilde{\mathbf{x}}(\mathbf{T} \mathbf{x}), \mathbf{v}). \end{aligned}$$

Applying  $\rightarrow$ E to this gives us:

$$(C_D^i(x_i, T_i x(T^*(\bar{T}x)\tilde{x}(Tx)v)))_{i=1}^n, (C_D^i(x_i, T_i(\bar{T}x)\tilde{x}(Tx)v))_{i=n+1}^m \vdash B_D(T'(\bar{T}x)\tilde{x}(Tx), v).$$

We can therefore take as realising sequences for (3.18) the following

$$\mathbf{S}_i := \begin{cases} \lambda \underline{x}v. T_i x(T^*(\bar{T}x)\tilde{x}(Tx)v), & \text{if } 1 \leq i \leq n, \\ \lambda \underline{x}v. T_i(\bar{T}x)\tilde{x}(Tx)v, & \text{if } n < i \leq m, \end{cases}$$

$$\mathbf{S} := \lambda \underline{x}. T'(\bar{T}x)\tilde{x}(Tx).$$

**Subcase 10.**  $\forall$ I. In  $\text{WE-HA}_{\text{ND}}^{\omega}$  we have

$$\frac{A(\mathbf{a}, b^\sigma)}{\forall z^\sigma A(\mathbf{a}, z)} \forall\text{I}$$

where  $b$  is not free in the assumptions which are  $C^1(\mathbf{c}_1), \dots, C^n(\mathbf{c}_n)$ . It is assumed that  $\text{FV}(C^i) = \{\mathbf{c}_i\}$  and  $\text{FV}(A) = \{\mathbf{a}, b\}$ , i.e. that  $b \notin \{\mathbf{a}\}$ . Write  $\mathbf{c}$  for  $\mathbf{c}_1 \dots \mathbf{c}_n$  and  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ . The IH is

$$(C_D^i(x_i, T_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{y}, \mathbf{a}, b),$$

and from the translation of  $C^1 \rightarrow \dots \rightarrow C^n \rightarrow \forall z A(z)$  we see that we have to find sequences  $\mathbf{S}_i, \mathbf{S}$  such that

$$(C_D^i(x_i, \mathbf{S}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y} \mathbf{z}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{S} \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{z}, \mathbf{y}, \mathbf{a}, z).$$

This is done simply by taking

$$\mathbf{S}_i := \lambda \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y} \mathbf{b}. T_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y}, \quad \mathbf{S} := \lambda \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{b}. \mathbf{T} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}.$$

Not surprisingly we see that free variables not free in assumptions are essentially treated *as if* they were universally quantified. Thus, term extraction treats free variables as if they would be universally quantified further down in the proof.

**Subcase 11.**  $\forall$ E.

$$\frac{\forall z^\sigma A(\mathbf{a}, z)}{A(\mathbf{a}, t^\sigma)} \forall\text{E}$$

Say, the assumptions equal  $C^1(\mathbf{c}_1), \dots, C^n(\mathbf{c}_n)$  and that

$$\text{FV}(C^i) = \{\mathbf{c}_i\}, \text{FV}(\forall z A(z)) = \{\mathbf{a}\} \text{ and } \text{FV}(t^\sigma) = \{\mathbf{b}\}.$$

In accordance with the usual abbreviations we write  $\mathbf{x}$  for  $\mathbf{x}_1 \dots \mathbf{x}_n$ ;  $\mathbf{c}$  for  $\mathbf{c}_1 \dots \mathbf{c}_n$ . The IH is

$$(C_D^i(x_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y} \mathbf{z}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{z}, \mathbf{y}, \mathbf{a}, z), \quad (3.19)$$

and we have to find sequences such that

$$(C_D^i(\mathbf{x}_i, \mathbf{S}_i \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{S} \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}, \mathbf{y}, \mathbf{a}, t).$$

Replacing the free variable  $z^\sigma$  by  $t^\sigma$  everywhere in (3.19) and  $\eta$ -expansion (that will also close with respect to the free variables  $\mathbf{b}$  in  $t$ ) gives the required sequences:

$$\mathbf{S}_i := \lambda \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{y}. \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y} t, \quad \mathbf{S} := \lambda \mathbf{c} \mathbf{a} \mathbf{b} \mathbf{x}. \mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x} t.$$

**Subcase 12.** The last rule of the derivation is the induction rule, Ind.

$$\frac{A(0^0) \quad \begin{array}{c} [u : A(b^0)], \\ \vdots \\ A(Sb^0) \end{array}}{A(z^0)} \text{Ind}, u$$

$z^0$  not free in any assumptions. The verification is considerably easier if we verify the following rule

$$\frac{\begin{array}{c} \emptyset \\ \vdots \\ A(0^0) \end{array} \quad \begin{array}{c} \emptyset \\ \vdots \\ A(b^0) \rightarrow A(Sb^0) \end{array}}{A(x^0)} \text{Ind}'$$

The rule is over natural deduction deductively equivalent to Ind. We see the equivalence by applying Ind' to  $B(y) \equiv A(0) \wedge \forall x(A(x) \rightarrow A(Sx)) \rightarrow A(y)$  which gives

$$A(0) \wedge \forall x(A(x) \rightarrow A(Sx)) \rightarrow \forall x A(x).$$

Assume therefore as IH

$$\begin{aligned} \text{WE-T}_{\text{ND}} \vdash A_D(\mathbf{T}_1 \mathbf{a}, \tilde{\mathbf{y}}, 0, \mathbf{a}), \\ \text{WE-T}_{\text{ND}} \vdash A_D(\mathbf{x}, \mathbf{T}_2 \mathbf{b} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{b}, \mathbf{a}) \rightarrow A_D(\mathbf{T}_3 \mathbf{b} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{S} \mathbf{b}, \mathbf{a}). \end{aligned}$$

We have to find sequences of terms  $\mathbf{S}$  such that

$$\text{WE-T}_{\text{ND}} \vdash A_D(\mathbf{S} \mathbf{b} \mathbf{a}, \mathbf{y}, \mathbf{b}, \mathbf{a}).$$

By simultaneous primitive recursion in higher types we define  $\mathbf{T}$  such that  $\mathbf{T} \mathbf{a} 0 = \mathbf{T}_1 \mathbf{a}$  and  $\mathbf{T} \mathbf{a} (\mathbf{S} \mathbf{b}) = \mathbf{T}_3 \mathbf{b} \mathbf{a} (\mathbf{T} \mathbf{a} \mathbf{b})$ . Then substitution of  $\mathbf{T} \mathbf{a} \mathbf{b}$  for  $\mathbf{x}$  and  $\mathbf{y}$  for  $\tilde{\mathbf{y}}$  give

$$\begin{aligned} A_D(\mathbf{T} \mathbf{a} 0, \mathbf{y}, 0, \mathbf{a}), \\ A_D(\mathbf{T} \mathbf{a} \mathbf{b}, \mathbf{T}_2 \mathbf{b} \mathbf{a} (\mathbf{T} \mathbf{a} \mathbf{b}) \mathbf{y}, \mathbf{b}, \mathbf{a}) \rightarrow A_D(\mathbf{T} \mathbf{a} (\mathbf{S} \mathbf{b}), \mathbf{y}, \mathbf{S} \mathbf{b}, \mathbf{a}). \end{aligned}$$

If we then put  $\tilde{\mathbf{T}} := \lambda \mathbf{b} \mathbf{y}. \mathbf{T}_2 \mathbf{b} \mathbf{a} (\mathbf{T} \mathbf{a} \mathbf{b}) \mathbf{y}$  it follows by the contraction lemma 3.4.3 that

$$A_D(\mathbf{T} \mathbf{a} \mathbf{b}, \mathbf{y}, \mathbf{b}, \mathbf{a})$$

is derivable in  $\text{WE-T}_{\text{ND}}$ . Define therefore  $\mathbf{S} \equiv \lambda b a. \mathbf{T} a b$  and we have, that

$$A_D(\mathbf{S} b a, \mathbf{y}, b, a)$$

is derivable without assumptions in  $\text{WE-T}_{\text{ND}}$ .

Left in the soundness proof of the D-translation are the rules for equality, the rule of weak extensionality (QF-ER), the rules for successor symbol, defining equations for the combinators  $\Pi_{\rho, \sigma}$ ,  $\Sigma_{\rho, \tau, \sigma}$ , and for recursion  $\mathbf{R}_{\sigma}$ . Essentially nothing happens when one Dialectica translates these formulas – the formulas are either quantifier free or universally closed – and the translations are therefore equivalent to the non-translated formulas.  $\dashv$

### 3.5 Relation between the Dialectica interpretation and the Diller-Nahm interpretation

J. Diller and W. Nahm (1974) gave a variant of Gödel's Dialectica interpretation where one does not need decidability of prime formulas. This decidability was needed for instance in the proof of the contraction lemma 3.4.2. In this respect the Diller-Nahm variant can be seen as a generalisation of Dialectica, but in some respects it is *another* interpretation. This should be clear from the following discussion.

Central to the Diller-Nahm translation<sup>4</sup> is that one translates from formulas of  $\text{WE-HA}^{\omega}$  into some fragment which contains the quantifier free formulas and formulas with bounded universal quantifiers. We therefore introduce bounded universal quantifiers for type 0 as a primitive notion. The following axioms define the syntax of bounded universal quantification, where all  $x$ 's and  $t$ 's are of type 0:

- |   |   |
|---|---|
| (i) $(\forall x < 0)A(x)$                     | (ii) $(\forall x < St)A(x) \rightarrow (\forall x < t)A(x)$               |
| (iii) $(\forall x < St)A(x) \rightarrow A(t)$ | (iv) $(\forall x < t)A(x) \wedge A(t) \rightarrow (\forall x < St)A(x)$ . |

In (iii) and (iv) we require that  $t$  is free for  $x$  in  $A$ , i.e. no free variables of  $t$  are quantified by the substitution. Normally, bounded quantification is introduced as a defined notion, but we have extended the language of  $\text{WE-HA}^{\omega}$  and the theory itself. However, the following equivalence

$$(\forall x < t)A(x) \leftrightarrow \forall x(x < t \rightarrow A(x))$$

is provable in the extended theory, thus showing that the bounded quantifier is definable in  $\text{WE-HA}^{\omega}$ . We will therefore not make any distinction between  $\text{WE-HA}^{\omega}$  and  $\text{WE-HA}^{\omega}$  extended by the axioms for the bounded quantifier.

The Diller-Nahm translation assigns to every formula  $A$  a formula  $A^{\wedge} \equiv \exists \mathbf{x} \forall \mathbf{y} A_{\wedge}(\mathbf{x}, \mathbf{y})$  where  $A_{\wedge}$  may contain bounded universal quantifiers and formulas from the quantifier free fragment. The inductive clauses for the definition of Diller-Nahm are the same as the Dialectica translation except for ' $\rightarrow$ '. The translation of bounded universal quantification and implication is defined by:

$$\begin{aligned} ((\forall z^0 < t)A(z))^{\wedge} &::= \exists \mathbf{X} \forall \mathbf{y} (\forall z^0 < t) A_{\wedge}(\mathbf{X}z, \mathbf{y}, z), \\ (A \rightarrow B)^{\wedge} &::= \exists \mathbf{W}, \mathbf{U}, \mathbf{Y} \forall \mathbf{x}, \mathbf{v} ((\forall w^0 < W \mathbf{x} \mathbf{v}) A_{\wedge}(\mathbf{x}, \mathbf{Y} w \mathbf{x} \mathbf{v}) \rightarrow B_{\wedge}(\mathbf{U} \mathbf{x}, \mathbf{v})). \end{aligned}$$

<sup>4</sup>For a short description of the Diller-Nahm variant see (Troelstra, 1973, 243–245).

Note, that if we require  $W\mathbf{xv} = 1$ , this would result in an interpretation equivalent to the Dialectica interpretation.

Diller gives a description of the idea underlying the translation of  $A \rightarrow B$ :

Given a deduction  $\Pi$  of a prenex formula  $\exists y\forall zB$  from an assumption  $\exists v\forall wA$ , we should be able to construct from  $\Pi$ , for any given  $v$ , an object  $y$  such that, for any  $z$ , every path in  $\Pi$  leading upwards to an occurrence of the assumption  $\exists v\forall wA$  gives us an object  $w$  which in this path is the reason for  $A$  to imply  $B$ ; as different paths in  $\Pi$  may produce different objects  $w$ , the deduction  $\Pi$  as a whole only gives us a set  $W$  of objects (reasons)  $w$  such that  $\Pi$  contains a proof of  $B$  from assumptions  $A$  for all  $w \in W$ .<sup>5</sup> (Diller, 1979, 149)

Diller says, in other words, that given the deduction  $\Pi$  of  $\exists y\forall zB$  from an assumption  $\exists v\forall wA$  that can occur as assumption at many places we collect all the witnesses for  $w$ ; then we discharge  $\exists v\forall wA$  and derive  $(A \rightarrow B)^\wedge$ . Note that Dialectica chooses all the time among the witnesses whereas Diller-Nahm collects them. When we look at contraction we see that the Diller-Nahm variant is an interpretation different from Dialectica: Two students of Diller – P. Rath and M. Stein, have in their theses given a Diller-Nahm interpretation of Heyting’s arithmetic formulated in natural deduction, where the language also contains set theoretical symbols, (Rath, 1978; Stein, 1977). But since the Diller-Nahm interpretation chooses a witness at the very end among a set of witnesses one is not in need of the contraction lemma. So at the very end one picks *globally* a witness by brute force, whereas the Dialectica chooses a witness *locally* all the way down the proof. From a methodological point of view this has important consequences when one wants to optimize a realiser: maybe local choices are preferable rather than one global choice. However, with respect to the principal epistemological matter whether there is a realiser or not, this has no influence.

### 3.6 Intuitionistic linear logic and the Dialectica interpretation

Girard (1987) presented linear logic as a refinement of classical logic. ‘Refinement’ should here be understood in the sense that the standard connectives are decomposed into more simple connectives but still, one can faithfully translate classical logic into linear logic such that derivability is preserved. Here we will consider only a fragment of the intuitionistic version of linear logic. Intuitionistic linear logic is denoted ILL.

Central to linear logic is that formulas are thought of as types of information—not as propositions. And each formula occurrence of, say  $A$  is a piece of information of type  $A$ . This implies that one occurrence of  $A$  is not equivalent to several occurrences of  $A$ . Naturally, we want to keep track of use of information and linear logic is designed to do this; in this respect linear logic can be seen as a ‘resource logic’. Since one occurrence of  $A$  is not equivalent to several occurrences of  $A$ , contraction and weakening are not allowed unrestricted in linear logic. Therefore, in contrast to the standard logics, all different formula occurrences will have different labels.

---

<sup>5</sup>This explanation makes good sense when we contrast it with the discussion of  $(\rightarrow^D)$  on page 24; the reason why we should be able to go upwards and produce an object  $w$  should be seen in connection with the interpretation of  $\forall xC_{\text{qf}}(x) \rightarrow \forall yD_{\text{qf}}(y)$ : from a counterexample to  $D_{\text{qf}}$  we can produce a counterexample to  $C_{\text{qf}}$ .

The notation for the natural deduction version of ILL is a little easier when we use the context version instead of the formula-tree version, which we used for the presentation of WE-HA<sub>ND</sub><sup>o</sup>.<sup>6</sup> Now the hypothesis are multisets and the structural rule ‘exchange’ (which is a rule we only need because of the context notation) is therefore implicitly a rule of linear logic. Linear implication,  $\multimap$ , is introduced and eliminated in the following way:

$$\frac{\Gamma, u:A \vdash B}{\Gamma \vdash A \multimap B} \multimap\text{I}, u \quad \frac{\Gamma \vdash A \quad \Delta \vdash A \multimap B}{\Gamma, \Delta \vdash B} \multimap\text{E}$$

In  $\multimap\text{I}$  we discharge one and only one occurrence of  $A$ .

The missing contraction and weakening give rise to two different notions of conjunction. Firstly, the multiplicative conjunction called ‘times’:

$$\frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash A \otimes B} \otimes\text{I} \quad \frac{\Gamma \vdash A \otimes B \quad \Delta, u:A, v:B \vdash C}{\Gamma, \Delta \vdash C} \otimes\text{E}, u, v$$

And secondly, the additive conjunction, ‘with’:

$$\frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \& B} \&\text{I} \quad \frac{\Gamma \vdash A_1 \& A_2}{\Gamma \vdash A_i} \&\text{E}_i, \quad i \in \{1, 2\}$$

Notice that the resources (or hypotheses) for obtaining  $A$  respectively  $B$  in the introduction rule for  $\&$  have to be exactly the same amount of the same formulas. This rule therefore says, that if the resources  $\Gamma$  allow us to infer  $A$  and if  $\Gamma$  also allow us to infer  $B$  then we can from one copy of  $\Gamma$  infer  $A \& B$ ; we have in other words to make an ‘internal choice’ pertaining to which one we actually want. We see that if contraction and weakening are allowed then the rules for  $\otimes$  and  $\&$  collapse to the usual rules for  $\wedge$ .

There are also two different connectives of disjunction; but here we will only consider one of them:  $\oplus$  (‘plus’):

$$\frac{\Gamma \vdash A_i}{\Gamma \vdash A_1 \oplus A_2} \oplus\text{I}, i \in \{1, 2\} \quad \frac{\Gamma \vdash A \oplus B \quad \Delta, u:A \vdash C \quad \Delta, v:B \vdash C}{\Gamma, \Delta \vdash C} \oplus\text{E}, u, v$$

Note that there is only one multiset of resources  $\Delta$  that establishes  $C$  from  $A$  or from  $B$ . This is so because only one such  $\Delta$  of resources will actually be used in establishing  $C$ , depending on how  $A \oplus B$  is established.

Now contraction is re-introduced in linear logic via the bang-operator “!”. One motivation for this is that it allows for an embedding of (standard) intuitionistic logic into ILL; see (Troelstra & Schwichtenberg, 1996, 242). Another reason is that in many circumstances contraction makes sense, e.g. when we know it is possible to establish a formula without any hypotheses (resources). We therefore introduce the new symbol ! (pronounced ‘bang’ or ‘of course’) which is an operator saying that the following formula can be used as often as one

<sup>6</sup>Proofs are therefore written as trees where the nodes are labelled with sequents (context notation) all the way down. Alternatively we could have written the formulas with proof terms in the sense of Curry-Howard. This would have the effect that a whole proof would be contained in one (compact) line.

wants. One can think of  $!$  in very much the same terms as  $\Box$  in the modal logic S4. Apparently there are different formulations of rules concerning this operator (see (Troelstra, 1995)); the specific formulation is not that interesting in our context. It is, however, important that with the modality  $!$  one can re-introduce weakening and contraction in a controlled way:

$$\frac{\Gamma \vdash !B \quad \Delta \vdash A}{\Gamma, \Delta \vdash A} \text{W} \quad \frac{\Gamma \vdash !B \quad u: !B, v: !B, \Delta \vdash A}{\Gamma, \Delta \vdash A} \text{C}, u, v$$

Notice that by specializing  $\Gamma \equiv !B$  and  $\Delta \equiv \Gamma$  we have

$$\frac{!B \vdash !B \quad \Gamma \vdash A}{\Gamma, !B \vdash A} \text{W}$$

since  $!B \vdash !B$  is an axiom.

The fragment of ILL that we now consider is the fragment of provable formulas built up from  $\otimes$ ,  $\oplus$ ,  $\multimap$ ,  $\exists$ ,  $\forall$  and  $!$  (quantifiers are in ILL treated in precisely the same way as in the standard logics). It is now easy to define a translation from this fragment into  $\mathcal{L}(WE\text{-}T)$  such that we get a functional interpretation. First we translate from our fragment of ILL under consideration into  $\mathcal{L}(WE\text{-}HA^\omega)$ . This is straight forward—the linear connectives go to their standard counterparts, and one just strips  $!$  off the formulas. It is now immediate that derivability in ILL implies derivability in  $WE\text{-}HA^\omega$ . Now we apply the Dialectica interpretation and we get an interpretation of the fragment.

The point about this long remark is that the only time we use the contraction lemma 3.4.2 is in the case of rule C of ILL. This shows that the contraction lemma is a resource lemma—it handles the witnesses for the existential quantifiers and when we apply the lemma we choose between different realisers. This also shows that if we do not have (or use) the contraction lemma then we can only interpret the pure part of the fragment of ILL (by pure is meant ILL without the Bang-operator). So, by the contraction lemma we separate linear logic from the standard logic.

### 3.7 Interpretation theorem for $WE\text{-}HA^\omega + MP^\omega + IP_{\forall}^\omega + AC$

We will now see that the Dialectica interpretation interprets the principles  $MP^\omega$ ,  $IP_{\forall}^\omega$  and AC. This displays one of the attractive features about Dialectica, namely that  $MP^\omega$  is validated.

A natural deduction formulation of the principles  $MP^\omega$ ,  $IP_{\forall}^\omega$  and AC consists of the rules:

$$\frac{\neg\neg\exists x^\sigma A_{\text{qf}}(x)}{\exists x^\sigma A_{\text{qf}}(x)} MP_{\text{R}}^\omega \quad \frac{\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow \exists y^\sigma B(y)}{\exists y^\sigma (\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow B(y))} IP_{\forall \text{R}}^\omega$$

$$\frac{\forall x^\sigma \exists y^\tau A(x, y)}{\exists Y^{\sigma\tau} \forall x^\sigma A(x, Yx)} AC_{\text{R}}$$

Regarding  $IP_{\forall \text{R}}^\omega$  there is the usual restriction that  $y^\sigma \notin \text{FV}(A_{\text{qf}}) \setminus \{\mathbf{x}\}$ , but there are no restrictions on the rules with respect to assumptions (as usual,  $\text{qf}$  means that the formulas are quantifier free). For the proof of the following theorem it will, nevertheless, be easier if we

work with the usual axiom schemes, which over  $\text{WE-HA}_{\text{ND}}^{\omega}$  are deductively equivalent to the rules. Let

$$\begin{aligned} \text{MP}^{\sigma} : & \quad \neg\neg\exists x^{\sigma}A_{\text{qf}}(x) \rightarrow \exists x^{\sigma}A_{\text{qf}}(x), \\ \text{IP}_{\forall}^{\omega} : & \quad (\forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow \exists y^{\sigma}B(y)) \rightarrow \exists y^{\sigma}(\forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow B(y)), \quad y \notin \text{FV}(A_{\text{qf}}) \setminus \{\mathbf{x}\}, \\ \text{AC}^{\sigma\tau} : & \quad \forall x^{\sigma}\exists y^{\tau}A(x, y) \rightarrow \exists Y^{\sigma\tau}\forall x^{\sigma}A(x, Yx). \end{aligned}$$

$\text{MP}^{\omega}$ ,  $\text{IP}_{\forall}^{\omega}$  and  $\text{AC}$  are the collections of these schemes, respectively, for all types. We can now formulate the extended interpretation theorem, which was first proved by Yasugi (1963).

**Theorem 3.7.1.** *If  $(C^i(\mathbf{c}_i))_{i=1}^n \vdash A(\mathbf{a})$  in  $\text{WE-HA}_{\text{ND}}^{\omega} + \text{MP}^{\omega} + \text{IP}_{\forall}^{\omega} + \text{AC}$ , then*

$$(C_D^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{c} \mathbf{a} \mathbf{x} \mathbf{y}, \mathbf{c}_i))_{i=1}^n \vdash A_D(\mathbf{T} \mathbf{c} \mathbf{a} \mathbf{x}, \mathbf{y}, \mathbf{a}),$$

in  $\text{WE-T}_{\text{ND}}$  for certain sequences of closed terms  $\mathbf{T}_1, \dots, \mathbf{T}_n, \mathbf{T}$  which can be extracted from a derivation of  $A(\mathbf{a})$ , where  $\mathbf{c} \equiv \mathbf{c}_1 \dots \mathbf{c}_n$  and  $\mathbf{x} \equiv \mathbf{x}_1 \dots \mathbf{x}_n$ .

**Proof.**

**Case 1.**  $\text{MP}^{\omega}$ . We can assume the quantifier free  $A_{\text{qf}}$  to be a formula without disjunction, since if it contained  $\vee$  it would be equivalent to a quantifier free formula without  $\vee$ , say  $t\mathbf{x} =_0 0$ , where  $t$  is the characteristic term for  $A_{\text{qf}}$ . Since  $A_{\text{qf}}$  is without  $\vee$  the translation is the identity. The translation of an instance of the principle is therefore

$$\exists Y \forall x (\neg\neg A_{\text{qf}}(x, \mathbf{a}) \rightarrow A_{\text{qf}}(Yx, \mathbf{a})).$$

Let  $T$  equal  $\lambda \mathbf{a}, x. x$ . Then we have

$$\text{WE-T}_{\text{ND}} \vdash \neg\neg A_{\text{qf}}(x, \mathbf{a}) \rightarrow A_{\text{qf}}(T\mathbf{a}x, \mathbf{a}).$$

Thus  $\text{MP}^{\omega}$  is interpretable because quantifier free formulas are stable.

**Case 2.**  $\text{IP}_{\forall}^{\omega}$ . The translation of the antecedent of an instance of  $\text{IP}_{\forall}^{\omega}$  is identical to the translation of the conclusion and the interpretation therefore reduces to  $C \rightarrow C$ . From the translation of an instance of the principle it is seen that we need to provide sequences of terms  $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$  and  $\mathbf{T}_4$  such that

$$\begin{aligned} (A_{\text{qf}}(\mathbf{X}(\mathbf{T}_1 \mathbf{c} \mathbf{X} \mathbf{y} \mathbf{p} \mathbf{q})) \rightarrow B_D(\mathbf{p}, \mathbf{T}_1 \mathbf{c} \mathbf{X} \mathbf{y} \mathbf{p} \mathbf{q}, \mathbf{y})) \rightarrow \\ (A_{\text{qf}}(\mathbf{T}_2 \mathbf{c} \mathbf{X} \mathbf{y} \mathbf{p} \mathbf{q}) \rightarrow B_D(\mathbf{T}_3 \mathbf{c} \mathbf{X} \mathbf{y} \mathbf{p}, \mathbf{q}, \mathbf{T}_4 \mathbf{c} \mathbf{X} \mathbf{y} \mathbf{p})) \end{aligned}$$

is provable in the quantifier free fragment, where  $\text{FV}(\forall \mathbf{x}A_{\text{qf}}(\mathbf{x})) \cup \text{FV}(\exists yB(y)) = \mathbf{c}$ . This is accomplished with

$$\begin{aligned} \mathbf{T}_1 &::= \lambda \mathbf{c}, \mathbf{X}, \mathbf{y}, \mathbf{p}, \mathbf{q}. \mathbf{q}, & \mathbf{T}_2 &::= \lambda \mathbf{c}, \mathbf{X}, \mathbf{y}, \mathbf{p}. \mathbf{X}, \\ \mathbf{T}_3 &::= \lambda \mathbf{c}, \mathbf{X}, \mathbf{y}, \mathbf{p}. \mathbf{p}, & \mathbf{T}_4 &::= \lambda \mathbf{c}, \mathbf{X}, \mathbf{y}, \mathbf{p}. \mathbf{y}. \end{aligned}$$

**Case 3.**  $\text{AC}$ . Analogous to case 2: The translation of the premise of an instance of  $\text{AC}$  is identical to the translation of the conclusion.  $\dashv$

### 3.8 The non-constructive theory $WE-HA^\omega + IP_{\neg\forall}^\omega + MP^\omega$

One could wonder whether or not the Dialectica interpretation of  $IP_{\neg\forall}^\omega$  is optimal; is it the best possible? It is as we shortly will see.

Let us restrict independence-of-premise to the case where the premise is a negated purely universal formula; call this principle  $IP_{\neg\forall}^\omega$ . In natural deduction the principle takes the form:

$$\frac{\neg\forall\mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow \exists yB(y)}{\exists y(\neg\forall\mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow B(y))} IP_{\neg\forall}^\omega,$$

where  $y \notin \text{FV}(A_{\text{qf}}) \setminus \{\mathbf{x}\}$ , and  $A_{\text{qf}}$  is quantifier free. We will see that Dialectica cannot interpret this principle. Note firstly, that

$$\neg\forall x\neg C(x) \rightarrow \neg\neg\exists xC(x), \quad (3.20)$$

is intuitionistically valid for any  $C$ . Let  $T$  be Kleene's  $T$ -predicate.

**Theorem 3.8.1.** *There is a  $\Pi_3^0$  sentence  $\forall x\exists y\forall z(Txxz \rightarrow Txy)$  provable in  $WE-HA^\omega + IP_{\neg\forall}^\omega + MP^\omega$  but no interpretation can provide a computable witness for the existential quantifier.*

**Proof.** Given  $IP_{\neg\forall}^\omega$  we would have the following derivation in  $WE-HA_{\text{ND}}^\omega + MP_{\text{R}}^\omega + IP_{\neg\forall}^\omega$  which has no assumptions:

$$\begin{array}{c} \frac{u : \neg\neg\exists zTxxz}{\exists zTxxz} MP_{\text{R}}^\omega \\ \hline \neg\neg\exists zTxxz \rightarrow \exists zTxxz \quad \rightarrow\text{I}, u \\ \hline \neg\forall z\neg Txxz \rightarrow \exists zTxxz \quad (3.20) \\ \hline \exists y(\neg\forall z\neg Txxz \rightarrow Txy) \quad IP_{\neg\forall}^\omega \end{array} \quad (3.21)$$

Since  $(\neg\forall x\neg B(x) \rightarrow C) \rightarrow \forall x(B(x) \rightarrow C)$  is intuitionistically valid for any  $B$  and  $C$  where  $x \notin \text{FV}(C)$  it follows that

$$\forall x\exists y\forall z(Txxz \rightarrow Txy).$$

If an interpretation could provide a computable witness  $f^1$  for  $\exists y$  then

$$\forall z(Txxz \rightarrow Txx(f^1x)),$$

and hence by logic

$$\exists zTxxz \rightarrow Txx(f^1x).$$

Since  $Txx(f^1x)$  is decidable given  $x$  this would solve the halting problem: if  $Txx(f^1x)$  then  $f^1x$  would witness  $\exists zTxxz$ ; but also the other way around: if  $\neg Txx(f^1x)$  then  $\neg\exists zTxxz$ .  $\neg$

It is, of course, also the case that when independence-of-premise is restricted to purely existential statements, then this principle is also not interpretable: From  $\exists y T xxy \rightarrow \exists y T xxy$  one could immediately derive a contradiction given such a principle.

We see that  $MP^\omega$  and  $IP_{\neg\forall}^\omega$  are incompatible with respect to computational content; that  $WE-HA^\omega + IP_{\neg\forall}^\omega + MP^\omega$  is a non-constructive theory. In the context of functional interpretation it shows that Dialectica cannot interpret  $IP_{\neg\forall}^\omega$ , since it interprets  $MP^\omega$ . Note, that in the above proof we only used  $MP^0$  and  $IP_{\neg\forall}^0$ ; thus the result actually applies already to HA together with these principles formulated in the language of HA. In chapter 4 we will define HA properly as a subsystem of  $WE-HA^\omega$ .

