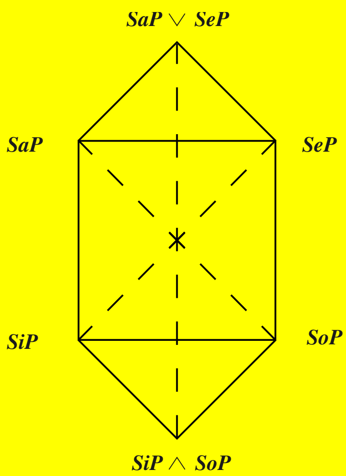
Editors

Vincent Hendricks
Fabian Neuhaus
Stig Andur Pedersen
Uwe Scheffler
Heinrich Wansing

First–Order Logic Revisited



Logische Philosophie

λογος

Die Open-Access-Stellung der Datei erfolgte mit finanzieller Unterstützung des Fachinformationsdiensts Philosophie (https://philportal.de/)

DOI: https://doi.org/10.30819/0475

Editors

Vincent Hendricks
Fabian Neuhaus
Stig Andur Pedersen
Uwe Scheffler
Heinrich Wansing

First-Order Logic Revisited

Logische Philosophie
Herausgeber:
H. Wessel, U. Scheffler, Y. Shramko, M. Urchs

λογος

Herausgeber der Reihe Logische Philosophie

Horst Wessel

Unter den Linden 61
D-14621 Schönwalde
Deutschland

WesselH@philosophie.hu-berlin.de

Uwe Scheffler

Institut für Philosophie
Humboldt-Universität zu Berlin
Unter den Linden 6
D-10099 Berlin
Deutschland

SchefflerU@philosophie.hu-berlin.de

Yaroslav Shramko

Lehrstuhl für Philosophie
Staatliche Pädagogische Universität
UA-324086 Kryvyj Rih
Ukraine

kff@kpi.dp.ua

Max Urchs

Fachbereich Philosophie
Universität Konstanz
D-78457 Konstanz
Deutschland

max.urchs@uni-konstanz.de

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISSN 1435-3415
ISBN 3-8325-0475-3

Logos Verlag Berlin
Gubener Str. 47, 10243 Berlin, Tel.: +49 030 42 85 10 90
INTERNET: http://www.logos-verlag.de

David Hilbert (1862–1943)

[http://www-gap.dcs.st-and.ac.uk/
history/PictDisplay/Hilbert.html]

Wilhelm Ackermann (1896–1962)

[Thanks to *Albert Röser, Steinfurt*]

Preface

First-Order Logic Revisited is the proceedings from the conference *FOL75 – 75 Years of First-Order Logic* held at Humboldt University, Berlin, Germany, September 18 – 21, 2003.

As the editors of this volume and the core group of organization for the conference we would like to express our gratitude to the following individuals for contributing to making *FOL75 – 75 Years of First-Order Logic* the successful conference it was:

J.v. Benthem, J.M. Dunn, H.-D. Ebbinghaus, D. M. Gabbay, G. Sandu, P.G. Hansen, and a group of students from Humboldt University

The conference was only made possible by the financial and organizational cooperation between the following institutions to which we would also like to extend our gratitude:

SHF – The Danish Research Council for the Humanities, DFG – Deutsche Forschungsgemeinschaft, Dresden University of Technology, Humboldt University, Logos-Verlag, ΦLOG – The Danish Network for Philosophical Logic and Its Applications, Roskilde University, Carl und Max Schneider Stiftung at the Institute of Philosophy of Humboldt University

It is our hope that these proceedings adequately convey the satisfactory fulfillment of the conference aim of reflecting upon and discussing first-order logic, its history, its wide range of applications, its extensions and alternatives in the light of the 75 years past since the publication of Hilbert and Ackermann's seminal *Grundzüge der Theoretischen Logik.*

Vincent F. Hendricks Fabian Neuhaus
Stig Andur Pedersen Uwe Scheffler Heinrich Wansing
June 2004

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004)

Contents

Introduction

V. Hendricks, F. Neuhaus, S.A. Pedersen, U. Scheffler, H. Wansing

In 1928 Hilbert and Ackermann published their famous *Grundzüge der Theoretischen Logik*. In the impressively short book they were able to cover the propositional calculus, the calculus of classes, the higher-order calculus of relations and, most importantly, present an axiomatic system of the first-order logic, which altogether may be viewed as the very starting point of modern mathematical logic. Today, 75 years later, first-order logic (FOL) is a powerful tool and an indispensable companion in a variety of fields ranging from philosophy over mathematics to computer science, linguistics and psychology.

This Anniversary called for a celebration. The aim *FOL75 – 75 Years of First-Order Logic* was to honor the 75 years since the publication of the seminal book, and accordingly dedicate an event to reflection upon and discussion of first-order logic, its history, its