On the other hand we will see later on, that a much stronger version of independence-of-premise where the premise does not contain  $\exists$  nor  $\forall$  is interpretable by modified realisability. Using a variant of modified realisability this leads to a proof of the fact that  $WE-HA^\omega$  alone is closed under a corresponding rule. However, the argument above shows that this closure property breaks down if  $MP^\omega$  is added:

**Theorem 3.8.2.** *There are quantifier free formulas  $A_{\text{qf}}$  and  $B_{\text{qf}}$  such that*

$$WE-HA^\omega + MP^\omega \vdash \neg\forall x^0 A_{\text{qf}}(x) \rightarrow \exists y^0 B_{\text{qf}}(y),$$

but

$$WE-HA^\omega + MP^\omega + IP_{\forall}^\omega + AC \not\vdash \exists y^0 (\neg\forall x^0 A_{\text{qf}}(x) \rightarrow B_{\text{qf}}(y)).$$

**Proof.** From (3.21) we see that  $WE-HA^\omega + MP^\omega \vdash \neg\forall x \neg T z z x \rightarrow \exists y T z z y$ . Now, assume that

$$WE-HA^\omega + MP^\omega + IP_{\forall}^\omega + AC \vdash \exists y (\neg\forall x \neg T z z x \rightarrow T z z y).$$

By theorem 3.7.1 it then follows that there exists a closed term  $f^1$  of WE-T such that

$$WE-HA^\omega \vdash \forall x (T z z x \rightarrow T z z (fz)),$$

which is absurd. Take therefore  $A_{\text{qf}}(x^0) \equiv \neg T z z x$  and  $B_{\text{qf}}(y^0) \equiv T z z y$ . ⊥

This displays the computational conflict between Markov's principle and independence-of-premise. But it also displays that two different interpretations, modified realisability and functional interpretation, give different views on what 'constructivity' is, and one will have to choose between the different interpretations in order to get a coherent view of what a 'constructive method' is. This is a theme we will come back to later on.

## Classical Arithmetic and Kuroda’s Negative Translation

We will now turn to classical arithmetic. Mathematically there are many reasons why we should study classical logic and arithmetic. One reason, above a lot of others, is that classical logic is *the* logic used by virtually every mathematician. This is probably so mostly due to the fact that classical logic provides short and often elegant proofs. The method of indirect proof is the essence of classical logic—if one works intuitionistically but suddenly allows for indirect proofs, one will get classical logic. As was discussed in chapter 1 Gödel saw his Dialectica interpretation as a continuation of Hilbert’s program, but generally also as a contribution to the discussions on constructivity and the foundations of mathematics. The Dialectica interpretation carries a lot of benefits in the context of classical arithmetic: Together with a negative translation it makes a characterisation of the class of provably total recursive functions possible, and as to Hilbert’s thoughts on a justification of classical logic it provides, among other results, the following: (i) With respect to a certain important class of formulas it is shown that the quantifier free type theory is just as good as classical arithmetic (the reflection principle, see (1.1) on page 4), and (ii) the interpretation makes it possible to show that classical arithmetic is consistent—or at least consistent relative to the quantifier free theory. This was the property Gödel referred to in the title “Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes”.

### 4.1 Formulation of WE-PA<sup>ω</sup>

**Definition 4.1.1.** Classical arithmetic in all finite types with weak extensionality based on natural deduction, WE-PA<sub>ND</sub><sup>ω</sup>, arises when we exchange in WE-HA<sub>ND</sub><sup>ω</sup> the intuitionistic  $\perp_I$  rule with the classical rule:

$$\frac{[u : \neg A] \quad \vdots \quad \perp}{A} \perp_C, u$$

◁

The rule  $\perp_C$  formalises in WE-PA<sub>ND</sub><sup>ω</sup> the method of indirect proof. If we want to show  $A$  we assume its negation,  $\neg A$ , – possibly among other assumptions – and then show that  $\neg A$  simply is not possible: it will lead to an inconsistency. On the basis of this we are allowed to conclude that  $A$ , indeed, is the case, and to discharge all  $\neg A$ ’s labelled by  $u$ . We see that the intuitionistic rule  $\perp_I$  is nothing but a special case of  $\perp_C$ , since if no occurrences of  $\neg A$  are labelled by  $u$  no formulas are discharged and the rule is precisely  $\perp_I$ . However, if we work intuitionistically and derive  $\perp$  from  $\neg A$  then we can discharge  $\neg A$  by  $\rightarrow I$  and conclude, only,  $\neg\neg A$ .

Another way of getting classical logic from intuitionistic logic is by adjoining “tertium non datur”—the principle of excluded middle. The principle says that every well-formed mathematical statement has a determinate truth value: It is either true or false, independently

of whether that truth value is accessible to us. The principle is formalised by the schema:

$$\text{TND:} \quad A \vee \neg A.$$

The method of indirect proof implies formally the principle of the excluded middle:

$$\frac{\frac{\frac{u : A}{A \vee \neg A} \vee I_1 \quad v : \neg(A \vee \neg A)}{\perp} \rightarrow E}{\frac{\frac{\perp}{\neg A} \rightarrow I, u}{A \vee \neg A} \vee I_2 \quad v : \neg(A \vee \neg A)}{\perp} \rightarrow E}{\frac{\perp}{A \vee \neg A} \perp_C, v} \rightarrow E$$

The implication also goes in the other direction. Using intuitionistic logic we get  $\perp_C$  as a derived rule from TND:

$$\frac{A \vee \neg A \quad u : A \quad \frac{[v : \neg A] \quad \vdots \quad \perp}{A} \perp_I}{A} \vee E, u, v$$

**Definition 4.1.2.** Classical arithmetic in all finite types with weak extensionality in Hilbert style,  $\text{WE-PA}_H^\omega$  is  $\text{WE-HA}_H^\omega$  plus  $(A \vee \neg A)$  for any formula  $A$ .  $\triangleleft$

When it is of no importance whether we work with  $\text{WE-PA}_{ND}^\omega$  or with  $\text{WE-PA}_H^\omega$  we simply write  $\text{WE-PA}^\omega$ .

## 4.2 From classical to intuitionistic logic: Kuroda's negative translation

There are different ways of reducing classical logic and arithmetic to their intuitionistic counterparts. We will here use one of the so-called negative translations. A negative translation takes a formula of a classical system and translates it into a formula of an intuitionistic system in such a way that provability is preserved. In our case we will translate every formula  $A$  of  $\mathcal{L}(\text{WE-PA}^\omega)$  into a formula  $A'$  of  $\mathcal{L}(\text{WE-HA}^\omega)$  such that  $\text{WE-PA}^\omega \vdash A$  implies  $\text{WE-HA}^\omega \vdash A'$ . One immediate consequence of this and the fact that  $\text{WE-HA}^\omega \vdash (0 = 1)' \leftrightarrow (0 = 1)$  is a (constructive) proof of the consistency of the classical system from the consistency of the intuitionistic system. But there are, as we shall see, many other interesting consequences of negative translations.

There are at least four different kinds of negative translations. The first ones were discovered independently by Kolmogorov (1925), Gentzen (1933) and Gödel (1933). We will in the following work with a negative translation due to Kuroda (1951).<sup>1</sup>

**Definition 4.2.1.** (Kuroda's negative translation). Let  $A$  be a formula of  $\mathcal{L}(\text{WE-PA}^\omega)$ . Then the translation of  $A$  is  $A' := \neg\neg A^*$ , where  $A^*$  is defined inductively as follows:

$$\begin{aligned} (P^*) \quad A^* &::= A, \text{ if } A \text{ is prime,} \\ (\wedge^*) \quad (A \wedge B)^* &::= A^* \wedge B^*, \\ (\vee^*) \quad (A \vee B)^* &::= A^* \vee B^*, \\ (\rightarrow^*) \quad (A \rightarrow B)^* &::= A^* \rightarrow B^*, \\ (\exists^*) \quad (\exists x A(x))^* &::= \exists x (A(x))^*, \\ (\forall^*) \quad (\forall x A(x))^* &::= \forall x \neg\neg (A(x))^*. \end{aligned}$$

◁

**Definition 4.2.2.** We say that a formula is *negative* if it is built up by negated prime formulas using only the symbols  $\wedge$ ,  $\rightarrow$ , and  $\forall$ . ▷

Since  $\text{WE-HA}^\omega \vdash P \leftrightarrow \neg\neg P$ , for any prime formula  $P$ , it can be proved by induction of the complexity of  $A$  that  $A'$  is equivalent within  $\text{WE-HA}^\omega$  to a negative formula, see (Luckhardt, 1973, 43). In fact this is the core of a negative translation: Essentially it translates into the negative fragment.

**Theorem 4.2.3.** For formulas  $C_1, \dots, C_n, A$  of  $\mathcal{L}(\text{WE-PA}^\omega)$  it is the case that

$$\text{if } C_1, \dots, C_n \vdash A \text{ in } \text{WE-PA}_{\text{ND}}^\omega \text{ then } C'_1, \dots, C'_n \vdash A' \text{ in } \text{WE-HA}_{\text{ND}}^\omega.$$

It could be noted that for the following proof we are not in need of the intuitionistic lemma  $A \leftrightarrow \neg\neg A$  for negative  $A$ . However, one needs this lemma when proving the corresponding theorem for the Gödel-Gentzen negative translation. But when the Kuroda translation is compared with other translations, e.g. the Gödel-Gentzen translation, it is seen that the Kuroda translation uses negations,  $\neg$ , at a minimum.

**Proof.** The theorem is proved by induction on the length of the proof in  $\text{WE-PA}_{\text{ND}}^\omega$ .<sup>2</sup> In case of the logic we will only verify two of the rules; one simple and one more complex.

**Case 1.**  $\wedge E_1$ . Assume that the last rule used in the classical proof is  $\wedge E_1$ :

$$\frac{A \wedge B}{A} \wedge E_1$$

By induction hypothesis we have an intuitionistic derivation of  $\neg\neg(A^* \wedge B^*)$ . The following

<sup>1</sup>There is a thorough treatment of the different kinds of negative translations in (Luckhardt, 1973, 41–50). See also (Murthy, 1990) for a discussion of the different negative translations in the context of extracting computational content from classical proofs. Different translations use double negations in different ways. Since negations raise the type of the extracted programs, different translations give rise to different programs.

<sup>2</sup>See (Luckhardt, 1973, 44) for the proof within Hilbert style.

derivation verifies the rule:

$$\begin{array}{c}
 \frac{u : A^* \wedge B^*}{A^*} \quad v : \neg A^* \\
 \hline
 \perp \\
 \hline
 \frac{\neg(A^* \wedge B^*) \quad u}{\neg(A^* \wedge B^*)} \\
 \hline
 \frac{\perp}{\neg\neg A^*} v
 \end{array}
 \quad
 \begin{array}{c}
 \text{IH} \\
 \vdots \\
 \neg\neg(A^* \wedge B^*)
 \end{array}$$

**Case 2.**  $\vee E$ . This is the only part of the proof which is a little tricky. We have by IH that  $\Gamma' \vdash \neg\neg(A^* \vee B^*)$  and  $\Delta', [\neg\neg A^*] \vdash \neg\neg C^*$  and  $\Xi', [\neg\neg B^*] \vdash \neg\neg C^*$  using intuitionistic logic where  $\Gamma', \Delta', \Xi', \neg\neg A^*$  and  $\neg\neg B^*$  are the translated assumptions. We have to prove  $\neg\neg C^*$  by intuitionistic logic. The proof is not that simple because we cannot intuitionistically conclude  $(\neg\neg A^* \vee \neg\neg B^*)$  from  $\neg\neg(A^* \vee B^*)$ . The idea is to replace every assumption  $\neg\neg A^*$  and  $\neg\neg B^*$  by a derivation of  $\neg\neg A^*$  and  $\neg\neg B^*$  from  $A^*$  and  $B^*$  respectively and then to work on towards  $\neg\neg C^*$ :

$$\begin{array}{c}
 \frac{v_0 : A^* \quad u : \neg A^*}{\perp} \quad \frac{v_1 : B^* \quad u : \neg B^*}{\perp} \\
 \hline
 \frac{\perp}{[\neg\neg A^*] \quad u} \quad \frac{\perp}{[\neg\neg B^*] \quad u} \\
 \vdots \quad \vdots \\
 \frac{w : A^* \vee B^*}{\neg\neg C^*} \quad v_0, v_1 \quad \tilde{v} : \neg C^* \\
 \hline
 \frac{\neg\neg C^*}{\neg\neg C^*} \\
 \hline
 \frac{\perp}{\neg(A^* \vee B^*)} w \quad \frac{\text{IH}}{\neg\neg(A^* \vee B^*)} \\
 \hline
 \frac{\perp}{\neg\neg C^*} \tilde{v}
 \end{array}$$

The verification of the classical use of  $\exists E$  follows along the same lines.

**Case 3.** Induction rule. The induction rule is verified by itself using that  $\neg\neg(A \rightarrow B) \leftrightarrow (\neg\neg A \rightarrow \neg\neg B)$  holds intuitionistically.

**Case 4.** Equality, arithmetical axioms and rules (except induction). For these it suffices to note that  $\text{WE-HA}^0$  proves

$$(i) A_{\text{qf}} \leftrightarrow \neg\neg A_{\text{qf}} \quad \text{and} \quad (ii) \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \leftrightarrow \neg\neg \forall \mathbf{x} \neg\neg A_{\text{qf}}(\mathbf{x}).$$

The last equivalence is due to

$$\text{WE-HA}^0 \vdash \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \leftrightarrow \forall \mathbf{x} \neg\neg A_{\text{qf}}(\mathbf{x}) \leftrightarrow \neg\neg \forall \mathbf{x} \neg\neg A_{\text{qf}}(\mathbf{x}). \quad (4.1)$$

As an example, take rule QF-ER. The Kuroda translation of the hypothesis is by (ii) equivalent to  $\forall \mathbf{x}(s^\sigma \mathbf{x} =_0 t^\sigma \mathbf{x})$ . Since the translated assumptions are quantifier free we use QF-ER to obtain  $\forall \mathbf{y}(r[s]\mathbf{y} =_0 r[t]\mathbf{y})$  which again is equivalent to its Kuroda translation.  $\dashv$

Theorem 4.2.3 is interesting in itself for many reasons. Firstly, we note that the theorem also holds when  $CL^\omega$  (typed classical predicate logic) is interpreted into  $IL^\omega$ —this is seen directly from the proof. Secondly, by the proof we are given an algorithm which takes any classical proof of a formula and transforms it into an intuitionistic proof of a corresponding formula. Consequently – and independently of whether one is an intuitionist or not – one has a constructive proof of the consistency of the classical logic relative to the intuitionistic.

### 4.3 Interpretation theorem for $WE-PA^\omega + QF-AC$ and consistency of the theory

In the context of functional interpretation a negative translation has important consequences. It makes it possible to interpret classical arithmetic plus quantifier free axiom of choice,  $QF-AC$ . Recall, in natural deduction it has the form:

$$\frac{\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)}{\exists Y^{\sigma\tau} \forall x^\sigma A_{\text{qf}}(x, Yx)} \text{QF-AC}_R,$$

where  $A_{\text{qf}}$  is quantifier free, but there are no restrictions on assumptions.

**Lemma 4.3.1.** For formulas  $C_1, \dots, C_n, A$  of  $\mathcal{L}(WE-PA^\omega)$ , if

$$\begin{array}{l} C_1, \dots, C_n \vdash A \quad \text{in } WE-PA_{ND}^\omega + QF-AC_R, \text{ then} \\ C'_1, \dots, C'_n \vdash A' \quad \text{in } WE-HA_{ND}^\omega + QF-AC_R + MP_R^\omega. \end{array}$$

The lemma is an example of a technique often used in reductive proof theory. First one reduces some classical system to the corresponding intuitionistic system plus Markov's rule or principle. Then, later on by some method or another one interprets this intermediate system in the pure intuitionistic system.

**Proof.** We have by theorem 4.2.3 an interpretation of  $WE-PA_{ND}^\omega$  in  $WE-HA_{ND}^\omega$ . To be shown is that

$$\frac{\neg\neg\forall x^\sigma \neg\neg\exists y^\tau A_{\text{qf}}(x, y)}{\neg\neg\exists Y^{\sigma\tau} \forall x^\sigma \neg\neg A_{\text{qf}}(x, Yx)} \quad (4.2)$$

with no restrictions on assumptions, is a derived rule in  $WE-HA_{ND}^\omega + QF-AC_R + MP_R^\omega$ .

We will need

$$\neg\neg\forall x \neg\neg C(x) \rightarrow \forall x \neg\neg C(x) \quad (4.3)$$

(fairly easy to prove intuitionistically) and use  $C \vdash \neg\neg C$ . The following derivation shows (4.2) to be a derived rule.

$$\begin{array}{c}
\frac{\neg\neg\forall x\neg\neg\exists yA_{\text{qf}}(x,y)}{\forall x\neg\neg\exists yA_{\text{qf}}(x,y)} \quad (4.3) \\
\frac{\forall x\neg\neg\exists yA_{\text{qf}}(x,y)}{\neg\neg\exists yA_{\text{qf}}(x,y)} \quad \forall E \\
\frac{\neg\neg\exists yA_{\text{qf}}(x,y)}{\exists yA_{\text{qf}}(x,y)} \quad \text{MP}_R^\omega \\
\frac{\exists y\neg\neg A_{\text{qf}}(x,y)}{\exists y\neg\neg A_{\text{qf}}(x,y)} \quad \exists I \\
\frac{\exists y\neg\neg A_{\text{qf}}(x,y)}{\forall x\exists y\neg\neg A_{\text{qf}}(x,y)} \quad \forall I \\
\frac{\forall x\exists y\neg\neg A_{\text{qf}}(x,y)}{\exists Y\forall x\neg\neg A_{\text{qf}}(x,Yx)} \quad \text{QF-AC}_R \\
\frac{\exists Y\forall x\neg\neg A_{\text{qf}}(x,Yx)}{\neg\neg\exists Y\forall x\neg\neg A_{\text{qf}}(x,Yx)} \quad \exists E, u
\end{array}$$

⊢

Now we are in a position where we can interpret  $\text{WE-PA}^\omega + \text{QF-AC}$ .

**Theorem 4.3.2.** *If  $(C^i(\mathbf{c}_i))_{i=1}^n \vdash A(\mathbf{a})$  in  $\text{WE-PA}_{\text{ND}}^\omega + \text{QF-AC}_R$ , then*

$$((C^i)_D(\mathbf{x}_i, \mathbf{T}_i \mathbf{cax}_y, \mathbf{c}_i))_{i=1}^n \vdash (A')_D(\mathbf{T} \mathbf{cax}, \mathbf{y}, \mathbf{a}),$$

in  $\text{WE-T}_{\text{ND}}$  for certain sequences of closed terms  $\mathbf{T}_1, \dots, \mathbf{T}_n, \mathbf{T}$  which can be extracted from a derivation of  $A(\mathbf{a})$ , where  $\mathbf{c} \equiv \mathbf{c}_1 \dots \mathbf{c}_n$  and  $\mathbf{x} \equiv \mathbf{x}_1 \dots \mathbf{x}_n$ .

**Proof.** The theorem follows from lemma 4.3.1 and theorem 3.7.1. ⊢

Due to the fact that the theorem is formulated within natural deduction it is a little heavy in notation. But the flexibility with respect to assumptions has for practical use advantages over Hilbert style. If there are no assumptions of the derivation we have the more simple formulation:

$$\text{WE-PA}^\omega + \text{QF-AC} \vdash A(\mathbf{a}) \Rightarrow \text{WE-T} \vdash (A')_D(\mathbf{T} \mathbf{a}, \mathbf{y}, \mathbf{a})$$

for closed extractable sequence  $\mathbf{T}$  of terms.

Our next theorem, a corollary to the forgoing theorem, is an important contribution to a generalised Hilbert program. The interpretations prescribe a purely finitistic and combinatorial way of interpreting  $\text{WE-PA}^\omega + \text{QF-AC}$  in  $\text{WE-T}$ . As such the proof of the next corollary is very close to Hilbert's prescriptions for consistency proofs. see also (Yasugi, 1963, Th. 3).

**Corollary 4.3.3.** *Consistency of  $\text{WE-T}$  implies consistency of  $\text{WE-PA}^\omega + \text{QF-AC}$ .*

**Proof.** If  $\text{WE-PA}^\omega + \text{QF-AC}$  is inconsistent then  $\text{WE-T} \vdash (0 = 1)'$ ; hence  $\text{WE-T} \vdash 0 = 1$ . ⊢

#### 4.4 Extraction theorem and conservativeness

We can now prove the important extraction theorem.

**Theorem 4.4.1.** (Extraction theorem). *Let  $\forall \mathbf{x}_1 C_{\text{qf}}^1(\mathbf{x}_1), \dots, \forall \mathbf{x}_n C_{\text{qf}}^n(\mathbf{x}_n)$  and  $\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$  be sentences of  $\mathcal{L}(\text{WE-PA}^\omega)$ , where  $C_{\text{qf}}^i, A_{\text{qf}}$  are all quantifier free. If*

$$\forall \mathbf{x}_1 C_{\text{qf}}^1(\mathbf{x}_1), \dots, \forall \mathbf{x}_n C_{\text{qf}}^n(\mathbf{x}_n) \vdash \forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$$

is provable in  $WE-PA_{ND}^\omega + QF-AC_R$  then

$$C_{qf}^1(\mathbf{t}_1 x^\sigma), \dots, C_{qf}^n(\mathbf{t}_n x^\sigma) \vdash A_{qf}(x^\sigma, t x^\sigma)$$

is provable in  $WE-T_{ND}$  for closed extractable terms  $\mathbf{t}_1, \dots, \mathbf{t}_n, t^{\sigma\tau}$ .

Compare the theorem with (the simplified) Herbrand's theorem on page 35 and note how this have been generalised.

**Proof.** Strictly speaking the theorem is a corollary of theorem 4.3.2 but if we prove it directly from this theorem the types of the extracted terms would raise unnecessarily in type due to the double negations. We therefore give a slightly different proof in order to keep the types of the terms down.

Suppose we have the derivation in  $WE-PA_{ND}^\omega + QF-AC_R$ . Via lemma 4.3.1 we arrive at the following in  $WE-HA_{ND}^\omega + MP_R^\omega + QF-AC_R$ :

$$\neg\neg\forall\mathbf{x}_1\neg\neg C_{qf}^1(\mathbf{x}_1), \dots, \neg\neg\forall\mathbf{x}_n\neg\neg C_{qf}^n(\mathbf{x}_n) \vdash \neg\neg\forall x\neg\neg\exists y A_{qf}(x, y).$$

The assumptions  $\neg\neg\forall\mathbf{x}_i\neg\neg C_{qf}^i(\mathbf{x}_i)$  are over  $WE-HA^\omega$  equivalent to  $\forall\mathbf{x}_i C_{qf}^i(\mathbf{x}_i)$ , see (4.1). Furthermore, we have the intuitionistic valid formula  $\neg\neg\forall z\neg\neg B(z) \rightarrow \forall z\neg\neg B(z)$ . This used together with an application of  $MP_R^\omega$  gives us

$$\forall\mathbf{x}_1 C_{qf}^1(\mathbf{x}_1), \dots, \forall\mathbf{x}_n C_{qf}^n(\mathbf{x}_n) \vdash \forall x\exists y A_{qf}(x, y).$$

Now we simply use the interpretation theorem for  $WE-HA_{ND}^\omega + MP_R^\omega + QF-AC_R$  (theorem 3.7.1) to get the desired terms.  $\dashv$

We note that the terms  $\mathbf{t}_1, \dots, \mathbf{t}_n$  referred to in the theorem are trivial realisers: They specify which *instances* of the universal assumptions we have used in the proof of the conclusion.

The extraction theorem is connected with an analysis of provably total recursive functionals of  $WE-PA^\omega + QF-AC$ . Now, any quantifier free formula of  $WE-PA^\omega + QF-AC$  is decidable—remember we have already in the intuitionistic theory that  $A_{qf} \vee \neg A_{qf}$  is provable for quantifier free formulas. Thus  $A_{qf}(x^\sigma, y^0)$  as a formula of  $\mathcal{L}(WE-PA^\omega)$  with just  $x$  and  $y$  as free variables defines a partial recursive functional  $\Phi$  of type  $\sigma 0$ :

$$\Phi(x^\sigma) = \begin{cases} \min y^0(A_{qf}(x, y)), & \text{if } \exists y^0 A_{qf}(x, y), \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

When  $\forall x^\sigma\exists y^0 A_{qf}(x, y)$  is proved in  $WE-PA^\omega + QF-AC$  the totality of  $\Phi$  is proved. From the extraction theorem we see that every recursive functional  $\Phi$  of type  $\sigma 0$  proved total in  $WE-PA^\omega + QF-AC$  is denoted by a term in  $WE-T$ . Furthermore we have a procedure for extracting such a term for  $\Phi$  from a classical proof of its totality. We have, in other words, that  $WE-PA^\omega + QF-AC$  does not prove more functionals of type  $\sigma 0$  to be total than  $WE-HA^\omega$ , since  $WE-T$  is the quantifier free fragment of  $WE-HA^\omega$ .

With respect to extraction of constructive information it is actually easier to consider the following theorem, which is stated and proved by Luckhardt (1973). For simplicity we take the Hilbert style version.

**Theorem 4.4.2.** *Let  $A(\mathbf{a})$  be a formula of  $\mathcal{L}(\text{WE-PA}^\omega)$ , then*

$$\text{WE-PA}^\omega + \text{QF-AC} \vdash A(\mathbf{a}) \Rightarrow \text{WE-HA}^\omega \vdash \forall \mathbf{y}(A')_D(\mathbf{T}\mathbf{a}, \mathbf{y}, \mathbf{a})$$

for closed extractable sequence  $\mathbf{T}$  of terms.

**Proof.** Since WE-T is the quantifier free fragment of WE-HA<sup>ω</sup> we have actually already proved this theorem. But the verification of induction is considerably easier when done in WE-HA<sup>ω</sup> instead of WE-T:

We will verify the rule of induction, which is deductively equivalent to the axiom schema, see page 47.

$$\frac{A(0^0, \mathbf{a}) \quad A(x^0, \mathbf{a}) \rightarrow A(Sx^0, \mathbf{a})}{A(x^0, \mathbf{a})}$$

Since we are *not* interpreting into the quantifier free fragment we will not need the induction lemma. Let  $B \equiv A'$ . Generally we have by intuitionistic logic  $\neg\neg(C \rightarrow D) \leftrightarrow (\neg\neg C \rightarrow \neg\neg D)$ . The induction hypothesis is therefore equivalent to

$$\forall \mathbf{z} B_D(\mathbf{T}_1 \mathbf{a}, \mathbf{z}, 0, \mathbf{a}) \quad \text{and} \\ \forall \mathbf{y}, \tilde{\mathbf{y}} (B_D(\mathbf{y}, \mathbf{T}_2 x \mathbf{a} \mathbf{y} \tilde{\mathbf{y}}, x, \mathbf{a}) \rightarrow B_D(\mathbf{T}_3 x \mathbf{a} \mathbf{y}, \tilde{\mathbf{y}}, Sx, \mathbf{a})).$$

We then define  $\mathbf{T}$  by simultaneous primitive recursion in higher types such that

$$\begin{aligned} \mathbf{T}\mathbf{a}0 &= \mathbf{T}_1 \mathbf{a}, \\ \mathbf{T}\mathbf{a}(Sx) &= \mathbf{T}_3 x \mathbf{a}(\mathbf{T}\mathbf{a}x). \end{aligned}$$

The second part of the induction hypothesis now implies

$$\forall \mathbf{z} B_D(\mathbf{T}\mathbf{a}x, \mathbf{z}, x, \mathbf{a}) \rightarrow \forall \mathbf{z} B_D(\mathbf{T}\mathbf{a}(Sx), \mathbf{z}, Sx, \mathbf{a}),$$

and we obtain

$$\forall \mathbf{z} B_D(\mathbf{T}\mathbf{a}x, \mathbf{z}, x, \mathbf{a})$$

by the induction rule. ⊢

From this theorem we also get (immediately) an extraction theorem as theorem 4.4.1: just replace WE-T with WE-HA<sup>ω</sup> and retain the universal quantifiers. This yields:

**Corollary 4.4.3.** *WE-PA<sup>ω</sup> + QF-AC is conservative over WE-HA<sup>ω</sup> with respect to sentences of the form  $\forall \mathbf{x} \exists \mathbf{y} A_{\text{qf}}(\mathbf{x}, \mathbf{y})$ , where  $A_{\text{qf}}(\mathbf{x}, \mathbf{y})$  quantifier free.*

#### 4.5 The philosophical significance of interpretation theorems

Theorems 4.3.2 and 4.4.1 have important consequences for the philosophy of mathematics. First of all, theorem 4.3.2 is an important contribution to a generalised Hilbert program—the consistency of classical arithmetic is conceptually reduced to a quantifier free type theory. That this type theory actually is a reasonable extension of Hilbert's finitism will depend on an argument for the constructivity of the functionals of the theory.

## 4.5.1 Support for Hilbert

Gödel's theory of functionals is a natural generalisation of Hilbert's finitary part of mathematics: It is primitive recursion generalised, only, to higher types. Primitive recursive functionals of higher types were originally introduced by Hilbert (1926) as means to prove the continuum hypothesis. Such a proof should of course not use disputable elements of ideal mathematics and this indicates that Hilbert considered the functionals of higher types to be a natural generalisation of finitism, though he does not state it explicitly.

There are, however, more direct arguments in favour of constructivity of the functionals. The more complicated functionals are defined inductively by a chain of definitions where each step defines a new functional in terms of previously defined ones. Now, the single steps describe simple calculations. But given a concrete well-formed closed term we cannot directly read off how many calculations we have to perform before the overall calculation is done. By mathematical reasoning we can, of course, give bounds for any term, but this is in the case at hand useless since it is the same mathematics we want to justify by our interpretation. However, we are given inductive rules in order to do the computation—intuitively we just do not know how many we will have to perform.

Gödel (1941, 17) argues for calculability of the functionals by the following argument:

So the schemes of definition are formally the same as in recursive number theory, the only difference being that the objects with which we are dealing now are not only numbers but also functions or, in other words, procedures for obtaining numbers out of given numbers (respectively, for obtaining procedures out of given procedures, ...). Accordingly, we have a new primitive operation, namely, the operation of applying the procedure to an object of appropriate type. But this operation is actually calculable since it is contained in the notion of a procedure that it can always be carried through.

Therefore, when we define new procedures (or functionals) out of previously defined ones by the schemes we get a calculable functional since the equations defining  $\Pi_{\sigma,\tau}$ ,  $\Sigma_{\rho,\tau,\sigma}$  and  $R_{\sigma}$  prescribe constructive operations. Epistemologically, to follow inductive rules can hardly be problematic and thus it seems justified to expand Hilbert's finitism by Gödel's functionals. We consider the ability to carry out such operations to be a part of the general human ability to reason and do science. On this general ability Hilbert says

Nun gebe ich zu, daß schon zum Aufbau der theoretischen Fachwerke gewisse apriorische Einsichten nötig sind und daß stets den Zustandekommen unserer Erkenntnis solche zugrund liegen. Ich glaube, daß auch die mathematische Erkenntnis letzten Endes auf einer Art solcher anschaulicher Einsicht beruht. ... Das Apriori ist dabei nichts mehr und nichts weniger als eine Grundeinstellung oder der Ausdruck für gewisse unerläßliche vorbedingungen des Denkens und Erfahrens. ... [Das Apriori] ist im wesentlichen die von mir in verschiedenen Abhandlungen charakterisierte finite Einstellung. (Hilbert, 1930, 383–385).

A mathematical approach to the constructiveness of Gödel's functionals is taken by defining a functional  $T$  of type  $\sigma_1 \cdots \sigma_n 0$  to be *calculable* if for arbitrary calculable  $t_1^{\sigma_1}, \dots, t_n^{\sigma_n}$  it

can be proved that  $Tt_1^{\sigma_1} \dots t_n^{\sigma_n}$  equals a number. In a series of lectures held at Princeton – also in 1941 – Gödel observes that every functional can be proved to be calculable in this sense.<sup>3</sup> However, the proof of this uses rules and axioms of mathematics which we want to justify by the interpretation. Therefore, the proof has strictly speaking no value foundationally.<sup>4</sup> Another mathematical approach to the problem involves assigning notations representing ordinals less than  $\varepsilon_0$ , i.e. constructive ordinals, to terms in such a way that a computation or reduction described by the equations for  $\Pi_{\sigma,\tau}$ ,  $\Sigma_{\rho,\tau,\sigma}$  and  $R_\sigma$  decreases the associated ordinal. Hence the well-foundedness of  $\varepsilon_0$  guarantees the calculability of the functionals. This idea is also present in Gödel's Princeton lectures. Over a quarter of a century later the idea was carried out by Hinata (1967), Diller (1968) and Howard (1970). For a recent reformulation of Howard's approach (also for various fragments of T) see Weiermann (1998).<sup>5</sup>

#### 4.5.2 Implications for constructive existence

The theorems on the forgoing pages also have consequences in a somewhat different direction. This direction concerns existence and computational content of classical proofs, rather than consistency. With respect to this, the extraction theorem says the following. If we from a set  $\Gamma$  of purely universal assumptions have proved  $\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$  using classical logic and quantifier free axiom of choice then this existence proof is not fraud: in fact we can extract constructively a program  $t$  of type  $\sigma\tau$  such that for any  $x$  of type  $\sigma$ ,  $A_{\text{qf}}(x, tx)$  is verified in a quantifier free calculus. But the theorem also says something important about the influence of the way we have established  $\Gamma$ . We could very well view  $\Gamma$  as a set of lemmata which we *know* to be true, or which we have proved earlier in order to keep the length of the proof of  $\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$  down. Now, the theorem says that if we have proved these lemmata earlier, then these proofs have absolutely no influence on our program  $t$  realising  $\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$ . We can in other words use as much classical logic as we want to prove purely universal lemmata—these proofs will have no consequences for the computational content of the final proof of  $\forall x^\sigma \exists y^\tau A_{\text{qf}}(x, y)$ .

We can also add universal *axioms*  $\Gamma$  to  $\text{WE-PA}^\omega + \text{AC}$  at no cost—this was stressed continuously by Kreisel in the 50s. Since universal sentences are trivially interpreted by themselves they do not influence on the computational content. Therefore, for a set  $\Gamma$  of purely universal axioms in the language of  $\text{WE-PA}^\omega$  we have that  $\text{WE-PA}^\omega + \Gamma + \text{QF-AC}$  has the same provably total recursive functionals as  $\text{WE-HA}^\omega + \Gamma$  has. There are many interesting axioms of this kind, such as universal sentences undecidable by  $\text{WE-PA}^\omega$ , say,  $\text{Con}_{\text{PA}}$ . But also conjectures of different kinds as for instance Goldbach's conjecture or the Riemann hypothesis.<sup>6</sup> Kohlenbach (1996), working within proof mining, extended this idea using so-called monotone functional interpretation to much more general lemmata. All this points to the fact, that the kind of questions Hilbert originally directed the foundational interest towards, namely consistency, are harmless with respect to computational content. So Hilbert's

<sup>3</sup>Troelstra (1995a, 188) mentions these lectures by Gödel.

<sup>4</sup>The idea that Gödel here mentions is similar to the idea of “convertibility” predicates that Tait (1967) uses in order to show normalisation of functionals. For an elegant version of this, see (Schwichtenberg, 2000, 149–151).

<sup>5</sup>We note that Diller's ordinal assignment is not optimal in the sense that ordinals above  $\varepsilon_0$  are used.

<sup>6</sup>That the Riemann hypothesis in fact is equivalent to a purely universal statement is proved in (Kreisel, 1958).

original program searching for consistency and Kreisel's program searching for computational content really go in different directions.

In any case, these remarks support Kreisel's view that in order to obtain constructive results we *do not* always have to restrict ourselves to constructive reasoning. Therefore, a closer analysis shows that mathematicians interested in constructive existence can in fact use classical logic in many situations.

#### 4.6 PA as a subsystem of WE-PA<sup>ω</sup>

In this text we have not defined PA (nor HA) explicitly. A completely standard way to do this is to take as language of the theory some (untyped) first order language with the standard logical symbols. There is an infinite list of number variables, the constant 0 and the unary function constant S. In the language we also have function symbols for addition, multiplication, some initial primitive recursive functions and a single binary predicate symbol =, which denotes equality between numbers. Terms and formulas are defined in the usual way and the logic is classical first order predicate logic with equality. There are defining equations for initial primitive recursive functions (0-function, successor-function and  $n$ -place projection functions); furthermore there are schemata for composition and recursion. The precise selection of these initial functions and the schemata do not really matter, but see (Troelstra, 1973, 18–19) for a simple example. The point is that from the initial functions and the composition and recursion schemata we can introduce in PA all primitive recursive (definitions of) functions.

We can now associate to each function constant  $f$  of PA a term  $T_f$  of WE-PA<sup>ω</sup> such that PA becomes a subsystem of WE-PA<sup>ω</sup>. The idea of the mapping  $F$  is to use the properties we know that the  $\lambda$ -operator and  $R_0$  have. For instance  $F(0) \equiv 0^0$ ,  $F(x) \equiv x^0$ ,  $F(S) \equiv S^1$ ; if  $U_n^i$  is a function symbol such that  $U_n^i(x_1, \dots, x_n) = x_i$  then  $F(U_n^i) \equiv \lambda x_1, \dots, x_n. x_i$ ; and so forth—see (Troelstra, 1973, 42) for details of the mapping.

In precisely the same way we see HA as a subsystem of WE-HA<sup>ω</sup>.

#### 4.7 The no-counterexample interpretation of Peano arithmetic

With respect to quantifier complexity the extraction theorem is the best possible. Let us consider the formula  $\forall x^0 \exists y^0 \forall z^0 (Txy \vee \neg Txz)$ , which is classically provable. In fact we have seen on page 53 that it is provable in WE-HA<sup>ω</sup> + IP<sup>ω</sup><sub>¬∇</sub> + MP<sup>ω</sup>. But a realisation of the existential quantifier cannot be recursive in  $x$ , since it would then solve the halting problem. However, we can extract a different kind of constructive information from classical proofs of formulas as this via the general Dialectica interpretation, theorem 4.3.2. When it comes to PA this connects with Kreisel's (1951) no-counterexample interpretation (n.c.i.) of Peano arithmetic: the n.c.i. is a corollary of theorem 4.3.2.

##### 4.7.1 Herbrand normal form and the n.c.i.

The n.c.i. of PA can be seen as a spin-off from the Dialectica interpretation as was noted by Kreisel (1959). First we need two definitions.

**Definition 4.7.1.** (Herbrand normal form). Let  $A$  be a formula on prenex normal form, i.e.  $\forall y_0 \exists x_1 \forall y_1 \dots \exists x_n \forall y_n A_{\text{qf}}(y_0, x_1, y_1, \dots, x_n, y_n)$ . The Herbrand normal form of  $A$  is then defined to be

$$A^H := \forall y_0, Z_1, \dots, Z_n \exists x_1, \dots, x_n A_{\text{qf}}(y_0, x_1, Z_1 x_1, \dots, x_n, Z_n x_1 \dots x_n),$$

where  $Z_1, \dots, Z_n$  are variables not free in  $A_{\text{qf}}$ .  $\triangleleft$

The transformation of  $A$  to  $A^H$  is intuitionistically provable and we therefore have intuitionistically  $A \rightarrow A^H$ . But it is a little awkward to state it in this way, since  $A$  has to be on prenex normal form. If  $A$  is not on prenex normal form then we have to get  $A$  on this which, generally, requires classical logic. Consequently, for any formula  $A$ :

$$\text{CL}^\omega \vdash A \rightarrow A^H.$$

However, to get a proof of  $A^H \rightarrow A$  we need classical logic + axiom of choice for numbers. This is proved by using

$$\frac{\forall x^0 \exists y^0 \neg B(x, y) \rightarrow \exists Y^1 \forall x^0 \neg B(x, Yx) \quad \exists Y \forall x \neg B(x, Yx) \rightarrow \perp}{\neg \forall x^0 \exists y^0 \neg B(x, y)}$$

and the classical equivalence  $\forall x C(x) \leftrightarrow \neg \exists x \neg C(x)$  and cancellation of double negations.

In PA we do not have the typed language and therefore no higher order quantifiers. If we want to express  $A^H$  we supply with new function symbols  $f_1, \dots, f_n$  and use  $f_i$  instead of  $Z_i$ .

**Definition 4.7.2.** (The no-counterexample interpretation). Let  $A \in \mathcal{L}(\text{PA})$  be a sentence on prenex normal form. If a sequence of closed terms  $\mathbf{T} \equiv T_1, \dots, T_n$  realises the Herbrand normal form of  $A$ , i.e.

$$\text{WE-T} \vdash A_{\text{qf}}(y_0, T_1 \mathbf{z}, z_1(T_1 \mathbf{z}), \dots, T_n \mathbf{z}, z_n(T_1 \mathbf{z}) \dots (T_n \mathbf{z}))$$

then we say that  $\mathbf{T}$  satisfies the n.c.i. of  $A$  in WE-T or just  $\mathbf{T}$  n.c.i.  $A$ .  $\triangleleft$

**Remark 4.7.3.** Note how the index functions  $z_i$  are of type level 1 and that the realising functionals  $T_i$  are of type level 2. These levels are consequently independent of  $n$ .

**Theorem 4.7.4.** For any PA-provable sentence  $A$  of  $\mathcal{L}(\text{PA})$  on prenex normal form there exist  $\mathbf{T}$  of WE-T such that  $\mathbf{T}$  n.c.i.  $A$ .

**Proof.** Let  $A$  be a sentence on prenex normal form and suppose  $\text{PA} \vdash A$ . By theorem 4.4.1 there exist closed terms  $\mathbf{T}$  such that  $\mathbf{T}$  n.c.i.  $A$ .  $\dashv$

The name of the n.c.i. is motivated by trying to find counterexamples to formulas. Let us for instance consider the closed  $\Sigma_3^0$  formula

$$\exists x \forall y \exists z A_{\text{qf}}(x, y, z). \quad (4.4)$$

The Herbrand normal form of this formula is  $\forall f \exists x, z A_{\text{qf}}(x, fx, z)$ . If we could come up with some Skolem function  $g$  such that for any  $x, z$

$$\neg A_{\text{qf}}(x, gx, z),$$

then we would have a counterexample to (4.4). Now we want to show constructively that in case (4.4) is classically provable this cannot be the case: There cannot exist such  $g$ —there is no counterexample. This would be to show that for any  $g$  there exist functionals  $T_1$  and  $T_2$  such that

$$A_{\text{qf}}(T_1g, g(T_1g), T_2g).$$

But this is precisely the Dialectica interpretation of  $\forall f \exists x, z A_{\text{qf}}(x, fx, z)$  and we have the n.c.i. of (4.4).

So generally, if we have proved  $A$  in PA, then by the n.c.i. we know constructively that there cannot be a counterexample to  $A$ . This also provides us with a proof of the consistency of PA. Since, if PA were inconsistent then certainly there would be a formula  $A$  on prenex normal form such that both  $A$  and  $\neg A$  were provable. But due to the n.c.i. we know that this cannot be the case. However, such a consistency proof does not provide new insight since we arrived at the n.c.i. via negative translation and Dialectica interpretation, which already yield consistency. Therefore, when we want to evaluate the significance of a proof of consistency by the n.c.i. it depends on the techniques which have been used to obtain the n.c.i.

#### 4.7.2 Remarks on n.c.i.

Kreisel originally introduced and proved the n.c.i. of PA by using the technically complicated method of  $\varepsilon$ -substitution.<sup>7</sup> This method of using  $\varepsilon$ -terms instead of quantifiers was originally introduced by Ackermann and Hilbert in the 20s. As such the n.c.i. is an interpretation different from negative translation + Dialectica interpretation, but the n.c.i. can be seen as a spin-off, as the proof of theorem 4.7.4 shows. There are similarities between the two interpretations but, certainly, there are also important differences.

First of all, with respect to  $\Pi_3^0$  formulas the n.c.i. coincides essentially with the interpretation by negative translation + Dialectica. However, they diverge with respect quantifier complexity greater than or equal  $\Sigma_3^0$ . If  $\exists x \forall y \exists z A_{\text{qf}}(x, y, z)$  is a closed PA-provable  $\Sigma_3^0$  formula then, by the n.c.i., there exist functionals  $T_1$  and  $T_2$  of type 2 such that for any  $g$  of type 1

$$A_{\text{qf}}(T_1g, g(T_1g), T_2g)$$

is provable in WE-T. Contrary to this, the negative translation + Dialectica interpretation of  $\exists x \forall y \exists z A_{\text{qf}}(x, y, z)$  is that

$$A_{\text{qf}}(T_3Y, Y(T_3Y)(T_4Y), T_4Y(Y(T_3Y)(T_4Y))),$$

is provable in WE-T for certain  $T_3$  of type (010)0 and  $T_4$  of type (010)1. Thus negative translation + Dialectica use functionals of higher types, whereas the functionals provided by the n.c.i. stay within type level 2. That this is only seemingly an advantage of the n.c.i. shows up in connection with the modus ponens rule. Kohlenbach (1999) has shown that the n.c.i. has no simple and uniform interpretation of modus ponens which stays within any finite *subsystem* of Gödel's functionals. Moreover, if PA-provability of  $A$  and  $A \rightarrow B$  is *not* assumed then

<sup>7</sup>That Kreisel's original proof indeed is non-trivial is witnessed by Feferman (1996) who says that he "never tried to wade through" the proof.

one needs bar recursion of type 0 for the n.c.i. of  $B$ . On the other hand negative translation + Dialectica provide uniformly realising functionals for  $(B')^D$  from any realisations of  $(A')^D$  and  $((A \rightarrow B)')^D$ .

Moreover, the n.c.i. is not faithful for fragments of PA where induction is only to be applied to formulas of a certain complexity, whereas Dialectica together with negative translation faithfully interprets various subsystems of PA as was shown by Parson (1972).

These points are due to the fact that the combination of negative translation and Dialectica interpretation of a formula  $A$  is much closer to the original  $A$  than the n.c.i. of  $A$  is. As observed by Kreisel (1959, 120), it is not difficult to prove that

$$\text{WE-PA}^\omega + \text{QF-AC} \vdash A \leftrightarrow (A')^D$$

However, to prove the equivalence between  $A$  and the n.c.i. of  $A$  one needs classical logic with choice (though only for numbers) for arbitrary arithmetical formulas whereas the combination of negative translation with Dialectica interpretation only needs quantifier-free choice (in higher types). Thus, negative translation + Dialectica use higher types in order to stay closer to the original formula.

Furthermore, Spector (1962) showed that if one adds bar recursion to WE-T then the Dialectica interpretation together with negative translation can be extended to subsystems of analysis. On the other hand, no such extension of the n.c.i. to systems of analysis is known.

All this points towards the fact that *both* mathematically and philosophically Dialectica + negative translation is as device preferable compared with n.c.i. Consequently, we will not pay more attention to the n.c.i. in this thesis.

## Modified Realisability, A-translation and Applications

We have in the foregoing chapter seen that the Dialectica interpretation together with negative translation provides a strong tool for extracting constructive information from classical proofs. One of the reasons that Dialectica is powerful is the way it treats universal quantifiers occurring negatively in formulas, as for instance  $\neg\forall xA_{\text{qf}}(x)$ . This quantifier is treated as if it were the existential quantifier in  $\exists x\neg A_{\text{qf}}(x)$ . Classically we have equivalence between these two formulas, but not intuitionistically. What makes Dialectica strong is that it is able to verify (or interpret) this classical equivalence. More generally, an implication between two universal sentences  $\forall x^\sigma A_{\text{qf}}(x) \rightarrow \forall y^\tau B_{\text{qf}}(y)$  is by Dialectica interpreted in the way that there is a primitive recursive functional  $T$  of type  $\tau\sigma$  such that, given any counterexample to  $B_{\text{qf}}$  this functional produces a counterexample to  $A_{\text{qf}}$ :

$$\forall y^\tau (\neg B_{\text{qf}}(y) \rightarrow \neg A_{\text{qf}}(Ty))$$

This interpretation is computationally meaningful, although the BHK interpretation does *not* interpret  $\forall xA_{\text{qf}}(x) \rightarrow \forall yB_{\text{qf}}(y)$  in this manner. But precisely this way of treating universal quantifiers has the effect that Dialectica can interpret Markov's principle, which again is not validated by BHK. That Markov's principle is not provable in intuitionistic arithmetic was demonstrated by Kreisel at the end of the 50s. To show that, Kreisel further developed numerical realisability – which was discovered by Kleene (1945) – into modified realisability. Modified realisability is, as Dialectica, a way of interpreting typed intuitionistic arithmetic. Thus, the Dialectica interpretation and modified realisability present two ways of interpreting mathematics where the former validates Markov's principle whereas the latter invalidates it. This is, of course, philosophically interesting and we will discuss it, but there are also mathematical reasons motivating a study of modified realisability as well.<sup>1</sup>

Modified realisability together with the so-called A-translation *and* negative translation can – as Dialectica and negative translation – be used to extract constructive content from classical proofs. Though modified realisability and Dialectica in certain respects behave similarly they are generally very different. For instance, the combination of negative translation, A-translation and modified realisability can *only* extract constructive information from proofs of  $\forall\exists$ -statements,<sup>2</sup> whereas negative translation + Dialectica interpretation provide constructive interpretations of proved formulas in general.<sup>3</sup> We will take a closer look at the differences.

Realisability by numbers was introduced by Kleene in order to scrutinize the connection between intuitionism and recursive functions. In a way one could say that Kleene's realisabil-

<sup>1</sup>The notion of realisability as developed by Kleene, uses partial and untyped operations as realisers, whereas modified realisability uses only total but typed realisers. Interestingly, Kleene's realisability validates Markov's principle in the context of HA. Hence the requirement of totality is in conflict with Markov's principle.

<sup>2</sup>Note that Berger and Schwichtenberg (2000) has slightly generalised this by introducing the notions of definite and goal formulas.

<sup>3</sup>Of course, if a formula is provable in PA then one can put it on Herbrand normal form and then apply negative translation, A-translation and modified realisability. This gives an interpretation in the style of the n.c.i., but there are, as we have just seen, problems connected with such an interpretation, and we will therefore not go deeper into this idea.

ity (in fact a variant of realisability, as we will see) shows that intuitionistic arithmetic really is constructive in the sense that if one proves  $A \vee B$  for sentences  $A$  and  $B$  then by realisability one can tell which one of them actually holds; realisability shows, in other words, that intuitionistic arithmetic possesses disjunction property. Likewise existence property: If we prove the sentence  $\exists x A(x)$  in intuitionistic arithmetic, then by realisability we can extract a number  $n$  such that  $A(n)$  holds. In many respects recursive realisability is very close to the BHK interpretation, but instead of being based on the notions of abstract ‘proofs’ and ‘constructions’ it is based on the concepts of recursive function and number. It is thus more concrete (at least in the context of arithmetic) and can be seen as giving a classically meaningful definition of intuitionistic truth.

As *Dialectica*, modified realisability (m.r.) is an interpretation of typed Heyting arithmetic. But m.r. only interprets into the negative fragment, i.e. the fragment containing  $\wedge, \rightarrow, \forall$ . Thus, m.r. is not a (conceptual) reduction to the quantifier free fragment as the *Dialectica* interpretation is. On the other hand m.r. interprets a stronger theory than *Dialectica*; m.r. namely interprets typed Heyting arithmetic with *full* extensionality.

### 5.1 Definition of $\text{E-HA}^\omega$ and modified realisability

**Definition 5.1.1.** Heyting arithmetic in all finite types with full extensionality,  $\text{E-HA}^\omega$ , is an extension of  $\text{WE-HA}^\omega$ . Pertaining to how the logic is formulated there are to possibilities:

1.  $\text{E-HA}_H^\omega$  is the extension of  $\text{WE-HA}_H^\omega$  where the quantifier free extensionality rule QF-ER is exchanged with

$$\forall z^\sigma, x_1^{\sigma_1}, y_1^{\sigma_1}, \dots, x_n^{\sigma_n}, y_n^{\sigma_n} \left( \bigwedge_{i=1}^n x_i =_{\sigma_i} y_i \rightarrow z\mathbf{x} =_0 z\mathbf{y} \right).$$

for all types  $\sigma$ , ( $\sigma \equiv \sigma_1 \dots \sigma_n 0$ ).

2.  $\text{E-HA}_{\text{ND}}^\omega$  is the extension of  $\text{WE-HA}_{\text{ND}}^\omega$  where the quantifier free rule of extensionality is exchanged with:

$$\frac{s^\sigma =_\sigma t^\sigma}{r[s^\sigma]^\tau =_\tau r[t^\sigma]^\tau} \text{ER}$$

The rule is thus the same as QF-ER for natural deduction, just without restrictions on assumptions.

◁

Now we turn to the definition of modified realisability as introduced by Kreisel (1962). Let  $()$  be the empty sequence of variables.

**Definition 5.1.2.** (Modified realisability). Let  $A(\mathbf{a})$  be a formula of  $\mathcal{L}(\text{E-HA}^\omega)$  with  $\text{FV}(A) = \{\mathbf{a}\}$ . Then  $\mathbf{x} \text{mr} A(\mathbf{a})$ , where  $\{\mathbf{x}\} \cap \{\mathbf{a}\} = \emptyset$  and  $\text{FV}(\exists \mathbf{x}(\mathbf{x} \text{mr} A(\mathbf{a}))) = \{\mathbf{a}\}$ , is de-

defined inductively:

$$\begin{aligned}
(P^{\underline{\text{mr}}}) \quad & () \underline{\text{mr}}A(\mathbf{a}) &::= & A(\mathbf{a}), \text{ if } A(\mathbf{a}) \text{ is prime,} \\
(\wedge^{\underline{\text{mr}}}) \quad & \mathbf{x}, \mathbf{y} \underline{\text{mr}}(A(\mathbf{a}) \wedge B(\mathbf{b})) &::= & \mathbf{x} \underline{\text{mr}}A(\mathbf{a}) \wedge \mathbf{y} \underline{\text{mr}}B(\mathbf{b}), \\
(\vee^{\underline{\text{mr}}}) \quad & z^0, \mathbf{x}, \mathbf{y} \underline{\text{mr}}(A(\mathbf{a}) \vee B(\mathbf{b})) &::= & (z^0 = 0 \rightarrow \mathbf{x} \underline{\text{mr}}A(\mathbf{a})) \wedge \\
& & & (z^0 \neq 0 \rightarrow \mathbf{y} \underline{\text{mr}}B(\mathbf{b})), \\
(\exists^{\underline{\text{mr}}}) \quad & y^\sigma, \mathbf{x} \underline{\text{mr}}(\exists z^\sigma A(z, \mathbf{a})) &::= & \mathbf{x} \underline{\text{mr}}A(y^\sigma, \mathbf{a}), \\
(\forall^{\underline{\text{mr}}}) \quad & \mathbf{x} \underline{\text{mr}}(\forall z^\sigma A(z, \mathbf{a})) &::= & \forall z^\sigma (\mathbf{x} z \underline{\text{mr}}A(z, \mathbf{a})), \\
(\rightarrow^{\underline{\text{mr}}}) \quad & \mathbf{x} \underline{\text{mr}}(A(\mathbf{a}) \rightarrow B(\mathbf{b})) &::= & \forall \mathbf{y} (\mathbf{y} \underline{\text{mr}}A(\mathbf{a}) \rightarrow \mathbf{x} \underline{\text{mr}}B(\mathbf{b})).
\end{aligned}$$

◁

Clause  $(\wedge^{\underline{\text{mr}}})$  requires that  $\{\mathbf{x}\} \cap \{\mathbf{y}\} = \emptyset$ .

Notice that  $\underline{\text{mr}}$  is a defined predicate symbol of  $\mathcal{L}(E\text{-HA}^\omega)$ , which holds between a sequence of terms and a formula. As an example we see that

$$\mathbf{x} \underline{\text{mr}} \neg \neg \exists y P(y)$$

for any prime formula  $P$  is the formula

$$\neg \forall y \neg P(y).$$

$\mathbf{x}$  is, in other words, the empty sequence.

An important property of  $\underline{\text{mr}}$  is that for any sequence  $\mathbf{t}$  of terms,  $\mathbf{t} \underline{\text{mr}}A$  is  $\exists$ - and  $\forall$ -free. If we say that  $\exists \mathbf{x}(\mathbf{x} \underline{\text{mr}}A)$  is the  $\underline{\text{mr}}$ -translation of  $A$  then the translation of any formula is  $\exists$  quantifiers followed by a negative formula. Another important property of  $\underline{\text{mr}}$  is that  $\mathbf{x} \underline{\text{mr}}A \equiv A$  for negative  $A$ .

Instead of using the notation,  $\mathbf{x} \underline{\text{mr}}A$ , we could have used a notation somewhat closer to the notation we used in case of Dialectica. In such a notation we would write  $A_{\underline{\text{mr}}}(\mathbf{x})$  instead of  $\mathbf{x} \underline{\text{mr}}A$  and the translation of  $A$  would be  $A^{\underline{\text{mr}}} := \exists \mathbf{x} A_{\underline{\text{mr}}}(\mathbf{x})$ . Using this notation, the translation of, say, an implication would be

$$(A \rightarrow B)^{\underline{\text{mr}}} := \exists \mathbf{X} \forall \mathbf{y} (A_{\underline{\text{mr}}}(\mathbf{y}) \rightarrow B_{\underline{\text{mr}}}(\mathbf{X}\mathbf{y})).$$

Our ‘official’ notation – used in the definition – is, however, the commonly used notation in the literature, and it is probably a little easier to use.

It is important to note that the  $\underline{\text{mr}}$  translation focuses on existential quantifiers occurring strictly positively in a formula. So, there is a difference between Dialectica and realisability translations. Classically, a universal quantifier occurring negatively in a formula is essentially an existential quantifier, in other words ( $\text{CL}^\omega$  denoting typed classical logic):

$$\text{CL}^\omega \vdash (\forall x^\sigma A(x) \rightarrow B) \leftrightarrow \exists x^\sigma (A(x) \rightarrow B), \quad x^\sigma \notin \text{FV}(B).$$

As we have seen the Dialectica interpretation will look for witnesses to (some of) these universal quantifiers. But this is not the case for  $\underline{\text{mr}}$ . This is clear from the definition of  $(\rightarrow^{\underline{\text{mr}}})$ : The sequence  $\mathbf{y}$  realising  $A \rightarrow B$  will only take *any* realiser  $\mathbf{x}$  of  $A$  and turn it into a realiser

$\mathbf{y}\mathbf{x}$  of  $B$ . Thus  $\mathbf{y}$  is not realising anything in  $A$ . Nested implications show this clearly.  $\mathbf{z}$  would be a realiser of  $(A \rightarrow B) \rightarrow C$ , if  $\mathbf{z}$  could do the following:

$$\forall \mathbf{y} (\forall \mathbf{x} (\mathbf{x} \underline{\text{mr}} A \rightarrow \mathbf{y}\mathbf{x} \underline{\text{mr}} B) \rightarrow \mathbf{z}\mathbf{y} \underline{\text{mr}} C).$$

Thus  $\forall \mathbf{x}$  which occur negatively are not analysed any further.<sup>4</sup>

In this respect,  $\underline{\text{mr}}$  is close to the BHK interpretation. If we go through the different components of the BHK interpretation we see that modified realisability can be motivated from this. Actually modified realisability corresponds to BHK—we only concentrate on arithmetical information in case of  $\exists$  and choices for disjunctions and the ‘constructions’ mentioned when  $\forall$  and  $\rightarrow$  are interpreted are considered to be computable operations.

## 5.2 Realisability interpretation of $\text{E-HA}^\omega + \text{AC} + \text{IP}_{\text{ef}}^\omega$

In the next definition we follow Troelstra (1998).

**Definition 5.2.1.** A formula  $A$  of  $\mathcal{L}(\text{E-HA}^\omega)$  is called  $\exists$ -free if it is built up from prime formulas using only the symbols  $\wedge$ ,  $\rightarrow$  and  $\forall$ . Let independence-of-premise for type  $\sigma$  restricted to  $\exists$ -free formulas be

$$\text{IP}_{\text{ef}}^\sigma : (A \rightarrow \exists x^\sigma B(x)) \rightarrow \exists x^\sigma (A \rightarrow B(x)),$$

where  $A$  is  $\exists$ -free and  $x \notin \text{FV}(A)$ . Then

$$\text{IP}_{\text{ef}}^\omega := \bigcup_{\sigma \in \mathcal{T}} \{\text{IP}_{\text{ef}}^\sigma\}.$$

◁

The motivation for calling this class of formulas  $\exists$ -free is, that  $\forall$  can be defined in terms of  $\exists$ ,  $\wedge$  and  $\rightarrow$ . Therefore, if one works with a defined notion of  $\forall$ ,  $\exists$ -free implies  $\forall$ -free. Note that any  $\exists$ -free formula is over  $\text{E-HA}^\omega$  equivalent to a negative formula.

Let axiom of choice be as previously defined. Whether one wishes to have  $\text{IP}_{\text{ef}}^\omega$  and AC expressed as rules (for natural deduction) or as schemes is not important. We can now state and prove the  $\underline{\text{mr}}$  interpretation theorem for  $\text{E-HA}^\omega$  plus the extra principles.

We state the interpretation theorem within natural deduction. Therefore: By considerations similar to those we had when discussing how to Dialectica translate derivations with assumptions, we translate any assumption  $A$  to  $\mathbf{x} \underline{\text{mr}} A$ , where no variable occurring in  $\mathbf{x}$  is free in the original derivation or among the other translated assumptions except those labelled by the same letter as  $A$ . Those assumptions (formula occurrences) labelled by the same letter  $u$  will be translated with the same string  $\mathbf{x}_u$  of variables. If  $\Gamma$  is a multiset of formulas occurring as assumptions of a derivation, then by  $\Gamma_{\underline{\text{mr}}}$  we mean  $\{\mathbf{x} \underline{\text{mr}} A \mid A \in \Gamma\}$ . The following theorem within Hilbert style is stated and proved in details in (Troelstra, 1973, 215-217). A proof within natural deduction of soundness for a refined version of modified realisability is found in (Schwichtenberg, 2000, 154–157).

<sup>4</sup>This is perhaps seen more clearly in our Dialectica style notation. With that  $((A \rightarrow B) \rightarrow C)^{\underline{\text{mr}}}$  would be  $\exists \mathbf{Z} \forall \mathbf{Y} (\forall \mathbf{x} (A_{\underline{\text{mr}}}(\mathbf{x}) \rightarrow B_{\underline{\text{mr}}}(\mathbf{Y}\mathbf{x})) \rightarrow C_{\underline{\text{mr}}}(\mathbf{Z}\mathbf{Y}))$ .

**Theorem 5.2.2.** (Soundness of  $\underline{\text{mr}}$ -translation) Let  $H^*$  be  $E\text{-HA}^\omega + IP_{\text{ef}}^\omega + AC$  and let  $A(\mathbf{a})$  be a formula of  $\mathcal{L}(E\text{-HA}^\omega)$  with  $\text{FV}(A) = \{\mathbf{a}\}$ .

If  $\Gamma \vdash A(\mathbf{a})$  in  $H^*$  then  $\Gamma_{\underline{\text{mr}}} \vdash \mathbf{t} \underline{\text{mr}}A(\mathbf{a})$  in  $E\text{-HA}_{\text{ND}}^\omega$ ,

for some sequence of terms  $\mathbf{t}$  with  $\text{FV}(\mathbf{t}) \subseteq \{\mathbf{a}\} \cup \{\mathbf{x} \mid \mathbf{x} \in \text{FV}(\mathbf{y} \underline{\text{mr}}B) \text{ and } B \in \Gamma\}$ , which can be extracted from a proof of  $A(\mathbf{a})$  from  $\Gamma$ .

**Proof.** The proof is by induction on the length of the derivation in  $H^*$  of  $A(\mathbf{a})$ . We give two examples.

**Case 1.**  $\rightarrow\text{I}$ .

$$\frac{\begin{array}{c} \Gamma, [u : A] \\ \vdots \\ B \end{array}}{A \rightarrow B} \rightarrow\text{I}, u$$

The induction hypothesis is that  $\Gamma_{\underline{\text{mr}}}$  and  $\mathbf{x}_u \underline{\text{mr}}A$  prove  $\mathbf{t} \underline{\text{mr}}B$ , where  $\mathbf{t}$  have free variables at most among the free variables of  $\mathbf{x}_u \underline{\text{mr}}A$ ,  $B$  and  $\Gamma_{\underline{\text{mr}}}$ . On the basis of this we get:

$$\frac{\begin{array}{c} \Gamma_{\underline{\text{mr}}}, [u : \mathbf{x}_u \underline{\text{mr}}A] \\ \vdots \\ \mathbf{t} \underline{\text{mr}}B \end{array}}{\mathbf{x}_u \underline{\text{mr}}A \rightarrow \mathbf{t} \underline{\text{mr}}B} \rightarrow\text{I}, u}{\forall \mathbf{x}_u (\mathbf{x}_u \underline{\text{mr}}A \rightarrow \mathbf{t} \underline{\text{mr}}B)} \forall\text{I}$$

From this we see that

$$\Gamma_{\underline{\text{mr}}} \vdash \lambda \mathbf{x}_u. \mathbf{t} \underline{\text{mr}}A \rightarrow B$$

in  $E\text{-HA}_{\text{ND}}^\omega$ , where  $\text{FV}(\lambda \mathbf{x}_u. \mathbf{t}) \subseteq \text{FV}(A \rightarrow B) \cup \{\mathbf{x} \mid \mathbf{x} \in \text{FV}(\mathbf{y} \underline{\text{mr}}C) \text{ and } C \in \Gamma\}$ .

**Case 2.**  $\exists\text{E}$ .

$$\frac{\begin{array}{c} \Gamma \\ \vdots \\ \exists z^\sigma A(z) \end{array} \quad \begin{array}{c} \Delta, [u : A(b^\sigma)] \\ \vdots \\ C \end{array}}{C} \exists\text{E}, u$$

From the induction hypothesis we get (displaying only the free variables of importance):

$$\frac{\begin{array}{c} \Gamma_{\underline{\text{mr}}} \\ \vdots \\ \mathbf{t}_1 \underline{\text{mr}}A(t_0^\sigma) \end{array} \quad \frac{\begin{array}{c} \Delta_{\underline{\text{mr}}}, [u : \mathbf{x}_u \underline{\text{mr}}A(b^\sigma)] \\ \vdots \\ \mathbf{t}_3[\mathbf{x}_u, b] \underline{\text{mr}}C \end{array}}{\mathbf{x}_u \underline{\text{mr}}A(b) \rightarrow \mathbf{t}_3[\mathbf{x}_u, b] \underline{\text{mr}}C} \rightarrow\text{I}, u}{\mathbf{t}_1 \underline{\text{mr}}A(t_0^\sigma) \rightarrow \mathbf{t}_3[\mathbf{t}_1, t_0^\sigma] \underline{\text{mr}}C} \forall\text{I}, \forall\text{E}}{\mathbf{t}_3[\mathbf{t}_1, t_0^\sigma] \underline{\text{mr}}C} \rightarrow\text{E}$$

Hence,  $t_3[t_1, t_0^\sigma]$  realise  $C$ , and the free variables of  $t_3[t_1, t_0^\sigma]$  are among the free variables of  $C$  and the free variables of the translated assumptions. (If the realising terms have free variables among the free variables of  $A$  replace them by zero functionals  $o$  of the corresponding type).

The rest of proof is similar to the soundness proof of Dialectica. For example  $\rightarrow E$  is realised by function application;  $\forall I$  is realised by lambda abstraction and so on.  $IP_{ef}^\omega$  is also easy to realise: just take the identity functional(s) of the right type(s).  $AC$  is also realised by the identity. Equality axioms (rules); extensionality; defining equations for combinators, recursors etc. are trivially realised, since they are negative. Induction is realised by the recursion operator.  $\dashv$

It should be noted that  $WE-HA^\omega$  is also  $\underline{mr}$ -interpretable.  $QF-ER$  is trivially realised, since it concerns only purely universal formulas which are (provably equivalent to purely universal formulas) not containing  $\vee$ .

### 5.3 Contraction and negatively occurring universal quantifiers

There are many similarities between the soundness proof of the Dialectica translation and the proof above. But there are also some important differences; most conspicuously, the way negatively occurring universal quantifiers are treated. Among other consequences, this is to the effect that for the proof of  $\underline{mr}$  soundness there is no need for a contraction lemma. This is an essential difference.

In case of an application of the contraction lemma the Dialectica interpretation always chooses the realiser to the left if this does the job, so to speak. But this can of course be loosened so that we are free to choose the most optimal realiser, as long as this realiser makes the deduction possible. In any case Dialectica has to consider all the different possible realisers, whereas modified realisability chooses a realiser from the beginning and uses that all the way. Accordingly, Dialectica is more flexible with respect to getting the most optimal realiser—but Dialectica is also more lengthy.

In order to compare Dialectica and modified realisability let us again take the example with the valid formula  $A \rightarrow A \wedge A$ . We saw on page 28 that the Dialectica realisation of this corresponds to definition by cases: In case  $A_D(\mathbf{T}_2\mathbf{x}, \mathbf{y}_1) \wedge A_D(\mathbf{T}_3\mathbf{x}, \mathbf{y}_2)$  is false  $\mathbf{T}_1$  will have to choose between  $\mathbf{y}_1$  and  $\mathbf{y}_2$  in order to falsify  $A_D(\mathbf{x}, \mathbf{T}_1\mathbf{x}\mathbf{y}_1\mathbf{y}_2)$ —this is done by checking which of the conjuncts is false. Consequently, the Dialectica interpretation needs decidability of prime formulas. The modified realisability realiser of  $A \rightarrow A \wedge A$  is on the other hand very simple; this is  $\lambda\mathbf{x}.\langle\mathbf{x}, \mathbf{x}\rangle$ .<sup>5</sup> This realiser copies any given realiser of  $A$  and turns it into a realiser of  $A \wedge A$ . Contraction is also present in Hilbert style by the axiom schema  $A \vee A \rightarrow A$ . Modified realisability interprets this by using definition by cases, but need not decidability of prime formulas. This is similar to the way Dialectica interprets the schema.

Another matter concerning the different behaviour on negatively occurring universal quantifiers shows up in case of Markov's principle. The  $\underline{mr}$ -translation of an instance of Markov's principle ( $A_{qf}$  assumed w.l.o.g. to be without  $\vee$ ) is

$$\exists y^\sigma (y^\sigma \underline{mr}(\neg\neg\exists x^\sigma A_{qf}(x) \rightarrow \exists x^\sigma A_{qf}(x))) \equiv \exists y^\sigma (\neg\forall x^\sigma \neg A_{qf}(x) \rightarrow A_{qf}(y^\sigma)).$$

<sup>5</sup>Recall,  $\lambda\mathbf{x}.\langle\mathbf{x}, \mathbf{x}\rangle$  is shorthand for  $\lambda\mathbf{x}.\mathbf{x}.\lambda\mathbf{x}.\mathbf{x}$ .

Seeing this it seems impossible that modified realisability can provide a realiser of every instance of Markov's principle. That modified realisability cannot do this we will see due to a counterexample. From this counterexample it is, via the mr soundness theorem, immediate to infer that  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC}$  cannot prove the principle.

But there are also some positive consequences of the treatment of universal quantifiers. First of all, mr validates  $\text{IP}_{\text{ef}}^\omega$  which is much stronger than  $\text{IP}_{\forall}^\omega$ . Furthermore we can have, not only purely universal lemmata, but  $\exists$ -free lemmata of an intuitionistic proof with no influence on an eventual realiser of a conclusion. This follows from soundness (theorem 5.2.2) since the translation of  $\exists$ -free formulas are literally themselves. Therefore, proofs of  $\exists$ -free lemmata (sentences) have no influence on the mr-realisers.

#### 5.4 Markov's principle intuitionistically unprovable

The rest of this chapter is devoted to different applications of modified realisability. The first results in the following theorem due to Kreisel (1959, 1962).

**Theorem 5.4.1.** *Markov's principle, even for type 0, is unprovable in  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC}$ .*

**Proof.** Assume  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} \vdash \text{MP}^0$ . Since any instance of  $\text{IP}_{\neg\forall}^\omega$  is provable from  $\text{IP}_{\text{ef}}^\omega$  it follows from theorem 3.8.1 that

$$E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} \vdash \forall x \exists y \forall z (T_{xxz} \rightarrow T_{xxy}),$$

where  $\exists y$  requires a non-computable realiser. This, however, contradicts soundness of the mr-translation, theorem 5.2.2.  $\dashv$

Again we see that Markov's principle and independence-of-premise for  $\exists$ -free formulas are computationally incompatible.

Assume we really want our functions to be recursive functions, in other words add some version of Church's thesis as an axiom. Let  $\text{CT}_0$ , in the language of  $\text{HA}$ , be a formal version of Church's thesis, i.e. if  $\forall x \exists y A(x, y)$  holds then there is total recursive function  $f$  realising  $\exists y$ . Now the following result by Troelstra (1973, 201) should not be a big surprise:<sup>6</sup>

$$\text{HA} + \text{CT}_0 + \text{IP}_{\text{ef}} + \text{MP} \text{ is inconsistent.}$$

Due to theorem 3.8.1 this result can be strengthened if  $\text{IP}_{\text{ef}}$  is replaced by  $\text{IP}_{\neg\forall}$ .

Philosophically and methodologically this has consequences for the mathematician who wants all functions to be recursive. For instance the Russian school of constructivity (Markov and followers) accepted both Church's thesis and Markov's principle. As a consequence they are not allowed to use  $\text{IP}_{\text{ef}}^0$ . However, the Dialectica interpretation shows that they can use  $\text{IP}_{\forall}^0$  in case they do not require full extensionality.

#### 5.5 $E\text{-HA}^\omega \pm \text{IP}_{\text{ef}}^\omega \pm \text{AC}$ not closed under Markov's rule

Extensionality is in general constructively problematic. However, modified realisability can in the context of  $E\text{-HA}^\omega$  extract terms that realise existential statements as  $\exists x^\sigma A(x)$ , even if

<sup>6</sup> $\text{IP}_{\text{ef}}$ ,  $\text{MP}$  are, respectively, formulations in the language of  $\text{HA}$  of  $\text{IP}_{\text{ef}}^\omega$ ,  $\text{MP}^0$ .

full extensionality is used. But the fact that the extracted terms actually realise the formula, i.e.  $t^\sigma, \mathbf{s} \underline{\text{mr}} \exists x^\sigma A(x)$  is, on the other hand, verified in  $\text{E-HA}^\omega$ .

With respect to extensionality one has to be careful. Theorem 5.4.1 showed that Markov's principle is unprovable in  $\text{E-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC}$ . But something stronger is in fact the case. The following theorem – an unpublished result of Ulrich Kohlenbach – shows that any of the theories  $\text{E-HA}^\omega \pm \text{IP}_{\text{ef}}^\omega \pm \text{AC}$  is not closed under Markov's *rule*, not even for type 0. Markov's rule is the special case of Markov's principle where there are no assumptions, thus for type  $\sigma$ , Markov's rule is

$$\vdash \neg\neg\exists x^\sigma A_{\text{qf}}(x, \mathbf{a}) \Rightarrow \vdash \exists x^\sigma A_{\text{qf}}(x, \mathbf{a}),$$

for quantifier free  $A_{\text{qf}}$ .

**Theorem 5.5.1.** *There is a quantifier free formula  $A_{\text{qf}}(x^0)$  of  $\mathcal{L}(\text{E-HA}^\omega)$  such that*

$$\text{E-HA}^\omega \vdash \neg\neg\exists x^0 A_{\text{qf}}(x), \text{ but } \text{E-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} \not\vdash \exists x^0 A_{\text{qf}}(x).$$

**Proof.** The simplest non-trivial case of the extensionality axioms of  $\text{E-PA}^\omega$  is

$$\forall z^2, f^1, g^1 (f =_1 g \rightarrow z f =_0 z g), \quad (5.1)$$

which is the same as  $\forall x^0 (f^1 x =_0 g^1 x) \rightarrow z^2 f =_0 z^2 g$ . From this follows by classical logic that

$$\text{E-PA}^\omega \vdash \exists x^0 (f^1 x =_0 g^1 x \rightarrow z^2 f =_0 z^2 g). \quad (5.2)$$

Now, Kuroda's negative translation also provides a reduction of  $\text{E-PA}^\omega$  to  $\text{E-HA}^\omega$ . One only has to extend the proof of theorem 4.2.3 by showing that the Kuroda translation of any instance of the extensionality axioms is provable in  $\text{E-HA}^\omega$ . But this is done by noting that  $\text{E-HA}^\omega \vdash \neg\neg\forall x\neg\neg A_{\text{qf}}(x) \leftrightarrow \forall x A_{\text{qf}}(x)$ , (recall the equivalences of (4.1) from page 58). Thus (5.2) implies

$$\text{E-HA}^\omega \vdash \neg\neg\exists x^0 (f^1 x =_0 g^1 x \rightarrow z^2 f =_0 z^2 g).$$

Assume (working towards a contradiction) that

$$\text{E-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} \vdash \exists x^0 (f^1 x =_0 g^1 x \rightarrow z^2 f =_0 z^2 g).$$

By soundness of  $\underline{\text{mr}}$ -translation (theorem 5.2.2) there would be a term  $t^0$  with  $\text{FV}(t) \subseteq \{z^2, f^1, g^1\}$  realising this in  $\text{E-HA}^\omega$ . By introducing universal quantifiers we get

$$\text{E-HA}^\omega \vdash \forall z^2, f^1, g^1 (f^1 t^0 =_0 g^1 t^0 \rightarrow z^2 f =_0 z^2 g).$$

But then  $\lambda z^2, f^1, g^1. t^0$  would satisfy the *Dialectica* translation of (5.1) which is contradicting the result of Howard (1973, 458, Th. 3.2) stating that there cannot be any such primitive recursive functional. Take therefore  $f^1 x^0 = g^1 x^0 \rightarrow z^2 f =_0 z^2 g$  for  $A_{\text{qf}}(x^0)$ .  $\dashv$

With respect to our analysis this is very interesting since it shows, among other things, that one has to be extremely careful when extensionality and Markov's principle/rule are combined. Also of interest is that Howard in the paper cited above shows that there are models of ZF (Zermelo Fraenkel set theory without axiom of choice) in which there are no functionals satisfying the Dialectica translation of the extensionality axiom of type 3. Full extensionality is, in other words, in the presence of  $MP^{\omega}$  a very strong principle.

### 5.6 The Friedman-Dragalin $A$ -translation for formulas of HA

We now come to the  $A$ -translation which was discovered independently by H. Friedman (1978) and A.G. Dragalin (1980). The  $A$ -translation makes it possible to extract computational content from PA-proofs of  $\Pi_2^0$ -formulas via negative translation and modified realisability. Using this device one will first have to apply negative translation, then  $A$ -translation and finally modified realisability. The goal of the  $A$ -translation is to show closure of HA under Markov's rule. The overall strategy in this approach is that, if  $PA \vdash \exists x A_{\text{qf}}(x, \mathbf{a})$  then by negative translation  $HA \vdash \neg\neg\exists x A_{\text{qf}}(x, \mathbf{a})$ . Closure under Markov's rule would then give  $HA \vdash \exists x A_{\text{qf}}(x, \mathbf{a})$  and one could then apply modified realisability. This strategy is in fact possible. But given the result above (theorem 5.5.1) it cannot be extended to  $E\text{-}PA^{\omega}$  and  $E\text{-}HA^{\omega}$ .

**Definition 5.6.1.** ( $A$ -translation). Let  $B$  and  $A$  be formulas of  $\mathcal{L}(HA)$  such that no free variables of  $A$  are bound in  $B$ . Then  $B^A$  is the formula that arises when every occurrence of a prime formula  $P$  in  $B$  is replaced by  $P \vee A$ .  $B^A$  is called the  $A$ -translation of  $B$ .  $\triangleleft$

The definition expresses the general idea of an  $A$ -translation. But the translation can be optimized in various ways. For instance, Troelstra & van Dalen (1988) define it similarly with the extra clause that  $\perp$  is replaced by  $A$  only. This is due to the fact that we have over intuitionistic logic  $A \leftrightarrow A \vee \perp$ . But this again suggests another optimization, namely, that every obvious false quantifier free sentence is replaced by just  $A$ .

**Remark 5.6.2.** Although the idea of the  $A$ -translation is very simple it does not seem to work for theories with a certain upper bound with respect to quantifier complexity on assumptions. Specifically, it does not work (at least directly) in case we have weak extensionality:  $WE\text{-}HA^{\omega}$ . As we will see after the following proof, the translation is problematic in case of the weak extensionality rule QF-ER.

If  $\Gamma$  is the multiset consisting of  $C_1, \dots, C_n$  then  $\Gamma^A$  denotes the corresponding  $A$ -translated multiset.

**Lemma 5.6.3.**

- (i) *Intuitionistic logic proves  $A \rightarrow B^A$ .*
- (ii) *If  $\Gamma \vdash B$  in HA then also  $\Gamma^A \vdash B^A$  in HA.*

**Proof.**

- (i). This is proved by formula induction. We show just the case where  $B$  is some prime

formula  $P$ .

$$\frac{\frac{u : P}{P \vee A} \vee I_1}{P \rightarrow P \vee A} \rightarrow I, u$$

(ii). This is easy by induction of the length of the proof in HA. We give two examples from the canonical proof. Case  $\perp_I$ . Here we need part (i) of the lemma. Say that  $\Gamma \vdash B$  and that the last rule used is  $\perp_I$ . Induction hypothesis is  $\Gamma^A \vdash \perp \vee A$ . By this we derive:

$$\frac{\frac{\frac{\Gamma^A}{\vdots} \perp \vee A}{A} \quad \frac{v : \perp}{A} \quad u : A}{A} \xrightarrow{v, u} A \rightarrow B^A}{B^A}$$

Case  $x = y \rightarrow fx = fy$ . We have to prove  $(x = y \vee A) \rightarrow (fx = fy \vee A)$ :

$$\frac{x = y \vee A \quad \frac{u : x = y \quad x = y \rightarrow fx = fy}{fx = fy} \quad \frac{v : A}{fx = fy \vee A}}{fx = fy \vee A} u, v$$

The pattern from the last derivation goes through in all the arithmetical rules and axioms of HA. The induction rule is verified by itself since it can be applied directly on the derivations given by the induction hypothesis.  $\dashv$

Let us continue remark 5.6.2 by looking at the problematic aspect of the  $A$ -translation in connection with WE-HA<sup>0</sup>. The translation is problematic in case of the weak extensionality rule QF-ER where the assumptions must be quantifier free: By induction hypothesis we would have

$$\frac{\Gamma_{\text{qf}}^A}{\vdots} \forall \mathbf{x} (s\mathbf{x} =_0 t\mathbf{x} \vee A)$$

but we cannot use the derivation pattern from above, and there seem no obvious way to prove  $\forall \mathbf{y} (r[s]\mathbf{y} =_0 r[t]\mathbf{y} \vee A)$ . On the other hand, if we had classical logic at our disposal then we could interpret QF-ER.<sup>7</sup> But of course we cannot use classical logic, but it seems necessary

<sup>7</sup>This is seen from the following equivalences which are classically provable:

$$\begin{aligned} ((P \rightarrow \forall \mathbf{x} (s\mathbf{x} = t\mathbf{x})) \rightarrow (P \rightarrow \forall \mathbf{y} (r[s]\mathbf{y} = r[t]\mathbf{y}))) &\stackrel{(\text{CL})}{\longleftrightarrow} ((P \wedge \neg \forall \mathbf{x} (s\mathbf{x} = t\mathbf{x})) \vee (\neg P \vee \forall \mathbf{y} (r[s]\mathbf{y} = r[t]\mathbf{y}))) \\ P \vee A \rightarrow \forall \mathbf{x} (s\mathbf{x} = t\mathbf{x} \vee A) &\stackrel{(\text{CL})}{\longleftrightarrow} (\neg P \wedge \neg A) \vee (\forall \mathbf{x} (s\mathbf{x} = t\mathbf{x}) \vee A) \\ \neg (P \vee A \rightarrow \forall \mathbf{y} (r[s]\mathbf{y} = r[t]\mathbf{y} \vee A)) &\stackrel{(\text{CL})}{\longleftrightarrow} P \wedge (\neg \forall \mathbf{y} (r[s]\mathbf{y} = r[t]\mathbf{y}) \wedge \neg A) \end{aligned}$$

By assuming the left-hand side of these formulas it is easy to conclude  $\perp$  and therefore, classically,  $P \vee A \rightarrow \forall \mathbf{y} (r[s]\mathbf{y} = r[t]\mathbf{y} \vee A)$  follows from QF-ER and the induction hypothesis.

for the verification to have  $\forall x(P(x) \vee A) \rightarrow \forall xP(x) \vee A$ , which is not intuitionistically valid. We therefore *conjecture* that the  $A$ -translation does not work for WE-HA<sup>o</sup>. However, there may be some special trick such that a refined  $A$ -translation can avoid this problem.

The strategy of applying negative translation +  $A$ -translation + modified realisability has another disadvantage: It cannot interpret QF-AC, since Markov's principle is used for this. A refined  $A$ -translation as given by Coquand & Hofmann (1999) can, nevertheless, remedy this problem.

### 5.6.1 HA is closed under Markov's rule

In the case of HA we can now show closure of Markov's rule. The first published proof of this is in (Kreisel, 1958, remark 6.1), but the proof we give is due to Friedman (1978) and Dragalin (1980).

**Theorem 5.6.4.** *HA is closed under Markov's rule, i.e. for quantifier free  $A_{\text{qf}}$  we have*

$$HA \vdash \neg\neg\exists x A_{\text{qf}}(x, \mathbf{a}) \Rightarrow HA \vdash \exists x A_{\text{qf}}(x, \mathbf{a}).$$

**Proof.** Suppose  $HA \vdash \neg\neg\exists x A_{\text{qf}}(x, \mathbf{a})$  and that  $FV(A_{\text{qf}}) = \{x, a_1, \dots, a_n\}$ . Let  $t$  be the characteristic term (i.e.  $n + 1$ -ary function) for  $A_{\text{qf}}(x, \mathbf{a})$ . We have  $HA \vdash \neg\neg\exists x(t\mathbf{x}\mathbf{a} = 0)$ . Now we use  $\exists x(t\mathbf{x}\mathbf{a} = 0)$  as translation formula. By soundness of translation (lemma 5.6.3 (ii)) we get

$$HA \vdash (\neg\neg\exists x(t\mathbf{x}\mathbf{a} = 0))^{\exists x(t\mathbf{x}\mathbf{a}=0)}$$

And since  $B \leftrightarrow \perp \vee B$  for any formula  $B$ :

$$HA \vdash (\exists x((t\mathbf{x}\mathbf{a} = 0) \vee \exists x(t\mathbf{x}\mathbf{a} = 0)) \rightarrow \exists x(t\mathbf{x}\mathbf{a} = 0)) \rightarrow \exists x(t\mathbf{x}\mathbf{a} = 0).$$

Intuitionistic logic proves  $\exists y(B(y) \vee \exists yB(y)) \rightarrow (\exists yB(y) \vee \exists yB(y))$  for any formula  $B(y)$ . This leads to

$$HA \vdash (\exists x(t\mathbf{x}\mathbf{a} = 0) \vee \exists x(t\mathbf{x}\mathbf{a} = 0) \rightarrow \exists x(t\mathbf{x}\mathbf{a} = 0)) \rightarrow \exists x(t\mathbf{x}\mathbf{a} = 0).$$

Now, for any  $B$  we have by logic  $B \vee B \rightarrow B$  and consequently,

$$HA \vdash \exists x(t\mathbf{x}\mathbf{a} = 0),$$

which is equivalent to the conclusion of the theorem.  $\dashv$

Essential in the proof is that we have characteristic terms for quantifier free formulas. This is what makes it possible to prove the theorem for quantifier free formulas. In case one works with a theory where this is not possible one gets a weaker theorem where the quantifier free  $A_{\text{qf}}$  is exchanged by any prime formula  $P$ .

Within natural deduction one could be tempted to formulate the theorem with assumptions  $\Gamma$ . These would then also be translated into  $\Gamma^{\exists x(t[x]=0)}$ . In case  $\Gamma$  consists of purely universal formulas this would be valid, since  $\Gamma^{\exists x(t[x]=0)}$  would follow from  $\Gamma$ . Accordingly we see (again) that proofs of universal lemmata have no influence on the computational content.

### 5.6.2 Extraction theorem for PA by negative translation, A-translation and m.r.

The idea of extracting programs realising  $\Pi_2^0$  formulas provable in PA is a part of Kreisel's program, and the following theorem displays a strategy in this direction that U. Berger and H. Schwichtenberg have developed and refined, see for instance (Berger & Schwichtenberg, 1995).

**Theorem 5.6.5.** (Program extraction for PA). *Let  $A_{\text{qf}}$  be a quantifier free formula of  $\mathcal{L}(\text{PA})$  containing only  $x$  and  $y$  free.*

$$\text{If } \text{PA} \vdash \forall x \exists y A_{\text{qf}}(x, y) \text{ then } \text{HA} \vdash \forall x A_{\text{qf}}(x, fx)$$

for a closed term  $f$  of type 1 of Gödel's system T.

**Proof.** Suppose  $\text{PA} \vdash \forall x \exists y A_{\text{qf}}(x, y)$ , by Kuroda's negative translation we get  $\text{HA} \vdash \neg \neg \forall x \neg \neg \exists y A_{\text{qf}}(x, y)$ . This is by intuitionistic logic the same as  $\text{HA} \vdash \forall x \neg \neg \exists y A_{\text{qf}}(x, y)$ . Using  $\exists y (t_{A_{\text{qf}}}.xy = 0)$ -translation, theorem 5.6.4 yields

$$\text{HA} \vdash \forall x \exists y A_{\text{qf}}(x, y).$$

Via soundness of  $\underline{\text{mr}}$ -translation (theorem 5.2.2) we extract by recursion on the last derivation a term  $f^1$  doing what is required.  $\dashv$

Note, that we also have a proof of conservativity for  $\Pi_2^0$ -sentences of PA over HA. It is negative translation and A-translation which are responsible for this.

## 5.7 Realisability with truth: Closure properties

We have just seen three applications of modified realisability. Firstly, we showed non-derivability of Markov's principle and thereafter – using also negative translation – that  $\text{E-HA}^\omega$  and various extensions are not closed under Markov's rule.<sup>8</sup> In our third application modified realisability together with negative translation and A-translation made it possible to unwind computational content from classical proofs in PA of  $\Pi_2^0$ -statements. In this section a fourth application of realisability will be introduced. By a “modified realisability with truth” translation closure of  $\text{H}^*$  under different rules will be shown, where  $\text{H}^*$  is any of the theories  $\text{E-HA}^\omega \pm \text{IP}_{\text{ef}}^\omega \pm \text{AC}$ . We will show:

- disjunction property (DP),
- existence property (EP),
- closure under rule of choice (ACR) and
- closure under the rule of independence-of-premise for  $\exists$ -free formulas ( $\text{IP}_{\text{ef}}^\omega$ )

for  $\text{H}^*$ .

---

<sup>8</sup>To be accurate, we used the notion of majorizability in that proof too, since this notion is used in Howard's proof.

For reasons which become obvious in connection with the next soundness theorem we will in this section work with Hilbert style. The following definition can be found in (Cook & Urquhart, 1993), but is an outgrowth of similar realisability translations. For a systematic approach to truth translations and historical notes see (Troelstra, 1998).

**Definition 5.7.1.** (Modified realisability with truth). Suppose that  $A(\mathbf{a})$  is a formula of  $\mathcal{L}(E\text{-}HA^\omega)$  with  $FV(A) = \{\mathbf{a}\}$ . Then  $\mathbf{x}\underline{\text{mrt}}A(\mathbf{a})$  is defined inductively, where  $\{\mathbf{x}\} \cap \{\mathbf{a}\} = \emptyset$  and  $FV(\exists \mathbf{x}(\mathbf{x}\underline{\text{mrt}}A(\mathbf{a}))) = \{\mathbf{a}\}$ .

The clauses for prime formulas,  $\wedge$ ,  $\vee$ ,  $\exists$  and  $\forall$  are the same as for  $\underline{\text{mr}}$ . Implication is translated in the following way:

$$(\rightarrow^{\underline{\text{mrt}}}) \quad \mathbf{x}\underline{\text{mrt}}(A(\mathbf{a}) \rightarrow B(\mathbf{b})) := \forall \mathbf{y}(\mathbf{y}\underline{\text{mrt}}A(\mathbf{a}) \rightarrow \mathbf{xy}\underline{\text{mrt}}B(\mathbf{b})) \wedge (A \rightarrow B).$$

◁

This truth translation has the crucial truth property, which also explains its name: An interpreted formula implies the original formula.

**Lemma 5.7.2.** (Truth property).

$$E\text{-}HA^\omega \vdash (\mathbf{t}\underline{\text{mr}}A) \rightarrow A,$$

for any sequence  $\mathbf{t}$  of terms.

**Proof.** The proof is straightforward by induction on the complexity of  $A$  and since the free variables do not play any role we omit them. For the *base case*, where  $A$  is a prime formula there is nothing to prove, since the translation leaves prime formulas unaffected. The *induction case* is also easy. We only show the case where  $A \equiv \exists x^\sigma B(x)$ .

First note that  $t^\sigma, \mathbf{s}\underline{\text{mrt}}\exists x^\sigma B(x)$  is literally  $\mathbf{s}\underline{\text{mrt}}B(t^\sigma)$ . From the induction hypothesis we get  $\mathbf{s}\underline{\text{mrt}}B(t^\sigma) \rightarrow B(t^\sigma)$ . Consequently:

$$\frac{(\mathbf{s}\underline{\text{mrt}}B(t^\sigma)) \rightarrow B(t^\sigma) \quad B(t^\sigma) \rightarrow \exists x^\sigma B(x)}{(\mathbf{s}\underline{\text{mrt}}B(t^\sigma)) \rightarrow \exists x^\sigma B(x)} \text{ Syl}$$

i.e.  $(t^\sigma, \mathbf{s}\underline{\text{mrt}}\exists x^\sigma B(x)) \rightarrow \exists x^\sigma B(x)$ .

The rest of the inductive cases are similar, except for implication which follows directly from the definition of  $\underline{\text{mrt}}$ . ◀

The following theorem is in its essence found in (Troelstra, 1973). See also (Cook & Urquhart, 1993, 155–158) for a full proof.

**Theorem 5.7.3.** (Soundness of  $\underline{\text{mrt}}$ -translation). Let  $H^*$  be any of the theories  $E\text{-}HA^\omega \pm IP_{\text{ef}}^\omega \pm AC$ , and let  $A$  be a formula of  $\mathcal{L}(E\text{-}HA^\omega)$ .

$$\text{If } H^* \vdash A \text{ then } H^* \vdash \mathbf{t}\underline{\text{mrt}}A$$

for some extractable sequence of terms  $\mathbf{t}$  with  $FV(\mathbf{t}) \subseteq FV(A)$ .

The proof is straightforward and is very much the same as the proof of the soundness theorem for  $\underline{\text{mrt}}$ . But as the translation of nested implications is a little complicated it is not a good idea to prove the theorem within natural deduction. In the case of Hilbert style one only has to be careful when considering the rules exportation and importation. No serious difficulties show up, though one will need lemma 5.7.2.

**Corollary 5.7.4.** (Closure properties). *Let  $H^*$  be any of the theories  $E\text{-HA}^\omega \pm \text{IP}_{\text{ef}}^\omega \pm AC$ , then:*

1.  $H^*$  has existence property, i.e. if  $H^* \vdash \exists x^\sigma A(x)$  then  $H^* \vdash A(t^\sigma)$  for some extractable term  $t$  with  $\text{FV}(t) \subseteq \text{FV}(A) \setminus \{x\}$ .
2.  $H^*$  has disjunction property, i.e. for closed  $A$  and  $B$ ,

$$\text{if } H^* \vdash A \vee B \text{ then } H^* \vdash A \text{ or } H^* \vdash B.$$

3.  $H^*$  is closed under the rule of choice (ACR):

$$\text{if } H^* \vdash \forall x^\sigma \exists y^\tau A(x, y) \text{ then } H^* \vdash \exists Y^{\sigma\tau} \forall x^\sigma A(x, Yx).$$

4.  $H^*$  is closed under the rule of independence-of-premise for  $\exists$ -free formulas ( $\text{IPR}_{\text{ef}}^\omega$ ):

$$H^* \vdash A \rightarrow \exists y^\sigma B(y) \text{ implies } H^* \vdash \exists y^\sigma (A \rightarrow B(y)),$$

$y \notin \text{FV}(A)$  and  $A$  is  $\exists$ -free.

It is, of course, not surprising that for instance  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega$  is closed under  $\text{IPR}_{\text{ef}}^\omega$ , but we have stated the theorem in the form above for the sake of uniformity.

**Proof.**

1. If  $H^*$  proves  $\exists x^\sigma A(x)$  then by soundness  $H^*$  proves  $t^\sigma$ ,  $\underline{\text{s mrt}} \exists x A(x)$ . This is  $\underline{\text{s mrt}} A(t^\sigma)$  and therefore by lemma 5.7.2,  $H^*$  proves  $A(t)$ .
2. This follows from 1 using the fact that every closed number term  $t^0$  of  $T$  can (provably in  $H^*$ ) be evaluated to a numeral.
3. This also follows from 1, but we give another proof. If  $H^*$  proves  $\forall x^\sigma \exists y^\tau A(x, y)$  then by soundness of  $\underline{\text{mrt}}$   $H^*$  proves  $\forall x^\sigma (\underline{\text{s x mrt}} A(x, t^{\sigma\tau x}))$ . Lemma 5.7.2 yields  $\forall x^\sigma A(x, t^{\sigma\tau x})$  and thus we introduce the existential quantifier.
4. This is also straightforward using soundness and truth condition.

□

The theorem has interesting applications together with the notion of majorizability. Combining these two Kohlenbach (1992a) has shown that systems like  $E\text{-HA}^\omega$  are closed under the so-called fan rule. For a proof of this application see (Troelstra, 1998, 434–436).

Theorem 5.7.4 shows that  $H^*$  for any of the theories above has a strong constructive character in the sense that it possesses the important constructive properties expressed by the

theorem. However, in case of the strongest theory these properties can also be proved by *modified realisability*, due to the observation:

**Lemma 5.7.5.** (Characterisation of  $\underline{\text{mr}}$ ).

$$E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} \vdash A \leftrightarrow \exists \mathbf{x}(\mathbf{x} \underline{\text{mr}} A)$$

The proof is by induction on the complexity of  $A$ , see (Troelstra, 1973, 217).

Using this lemma we can prove the following theorem, where  $\Gamma$  denotes any set of true  $\exists$ -free sentences.

**Theorem 5.7.6.** *The theory  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} + \Gamma$  has disjunction property, existence property and is closed under the rules  $\text{ACR}$  and  $\text{IPR}_{\text{ef}}^\omega$ .*

**Proof.** The proof is immediate from lemma 5.7.5 and theorem 5.2.2. That any set of true  $\exists$ -free sentences can be added freely is due to the observation that such a set has no influence on the  $\underline{\text{mr}}$ -realisers, see page 75.  $\dashv$

All this shows that modified realisability with truth can be used to show that a whole sequence of theories from  $E\text{-HA}^\omega$  to  $E\text{-HA}^\omega + \text{IP}_{\text{ef}}^\omega + \text{AC} + \Gamma$  have a strong constructive flavour. But note, that none of the theories are closed under Markov's rule—not even for type 0.

It is, however, natural to ask whether functional interpretation can be used to show closure properties in the same way as realisability. In the next chapter we will see that this is fact the case.

## Closure under Rules by Functional Interpretations

We will now come back to some of the questions Gödel (1941, 1938) originally pursued when he developed the Dialectica interpretation. Some of these were – as the title of the 1941 lecture says – on the constructivity of intuitionistic logic. One of the applications Gödel had in mind (see also chapter 1, page 11 of this text) was

to answer the question in which sense intuitionistic logic as applied in number theory, or more generally in any theory with decidable primitive terms, is constructive. Namely, if you are able to derive in intuitionistic number theory an existential proposition  $(\exists x)\varphi(x)$ , then ... we can find a term  $t$  composed of the functions of the system  $\Sigma$  such that  $\varphi'(t)$ . (Gödel, 1941, 26–27).

The notation used by Gödel here is that  $\varphi'$  corresponds to our  $\varphi_D$  and  $\Sigma$  is almost the system T of (Gödel, 1958), except that the statements are in  $\exists\forall$  form. That  $\Sigma$  proves  $\varphi'(t)$  corresponds to the case where the quantifier free fragment of typed Heyting arithmetic proves  $\varphi_D(t, y)$ . It seems as if Gödel wants to show that intuitionistic number theory has the existence property, but is able to show only the weaker result where the translated formula is witnessed. It is weaker since  $\varphi$  does not follow intuitionistically from  $\varphi^D$ .

Until now it has been an open question whether or not interpretations in the style of Dialectica can be used to show existence property and disjunction property for Heyting arithmetic. In the foregoing chapter we have seen that modified realisability with truth shows elegantly  $E\text{-HA}^\omega$  to have these properties. But can this also be done by Dialectica (in the case of  $WE\text{-HA}^\omega$ ) or is there a principle obstacle inherent in the D-translation?

### 6.1 There is no ‘Dialectica-with-truth’ interpretation of $WE\text{-HA}^\omega$

The simplicity of  $\underline{\text{mrt}}$  is striking; likewise the effective application. It is therefore natural to ask whether or not there is a corresponding ‘Dialectica-with-truth’,  $(\cdot)^{Dt}$ . As we will see below, due to a simple counterexample, there is not. But apart from showing that there is no such  $Dt$ -interpretation, the counterexample will also display something about the differences between modified realisability and Dialectica: Quantifiers occurring negatively in a formula are forgotten, so to speak, by realisability whereas Dialectica is requested to provide witnesses. This is the reason why the translation ‘with-truth’ does not work in connection with Dialectica.

A Dialectica-with-truth would be defined as the standard Dialectica translation except for implication. A formula with “ $\rightarrow$ ” as the outermost symbol would be translated according to:

$$(A \rightarrow B)^{Dt} := \exists UY \forall xv ((A_{Dt}(x, Yxv) \rightarrow B_{Dt}(Ux, v)) \wedge (A \rightarrow B)).$$

But a soundness theorem for this translation cannot be established.

**Theorem 6.1.1.** *The exportation rule is not sound under Dialectica-with-truth translation.*

**Proof.** Let  $A(x^0)$  be the quantifier free formula  $Tzzx$  where  $T$  is Kleene’s  $T$ -predicate. Then certainly we can derive  $\forall x\neg A(x) \wedge \exists xA(x) \rightarrow \perp$ , and therefore

$$\frac{\forall x\neg A(x) \wedge \exists xA(x) \rightarrow \perp}{\forall x\neg A(x) \rightarrow \neg\exists xA(x)} \text{Expo} \quad (6.1)$$

is provable within Hilbert style. The  $Dt$ -translation of the premise is (equivalent to)

$$\exists f\forall y((\neg A(fy) \wedge A(y) \rightarrow \perp) \wedge (\forall x\neg A(x) \wedge \exists xA(x) \rightarrow \perp)).$$

This is interpretable by taking, say,  $\lambda z, y.y$ . But the  $Dt$ -translation of the conclusion is (equivalent to)

$$\exists f\forall y((\neg A(fy) \rightarrow (\neg A(y) \wedge \neg\exists xA(x))) \wedge (\forall x\neg A(x) \rightarrow \neg\exists xA(x))).$$

If this were interpretable then there would be some closed term  $F$  of type level 1 witnessing  $f$  taking the free variable  $z$  and  $y$  as arguments. This would imply provability of

$$\forall y, z(\neg Tzz(Fzy) \rightarrow \neg Tzzy \wedge \neg\exists xTzzx).$$

But then the closed term  $g := \lambda z.Fz0$  of type 1 would decide  $\exists xTzzx$  for any given  $z$  which is absurd. Thus, the exportation rule is not sound under the  $Dt$ -translation.  $\dashv$

On the other hand, if we compare this with mrt, both the premise and the conclusion of (6.1) are easy to mrt interpret since we have for any  $B$  that  $(\ )_{\text{mrt}} \neg B$ . Therefore intuitionistic logic shows  $(\ )_{\text{mrt}} \forall x\neg P(x) \wedge \exists xP(x) \rightarrow \perp$ . From this the mrt translation of the conclusion of (6.1) follows.<sup>1</sup> Hence, the truth translation works for modified realisability just because it ‘forgets’ the negatively occurring universal quantifiers.

This points towards:

1. The innocent looking exportation rule is indeed non-trivial. The soundness of mrt is straightforward to verify, whereas there is no ‘Dialectica-with-truth’ because exportation is unsound for this translation.
2. We have seen that modified realisability and Dialectica validate different principles; basically because they interpret implication differently. This difference has consequences both philosophically and mathematically. It could, however, still be that they both can be seen as different parts of a more general method for extracting computational content from proofs. But this is not a plausible view. The counterexample to a Dialectica-with-truth shows that Dialectica and (modified) realisability are interpretations which

<sup>1</sup>The definition of mrt implies that  $(\ )_{\text{mrt}} \forall x\neg P(x) \wedge \exists xP(x) \rightarrow \perp$  is equivalent to

$$\forall y(\neg(\forall x\neg P(x) \wedge P(y)) \wedge \neg(\forall x\neg P(x) \wedge \exists xP(x))),$$

and  $(\ )_{\text{mrt}} \forall x\neg P(x) \rightarrow \neg\exists xP(x)$  is equivalent to

$$(\forall x\neg P(x) \rightarrow \forall y\neg P(y) \wedge \neg\exists zP(z)) \wedge (\forall x\neg P(x) \rightarrow \neg\exists xP(x)).$$

The last formula can be further reduced to its second conjunct.

are structurally different in an essential way: Every important notion of realisability has a truth variant (see (Troelstra, 1998)), which is contrary to Dialectica. We therefore reject the view that the Dialectica interpretation is by the end of the day just another type of realisability.

The question whether or not Dialectica can be used to show existence property, disjunction property and so on for intuitionistic arithmetic is still unanswered. There is, however, another problem which the counterexample above did not show. For the Dialectica interpretation we need decidability of prime formulas. From this follows decidability of  $A_D$  for any  $A$ , which is needed in case contraction is involved. But  $A_{D_t}$  is in general not quantifier free and therefore not decidable. Accordingly, in our further investigation of the possibilities of showing closure under rules we will turn to the Diller-Nahm variant of the Dialectica interpretation which does not need decidability of prime formulas. Now, with a variant of the Diller-Nahm interpretation we can show existence property, disjunction property and closure under different rules for intuitionistic arithmetic.

## 6.2 Definition and soundness of $Q$ -translation

Kleene (1969) used a q-variant of realisability to obtain derived rules of intuitionistic analysis with function variables. Later Troelstra (1973) developed and applied q-variants for different kinds of realisability – including modified realisability – in order to show closure properties. Accordingly, the closure properties of  $E\text{-HA}^\omega$  which we showed in the foregoing chapter using  $\text{mrt}$  were originally shown using q-realisability. But q-realisability is not closed under deductions (more on this below). Therefore the q-variant is – with respect to realisability – now widely replaced by the ‘truth’-variant.

Contrary to the truth variant, it is possible to transfer the idea of a q-variant of realisability into a q-variant of the Diller-Nahm interpretation. We will call this variant  $Q$ . For the  $Q$ -translation we introduce bounded universal quantification as a *defined* notion. For the Diller-Nahm interpretation (see page 48) we had the bounded quantifier as a primitive notion, since we wanted an interpretation in  $\text{WE-T}$  extended by bounded quantification. This time, however, we interpret only into  $\text{WE-HA}^\omega$  and therefore we take the bounded universal quantifier as a defined notion:

$$(\forall x < t)A(x) ::= \forall x(x < t \rightarrow A(x)).$$

In the following definition we will omit the free variables. They do not play any significant role and are treated as under the standard Dialectica translation. To each formula  $A$  of  $\mathcal{L}(\text{WE-HA}^\omega)$  we associate its  $Q$ -translation  $A^Q$

$$A^Q ::= \exists \mathbf{x} \forall \mathbf{y} A_Q(\mathbf{x}, \mathbf{y}), \quad (6.2)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are sequences of fresh variables; but in contrast to the Dialectica translation  $A_Q$  is *not* quantifier free.

**Definition 6.2.1.** ( $Q$ -translation).

$$\begin{aligned}
(P^Q) \quad A^Q &::= A_Q ::= A, \text{ if } A \text{ is prime,} \\
(\wedge^Q) \quad (A \wedge B)^Q &::= \exists \mathbf{x}, \mathbf{u} \forall \mathbf{y}, \mathbf{v} (A_Q(\mathbf{x}, \mathbf{y}) \wedge B_Q(\mathbf{u}, \mathbf{v})), \\
(\vee^Q) \quad (A \vee B)^Q &::= \exists z^0, \mathbf{x}, \mathbf{u} \forall \mathbf{y}, \mathbf{v} ((z = 0 \rightarrow A_Q(\mathbf{x}, \mathbf{y}) \wedge A) \wedge \\
&\quad (z \neq 0 \rightarrow B_Q(\mathbf{u}, \mathbf{v}) \wedge B)), \\
(\exists^Q) \quad (\exists z^\sigma A(z))^Q &::= \exists z^\sigma, \mathbf{x} \forall \mathbf{y} (A_Q(\mathbf{x}, \mathbf{y}, z) \wedge A(z)), \\
(\forall^Q) \quad (\forall z^\sigma A(z))^Q &::= \exists \mathbf{X} \forall \mathbf{y}, z^\sigma A_Q(\mathbf{X}z, \mathbf{y}, z), \\
(\rightarrow^Q) \quad (A \rightarrow B)^Q &::= \exists W, \mathbf{U}, \mathbf{Y}, \forall \mathbf{x}, \mathbf{v} ((\forall w^0 < W \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, \mathbf{Y} w \mathbf{x} \mathbf{v}) \wedge A \rightarrow B_Q(\mathbf{U} \mathbf{x}, \mathbf{v})).
\end{aligned}$$

◁

The crucial clauses in the definition are  $\exists^Q$  and  $\rightarrow^Q$  (since  $\vee$  and  $\exists, \wedge$  are inter-definable). Among other things, we aim at the rule: if  $\exists x A(x)$  is provable then  $A(t)$  is provable for some term  $t$ . This is the motivation behind  $\exists^Q$ : Given that we have soundness for the  $Q$ -translation the argument would be: If  $\exists x A(x)$  is provable then by soundness there are terms such that  $\forall \mathbf{y} A_Q(\mathbf{t}, \mathbf{y}, t_0) \wedge A(t_0)$ ; by taking the second conjunct we have the required. The definition of  $\rightarrow^Q$  is designed to make this possible. Compared to the standard Diller-Nahm translation (see page 48) the  $Q$ -translation of implication makes the premise stronger by adding ‘ $\wedge A$ ’. But precisely this modification makes it possible to  $Q$ -interpret  $\exists$ -introduction.<sup>2</sup>

We will need some basic arithmetic in order to show soundness of  $Q$ -translation. Recall from page 20 the definitions of  $\dot{+}$ ;  $+$ ;  $\text{sg}$  and so on. We will furthermore need a primitive recursive functional of type  $(00)00$  taking a finite list of terms and giving the maximum of that list. We will write the list as  $\{t[x]^0 \mid x < s^0\}$ . The functional has the following properties:

$$\max\{t[x] \mid x < 0\} =_0 0, \quad \max\{t[x] \mid x < Ss\} =_0 \max(\max\{t[x] \mid x < s\}, t[s])$$

The functional is defined by iterating the maximum of two numbers defined earlier. We will also need a monotone pairing functional  $\langle \cdot, \cdot \rangle$  of type  $000$  with inverses  $j_1$  and  $j_2$  of type  $00$  having the following properties:

$$\langle 0, 0 \rangle =_0 0, \quad \langle 0, St \rangle =_0 \langle 0, t \rangle + St, \quad \langle s, t \rangle =_0 \langle 0, s+t \rangle + s$$

and

$$j_1 \langle s, t \rangle =_0 s, \quad j_2 \langle s, t \rangle =_0 t, \quad \langle j_1 r, j_2 r \rangle =_0 r.$$

These are not difficult to define in WE-T, see e.g. (Diller & Nahm, 1974, 51–52) or (Kleene, 1952, 222–223).

**Lemma 6.2.2.**

(i) If  $x \notin \text{FV}\{A\}$ , then

$$\text{WE-HA}^\omega \vdash A \wedge B(x) \rightarrow C(x) \Rightarrow \text{WE-HA}^\omega \vdash A \wedge (\forall x < t) B(x) \rightarrow (\forall x < t) C(x).$$

<sup>2</sup>We note that the  $q$ -variant is in general different from the truth-variant. We do not need to, neither can, show the truth condition (lemma 5.7.2) in order to prove the soundness theorem or the closure properties.

(ii)

$$\begin{aligned}
& WE-HA^\omega \vdash (\forall x < t) Sx \dot{-} t =_0 0, \\
& WE-HA^\omega \vdash s \dot{-} t =_0 0 \wedge (\forall x < t) A(x) \rightarrow (\forall x < s) A(x), \\
& WE-HA^\omega \vdash Ss \dot{-} t =_0 0 \wedge (\forall x < t) A(x) \rightarrow A(s), \\
& WE-HA^\omega \vdash (\forall x < s+t) A(x) \rightarrow (\forall x < t) A(s+x), \\
& WE-HA^\omega \vdash (\forall x < t) (A(x) \wedge B(x)) \rightarrow ((\forall x < t) A(x) \wedge (\forall x < t) B(x)).
\end{aligned}$$

(iii) If  $x \notin FV\{A\}$ , then  $WE-HA^\omega$  proves

$$(\forall x < \langle s, \max\{t[y] \mid y < s\} \rangle) A(j_1x, j_2x) \rightarrow (\forall y < s)(\forall z < t[y]) A(y, z).$$

**Proof.** See (Diller & Nahm, 1974, 52). ⊣

Soundness is proved within Hilbert style:

**Theorem 6.2.3.** (Soundness of  $Q$ -translation). Let  $H^\omega$  be  $WE-HA^\omega \pm IP_\forall^\omega$ . If  $A$  is a formula of  $\mathcal{L}(WE-HA^\omega)$  with  $FV(A) = \{\mathbf{a}\}$ , then

$$H^\omega \vdash A(\mathbf{a}) \text{ implies } H^\omega \vdash \forall \mathbf{y} A_Q(\mathbf{T}\mathbf{a}, \mathbf{y}, \mathbf{a}),$$

for suitable sequence  $\mathbf{T}$  of closed terms, which can be extracted from a  $H^\omega$ -proof of  $A(\mathbf{a})$ .**Proof.** We give the proof by verifying axioms and rules from the system given by Spector (1962) (as presented here in chapter 2). There are no particular obstacles in the proof when compared to the other soundness proofs we have seen and it has many similarities with the proof of soundness for the Diller-Nahm translation; see for instance (Diller & Nahm, 1974, 57–59). However, there is a special phenomenon which all kinds of q-variants share. In order to interpret modus ponens, syllogism and induction one needs sub-derivations from the original derivations in order to prove the interpretation. This phenomenon is due to the weakening of the premise which occurs when an implication is translated. Consequently, we need every now and then the original derivations.

In the proof we will often use without mentioning the following rule (in Hilbert style):

$$\frac{A_{\text{qf}} \rightarrow s =_\sigma t \quad B(s)}{A_{\text{qf}} \rightarrow B(t)} \quad (6.3)$$

where  $A_{\text{qf}}$  is quantifier free. The rule is a generalisation of lemma 3.4.1 and is derivable from QF-ER by an induction argument on the complexity of  $B$ . In case  $B$  is a prime formula the assertion is immediate from QF-ER. In the general case one uses

$$(C \leftrightarrow D) \rightarrow (B[A := C] \leftrightarrow B[A := D])$$

where  $B[A := C]$  results by replacing some sub-formula occurrence of  $A$  in  $B$  by  $C$ .**Case 1.**  $A(\mathbf{a}) \rightarrow A(\mathbf{a})$ . The translation is

$$\exists W, \mathbf{Y}, \mathbf{U} \forall \mathbf{x}, \mathbf{v} ((\forall w < W \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, \mathbf{Y} w \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(\mathbf{U} \mathbf{x}, \mathbf{v}, \mathbf{a})).$$

Define the following three sequences of closed terms:

$$T_1 := \lambda \mathbf{a}, \mathbf{x}, \mathbf{v}. 1^0, \quad T_2 := \lambda \mathbf{a}, w, \mathbf{x}, \mathbf{v}. \mathbf{v}, \quad T_3 := \lambda \mathbf{a}, \mathbf{x}. \mathbf{x}.$$

With these definitions we start from an ‘axiom of weakening’; use (6.3) and  $(\forall x < 1)B(x) \rightarrow B(0)$ . The following can easily be transformed into a proof within WE-HA<sup>0</sup>:

$$\frac{\frac{\frac{A_Q(\mathbf{x}, \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(\mathbf{x}, \mathbf{v}, \mathbf{a})}{A_Q(\mathbf{x}, T_2 \mathbf{a} 0 \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(T_3 \mathbf{a} \mathbf{x}, \mathbf{v}, \mathbf{a})}}{(\forall w < 1) A_Q(\mathbf{x}, T_2 \mathbf{a} w \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(T_3 \mathbf{a} \mathbf{x}, \mathbf{v}, \mathbf{a})}}{(\forall w < T_1 \mathbf{a} \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, T_2 \mathbf{a} w \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(T_3 \mathbf{a} \mathbf{x}, \mathbf{v}, \mathbf{a})}}{\forall \mathbf{x}, \mathbf{v} (\forall w < T_1 \mathbf{a} \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, T_2 \mathbf{a} w \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A \rightarrow A_Q(T_3 \mathbf{a} \mathbf{x}, \mathbf{v}, \mathbf{a})}$$

Then  $T_1$ ,  $T_2$  and  $T_3$  witness the three existential quantifiers, respectively.

**Case 2.**  $A(\mathbf{a}) \rightarrow A(\mathbf{a}) \vee B(\mathbf{b})$ . Without writing the free variables the translation is:

$$\exists W, \mathbf{Y}, Z, \mathbf{X}, \mathbf{U} \forall \mathbf{x}, \mathbf{y}, \mathbf{v} (\forall w < W \mathbf{x} \mathbf{y} \mathbf{v}) A_Q(\mathbf{x}, \mathbf{Y} w \mathbf{x} \mathbf{y} \mathbf{v}) \wedge A \rightarrow (Z \mathbf{x} = 0 \rightarrow A_Q(\mathbf{X} \mathbf{x}, \mathbf{y}) \wedge A) \wedge (Z \mathbf{x} \neq 0 \rightarrow B_Q(\mathbf{U} \mathbf{x}, \mathbf{v}) \wedge B).$$

The following derivation can clearly be transformed into a derivation in Hilbert style:

$$\frac{\frac{\frac{u : A_Q(\mathbf{x}, \mathbf{y}) \wedge A}{0 = 0 \rightarrow A_Q(\mathbf{x}, \mathbf{y}) \wedge A} \quad \frac{\frac{v : 0 \neq 0}{B_Q(\mathbf{o}, \mathbf{v}) \wedge B}}{0 \neq 0 \rightarrow B_Q(\mathbf{o}, \mathbf{v}) \wedge B} v}{(0 = 0 \rightarrow A_Q(\mathbf{x}, \mathbf{y}) \wedge A) \wedge (0 \neq 0 \rightarrow B_Q(\mathbf{o}, \mathbf{v}) \wedge B)}}{A_Q(\mathbf{x}, \mathbf{y}) \wedge A \rightarrow (0 = 0 \rightarrow A_Q(\mathbf{x}, \mathbf{y}) \wedge A) \wedge (0 \neq 0 \rightarrow B_Q(\mathbf{o}, \mathbf{v}) \wedge B)} u$$

Let  $\mathbf{c}$  equal  $\mathbf{a}, \mathbf{b}$  and define

$$\begin{aligned} T_1 &:= \lambda \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{v}. 1^0, & T_2 &:= \lambda \mathbf{c}, w, \mathbf{x}, \mathbf{y}, \mathbf{v}. \mathbf{y}, \\ T_3 &:= \lambda \mathbf{c}, \mathbf{x}. 0^0, & T_4 &:= \lambda \mathbf{c}, \mathbf{x}, \mathbf{x}, \\ T_5 &:= \tilde{\mathbf{o}}, \end{aligned}$$

where  $\tilde{\mathbf{o}}$  are the zero functionals fitting in type. Then by using (6.3) and the properties of the bounded universal quantifier the following is derivable in WE-HA<sup>0</sup>:

$$\forall \mathbf{x}, \mathbf{y}, \mathbf{v} (\forall w < T_1 \mathbf{c} \mathbf{x} \mathbf{y} \mathbf{v}) A_Q(\mathbf{x}, T_2 \mathbf{c} w \mathbf{x} \mathbf{y} \mathbf{v}) \wedge A \rightarrow (T_3 \mathbf{c} \mathbf{x} = 0 \rightarrow A_Q(T_4 \mathbf{c} \mathbf{x}, \mathbf{y}) \wedge A) \wedge (T_3 \mathbf{c} \mathbf{x} \neq 0 \rightarrow B_Q(T_5 \mathbf{c} \mathbf{x}, \mathbf{v}) \wedge B).$$

**Case 3.**  $B(\mathbf{a}) \rightarrow A(\mathbf{a}) \vee B(\mathbf{b})$ . This is treated symmetrically to the foregoing case.

**Case 4.**  $A(\mathbf{a}) \wedge B(\mathbf{b}) \rightarrow A(\mathbf{a})$ . The translation is

$$\exists W, \mathbf{Y}, \mathbf{V}, \mathbf{X} \forall \mathbf{x}, \mathbf{u}, \mathbf{y} (\forall w < W \mathbf{x} \mathbf{u} \mathbf{y}) (A_Q(\mathbf{x}, \mathbf{Y} w \mathbf{x} \mathbf{u} \mathbf{y}, \mathbf{a}) \wedge B_Q(\mathbf{u}, \mathbf{V} w \mathbf{x} \mathbf{u} \mathbf{y}, \mathbf{b})) \wedge (A \wedge B) \rightarrow A_Q(\mathbf{X} \mathbf{x} \mathbf{u}, \mathbf{y}, \mathbf{a}).$$

Clearly we have

$$(A_Q(\mathbf{x}, \mathbf{y}, \mathbf{a}) \wedge B_Q(\mathbf{u}, \mathbf{o}, \mathbf{b})) \wedge (A \wedge B) \rightarrow A_Q(\mathbf{x}, \mathbf{y}, \mathbf{a}).$$

Write  $\mathbf{c}$  for  $\mathbf{a}, \mathbf{b}$ ; if we define

$$\begin{array}{ll} T_1 & :\equiv \lambda \mathbf{c}, \mathbf{x}, \mathbf{u}, \mathbf{y}. 1^0, & T_2 & :\equiv \lambda \mathbf{c}, w, \mathbf{x}, \mathbf{u}, \mathbf{y}. \mathbf{y}, \\ T_3 & :\equiv \tilde{\mathbf{o}}, & T_4 & :\equiv \lambda \mathbf{c}, \mathbf{x}, \mathbf{u}. \mathbf{x}, \end{array}$$

then we have

$$\forall \mathbf{x}, \mathbf{u}, \mathbf{y} \left( (\forall w < T_1 \mathbf{c} \mathbf{x} \mathbf{u} \mathbf{y}) (A_Q(\mathbf{x}, T_2 \mathbf{c} w \mathbf{x} \mathbf{u} \mathbf{y}, \mathbf{a}) \wedge B_Q(\mathbf{u}, T_3 \mathbf{c} w \mathbf{x} \mathbf{u} \mathbf{y}, \mathbf{b})) \wedge (A \wedge B) \rightarrow A_Q(T_4 \mathbf{c} \mathbf{x} \mathbf{u}, \mathbf{y}, \mathbf{a}) \right).$$

**Case 5.**  $A(\mathbf{a}) \wedge B(\mathbf{b}) \rightarrow B(\mathbf{b})$ . Symmetric to the foregoing case.

**Case 6.**  $\perp \rightarrow A(\mathbf{a})$ . The translation is equivalent to

$$\exists \mathbf{x} \forall \mathbf{y} (\perp \rightarrow A_Q(\mathbf{x}, \mathbf{y}, \mathbf{a})).$$

Any closed term of right type will witness  $\exists \mathbf{x}$ , e.g. the corresponding zero functionals.

**Case 7.** Assume that the last rule of the proof in  $H^\omega$  is modus ponens:

$$\frac{A(\mathbf{a}) \quad A(\mathbf{a}) \rightarrow B(\mathbf{b})}{B(\mathbf{b})} \text{MP}$$

The induction hypothesis is, when we write  $\mathbf{c}$  for  $\mathbf{a} \mathbf{b}$

- (i)  $H^\omega \vdash \forall \mathbf{y} A_Q(T_1 \mathbf{a}, \mathbf{y}, \mathbf{a})$ ,
- (ii)  $H^\omega \vdash \forall \mathbf{x}, \mathbf{v} \left( (\forall w < T_2 \mathbf{c} \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, T_3 \mathbf{c} w \mathbf{x} \mathbf{v}, \mathbf{a}) \wedge A(\mathbf{a}) \rightarrow B_Q(T_4 \mathbf{c} \mathbf{x}, \mathbf{v}, \mathbf{b}) \right)$ ,

for given  $T_1, T_2, T_3$ , and  $T_4$ . We need to find  $T_5$  such that  $H^\omega \vdash \forall \mathbf{v} B_Q(T_5 \mathbf{b}, \mathbf{v}, \mathbf{b})$ . Eliminate  $\forall \mathbf{x}$  in (ii) by  $T_1 \mathbf{a}$  and likewise eliminate  $\forall \mathbf{y}$  in (i) by  $T_3 \mathbf{c} w(T_1 \mathbf{a}) \mathbf{v}$ . Since modus ponens is the last rule of the proof we have a derivation of  $A(\mathbf{a})$  in  $H^\omega$  and this together with lemma 6.2.2 gives us with the instantiated (i)

$$(\forall w < T_2 \mathbf{c}(T_1 \mathbf{a}) \mathbf{v}) A_Q(T_1 \mathbf{a}, T_3 \mathbf{c} w(T_1 \mathbf{a}) \mathbf{v}, \mathbf{a}) \wedge A(\mathbf{a}).$$

By using modus ponens and introducing universal quantifiers we obtain

$$H^\omega \vdash \forall \mathbf{v} B_Q(T_4 \mathbf{c}(T_1 \mathbf{a}), \mathbf{v}, \mathbf{b}).$$

Let  $\mathbf{o}^\sigma$  be the zero functional of type  $\sigma$ ; substitute  $\mathbf{o}$  of the corresponding type for any free variable in  $T_4 \mathbf{c}(T_1 \mathbf{a})$  that do not occur in  $\mathbf{b}$  and name the result  $\mathbf{T}$ . If we define  $T_5 :\equiv \lambda \mathbf{b}. \mathbf{T}$  we have the required result.

**Case 8.** The last rule of the derivation in  $H^\omega$  is syllogism:

$$\frac{A(\mathbf{a}) \rightarrow B(\mathbf{b}) \quad B(\mathbf{b}) \rightarrow C(\mathbf{c})}{A(\mathbf{a}) \rightarrow C(\mathbf{c})} \text{Syl}$$

Due to heavy notation we will omit explicit treatment of the free variables in this case. However, these are easily incorporated from the induction hypothesis. One will only have to substitute  $\mathfrak{o}$  for any free variable that occurs in  $B$  but not in  $A$  nor in  $C$  and proceed as below.

Induction hypothesis is that  $H^\omega$  proves:

$$\forall \mathbf{x}, \mathbf{v} ( (\forall w_1 < T_1 \mathbf{xv}) A_Q(\mathbf{x}, T_2 w_1 \mathbf{xv}) \wedge A \rightarrow B_Q(T_3 \mathbf{x}, \mathbf{v}) ), \quad (6.4)$$

$$\forall \mathbf{u}, \mathbf{q} ( (\forall w_2 < T_4 \mathbf{uq}) B_Q(\mathbf{u}, T_5 w_2 \mathbf{uq}) \wedge B \rightarrow C_Q(T_6 \mathbf{u}, \mathbf{q}) ), \quad (6.5)$$

and we have to find  $T_7, T_8$  and  $T_9$  such that

$$\forall \mathbf{x}, \mathbf{q} ( (\forall w < T_7 \mathbf{xq}) A_Q(\mathbf{x}, T_8 w \mathbf{xq}) \wedge A \rightarrow C_Q(T_9 \mathbf{x}, \mathbf{q}) ).$$

We eliminate  $\forall \mathbf{v}$  in (6.4) by  $T_5 w_2(T_3 \mathbf{x}) \mathbf{q}$  and  $\forall \mathbf{u}$  in (6.5) by  $T_3 \mathbf{x}$  and write  $\mathbf{T}[w_2, \mathbf{x}, \mathbf{q}]$  for  $T_5 w_2(T_3 \mathbf{x}) \mathbf{q}$  and  $t[\mathbf{x}, \mathbf{q}]$  for  $T_4(T_3 \mathbf{x}) \mathbf{q}$ . Then we use lemma 6.2.2 (i) and obtain

$$(\forall w_2 < t[\mathbf{x}, \mathbf{q}]) (\forall w_1 < T_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}]) A_Q(\mathbf{x}, T_2 w_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}]) \wedge A \rightarrow (\forall w_2 < t[\mathbf{x}, \mathbf{q}]) B_Q(T_3 \mathbf{x}, \mathbf{T}[w_2, \mathbf{x}, \mathbf{q}]), \quad (6.6)$$

$$(\forall w_2 < t[\mathbf{x}, \mathbf{q}]) B_Q(T_3 \mathbf{x}, \mathbf{T}[w_2, \mathbf{x}, \mathbf{q}]) \wedge B \rightarrow C_Q(T_6(T_3 \mathbf{x}), \mathbf{q}). \quad (6.7)$$

From the left sub-derivation of the non-interpreted derivation we have  $A \rightarrow B$ . Let  $A^*$  be the first conjunct of the premise of (6.6). By using syllogism on the axiom  $A^* \wedge A \rightarrow A$  and  $A \rightarrow B$  and subsequently the contraction rule on this and (6.6) we obtain:

$$(\forall w_2 < t[\mathbf{x}, \mathbf{q}]) (\forall w_1 < T_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}]) A_Q(\mathbf{x}, T_2 w_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}]) \wedge A \rightarrow (\forall w_2 < t[\mathbf{x}, \mathbf{q}]) B_Q(T_3 \mathbf{x}, \mathbf{T}[w_2, \mathbf{x}, \mathbf{q}]) \wedge B.$$

By applying syllogism to this formula and (6.7) and using lemma 6.2.2 (iii) we get

$$(\forall w < \langle t[\mathbf{x}, \mathbf{q}], \max\{T_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}] \mid w_2 < t[\mathbf{x}, \mathbf{q}]\} \rangle) A_Q(\mathbf{x}, T_2(j_2 w) \mathbf{xT}[j_1 w, \mathbf{x}, \mathbf{q}]) \wedge A \rightarrow C_Q(T_6(T_3 \mathbf{x}), \mathbf{q}).$$

Accordingly we define

$$\begin{aligned} T_7 &::= \lambda \mathbf{x}, \mathbf{q}. \langle t[\mathbf{x}, \mathbf{q}], \max\{T_1 \mathbf{xT}[w_2, \mathbf{x}, \mathbf{q}] \mid w_2 < t[\mathbf{x}, \mathbf{q}]\} \rangle, \\ T_8 &::= \lambda w, \mathbf{x}, \mathbf{q}. T_2(j_2 w) \mathbf{xT}[j_1 w, \mathbf{x}, \mathbf{q}], \\ T_9 &::= \lambda \mathbf{x}. T_6(T_3 \mathbf{x}). \end{aligned}$$

These terms suffice to prove the needed, due to (6.3).

**Case 9.** The contraction rule:

$$\frac{A(\mathbf{a}) \rightarrow B(\mathbf{b}) \quad A(\mathbf{a}) \rightarrow C(\mathbf{c})}{A(\mathbf{a}) \rightarrow B(\mathbf{b}) \wedge C(\mathbf{c})} \text{Con}$$

The induction hypothesis is

$$\begin{aligned} \forall \mathbf{x}, \mathbf{v} ( (\forall w < T_1 \mathbf{abxv}) A_Q(\mathbf{x}, T_2 \mathbf{abw} \mathbf{xv}, \mathbf{a}) \wedge A \rightarrow B_Q(T_3 \mathbf{abx}, \mathbf{v}, \mathbf{b}) ), \\ \forall \mathbf{x}, \mathbf{q} ( (\forall w < T_4 \mathbf{acxq}) A_Q(\mathbf{x}, T_5 \mathbf{acw} \mathbf{xq}, \mathbf{a}) \wedge A \rightarrow C_Q(T_6 \mathbf{acx}, \mathbf{q}, \mathbf{c}) ). \end{aligned}$$

Let  $\mathbf{d}$  equal the sequence consisting of  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ . We have to provide closed terms  $T_7, \mathbf{T}_8, \mathbf{T}_9$  and  $\mathbf{T}_{10}$  such that

$$\forall \mathbf{x}, \mathbf{v}, \mathbf{q} \left( (\forall w < T_7 \mathbf{d} \mathbf{x} \mathbf{v} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_8 \mathbf{d} w \mathbf{x} \mathbf{v} \mathbf{q}, \mathbf{a}) \wedge A \rightarrow B_Q(\mathbf{T}_9 \mathbf{d} \mathbf{x}, \mathbf{v}, \mathbf{b}) \wedge C_Q(\mathbf{T}_{10} \mathbf{d} \mathbf{x}, \mathbf{q}, \mathbf{c}) \right).$$

Take<sup>3</sup>

$$\mathbf{T}_8 := \lambda \mathbf{d}, w, \mathbf{x}, \mathbf{v}, \mathbf{q}. \text{Cond}(\mathbf{T}_2 \mathbf{a} \mathbf{b} w \mathbf{x} \mathbf{v}, \mathbf{T}_5 \mathbf{a} \mathbf{c} (w \div \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v}) \mathbf{x} \mathbf{q}, S w \div \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v}).$$

From lemma 6.2.2 (ii; 2. and 4. line) it follows that

$$\begin{aligned} (\forall w < \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v} + \mathbf{T}_4 \mathbf{a} \mathbf{c} \mathbf{x} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_8 \mathbf{d} w \mathbf{x} \mathbf{v} \mathbf{q}) &\rightarrow (\forall w < \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, \mathbf{T}_2 \mathbf{a} \mathbf{b} w \mathbf{x} \mathbf{v}), \\ (\forall w < \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v} + \mathbf{T}_4 \mathbf{a} \mathbf{c} \mathbf{x} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_8 \mathbf{d} w \mathbf{x} \mathbf{v} \mathbf{q}) &\rightarrow (\forall w < \mathbf{T}_4 \mathbf{a} \mathbf{c} \mathbf{x} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_5 \mathbf{a} \mathbf{c} w \mathbf{x} \mathbf{q}), \end{aligned}$$

and we therefore define

$$\begin{aligned} T_7 &:= \lambda \mathbf{d}, \mathbf{x}, \mathbf{v}, \mathbf{q}. \mathbf{T}_1 \mathbf{a} \mathbf{b} \mathbf{x} \mathbf{v} + \mathbf{T}_4 \mathbf{a} \mathbf{c} \mathbf{x} \mathbf{q}, & \mathbf{T}_9 &:= \lambda \mathbf{d}, \mathbf{x}. \mathbf{T}_3 \mathbf{a} \mathbf{b} \mathbf{x}, \\ \mathbf{T}_{10} &:= \lambda \mathbf{d}, \mathbf{x}. \mathbf{T}_6 \mathbf{a} \mathbf{c} \mathbf{x}. \end{aligned}$$

Using syllogism and contraction gives the required.

**Case 10.** Assume the last rule of the proof in  $H^\omega$  to be exportation:

$$\frac{A(\mathbf{a}) \wedge B(\mathbf{b}) \rightarrow C(\mathbf{c})}{A(\mathbf{a}) \rightarrow (B(\mathbf{b}) \rightarrow C(\mathbf{c}))} \text{Expo}$$

Since the free variables play no essential role here we do not write them explicitly. Then the induction hypothesis is that  $H^\omega$  proves

$$\forall \mathbf{x}, \mathbf{u}, \mathbf{q} \left( (\forall w < \mathbf{T}_1 \mathbf{x} \mathbf{u} \mathbf{q}) (A_Q(\mathbf{x}, \mathbf{T}_2 w \mathbf{x} \mathbf{u} \mathbf{q}) \wedge B_Q(\mathbf{u}, \mathbf{T}_3 w \mathbf{x} \mathbf{u} \mathbf{q})) \wedge (A \wedge B) \rightarrow C_Q(\mathbf{T}_4 \mathbf{x} \mathbf{u}, \mathbf{q}) \right).$$

We have to find terms  $T_5, T_6, \mathbf{T}_7, \mathbf{T}_8, \mathbf{T}_9$  such that

$$\forall \mathbf{x}, \mathbf{u}, \mathbf{q} \left( (\forall w_1 < T_5 \mathbf{x} \mathbf{u} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_7 w_1 \mathbf{x} \mathbf{u} \mathbf{q}) \wedge A \rightarrow \right. \\ \left. ((\forall w_2 < T_6 \mathbf{x} \mathbf{u} \mathbf{q}) B_Q(\mathbf{u}, \mathbf{T}_8 w_2 \mathbf{x} \mathbf{u} \mathbf{q}) \wedge B \rightarrow C_Q(\mathbf{T}_9 \mathbf{x} \mathbf{u}, \mathbf{q})) \right).$$

Let

$$\begin{aligned} A^* &\equiv (\forall w < \mathbf{T}_1 \mathbf{x} \mathbf{u} \mathbf{q}) A_Q(\mathbf{x}, \mathbf{T}_2 w \mathbf{x} \mathbf{u} \mathbf{q}), \\ B^* &\equiv (\forall w < \mathbf{T}_1 \mathbf{x} \mathbf{u} \mathbf{q}) B_Q(\mathbf{u}, \mathbf{T}_3 w \mathbf{x} \mathbf{u} \mathbf{q}), \\ (A \wedge B)^* &\equiv (\forall w < \mathbf{T}_1 \mathbf{x} \mathbf{u} \mathbf{q}) (A_Q(\mathbf{x}, \mathbf{T}_2 w \mathbf{x} \mathbf{u} \mathbf{q}) \wedge B_Q(\mathbf{u}, \mathbf{T}_3 w \mathbf{x} \mathbf{u} \mathbf{q})). \end{aligned}$$

Generally we have in WE-HA<sup>ω</sup> that

$$(\forall z < t) D_1(z) \wedge (\forall z < t) D_2(z) \rightarrow (\forall z < t) (D_1(z) \wedge D_2(z)).$$

<sup>3</sup>See page 21 for the definition of Cond.

Thus we can transform the following derivation into a derivation of  $H^0$

$$\frac{\frac{\frac{A^* \wedge A}{A^*} \quad \frac{B^* \wedge B}{B^*}}{(A \wedge B)^*} \quad \frac{\frac{A^* \wedge A}{A} \quad \frac{B^* \wedge B}{B}}{A \wedge B}}{(A \wedge B)^* \wedge (A \wedge B)}$$

This together with the induction hypothesis gives us  $C_Q(\mathbf{T}_4 \mathbf{xu}, \mathbf{q})$ . Then we introduce  $\rightarrow$  two times (firstly by discharging  $B^* \wedge B$  and secondly by discharging  $A^* \wedge A$ ); accordingly:

$$\begin{aligned} & (\forall w < T_1 \mathbf{xuq}) A_Q(\mathbf{x}, T_2 w \mathbf{xuq}) \wedge A \rightarrow \\ & ((\forall w < T_1 \mathbf{xuq}) B_Q(\mathbf{u}, T_3 w \mathbf{xuq}) \wedge B \rightarrow C_Q(\mathbf{T}_4 \mathbf{xu}, \mathbf{q})). \end{aligned}$$

Now defining

$$\begin{aligned} T_5 & := T_6 := T_1, & T_7 & := T_2, \\ T_8 & := \lambda \mathbf{x}, w, \mathbf{u}, \mathbf{q}. T_3 w \mathbf{xuq}, & T_9 & := T_4, \end{aligned}$$

will give the required terms.

**Case 11.** Last rule is importation:

$$\frac{A(\mathbf{a}) \rightarrow (B(\mathbf{b}) \rightarrow C(\mathbf{c}))}{A(\mathbf{a}) \wedge B(\mathbf{b}) \rightarrow C(\mathbf{c})} \text{Impo}$$

Again we leave out the free variables. Induction hypothesis is that  $H^0$  proves

$$\forall \mathbf{x}, \mathbf{u}, \mathbf{q} \left( (\forall w_1 < T_1 \mathbf{xuq}) A_Q(\mathbf{x}, T_3 w_1 \mathbf{xuq}) \wedge A \rightarrow \right. \\ \left. ((\forall w_2 < T_2 \mathbf{xuq}) B_Q(\mathbf{u}, T_4 w_2 \mathbf{xuq}) \wedge B \rightarrow C_Q(\mathbf{T}_5 \mathbf{xu}, \mathbf{q})) \right).$$

We are looking for terms  $T_6, \dots, T_9$  such that  $H^0$  proves

$$\forall \mathbf{x}, \mathbf{u}, \mathbf{q} \left( (\forall w < T_6 \mathbf{xuq}) (A_Q(\mathbf{x}, T_7 w \mathbf{xuq}) \wedge B_Q(\mathbf{u}, T_8 w \mathbf{xuq})) \wedge (A \wedge B) \right. \\ \left. \rightarrow C_Q(\mathbf{T}_9 \mathbf{xu}, \mathbf{q}) \right).$$

Let

$$\begin{aligned} A^* & \equiv (\forall w_1 < T_1 \mathbf{xuq}) A_Q(\mathbf{x}, T_3 w_1 \mathbf{xuq}), \\ B^* & \equiv (\forall w_2 < T_2 \mathbf{xuq}) B_Q(\mathbf{u}, T_4 w_2 \mathbf{xuq}). \end{aligned}$$

From  $(A^* \wedge B^*) \wedge (A \wedge B)$  we easily derive

$$A^* \wedge A \text{ and } B^* \wedge B.$$

From this and the induction hypothesis we have that  $H^0$  proves

$$\begin{aligned} & ((\forall w_1 < T_1 \mathbf{xuq}) A_Q(\mathbf{x}, T_3 w_1 \mathbf{xuq}) \wedge (\forall w_2 < T_2 \mathbf{xuq}) B_Q(\mathbf{u}, T_4 w_2 \mathbf{xuq})) \wedge \\ & (A \wedge B) \rightarrow C_Q(\mathbf{T}_5 \mathbf{xu}, \mathbf{q}). \end{aligned}$$

Now we write  $\underline{x}$  for  $\underline{xuq}$  and define

$$\begin{aligned} T_6 &::= \lambda \underline{x}. T_1 \underline{x} + T_2 \underline{x}, & T_7 &::= \lambda w_1, \underline{x}. T_3 w_1 \underline{x}, \\ T_8 &::= \lambda w_2, \underline{x}. T_4 \underline{x}(w_2 \div T_1 \underline{x}) uq. \end{aligned}$$

Then it follows from lemma 6.2.2 (ii; 2., 4. and 5. line) by using syllogism, exportation and importation

$$(\forall w < T_6 \underline{x})(A_Q(\underline{x}, T_7 w \underline{x}) \wedge B_Q(u, T_8 w \underline{x})) \wedge (A \wedge B) \rightarrow C_Q(T_5 \underline{xu}, q).$$

We have the required terms if we define

$$T_9 ::= \lambda \underline{x}, u. T_5 \underline{xu}.$$

**Case 12.** Last rule is the disjunction rule:

$$\frac{A(\mathbf{a}) \rightarrow C(\mathbf{c}) \quad B(\mathbf{b}) \rightarrow C(\mathbf{c})}{A(\mathbf{a}) \vee B(\mathbf{b}) \rightarrow C(\mathbf{c})} \text{Dis}$$

We do not write free variables. The induction hypothesis is therefore

$$\forall \underline{x}, q \left( (\forall w < T_1 \underline{xq}) A_Q(\underline{x}, T_2 w \underline{xq}) \wedge A \rightarrow C_Q(T_3 \underline{x}, q) \right), \quad (6.8)$$

$$\forall \underline{u}, q \left( (\forall w < T_4 \underline{uq}) B_Q(\underline{u}, T_5 w \underline{uq}) \wedge B \rightarrow C_Q(T_6 \underline{u}, q) \right). \quad (6.9)$$

We want to find terms  $T_7, \dots, T_{10}$  such that (if we write  $\underline{x}$  for  $\underline{zxuq}$ )

$$\forall z, \underline{x}, \underline{u}, q \left( (\forall w < T_7 \underline{x}) \left( (z = 0 \rightarrow A_Q(\underline{x}, T_8 w \underline{x}) \wedge A) \wedge (z \neq 0 \rightarrow B_Q(\underline{u}, T_9 w \underline{x}) \wedge B) \right) \wedge (A \vee B) \rightarrow C_Q(T_{10} z \underline{xu}, q) \right).$$

For any  $z^0$  we have  $z = 0$  or  $z \neq 0$ .

Assume  $z$  equals 0. From

$$(\forall w < T_1 \underline{xq}) \left( (z = 0 \rightarrow A_Q(\underline{x}, T_2 w \underline{xq}) \wedge A) \wedge (z \neq 0 \rightarrow B_Q(\underline{u}, T_5 w \underline{uq}) \wedge B) \right) \wedge (A \vee B)$$

we derive the premise of (6.8) using lemma (6.2.2) (ii) and consequently

$$\begin{aligned} &(\forall w < T_1 \underline{xq}) \left( (z = 0 \rightarrow A_Q(\underline{x}, T_2 w \underline{xq}) \wedge A) \wedge \right. \\ &\quad \left. (z \neq 0 \rightarrow B_Q(\underline{u}, T_5 w \underline{uq}) \wedge B) \right) \wedge (A \vee B) \rightarrow C_Q(T_3 \underline{x}, q), \end{aligned}$$

is derivable in  $H^0$ .

Similarly if  $z \neq 0$  we derive

$$\begin{aligned} &(\forall w < T_4 \underline{uq}) \left( (z = 0 \rightarrow A_Q(\underline{x}, T_2 w \underline{xq}) \wedge A) \wedge \right. \\ &\quad \left. (z \neq 0 \rightarrow B_Q(\underline{u}, T_5 w \underline{uq}) \wedge B) \right) \wedge (A \vee B) \rightarrow C_Q(T_6 \underline{u}, q). \end{aligned}$$

The following terms therefore suffice

$$\begin{aligned} T_7 &::= \lambda \underline{x}. \text{Cond}(T_1 \underline{x} \mathbf{q}, T_4 \mathbf{u} \mathbf{q}, z), \\ T_8 &::= \lambda w, \underline{x}. T_2 w \underline{x} \mathbf{q}, \\ T_9 &::= \lambda w, \underline{x}. T_5 w \mathbf{u} \mathbf{q}, \\ T_{10} &::= \lambda z, \underline{x}, \mathbf{q}. \text{Cond}(T_3 \underline{x}, T_6 \mathbf{u}, z). \end{aligned}$$

**Case 13.** The last rule is Q1:

$$\frac{B(\mathbf{b}) \rightarrow A(\mathbf{a}, z)}{B(\mathbf{b}) \rightarrow \forall x A(\mathbf{a}, x)} \text{ Q1}$$

where  $z \notin \{\mathbf{b}\}$ . When  $\mathbf{c}$  denotes the sequence consisting of  $\mathbf{b}$  and  $\mathbf{a}$  the induction hypothesis is

$$\forall \mathbf{u}, \mathbf{y}, x \left( (\forall w < T_1 \mathbf{c} z \mathbf{u} \mathbf{y}) B_Q(\mathbf{u}, T_2 \mathbf{c} z w \mathbf{u} \mathbf{y}, \mathbf{c}) \wedge B \rightarrow A_Q(T_3 \mathbf{c} z \mathbf{u}, \mathbf{y}, \mathbf{a}, z) \right).$$

for certain sequences of closed terms  $T_1$ ,  $T_2$  and  $T_3$ . We should find sequences of closed terms,  $T_4$ ,  $T_5$  and  $T_6$  such that

$$\forall \mathbf{u}, \mathbf{y}, x \left( (\forall w < T_4 \mathbf{c} \mathbf{u} \mathbf{y} \mathbf{x}) B_Q(\mathbf{u}, T_5 \mathbf{c} w \mathbf{u} \mathbf{y} \mathbf{x}, \mathbf{c}) \wedge B \rightarrow A_Q(T_6 \mathbf{c} \mathbf{u} \mathbf{x}, \mathbf{y}, \mathbf{a}, x) \right).$$

Take the terms

$$\begin{aligned} T_4 &::= \lambda \mathbf{c}, \mathbf{u}, \mathbf{y}, z. T_1 \mathbf{c} z \mathbf{u} \mathbf{y}, & T_5 &::= \lambda \mathbf{c}, w, \mathbf{u}, \mathbf{y}, z. T_2 \mathbf{c} z w \mathbf{u} \mathbf{y}, \\ T_6 &::= \lambda \mathbf{c}, \mathbf{u}, z. T_3 \mathbf{c} z \mathbf{u}. \end{aligned}$$

**Case 14.** The axiom Q2:

$$\forall z^\sigma A(z, \mathbf{a}) \rightarrow A(t^\sigma, \mathbf{a}).$$

Say  $\text{FV}(t^\sigma) = \{\mathbf{b}\}$ . We should find terms  $T_1$ ,  $T_2$ ,  $T_3$  and  $T_4$  such that if we write  $\underline{x}$  for  $\mathbf{a} \mathbf{b} w \mathbf{X} \mathbf{y}$

$$\forall \mathbf{X}, \mathbf{y} \left( (\forall w < T_1 \mathbf{a} \mathbf{b} \mathbf{X} \mathbf{y}) A_Q(\mathbf{X}(T_3 \underline{x}), T_2 \underline{x}, T_3 \underline{x}, \mathbf{a}) \wedge \forall z A(z, \mathbf{a}) \rightarrow A_Q(T_4 \mathbf{a} \mathbf{b} \mathbf{X}, \mathbf{y}, t, \mathbf{a}) \right).$$

Now,

$$A_Q(\mathbf{X} t, \mathbf{y}, t, \mathbf{a}) \wedge \forall z A(z, \mathbf{a}) \rightarrow A_Q(\mathbf{X} t, \mathbf{y}, t, \mathbf{a}),$$

is an axiom. Hence we take

$$\begin{aligned} T_1 &::= \lambda \mathbf{a}, \mathbf{b}, \mathbf{X}, \mathbf{y}. 1^0 & T_2 &::= \lambda \mathbf{a}, \mathbf{b}, w, \mathbf{X}, \mathbf{y}. \mathbf{y} \\ T_3 &::= \lambda \mathbf{a}, \mathbf{b}, w, \mathbf{X}, \mathbf{y}. t^\sigma & T_4 &::= \lambda \mathbf{a}, \mathbf{b}, \mathbf{X}. \mathbf{X} t^\sigma \end{aligned}$$

we obtain the required, in the same way as in the previous cases.

**Case 15.** Axiom Q3.  $A(t^\sigma, \mathbf{a}) \rightarrow \exists z^\sigma A(z, \mathbf{a})$ . The  $Q$ -translation is

$$\exists W, \mathbf{X}, \mathbf{Y}, Z \forall x, \mathbf{y} \left( (\forall w < W \mathbf{x} \mathbf{y}) A_Q(x, \mathbf{Y} w \mathbf{x} \mathbf{y}, t, \mathbf{a}) \wedge A(t, \mathbf{a}) \rightarrow A_Q(\mathbf{X} \mathbf{x}, \mathbf{y}, Z \mathbf{x}, \mathbf{a}) \wedge A(Z \mathbf{x}, \mathbf{a}) \right).$$

Say,  $FV(t^\sigma) = \{\mathbf{b}\}$ . Let

$$\begin{aligned} T_1 &::= \lambda \mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y}. 1^0, & T_2 &::= \lambda \mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{x}, \\ T_3 &::= \lambda \mathbf{a}, \mathbf{b}, w, \mathbf{x}, \mathbf{y}, \mathbf{y}, & T_4 &::= \lambda \mathbf{a}, \mathbf{b}, \mathbf{x}, t^\sigma, \end{aligned}$$

witness  $\exists W, \mathbf{X}, \mathbf{Y}, Z$ , respectively. Note, how the translation of  $\rightarrow$  is designed to meet the requirements of  $(\exists^Q)$ .

**Case 16.** Last rule in the proof is Q4:

$$\frac{A(\mathbf{a}, z) \rightarrow B(\mathbf{b})}{\exists \tilde{x} A(\mathbf{a}, \tilde{x}) \rightarrow B(\mathbf{b})} \text{Q4}$$

where  $z$  is not free in  $B(\mathbf{b})$ . Induction hypothesis is

$$\forall \mathbf{x}, \mathbf{v} ( (\forall w < T_1 \mathbf{a} z \mathbf{b} \mathbf{x} \mathbf{v}) A_Q(\mathbf{x}, T_2 \mathbf{a} z \mathbf{b} w \mathbf{x} \mathbf{v}, \mathbf{a}, z) \wedge A(\mathbf{a}, z) \rightarrow B_Q(T_3 \mathbf{a} z \mathbf{b} \mathbf{x}, \mathbf{v}, \mathbf{b}) ). \quad (6.10)$$

We should provide terms  $T_4$ ,  $T_5$  and  $T_6$  such that

$$\forall \tilde{x}, \mathbf{x}, \mathbf{v} ( (\forall w < T_4 \mathbf{a} \mathbf{b} \tilde{x} \mathbf{v}) (A_Q(\mathbf{x}, T_5 \mathbf{a} \mathbf{b} w \tilde{x} \mathbf{v}, \mathbf{a}, \tilde{x}) \wedge A(\mathbf{a}, \tilde{x})) \wedge \exists \tilde{y} A(\mathbf{a}, \tilde{y}) \rightarrow B_Q(T_6 \mathbf{a} \mathbf{b} \tilde{x} \mathbf{v}, \mathbf{v}, \mathbf{b}) ).$$

Assume

$$(\forall w < T_1 \mathbf{a} z \mathbf{b} \mathbf{x} \mathbf{v}) (A_Q(\mathbf{x}, T_2 \mathbf{a} z \mathbf{b} w \mathbf{x} \mathbf{v}, \mathbf{a}, z) \wedge A(\mathbf{a}, z)) \wedge \exists \tilde{y} A(\mathbf{a}, \tilde{y})$$

By eliminating the last conjunct of this formula and using 6.2.2 (ii) we obtain the premise of (6.10). Accordingly we have that  $H^0$  proves

$$(\forall w < T_1 \mathbf{a} z \mathbf{b} \mathbf{x} \mathbf{v}) (A_Q(\mathbf{x}, T_2 \mathbf{a} z \mathbf{b} w \mathbf{x} \mathbf{v}, \mathbf{a}, z) \wedge A(\mathbf{a}, z)) \wedge \exists \tilde{y} A(\mathbf{a}, \tilde{y}) \rightarrow B_Q(T_3 \mathbf{a} z \mathbf{b} \mathbf{x}, \mathbf{v}, \mathbf{b}).$$

We therefore define:

$$\begin{aligned} T_4 &::= \lambda \mathbf{a}, \mathbf{b}, z, \mathbf{x}, \mathbf{v}. T_1 \mathbf{a} z \mathbf{b} \mathbf{x} \mathbf{v}, & T_5 &::= \lambda \mathbf{a}, \mathbf{b}, w, z, \mathbf{x}, \mathbf{v}. T_2 \mathbf{a} z \mathbf{b} w \mathbf{x} \mathbf{v}, \\ T_6 &::= \lambda \mathbf{a}, \mathbf{b}, z, \mathbf{x}. T_3 \mathbf{a} z \mathbf{b} \mathbf{x}. \end{aligned}$$

By introducing universal quantifiers for the free variables  $\mathbf{x}$ ,  $z$  and  $\mathbf{v}$  we obtain the required.

**Case 17.** The last rule of the proof in  $H^0$  is induction:

$$\frac{A(0^0, \mathbf{a}) \quad A(x^0, \mathbf{a}) \rightarrow A(Sx^0, \mathbf{a})}{A(x^0, \mathbf{a})}$$

Assume as induction hypothesis that  $H^0$  proves

$$\forall z A_Q(T_1 \mathbf{a}, z, 0, \mathbf{a}) \quad \text{and} \\ \forall \mathbf{y}, \tilde{\mathbf{y}} ( (\forall w < T_2 \mathbf{x} \mathbf{a} \mathbf{y} \tilde{\mathbf{y}}) A_Q(\mathbf{y}, T_3 \mathbf{x} \mathbf{a} w \mathbf{y} \tilde{\mathbf{y}}, \mathbf{x}, \mathbf{a}) \wedge A(\mathbf{x}, \mathbf{a}) \rightarrow A_Q(T_4 \mathbf{x} \mathbf{a} \mathbf{y}, \tilde{\mathbf{y}}, Sx, \mathbf{a}) ).$$

We should provide terms  $\mathbf{S}$  such that

$$H^0 \vdash \forall z A_Q(\mathbf{S} \mathbf{x} \mathbf{a}, z, \mathbf{x}, \mathbf{a}).$$

Define  $\mathbf{T}$  (as we did on page 62) by simultaneous primitive recursion in higher types such that  $\mathbf{T}\mathbf{a}0 = \mathbf{T}_1\mathbf{a}$  and  $\mathbf{T}\mathbf{a}(Sx) = \mathbf{T}_4x\mathbf{a}(\mathbf{T}\mathbf{a}x)$ . With this definition the first part of the induction hypothesis equals  $\forall zA_Q(\mathbf{T}\mathbf{a}0, z, 0, \mathbf{a})$  and with the left sub-proof of the non-interpreted proof we have a proof in  $H^0$  of

$$\forall zA_Q(\mathbf{T}\mathbf{a}0, z, 0, \mathbf{a}) \wedge A(0, \mathbf{a}).$$

The second part of the induction hypothesis implies

$$\forall zA_Q(\mathbf{T}\mathbf{a}x, z, x, \mathbf{a}) \wedge A(x, \mathbf{a}) \rightarrow \forall zA_Q(\mathbf{T}\mathbf{a}(Sx), z, Sx, \mathbf{a}). \quad (6.11)$$

From the non-interpreted proof we have  $A(x, \mathbf{a}) \rightarrow A(Sx, \mathbf{a})$ . This implies together with (6.11)

$$\forall zA_Q(\mathbf{T}\mathbf{a}x, z, x, \mathbf{a}) \wedge A(x, \mathbf{a}) \rightarrow \forall zA_Q(\mathbf{T}\mathbf{a}(Sx), z, Sx, \mathbf{a}) \wedge A(Sx, \mathbf{a}).$$

By applying the induction rule we obtain  $\forall zA_Q(\mathbf{T}\mathbf{a}x, z, x, \mathbf{a}) \wedge A(x, \mathbf{a})$ . This yields

$$H^0 \vdash \forall zA_Q(\mathbf{T}\mathbf{a}x, z, x, \mathbf{a}).$$

Thus, define  $\mathbf{S} := \lambda x, \mathbf{a}. \mathbf{T}\mathbf{a}x$ .

**Case 18.** The quantifier free rule of extensionality:

$$\frac{A_{\text{qf}} \rightarrow s^\sigma = t^\sigma}{A_{\text{qf}} \rightarrow r[s^\sigma] =_\tau r[t^\sigma]} \text{QF-ER}$$

The translation of the hypothesis of the rule is equivalent to itself—likewise the conclusion. The rule is therefore interpreted by itself.

**Case 19.** The axiom schema for independence-of-premise for purely universal formulas:

$$(\forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow \exists z^\sigma B(z)) \rightarrow \exists z^\sigma (\forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow B(z)), \quad (6.12)$$

where  $A_{\text{qf}}(\mathbf{x})$  is quantifier free and  $\forall \mathbf{x}A_{\text{qf}}(\mathbf{x})$  does not contain  $z$  free. We assume w.l.o.g. that  $A_{\text{qf}}$  is  $\forall$ - and  $\rightarrow$ -free. The translation of the hypothesis is

$$\exists W, \mathbf{X}, z, \mathbf{u}\forall \mathbf{v}((\forall w < W\mathbf{v})A_{\text{qf}}(\mathbf{X}w\mathbf{v}) \wedge \forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow (B_Q(\mathbf{u}, \mathbf{v}, z) \wedge B(z))).$$

And the translation of the conclusion is

$$\exists W, \mathbf{X}, z, \mathbf{u}\forall \mathbf{v}(((\forall w < W\mathbf{v})A_{\text{qf}}(\mathbf{X}w\mathbf{v}) \wedge \forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow B_Q(\mathbf{u}, \mathbf{v}, z)) \wedge (\forall \mathbf{x}A_{\text{qf}}(\mathbf{x}) \rightarrow B(z))).$$

When these two are compared it is clear by witnessing the existential quantifiers by the obvious projection functionals that (6.12) is interpretable by functionals of WE-HA<sup>0</sup> *only*.  $\dashv$

**Remark 6.2.4.** MP<sup>0</sup> and AC cannot be  $Q$ -interpreted, since certain important quantifiers are ‘left behind’:

There is an instance of MP<sup>0</sup> such that the  $Q$ -translation of this instance is not satisfiable by any term of WE-T. Let  $P(x)$  be the prime formula  $t_Tyyx = 0$ , where  $t_T$  is the characteristic

term for Kleene's  $T$ . When we replace  $P(x) \wedge P(x)$  by the equivalent  $P(x)$  and delete bounded quantifiers not referring to anything, the translation  $(\neg\neg\exists xP(x) \rightarrow \exists xP(x))^Q$  becomes:

$$\exists X \forall Y^1, z^0 ( \neg ( (\forall w < z) \neg (P(Yw) \wedge \exists xP(x)) \wedge \neg\exists xP(x) ) \wedge \neg\neg\exists xP(x) \rightarrow P(XzY) ).$$

Assume there is a closed term  $F$  of type 0010 interpreting this. Now, since we have intuitionistically

$$((\forall x < t) \neg (B(x) \wedge A) \wedge \neg A) \leftrightarrow \neg A,$$

we derive by eliminating the universal quantifiers by  $0^0$  and  $o^1$ :

$$\neg\neg\exists xP(x) \rightarrow P(Fy0o).$$

This is of course a contradiction, since  $\lambda y.Fy0o$  of type 1 would then decide the halting problem.

AC is likewise problematic. The translation of an instance of AC is (when  $\underline{x}$  is  $w\tilde{Y}\tilde{U}\mathbf{x}\mathbf{v}$  and  $\underline{y}$  is  $\tilde{Y}\tilde{U}\mathbf{x}$ )

$$\begin{aligned} \exists W, X, \mathbf{V}, Y, \mathbf{U} \forall \tilde{Y}, \tilde{U}, x, \mathbf{v} ( (\forall w < W\mathbf{y}\mathbf{v}) (A_Q(\tilde{U}(X\underline{x}), \mathbf{V}\underline{x}, X\underline{x}, \tilde{Y}(X\underline{x})) \\ \wedge A(X\underline{x}, \tilde{Y}(X\underline{x}))) \wedge \forall \tilde{x} \exists y A(\tilde{x}, y) \rightarrow A_Q(\mathbf{U}\underline{y}, \mathbf{v}, x, Y\underline{y}) \wedge \forall \tilde{x} A(\tilde{x}, Y\tilde{Y}\tilde{U}\tilde{x}) ). \end{aligned}$$

If we take standard projections then this turns into

$$(A_Q(\mathbf{U}x, \mathbf{v}, x, Yx) \wedge A(x, Yx)) \wedge \forall \tilde{x} \exists y A(\tilde{x}, y) \rightarrow A_Q(\mathbf{U}x, \mathbf{v}, x, Yx) \wedge \forall \tilde{x} A(\tilde{x}, Y\tilde{x})$$

for any  $\mathbf{U}, Y, x, \mathbf{v}$ . But this certainly is not provable.

Because of these problems the closure properties, which are shown in the next section as an application of the  $Q$ -interpretation, cannot be proved in this way with respect to AC and  $MP^0$ . However, as we shall see in section 6.5, this can partly be remedied due to the ordinary Dialectica interpretation.

### 6.3 Closure properties of intuitionistic arithmetic showed by $Q$ -translation

The following theorem – displaying various important properties of intuitionistic arithmetic plus/minus  $IP_{\forall}^0$  – follows immediately from the soundness of the  $Q$ -translation. Note, however, that all properties except closure under Markov's rule follow from theorem 5.7.4.

**Theorem 6.3.1.** (Closure properties). *Let  $H^0$  be  $WE-HA^0 \pm IP_{\forall}^0$ . Then:*

1.  $H^0$  has existence property, i.e. if  $H^0 \vdash \exists x^\sigma A(x)$ , then  $H^0 \vdash A(t^\sigma)$  for extractable term  $t$  with  $FV(t) \subseteq FV(A) \setminus \{x\}$ .
2.  $H^0$  has disjunction property, i.e. for closed  $A$  and  $B$ ,

$$\text{if } H^0 \vdash A \vee B, \text{ then } H^0 \vdash A \text{ or } H^0 \vdash B.$$

3.  $H^0$  is closed under the rule of choice (ACR):

$$\text{if } H^0 \vdash \forall x^\sigma \exists y^\tau A(x, y), \text{ then } H^0 \vdash \exists Y^{\sigma\tau} \forall x^\sigma A(x, Yx).$$

4.  $H^0$  is closed under the rule of independence-of-premise for purely universal formulas ( $IPR_{\forall}^0$ ):

$$H^0 \vdash \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow \exists y^\sigma B(y) \text{ implies } H^0 \vdash \exists y^\sigma (\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow B(y)),$$

$y \notin \text{FV}(\forall \mathbf{x} A_{\text{qf}})$  and  $A_{\text{qf}}$  is quantifier free.

5.  $H^0$  is closed under Markov's rule:

$$\text{If } H^0 \vdash \neg\neg \exists x^\sigma A_{\text{qf}}(x, \mathbf{a}), \text{ then } H^0 \vdash \exists x^\sigma A_{\text{qf}}(x, \mathbf{a}).$$

**Proof.**

1. Assume  $H^0 \vdash \exists x A(x, \mathbf{a})$ . By soundness of  $Q$ -interpretation there are closed terms  $\mathbf{T}$  and  $T_0$  such that

$$H^0 \vdash \forall \mathbf{y} A_Q(\mathbf{T}\mathbf{a}, \mathbf{y}, T_0\mathbf{a}, \mathbf{a}) \wedge A(T_0\mathbf{a}, \mathbf{a}).$$

Take the second conjunct of this.

2. Assume  $H^0 \vdash A \vee B$  for closed  $A$  and  $B$ . By soundness of  $Q$ -interpretation and computability of type 0 terms there are closed sequences of terms  $\mathbf{t}_1, \mathbf{t}_2$  and a number term  $n$  such that

$$H^0 \vdash (n = 0 \rightarrow \forall \mathbf{y} A_Q(\mathbf{t}_1, \mathbf{y}) \wedge A) \wedge (n \neq 0 \rightarrow \forall \mathbf{v} B_Q(\mathbf{t}_2, \mathbf{v}) \wedge B).$$

If  $n$  equals 0 then  $H^0 \vdash A$ , otherwise  $H^0 \vdash B$ .

3. Similar to 1.

4. Let a proof in  $H^0$  of  $\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow \exists y^\sigma B(y)$  be given and assume w.l.o.g. that  $A_{\text{qf}}$  is  $\rightarrow$ -free and  $\vee$ -free. By soundness we have

$$H^0 \vdash \forall \mathbf{v} ((\forall w < T\mathbf{v}) A_{\text{qf}}(\mathbf{T}_1 w \mathbf{v}) \wedge \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow B_Q(\mathbf{t}_1, \mathbf{v}, t_0^\sigma) \wedge B(t_0^\sigma)).$$

This implies

$$H^0 \vdash (\forall w < T\mathbf{v}) A_{\text{qf}}(\mathbf{T}_1 w \mathbf{v}) \wedge \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow B(t_0^\sigma).$$

And since we have  $\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow (\forall w < T\mathbf{v}) A_{\text{qf}}(\mathbf{T}_1 w \mathbf{v})$  we also have

$$\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow (\forall w < T\mathbf{v}) A_{\text{qf}}(\mathbf{T}_1 w \mathbf{v}) \wedge \forall \mathbf{x} A_{\text{qf}}(\mathbf{x}),$$

and from this follows

$$\forall \mathbf{x} A_{\text{qf}}(\mathbf{x}) \rightarrow B(t_0^\sigma)$$

by using syllogism. This implies by axiom Q3 and modus ponens the desired conclusion.

5. Similar to 1 using that  $H^0 \vdash \neg\neg A_{\text{qf}} \leftrightarrow A_{\text{qf}}$ .  $\dashv$

#### 6.4 $Q$ -interpretation is not closed under deductions

The following phenomenon is due to a combination of (i) the weakening which occurs when an implication is translated and (ii) the fact that  $\text{IP}_{\forall}^{\omega}$  is  $Q$ -interpretable in  $\text{WE-HA}^{\omega}$ .

The  $Q$ -interpretation is not closed under deductions. This means that there exist formulas  $A$  and  $B$  such that:

1.  $B$  follows logically from  $A$ ,
2.  $A$  can be  $Q$ -interpreted in  $\text{WE-HA}^{\omega}$ , but
3.  $B$  is not  $Q$ -interpretable in  $\text{WE-HA}^{\omega}$ .

To show this, let  $C$  be some closed instance of  $\text{IP}_{\forall}^{\omega}$ , such that  $C$  is not provable in  $\text{WE-HA}^{\omega}$ . It is not that simple to show that there exists such an instance, but see (Smorynski, 1973, 369–370) for an example. However, in virtue of soundness of  $Q$ -translation we know that

$$\text{WE-HA}^{\omega} \vdash \forall \mathbf{y} C_Q(\mathbf{t}, \mathbf{y}), \quad (6.13)$$

for some closed  $\mathbf{t}$ . On the other hand, it is also the case that  $C \rightarrow C \vee C$  is interpretable in  $\text{WE-HA}^{\omega}$ . But  $C \vee C$  is not. If it were, we would – due to the disjunction property – be able to derive  $C$  in  $\text{WE-HA}^{\omega}$ , but we know this is not possible. Now, a derivation of  $C \vee C$  could have the following form in  $\text{WE-HA}^{\omega} + \text{IP}_{\forall}^{\omega}$ :

$$\frac{C \quad C \rightarrow C \vee C}{C \vee C} \text{MP}$$

But as was shown in the proof of theorem 6.2.3 we need the left sub-derivation of the non-interpreted derivation in order to interpret the conclusion of modus ponens. Although both  $C$  and  $C \rightarrow C \vee C$  are interpretable in  $\text{WE-HA}^{\omega}$ ,  $C \vee C$  is only interpretable in  $\text{WE-HA}^{\omega} + \text{IP}_{\forall}^{\omega}$ .

In the general case this shows, by taking

$$A \equiv C \wedge (C \rightarrow C \vee C) \quad \text{and} \quad B \equiv C \vee C,$$

that

$$\text{WE-HA}^{\omega} \not\vdash (A \rightarrow B) \rightarrow (A^Q \rightarrow B^Q).$$

#### 6.5 Closure properties of $\text{WE-HA}^{\omega} + \text{IP}_{\forall}^{\omega} + \text{MP}^{\omega} + \text{AC} + \Gamma$ by Dialectica

We have seen that the  $Q$ -interpretation cannot be used to show closure properties of theories based on  $\text{MP}^{\omega}$  or  $\text{AC}$ , but – as will become clear in this section – the ordinary Dialectica interpretation can be used for this.

Strictly speaking it is not necessary to develop the  $Q$ -translation in order to derive part 5 of theorem 6.3.1. This fact can be proved by the Dialectica interpretation alone, due to the equivalence

$$(\neg \neg \exists x A_{\text{qf}}(x))^D \leftrightarrow \exists x \neg \neg A_{\text{qf}}(x).$$

Actually, something much stronger can be proved. For the rest of this chapter let  $\Gamma$  be any set of true purely universal sentences. It follows from theorem 4.3.2 that

if  $WE-PA^\omega + QF-AC \vdash \neg\neg\exists x A_{qf}(x)$  from  $\Gamma$ , then  $WE-HA^\omega \vdash \exists x A_{qf}(x)$  from  $\Gamma$ .

In fact this is responsible for the success of Dialectica + negative translation in the context of classical arithmetic.<sup>4</sup>

We will now use the Dialectica interpretation to show closure properties of  $WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC + \Gamma$ . As was shown in chapter 2 we have

$$WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC \vdash A \leftrightarrow A^D. \quad (6.14)$$

This observation leads to the following theorem which can be found in (Troelstra, 1973, 260).

**Theorem 6.5.1.** (Closure properties).  *$WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC + \Gamma$  has disjunction property, existence property, is closed under  $ACR$ ,  $IPR_\forall^\omega$  and Markov's rule.*

**Proof.** The theorem is proved in the same way as the closure properties were proved by using  $\underline{mrt}$ , now using property (6.14) as truth property. That  $\Gamma$  can be added is due to the fact that  $\Gamma$  does not contribute to the computational content, as was noted on page 64.  $\dashv$

### 6.6 Constructiveness of $WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC + \Gamma$

Let us now see what the result obtained in the last paragraph has to say about constructivity, and hence pursue the motives of Gödel, as cited in the beginning of the chapter. We will in a broader context try “to answer the question in which sense intuitionistic logic ... is constructive”. It will be broader since we will also consider non-intuitionistic principles.

We denote the theory  $WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC + \Gamma$  as  $H^\Gamma$ . For any such  $\Gamma$ , which we happen to *know*, for some reason or another, the truth of, we will argue that  $H^\Gamma$  is constructive. Certainly, it is a strong and interesting theory. In our view theorem 6.5.1 together with the general reduction (theorem 3.7.1) of the theory to the quantifier free fragment of  $WE-HA^\omega$  together with  $\Gamma$  shows that  $WE-HA^\omega + IP_\forall^\omega + MP^\omega + AC + \Gamma$  really is a constructive theory. This is surely in conflict with the BHK interpretation, since already  $MP^\omega$  and  $IP_\forall^\omega$  are not validated under that interpretation. But we see no reason why BHK – which is a very *general* interpretation independent of any specific mathematical theory – should have *monopoly* on what is constructive in our specific context of arithmetic. On the contrary, we find that the Dialectica interpretation provides conclusive evidence for the constructivity of  $H^\Gamma$ :

Firstly, theorem 6.5.1 shows that  $H^\Gamma$  has the pleasant and desirable properties one would be inclined to require of a constructive theory. For instance, if  $H^\Gamma$  proves a sentence  $\exists x^\sigma A(x)$ , where  $A$  is of arbitrary complexity, then in fact one can provide constructively a closed primitive recursive functional  $t$  of type  $\sigma$  such that  $H^\Gamma$  proves  $A(t^\sigma)$ . Thus, the principles are computationally meaningful. Of course, that fact that  $t$  actually has property  $A$  is proved

<sup>4</sup>On the other hand, if we want to extract computational content from classical proofs and work with modified realisability we need some device in order to show closure under Markov's rule. But as we have seen there is no such result for  $E-HA^\omega$ . However,  $A$ -translation can be used to show closure of  $HA$  under Markov's rule.

using the whole theory. But we see no specific problem in that  $\dashv\vdash^{\sigma}$  is still computable. However, if one is inclined for having a complete reduction to some theory, that even orthodox intuitionists regard as constructive, then the following should convince them:

Secondly, we have in the Hilbertian sense a reduction of  $H^{\Gamma}$  to the quantifier free part of  $WE-HA^{\omega}$  together with  $\Gamma$ . Theorem 3.7.1 shows this. Due to this reduction we have that if  $H^{\Gamma}$  proves a  $\forall\exists$ -sentence such as  $\forall x^{\sigma}\exists y^{\tau}A_{\text{qf}}(x, y)$ , then indeed there is a constructive procedure given by the very proof of theorem 3.7.1 which provides a closed primitive recursive functional  $t$  of type  $\sigma\tau$  such that for any  $x^{\sigma}$ ,  $A_{\text{qf}}(x, tx)$  holds, and this is verified in an essentially quantifier free way.

## CHAPTER 7

# Conclusions

Coming to the end of our investigation we will try to sum up the mathematical and philosophical significance of the results obtained. But first we will make clear what parts of the mathematical discipline proof theory we have studied and why.

Proof theory is not exclusively a study of the foundations of mathematics. But large parts of proof theory are foundational studies in different ways. Mathematical theories, and thereby also various programs,<sup>1</sup> can be formalised and studied in several ways. We will mention three such ways:

1. The theories can be studied with respect to ordinal strength. This part of proof theory is called ordinal analysis. It can be seen as a natural outgrowth of Hilbert's focus on consistency, since typically in ordinal analysis one measures a formal theory  $S$  in terms of the least constructive ordinal  $\alpha$  such that primitive recursive arithmetic plus quantifier free transfinite induction up to  $\alpha$  proves consistency of  $S$ :

$$\text{PRA} + \text{QF-TI}(\alpha) \vdash \text{Cons}_S.$$

The paradigm for ordinal analysis is, without any doubt, Gentzen's work (1936; 1943) on Peano arithmetic. We have not in this thesis been concerned with his part of proof theory.

2. The derivations of theories can be studied with respect to computational content. This part of proof theory studies the constructive aspects of certain *prima facie* non-constructive mathematical principles, as for instance weak König's lemma. When applied to specific proofs using non-constructive notions this part of proof theory is called proof mining. The program was initiated by G. Kreisel, but is also related to:
3. Reductive proof theory. In general one tries in reductive proof theory to understand and justify ideal aspects of mathematics in more constructive terms. Non-constructive theories can be reduced to constructive theories; theories using quantifiers can be reduced – as we have seen – to theories without quantifiers, and so forth. This part of proof theory can be understood as a generalised form of Hilbert's program, where one seeks *local* reductions, instead of global reductions. S. Feferman has contributed greatly in this area.

These three branches are of course related, but their interests and purposes are somewhat different. In ordinal analysis the main interest is a mathematical calibration of theories in terms of ordinals. Proof mining unwinds the constructive content of certain proofs: this has

---

<sup>1</sup>Examples of such programs are: 1. E. Bishop's constructive mathematics. The philosophy behind this program is expressed in (Bishop, 1970), and this could roughly be said to be comprised by  $\text{IL}^\omega + \mathbb{N} + \text{MP}^\omega + \text{IP}_V^\omega + \text{QF-AC}$ , where  $\mathbb{N}$  is the set of natural numbers. 2. The intuitionism of L.E.J. Brouwer is associated with intuitionistic logic together with the theory of choice sequences and bar induction. 3. The mathematics of H. Weyl (1918) is a predicative development of mathematics taking  $\mathbb{N}$  for granted. This is formalised by system W, see Feferman (1988).

both mathematical and philosophical consequences. Whereas in reductive proof theory, conceptual reductions of specific theories are found, and the investigations are philosophically and mathematically motivated. The branches are, however, related: Ordinal analysis can, for instance, be one route into reductive proof theory, but certainly methods from reductive proof theory can also provide results in ordinal analysis.<sup>2</sup> However, some directions express very different motives. As an example of this we have seen that precisely those statements which are central to ordinal analysis – purely universal sentences expressing consistency – are on the contrary harmless with respect to computational content when used as assumptions. This was noted repeatedly by Kreisel in the 50s and it probably sharpened his critique (Kreisel, 1958, 155) of Hilbert’s insistence on consistency proofs.<sup>3</sup>

The proof theory treated in this thesis has been a combination of 2 and 3.

### A recapitulation of our motives and goals

As mathematicians we cannot restrict ourself to strictly constructive methods, where strictly is understood in the sense that the methods are constructive in any given situation. Such methods which are globally constructive are indeed pleasant to have, but not sufficient. Ideal elements, in the sense of Hilbert, are in general indispensable but they need to be analysed. One of our goals has been to analyse classical principles – primarily extensionality, Markov’s principle, independence-of-premise and axiom of choice – locally in the framework of type theory. We wanted to examine these concepts and principles in an un-biased way in order to get some constructive understanding of them. In our proof theoretical analysis we have used the Dialectica interpretation, modified realisability and variants hereof together with Kuroda’s negative translation and  $A$ -translation. We have thereby examined the principles and calibrated their computational contribution. But we have also used the Dialectica interpretation as a general tool in reductive proof theory when we reduced  $\text{WE-PA}^\omega + \text{QF-AC}$  to  $\text{WE-T}$ .

### 7.1 Interpreting mathematics by methods from proof theory

The interpretation of “ $\rightarrow$ ” turns out to be crucial. This is where realisability and functional interpretations really differ. Realisability only goes from left to right in an implication, so to speak. This is clear from chapter 5 when we defined modified realisability. The first pages of chapter 6 also showed this when we discovered that there is no Dialectica-with-truth interpretation. In this respect realisability is quite close to the BHK interpretation. The Dialectica interpretation of implication, on the other hand, goes in both directions, thereby verifying the classical equivalence  $\neg\forall x A_{\text{qf}}(x) \leftrightarrow \exists x \neg A_{\text{qf}}(x)$ , i.e. verifies Markov’s principle. All this points towards the fact that modified realisability and Dialectica interpretation are structurally different: They present different ways of interpreting mathematics or – more precisely – of interpreting the logical operations, most notably implication. These differences show up as

---

<sup>2</sup>See Feferman (2000) for a thorough survey, where reductive proof theory are divided into three separate branches.

<sup>3</sup>For a clear account of this critique and a formulation of Kreisel’s own program see (Kreisel, 1987).

differences in the interpretations of extensionality, Markov's principle and independence-of-premise.

### 7.1.1 Extensionality and Markov's rule

In general, the notion of equality as an extensional notion is constructively problematic. However, in most circumstances a mathematician would regard two functions  $f$  and  $g$  to be equal if for any  $x$ ,  $f(x) = g(x)$ ; likewise we regard two sets to be equal if they have the same elements. Although natural, extensional equality is certainly not decidable. Two Turing machines  $M_n$  and  $M_m$  can be given by two different descriptions. However, they compute the same recursive function if for any  $x$ ,  $M_n$  halts on  $x$  iff  $M_m$  halts, and if they halt they have the same output—but this is not decidable. Extensionality is globally a non-constructive notion, but how much of it can we retain in the framework of typed arithmetic. We have seen that we can use full extensionality within  $\text{HA}^\omega$  when we prove an existence formula: modified realisability validates trivially extensionality. However, we do not get a reduction in the sense of Hilbert, since the fact that the extracted realiser actually does what is required is proved by applying to the full theory. But still, it is interesting and mathematical useful that modified realisability can validate full extensionality. However, if we want any notion of Markov's principle – if only Markov's rule for numbers – then the system immediately gets non-constructive and eventual realisers cannot be any of the functionals in the sense of Gödel's T: Due to theorem 5.5.1 we see that  $\text{E-HA}^\omega$  is not closed under Markov's rule and that there cannot be a modified realisability interpretation of  $\text{E-HA}^\omega + \text{Markov's rule}$ .

When we in general want to extract computational content from *classical* proofs, closure under Markov's rule is necessary. Suppose  $A_{\text{qf}}$  is quantifier free and that  $\exists x A_{\text{qf}}(x)$  is provable in some classical theory  $\text{T}_c$ . Then by negative translation  $\neg\neg\exists x A_{\text{qf}}(x)$  is provable in the intuitionistic counterpart  $\text{T}_i$ . Now we need closure under Markov's rule. But if that is impossible then we cannot use realisability nor Dialectica for extracting constructive content. Therefore, if we want to unwind classical proofs we have to abandon the full theory  $\text{E-PA}^\omega$ . This shows that the application – mentioned as item 4 on page 11 – that Gödel had in mind when developing Dialectically is principally unattainable in the case of  $\text{E-PA}^\omega$ .

We can, however, with respect to certain sentences provable in  $\text{E-PA}^\omega$  eliminate extensionality in the sense of Luckhardt (1973). But in the general case we have to stick to what is next best: weak extensionality. From the Dialectica interpretation we see that weak extensionality together with Markov's rule, in fact the much stronger Markov's principle, can be validated. Therefore, weak extensionality together with Markov's principle can be given a constructive interpretation.

### 7.1.2 Markov's principle and independence-of-premise

Independence-of-premise,  $(A \rightarrow \exists y B(y)) \rightarrow \exists y (A \rightarrow B(y))$ ,  $y \notin \text{FV}(A)$ , is another globally non-constructive principle. But a particularly strong instance of a restricted principle is interpretable by modified realisability, namely the case where  $A$  is  $\exists$ -free:  $\text{IP}_{\text{ef}}^\omega$ . Due to mrt-interpretability of the this principle we can use it when we prove existence theorems and still extract a realiser. However, theorem 3.8.1 showed computational incompatibility between Markov's principle and  $\text{IP}_{\text{ef}}^\omega$ : there cannot be given a computational interpretation of any

arithmetical theory containing  $MP^\omega$  and any independence-of-premise principle as strong as  $IP_{\neg\forall}^\omega$ . The theorem showed that such a theory would be strongly non-constructive, and we even saw (theorem 3.8.2) that  $WE-HA^\omega + MP^\omega$  is not closed under the rule of independence-of-premise for negated purely universal formulas :

$$IPR_{\neg\forall}^\omega : \quad \vdash \neg\forall x A_{qf}(x) \rightarrow \exists y B(y) \Rightarrow \vdash \exists y (\neg\forall x A_{qf}(x) \rightarrow B(y)),$$

though  $WE-HA^\omega$  is.

But in case of  $IP_{\forall}^\omega$ , Dialectica validates both  $MP^\omega$  and  $IP_{\forall}^\omega$ . This poses a real choice between the two interpretations. We see that Markov's principle and independence-of-premise are in conflict, precisely as Markov's rule and full extensionality are.

Markov's principle is a strong principle with appealing mathematical consequences. It is precisely due to the validation of Markov's principle that functional interpretation (together with negative translation) can be used for extracting constructive content from proofs in  $WE-PA^\omega + QF-AC$  and many other classical theories. On the other hand, Markov's principle is not validated (theorem 5.4.1) by our second extraction method: negative translation +  $A$ -translation + modified realisability. Therefore, this strategy does not interpret  $QF-AC$  on top of classical arithmetic. In fact, due to the limitations of  $A$ -translation, this device can only be applied to proofs of  $\Pi_2^0$  formulas in  $PA$ . These problems can partly be remedied using refined versions of  $A$ -translations, see (Coquand & Hofmann, 1999; Schwichtenberg, 2000; Berger et al., n.d.).

## 7.2 Evaluation of modified realisability and Dialectica interpretation

Modified realisability has certain advantages. In an intuitionistic context it interprets full extensionality and  $IP_{ef}^\omega$ . But because extensionality and  $IP_{ef}^\omega$  are in conflict with Markov's principle certain problems show up. Foremost, negative translation +  $A$ -translation + modified realisability cannot be used directly for interpreting classical type theory with full extensionality, only proofs in  $PA$  of  $\Pi_2^0$  formulas allow for program extraction. This is a weak spot. Moreover, the combination of the three translations is perhaps not that elegant and uniform as negative translation + functional interpretation. However, there are many 'parameters' – so to speak – in the overall interpretation which allow for optimisation. This is what U. Berger and H. Schwichtenberg use in their paradigm of program extraction; see for instance (Berger et al., n.d.). They treat equality on the meta-level and consider terms with the same normal form to be equal. In this way they get around some of the problems occurring when one uses realisability for extracting constructive content from typed classical arithmetic, though they have to deal with normalisation.

On the other hand functional interpretation allows for a uniform approach to the problem of extracting constructive content from proofs. It gives an optimal interpretation of the combination of extensionality, Markov's principle and independence-of-premise. It is remarkable, for instance, that Markov's principle is validated even in higher types. These properties of Dialectica is to the effect, as we will indicate below, that it is a strong tool for extracting computational content from proofs in analysis.

### 7.2.1 Closure under rules

Before the  $Q$ -interpretation it has been unknown whether the Dialectica idea could be used to show closure under various rules, existence property and disjunction property of intuitionistic arithmetic. Recall, that it was one of the motives for Gödel (1941) to show such properties, but that he was only able to show realisations of the *interpreted* formulas. Modified realisability with truth has, on the other hand, been an effective tool in showing such results. We have, however, shown that with respect to functional interpretations there is no principle obstacle in these matters, and that the  $Q$ -interpretation can be used just as well to show these crucial properties of intuitionistic arithmetic (having only weak extensionality of course).

## 7.3 Two strong constructive theories

Depending on which interpretation one chooses, different theories are validated as constructive. This shows how crucial the interpretation of implication is.

### 7.3.1 $WE-HA^\omega + IP_{\forall}^\omega + MP^\omega + AC + \Gamma$

If  $\Gamma$  is some set of universal sentences which we happen to know the truth of then the Dialectica interpretation validates  $WE-HA^\omega + IP_{\forall}^\omega + MP^\omega + AC + \Gamma$  as a constructive theory. This was discussed extensively at the end of chapter 6. It is constructive in the sense that it has existence property, disjunction property and is closed under various rules. Although the theory is based on principles which are *prima facie* non-constructive (i.e. in conflict with the BHK interpretation) a closer analysis shows that the principles are locally constructive. They do carry computational content and we also have a method for extracting it. This shows that the BHK interpretation is a very general (global) rule of thumb regarding constructivity, but when it comes to specific theories one can constructively allow for more than what BHK validates.

### 7.3.2 $E-HA^\omega + IP_{\text{ef}}^\omega + AC + \Gamma$

Modified realisability interprets  $E-HA^\omega + IP_{\text{ef}}^\omega + AC + \Gamma$ , where  $\Gamma$  is any set of true  $\exists$ -free sentences. Theorem 5.7.6 shows that this theory has existence property, disjunction property and is closed under the different rules, except Markov's rule. This last exception is, however, a big disadvantage in connection with program extraction from classical proofs—as discussed above. And *if* one requires of a constructive theory that it is closed under Markov's rule, then this theory fails to be constructive.

## 7.4 Dialectica as a tool in proof mining and reductive proof theory

In a continuation of Kreisel's program – the program of proof mining – the Dialectica interpretation and variants hereof have interesting applications in classical analysis. Dialectica is a local transformation which means that the form of the proof is not changed essentially under the interpretation. This makes it easy to apply to real proofs.

In contrast to the situation just five or ten years ago (see e.g. Feferman (1996)) it seems today as if applications of proof theory in analysis has a promising future. Especially Kohlen-

bach has by now given many interesting clear cut applications of the Dialectica interpretation thereby providing specific information with respect to non-effective proofs in analysis. That is, information actually sought by mathematicians of analysis.

Kohlenbach (1993) shows how to represent real numbers as Cauchy sequences of rational numbers with fixed rate of convergence in *extensions* of systems we have studied here and how to represent also complete separable metric spaces and compact metric spaces.<sup>4</sup> Using these representations it is shown how many important theorems of analysis have forms which allow for constructive interpretations, thus making it possible to extract constructive information from non-constructive proofs in analysis. This turns out to be particularly successful in connection with uniqueness proofs in best approximation theory. Specifically, Kohlenbach's work yields improvement of known estimates in connection with Chebycheff approximation. This work is continued in (Kohlenbach, 1993a). Another result obtained recently in (Kohlenbach, n.d.) also gives an example of how successful the idea of analysing proofs in functional analysis can be. Kohlenbach has by use of functional interpretation and majorization improved bounds considerably in connection with the so-called Krasnoselski-Mann iteration.

One of the reasons why Dialectica is powerful in these interpretations is due to its interpretation of  $\forall x A_{\text{qf}}(x) \rightarrow \forall y B_{\text{qf}}(y)$ :

$$\exists X \forall y (A_{\text{qf}}(Xy) \rightarrow \forall y B_{\text{qf}}(y)).$$

Because of this interpretation Dialectica provides computational information in the case where  $\forall y B_{\text{qf}}(y)$  is false. This is in contrast to realisability.

Proof mining is interesting for mathematical and philosophical reasons concerning computational realisations. Reductive proof theory is connected to this, but the philosophical emphasis is closer to Hilbert's ideas with respect to reducing complex theories to simpler theories. Functional interpretation is a very important tool in reductive proof theory seeking partial realisations of Hilbert's program. In chapter 4 we saw Dialectica at work and it *reduces*

$$\text{WE-PA}^{\omega} + \text{QF-AC to WE-T}$$

in the sense that for any sentence  $\forall x^{\sigma} \exists y^{\tau} A_{\text{qf}}(x, y)$  of  $\mathcal{L}(\text{WE-PA}^{\omega})$  we have a constructive procedure  $(\cdot)^*$  which takes any proof  $p$  in  $\text{WE-PA}^{\omega} + \text{QF-AC}$  of  $\forall x^{\sigma} \exists y^{\tau} A_{\text{qf}}(x, y)$  into a proof  $p^*$  in WE-T of  $A_{\text{qf}}(x^{\sigma}, T x)$ , where  $T^{\sigma\tau}$  is some closed primitive recursive functional given by the proof transformation.  $p^*$  is not much longer than  $p$  and has essentially the same structure, due to the locality of the interpretation. This is a clear cut contribution to a generalised Hilbert program, since it shows that all ideal elements – such as tertium non datur, quantifier free axiom of choice and induction on complex formulas – used in the proof of  $\forall x^{\sigma} \exists y^{\tau} A_{\text{qf}}(x, y)$  can be eliminated. The reflection principle – as stated on page 4 as the essence of Hilbert's program – is provable for this class of formulas. Moreover, we can extract for these formulas a realiser for  $y$ . Finally, the reduction shows consistency of  $\text{WE-PA}^{\omega} + \text{QF-AC}$  relative to a natural generalisation of Hilbert's finitism, as discussed in section 4.5.1.

<sup>4</sup>The typed language is very useful and allows for simplifications when compared with second order systems as for instance  $\text{WKL}_0$  as used by Simpson (1999).

This reduction was extended (also in chapter 4) in order to get a reduction of

$$\text{WE-PA}^\omega + \text{QF-AC} + \Gamma \text{ to } \text{WE-T} + \Gamma,$$

where  $\Gamma$  is any set of true purely universal sentences. This observation generalises to many different theorems of this kind based on functional interpretation. Both for stronger and weaker systems and for systems not comparable to those studied in this thesis. For instance, let WKL be a formalisation of a binary version of König's lemma. Utilizing the notion of majorization Kohlenbach has shown that in systems of the kind studied in this thesis *extended* by WKL and restricted forms of comprehension one can add arbitrary axioms of the form

$$\forall x^1 \exists y \leq_1 .sx \forall z^0 A_{\text{qf}}(x, y, z),$$

without having to consider their proves. This covers as a special case WKL. Also, let  $\text{E-PRA}^\omega$  denote the restriction of  $\text{E-PA}^\omega$  with (i) quantifier free induction and (ii) primitive recursion in type 0 only, but with parameters of arbitrary types. Now, Kohlenbach (1992) shows that

$$\text{E-PRA}^\omega + \text{QF-AC}^{1,0} + \text{QF-AC}^{0,1} + \text{WKL} \text{ is conservative over PRA}$$

for  $\Pi_2^0$  formulas. Note that already  $\text{QF-AC}^{0,0}$  suffices to get the system  $\text{WKL}_0$  as a subsystem. This reduction is particularly interesting since  $\text{E-PRA}^\omega + \text{QF-AC}^{1,0} + \text{QF-AC}^{0,1} + \text{WKL}$  is a strong theory where a good deal of analysis can be carried out. In the theory we can prove, for instance, the Heine-Borel covering lemma: Every covering of the closed interval  $[0, 1]$  by a sequence of open intervals has a finite sub-covering. Within the theory it is also provable that any continuous real-valued function on  $[0, 1]$ , or on any compact metric space, is bounded; see (Simpson, 1999, 36) for a list of theorems provable in  $\text{WKL}_0$ . The theory is, nevertheless, proof theoretical weak in the sense that it can be reduced to PRA, which essentially corresponds to Hilbert's finitism.

WKL expresses compactness of the Cantor space. It is thus interesting to see that functional interpretation is capable of ascribing constructive meaning to a notion such as compactness, which is a very fruitful concept in classical (ideal) mathematics.

We thus conclude that functional interpretation as a tool for analysing mathematics, for justifying ideal elements and for unwinding constructive content from classical proofs is a particularly strong tool, which provides deep insights with relevant conclusions both for mathematics and for philosophy of mathematics.

### 7.5 Different interpretations—different validations: Mathematics and language

Finally, we will come back to the fact that different interpretations validate different principles; the fact that there is no unambiguous characterisation of constructivism. We see that it is somehow up to the individual to choose the interpretation which fits the mathematical problem at hand, or the one which – in his view – gives the broadest and most coherent interpretation. Why is this? This certainly is a difficult question. But our generalised Hilbertian view on mathematics – as outlined in the first chapter – has something to say on these matters. There is some finitary and absolutely objective part of mathematics, and as such there are no problems with respect to this. But as mathematics evolves, ideal objects are introduced—the

set of all natural numbers for instance. But given this set it is natural to ask for the set of all subsets of natural numbers. And in this way many other abstract ideal elements are introduced, such as compactness, transfinite induction, and so on. However, as the progression takes place, mathematics exceeds the power of any language, since – due to the following argument – we cannot have names for all subsets of the natural numbers.

Any alphabet must be countable and any name must be some finite combination of letters from an alphabet. Since all names from any given language can be ordered lexicographically there can be only countable many names. Therefore no language can capture all aspects about ideal mathematics. This is again contrary to Hilbert who thought that we could investigate all aspects and details concerning mathematical concepts and objects by the axiomatic method. But since language is what we have for describing mathematics, the failure of language to capture all aspects about mathematics explains why we do not get a clear cut global interpretation of mathematics, since any interpretation must be relative to some specific language. However, as we have seen in numerous examples of this thesis, there are definitely local interpretations which carry important mathematical and philosophical consequences. The Dialectica interpretation is one of them.

## Bibliography

- Andersen, G., Jeppesen, L. M., Jørgensen, K. F. & Zeck, I. P. (1996). Bevisteori: Eksemplificeret ved Gentzens bevis for konsistensen af teorien om de naturlige tal, *Tekster fra IMFUFA* 326, Roskilde. The text can also be downloaded at <http://akira.ruc.dk/~frovin/bevist.pdf>.
- Andersen, J. H. (2000). *Hilberts matematikfilosofi*, Master's thesis, Roskilde.
- Artëmov, S. N. (2001). Explicit provability and constructive semantics, *Bulletin of Symbolic Logic* 6(1): 1–36.
- Audi, R. (ed.) (1995). *The Cambridge Dictionary of Philosophy*, Cambridge University Press, Cambridge.
- Avigad, J. & Feferman, S. (1998). Gödel's Functional ("Dialectica") Interpretation, in (Buss, 1998, 337–406).
- Barwise, J. (1977). *Handbook of Mathematical Logic*, North-Holland, Amsterdam.
- Bauer, F. L. & Steinbrüggen, R. (eds) (2000). *Foundations of Secure Computation*, Vol. 175 of *Series F: Computer and Systems Sciences*, IOS Press, Amsterdam.
- Berger, U., Buchholz, W. & Schwichtenberg, H. (n.d.). Refined program extraction from classical proofs, to appear in *Annals of Pure and Applied Logic*.
- Berger, U. & Schwichtenberg, H. (1995). Program extraction from classical proofs, in (Leivant, 1995, 77–97).
- Bernays, P. (1922). Über Hilberts Gedanken zur Grundlegung der Mathematik, *Jahresberichte DMV* 31: 10–19.
- Bernays, P. (1928). Über Nelsons Stellungnahme in der Philosophie der Mathematik, *Die Naturwissenschaften, Wochenschrift für die Fortschritte der Naturwissenschaft, der Medizin und der Technik* 9: 142–145.
- Bernays, P. (1967). David Hilbert, in (Edwards, 1967, 496–504).
- Bernays, P. (1976). *Abhandlungen zur Philosophie der Mathematik*, Wissenschaftliche Buchgesellschaft, Darmstadt.
- Bernays, P. & Hilbert, D. (1939). *Grundlagen der Mathematik*, Vol. 2, Springer.
- Bezem, M. (1988). Equivalence of bar recursors in the theory of functionals of finite type, *Arch. Math. Logic* 27: 149–160.
- Bishop, E. (1970). Mathematics as a numerical language, in (Kino et al., 1970, 53–71).
- Brouwer, L. E. J. (1907). *Over de grondslagen der wiskunde*, PhD thesis, Amsterdam.

- Buss, S. (ed.) (1998). *Handbook of Proof Theory*, Elsevier, Amsterdam.
- Cook, S. A. & Urquhart, A. (1993). Functional interpretations of feasibly constructive arithmetic, *Annals of Pure and Applied Logic* **63**: 103–200.
- Coquand, T. & Hofmann, M. (1999). A new method for establishing conservativity of classical systems over their intuitionistic version, *Math. Struct. in Comp. Science* **9**: 323–333.
- Detlefsen, M. (1995). Hilbert's Program, in (Audi, 1995, 327–328).
- Diller, J. (1968). Zur Berechenbarkeit primitiv-rekursiver Funktionale endlicher Typen, in (Schmidt et al., 1968, 109–120).
- Diller, J. (1979). Functional interpretations of Heyting's arithmetic in all finite types, *Proceedings Bicentennial Congress Wiskundig Genootschap*, pp. 149–176, part 3.
- Diller, J. & Nahm, W. (1974). Eine Variante zur Dialecetica—Interpretation der Heyting-Arithmetik endlicher Typen, *Archiv für mathematische Logik und Grundlagenforschung* **16**: 49–66.
- Dragalin, A. G. (1980). New forms of realizability and Markov's rule (Russian), *Dokl. Akad. Nauk. SSSR* **251**: 534–537. English translation in *Soviet Math. Dokl* **21**: 461–464, (1980).
- Dummett, M. (1977). *Elements of Intuitionism*, Clarendon Press, Oxford.
- Edwards, P. (ed.) (1967). *Encyclopedia of Philosophy*, Vol. 3, Macmillan and Free Press, New York.
- Feferman, S. (1988). Weyl Vindicated: *Das Kontinuum* Seventy Years Later, reprint in (Feferman, 1998, 249–283).
- Feferman, S. (1996). Kreisel's "unwinding" program, in (Odifreddi, 1996, 247–273).
- Feferman, S. (1998). *In the Light of Logic*, Oxford University Press, Oxford.
- Feferman, S. (2000). Does reductive proof theory have a viable rationale?, *Erkenntnis* **53**: 63–96.
- Friedman, H. (1978). Classically and intuitionistically provably recursive functions, in (Müller & Scott, 1978, 21–27).
- Friedrich, W. (1985). Gödelsche Funktionalinterpretation für eine Erweiterung der Klassischen Analysis, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **31**: 3–29.
- Gentzen, G. (1933). Über das Verhältnis zwischen intuitionistischer und klassischer Logik. Originally to appear in the *Mathematische Annalen*, reached the stage of galley proofs but was withdrawn. It was finally published in *Archiv für Mathematische Logik und Grundlagenforschung* **16**: 119–132, 1974, translation in (Gentzen, 1969, 53–67).

- Gentzen, G. (1935). Untersuchungen über das logische Schließen, *Mathematische Zeitschrift* **39**: 176–210.
- Gentzen, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie, *Mathematische Annalen* **112**: 493–565. English translation in (Gentzen, 1969, 132–213).
- Gentzen, G. (1943). Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, *Mathematische Annalen* **119**: 140–161.
- Gentzen, G. (1969). *The Collected Papers of Gerhard Gentzen*, North Holland, Amsterdam. English translations of Gentzen's papers, edited and introduced by M. E. Szabo.
- Girard, J.-Y. (1987). Linear logic, *Theoretical Computer Science* **50**: 1–102.
- Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik* **38**: 173–198. Reprinted with an English translation in (Gödel, 1986, 144–195).
- Gödel, K. (1933). Zur intuitionistischen Arithmetik und Zahlentheorie, *Ergebnisse eines mathematischen Kolloquiums* **4**: 34–38. With translation also in (Gödel, 1986, 300–303).
- Gödel, K. (1933a). The present situation in the foundations of mathematics, in (Gödel, 1995, 45–53).
- Gödel, K. (1938). Vortrag bei Züri, in (Gödel, 1995, 86–113).
- Gödel, K. (1941). In what sense is intuitionistic logic constructive?, in (Gödel, 1995, 189–200).
- Gödel, K. (1958). Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica* **12**: 280–287. Reprinted with English translation and introduction in (Gödel, 1990, 241–251).
- Gödel, K. (1986). *Collected Works*, Vol. I, Oxford University Press, New York.
- Gödel, K. (1990). *Collected Works*, Vol. II, Oxford University Press, New York.
- Gödel, K. (1995). *Collected Works*, Vol. III, Oxford University Press, New York.
- Goodman, N. D. (1970). A theory of constructions equivalent to arithmetic, in (Kino et al., 1970, 101–120).
- Hendricks, V. F., Pedersen, S. A. & Jørgensen, K. F. (eds) (2000). *Proof Theory: History and Philosophical Significance*, Vol. 292 of *Synthese Library*, Kluwer Academic Publishers, Dordrecht.
- Heyting, A. (1930). Die formalen Regeln der intuitionistischen Logik, *Sitzungsber. Preuss. Akad. Wiss. Phys.-Math.* pp. 42–56.

- Heyting, A. (1930a). Die formalen Regeln der intuitionistischen Mathematik, *Sitzungsber. Preuss. Akad. Wiss. Phys.-Math.* pp. 57–71, 158–169.
- Heyting, A. (1934). *Mathematische Grundlagenforschung. Intuitionismus, Beweistheorie*, Springer, Berlin.
- Heyting, A. (ed.) (1959). *Constructivity in mathematics*, North-Holland, Amsterdam.
- Hilbert, D. (1918). Axiomatisches Denken, in (Hilbert, 1935, 146–156).
- Hilbert, D. (1922). Neubegründung der Mathematik, *Abhandlungen aus dem mathematischen Seminar der Hamburgerischen Universität* **1**: 157–177. Also published in (Hilbert, 1935, 157–178).
- Hilbert, D. (1926). Über das Unendliche, *Mathematische Annalen* **95**: 161–190.
- Hilbert, D. (1927). Die Grundlagen der Mathematik, *Abhandlungen aus dem mathematischen Seminar der Hamburgerischen Universität* **6**: 65–85.
- Hilbert, D. (1930). Naturerkennen und Logik, in (Hilbert, 1935, 378–388).
- Hilbert, D. (1935). *Gesammelte Abhandlungen*, Vol. 3, Springer, Berlin.
- Hilbert, D. & Ackermann, W. (1928). *Grundzüge der theoretischen Logik*, Springer, Berlin.
- Hinata, S. (1967). Calculability of primitive recursive functionals of finite type, *Science reports of the Tokyo Kyoiku Daigaku, A* **9**: 218–235.
- Hodges, W., Hyland, M., Steinhorn, C. & Truss, J. (eds) (1996). *Logic: From Foundations to Applications*, Oxford University Press, Oxford.
- Howard, W. (1970). Assignment of ordinals to terms for primitive recursive functionals of finite type, in (Kino et al., 1970, 453–468).
- Howard, W. (1973). Hereditarily majorizable functionals of finite type, in (Troelstra, 1973, 454–461).
- Jørgensen, K. F. & Pedersen, S. A. (2000). Matematik i et filosofisk perspektiv, *Kvan* **56**: 111–122.
- Kant, I. (1781/87). *Kritik der Reinen Vernunft*, Felix Meiner Verlag, Hamburg. Both first and second edition are published by Raymund Schmidt (1990).
- Kino, A., Myhill, J. & Vesley, R. E. (eds) (1970). *Intuitionism and Proof Theory*, North Holland, Amsterdam.
- Kirby, L. A. S. & Paris, J. B. (1982). Accessible independence results for Peano Arithmetic, *Bulletin of the London Mathematical Society* **14**: 285–293.
- Kleene, S. C. (1945). On the interpretation of intuitionistic number theory, *Journal of Symbolic Logic* **10**: 109–124.

- Kleene, S. C. (1952). *Introduction to Metamathematics*, North Holland, Amsterdam.
- Kleene, S. C. (1967). *Mathematical Logic*, John Wiley & Sons, New York—London—Sydney.
- Kleene, S. C. (1969). *Formalized recursive functionals and formalized realizability*, Vol. 89 of *Memoirs of the American Mathematical Society*, American Mathematical Society, Providence, Rhode Island.
- Kohlenbach, U. (1992). Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization, *Journal of Symbolic Logic* **57**: 1239–1273.
- Kohlenbach, U. (1992a). Pointwise hereditary majorization and some applications, *Archive for Mathematical Logic* **31**: 227–241.
- Kohlenbach, U. (1993). Effective moduli from ineffective uniqueness proofs. An unwinding of de La Vallee Poussin's proof for Chebycheff approximation, *Annals of Pure and Applied Logic* **64**: 27–94.
- Kohlenbach, U. (1993a). New effective moduli of uniqueness and uniform apriori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory, *Numer. Funct. Anal. and Optimiz.* **14**: 581–606.
- Kohlenbach, U. (1996). Analysing proofs in analysis, in (Hodges et al., 1996).
- Kohlenbach, U. (1998). Relative constructivity, *Journal of Symbolic Logic* **63**: 1218–1238.
- Kohlenbach, U. (1998a). Proof Interpretations, *BRICS Lecture Series 98-1*, BRICS.
- Kohlenbach, U. (1999). On the no-counterexample interpretation, *Journal of Symbolic Logic* **64**: 1491–1511.
- Kohlenbach, U. (2001). A note on Spector's quantifier-free rule of extensionality, *Archive for Mathematical Logic* **40**: 89–92.
- Kohlenbach, U. (n.d.). On the computational content of the Krasnoselski and Ishikawa fixed point theorems, to appear in: Blanck, J., Brattka, V., Hertling, P., Weihrauch, K. (eds), *Proceedings of the Fourth Workshop on Computability and Complexity in Analysis*, Springer Lecture Notes in Computer Science.
- Kolmogorov, A. N. (1925). On the principle of the excluded middle (Russian), *Matematicheskij Sbornik. Akademiya Nauk SSSR i Moskovskoe Matematicheskoe Obshchesto* **32**: 646–667. Translation in (van Heijenoort, 1967, 414–437).
- Kreisel, G. (1951). On the interpretation of non-finitist proofs, part I, *Journal of Symbolic Logic* **16**: 241–267.
- Kreisel, G. (1958). Mathematical significance of consistency proofs, *Journal of Symbolic Logic* **23**: 155–182.

- Kreisel, G. (1959). Interpretation of analysis by means of constructive functionals of finite type, in (Heyting, 1959, 101–128).
- Kreisel, G. (1962). On weak completeness of intuitionistic predicate logic, *Journal of Symbolic Logic* **27**: 139–158.
- Kreisel, G. (1962a). Foundations of intuitionistic logic, in (Nagel et al., 1962, 198–210).
- Kreisel, G. (1987). Proof theory: Some personal recollections, in (Takeuti, 1987, 395–405).
- Kuroda, S. (1951). Intuitionistische Untersuchungen der formalistischen Logik, *Nagoya Math. J.* **2**: 35–47.
- Leivant, D. (ed.) (1995). *Logic and Computational Complexity. International Workshop LCC '94*, Vol. 960 of *Lecture Notes in Computer Science*, Springer, Berlin.
- Luckhardt, H. (1973). *Extensional Gödel Functional Interpretation: A Consistency Proof of Classical Analysis*, Springer, Berlin.
- Majer, U. (1993). Hilberts Methode der Idealen Elemente und Kants regulativer Gebrauch der Ideen, *Kant-studien* **84**(1): 50–77.
- Mancuso, P. (n.d.). On the constructivity of proofs. A debate among Behmann, Bernays, Gödel, and Kaufmann, forthcoming in Sieg, W., Sommer, R. & Talcott, C. (eds), *Reflections. Essays in honour of Solomon Feferman*, A. K. Peters.
- Müller, G. H. & Scott, D. S. (eds) (1978). *Higher Set Theory*, Springer, Berlin.
- Murthy, C. (1990). *Extracting Constructive Content from Classical Proofs*, PhD thesis, Cornell University.
- Nagel, E., Suppes, P. & Tarski, A. (eds) (1962). *Proc. Logic Methodology and Philosophy of Science*, Stanford University Press, Stanford.
- Odifreddi, P. (ed.) (1996). *Kreiseliana: About and Around George Kreisel*, A. K. Peters.
- Paris, J. B. & Harrington, L. (1977). A mathematical incompleteness in Peano arithmetic, in (Barwise, 1977, 1133–1142).
- Parson, C. (1972). On  $n$ -quantifier induction, *Journal of Symbolic Logic* **37**: 466–482.
- Posy, C. J. (1995). Philosophy of mathematics, in (Audi, 1995, 594–597).
- Rath, P. (1978). *Eine verallgemeinerte Funktionalinterpretation der Heyting-Arithmetik*, PhD thesis, Universität Münster.
- Rowe, D. (2000). The calm before the storm: Hilbert's early views on foundations, in (Hendricks et al., 2000, 55–93).
- Schmidt, H., Schütte, K. & Thiele, H. (eds) (1968). *Contributions to Mathematical Logic*, North Holland, Amsterdam.

- Schwichtenberg, H. (2000). Refined Program Extraction from Classical Proofs: Some Case Studies, in (Bauer & Steinbrüggen, 2000, 147–167).
- Simpson, S. G. (1999). *Subsystems of Second Order Arithmetic*, Springer, Berlin.
- Smorynski, C. (1973). Applications of Kripke models, in (Troelstra, 1973, 324–391).
- Smorynski, C. (1977). The Incompleteness Theorems, in (Barwise, 1977, 821–867).
- Spector, C. (1962). Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, *Proceedings of the Symposia in Pure Mathematics* **5**: 1–27.
- Stein, M. (1977). *Interpretationen der Heyting-Arithmetik endlicher Typen*, PhD thesis, Universität Münster.
- Tait, W. W. (1967). Intensional interpretations of functionals of fine type I, *Journal of Symbolic Logic* **32**: 198–212.
- Tait, W. W. (n.d.). Remarks on Finitism, forthcoming in Sieg, W., Sommer, R. & Talcott, C. (eds), *Reflections. Essays in honour of Solomon Feferman*, A. K. Peters.
- Takeuti, G. (1987). *Proof Theory*, 2. edn, North Holland, Amsterdam.
- Troelstra, A. S. (1990). Introductory note to 1958 and 1972, in (Gödel, 1990, 217–241).
- Troelstra, A. S. (1995). Natural deduction for intuitionistic linear logic, *Annals of Pure and Applied Logic* **73**(1): 79–108.
- Troelstra, A. S. (1995a). Introductory note to \*1941, in (Gödel, 1995, 186–189).
- Troelstra, A. S. (1998). Realizability, in (Buss, 1998, 407–475).
- Troelstra, A. S. (ed.) (1973). *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, Springer, Berlin—Heidelberg—New York.
- Troelstra, A. S. & Schwichtenberg, H. (1996). *Basic Proof Theory*, Cambridge University Press, Cambridge.
- Troelstra, A. S. & van Dalen, D. (1988). *Constructivism in Mathematics: An Introduction*, North Holland, Amsterdam—New York—Oxford—Tokyo.
- van Heijenoort, J. (ed.) (1967). *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*, Harvard University Press, Cambridge M.A.
- Weiermann, A. (1998). How is it that infinitary methods can be applied to finitary mathematics? Gödel's *T*: a case study, *Journal of Symbolic Logic* **63**: 1348–1378.
- Weyl, H. (1918). *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*, Veit, Leipzig.

Whitehead, A. N. & Russell, B. (1910–13). *Principia Mathematica*, Vol. I-III, 1. edn, Cambridge University Press, Cambridge.

Yasugi, M. (1963). Intuitionistic analysis and Gödel's interpretation, *Journal of the Mathematical Society of Japan* **15**(2): 101–112. Review with corrections in *J.S.L.* **37**:104, 1972.

## Index

- $\Gamma_{\underline{mr}}$ , 72
- $(C^i(\mathbf{x}_i, \mathbf{T}_i \mathbf{x}))_{i=1}^n$ , 37
- $\perp_C$ , 55
- $:\equiv$ , 16
- $H^\Gamma$ , 101
- $\langle \cdot, \cdot \rangle$ , 87
- $\underline{mr}$ , 70
- $\underline{mrt}$ , 81
  
- A-translation, 69, 77
- assumption class
  - $[A]$ , 29
- axiom of choice, 13
  - AC, 23
  - $AC^{\sigma, \tau}$ , 24
  - AC, 51
  - $AC_R$ , 51
  
- Bernays, Paul, 3
- $\beta$ -contraction, 20
- BHK interpretation, 7, 69, 72, 101, 107
- bounded universal quantifier, 48, 86
- Brouwer, L.E.J., 2, 103
  
- characterisation of  $\underline{mr}$ , 83
- characteristic term, 22
- Church's thesis, 13, 75
- $CL^\omega$ , 59, 71
- closure properties
  - of  $E\text{-}HA^\omega \pm IP_{\text{ef}}^\omega \pm AC$ , 82
  - of  $WE\text{-}HA^\omega + IP_{\forall}^\omega + MP^\omega + AC + \Gamma$ , 101
  - of  $WE\text{-}HA^\omega \pm IP_{\forall}^\omega$ , 98
- combinator
  - $\Sigma_{p, \tau, \sigma}$ , 19
- compactness, 108, 109
- $\text{Con}_{PA}$ , 34
- constructive content, 12
- constructive existence, 64
- contraction, 31
  - $A \rightarrow A \wedge A$ , 28, 74
  - in natural deduction, 31
- contraction lemma, 29, 38, 51, 74
  
- deduction theorem, 33
  - with assumptions, 33
  - with axioms, 33
- definition of
  - formula, 16
  - term, 16
- Dialectica interpretation of
  - $WE\text{-}HA_{ND}^\omega + MP^\omega + IP_{\forall}^\omega + AC$ , 52
  - $WE\text{-}PA_{ND}^\omega + QF\text{-}AC_R$ , 60
  - $WE\text{-}HA_H^\omega$ , 27
  - $WE\text{-}HA_{ND}^\omega$ , 39
- Dialectica translation
  - definition of, 23
  - of derivation, 37
  - soundness of, 27, 39
- Dialectica with truth translation
  - definition of, 84
- Diller-Nahm variant, 86
- disjunction property, 8, 70, 80
  
- $E\text{-}HA^\omega$ , 18, 70
- eigenvariable, 17
- $\exists$ -free
  - formula, 72
- existence property, 8, 70, 80
- extensionality, 13, 18, 105
- extraction theorem, 60
  
- Feferman, Solomon, 12, 103
- finitary mathematics, 3
  
- Gödel, Kurt, 3, 63, 84
- Gentzen, Gerhard, 9, 103
  
- $HA^\omega$ , 15
- $HA_H^\omega$ , 15
- $HA_{ND}^\omega$ , 15
- Herbrand's theorem, 35, 61
- Heyting arithmetic, 15
  - HA, 15, 65, 79
- Heyting, Arend, 3
- higher type equations, 17
- Hilbert's program, 3, 62, 108
- Hilbert, David, 2
  
- ideal elements, 3, 12, 108, 109
- ideal mathematics, 3, 109

- IL<sup>ω</sup>, 59
- incompleteness, 6
- independence-of-premise, 13, 25, 105
  - IP<sub>∇</sub><sup>ω</sup>, 25, 51
  - IP<sub>ef</sub><sup>ω</sup>, 72
  - IP<sub>∇</sub><sup>ω</sup>, 25, 53
  - IP<sup>ω</sup>, 25
  - IP<sub>∇R</sub><sup>ω</sup>, 51
- induction lemma, 39
- j<sub>i</sub>*, 87
- König's lemma, 13
- Kant, Immanuel, 3
- Kleene's *T*-predicate, 26, 53, 85
- Kreisel, Georg, 8, 12, 103
- λ-abstraction, 19
- $\mathcal{L}(\text{WE-HA}^\omega)$ , 16
- linear logic, 31, 49
- Markov's principle, 13, 25, 74, 75
  - MP<sup>ω</sup>, 25, 51
  - MP<sup>σ</sup>, 25
  - MP<sub>R</sub><sup>ω</sup>, 51
- Markov's rule, 76, 79, 99, 105
- modified realisability translation
  - definition of, 70
  - soundness of, 73
- modified realisability with truth translation
  - definition of, 81
  - soundness of, 81
- natural deduction, 30
- negative
  - formula, 57
  - translation, 9, 56
- non-constructive
  - proof, 2
- normal form
  - Herbrand, 66
  - prenex, 66
- ordinal analysis, 103
- Peano arithmetic, 6, 15
  - PA, 65
- PRA, 3
- program extraction
  - for PA, 80
  - for WE-PA<sub>ND</sub><sup>ω</sup> + QF-AC, 60
- projector
  - Π<sub>σ,τ</sub>, 19
- proof mining, 12, 64, 103, 107
- Q*-translation
  - definition of, 87
  - soundness of, 88
- q-realisability, 86
- QF-ER, 18, 32
- QF-AC, 59
- recursor
  - R<sub>σ</sub>, 19
- reduction
  - global, 6
  - local, 6, 103
- reductive proof theory, 103
- reflection principle, 4
- rule of
  - choice (ACR), 80, 99
  - independence-of-premise for purely universal formulas (IPR<sub>∇</sub><sup>ω</sup>), 99
  - independence-of-premise for ∃-free formulas (IPR<sub>ef</sub><sup>ω</sup>), 80
  - induction, 32
- Russell, Bertrand, 2
- Spector, Clifford, 68
- T, 15
- tertium non datur, 8, 55
- the no-counterexample interpretation, 66
- transfinite induction, 10, 103
- translation
  - A*, 77
  - Q*, 87
  - Dialectica, 23
  - Dialectica with truth, 84
  - Diller-Nahm, 48
  - Kuroda's negative, 57
  - modified realisability, 70
  - modified realisability with truth, 81
  - negative, 56
- truth property, 81
- type assignment, 15

type level, 15

WE-HA<sub>H</sub><sup>ω</sup>, 17

WE-HA<sub>ND</sub><sup>ω</sup>, 30

WE-T<sub>ND</sub>, 32

WE-T, 21

WE-T<sub>H</sub>, 21

WE-HA<sup>ω</sup>, 15

weak König's lemma, 103

WKL, 109

WKL<sub>0</sub>, 11, 109

WE-PA<sup>ω</sup>, 55

WE-PA<sub>H</sub><sup>ω</sup>, 56

WE-PA<sub>ND</sub><sup>ω</sup>, 55

Weyl, Hermann, 2

zero-functional

$\circ^\sigma$ , 28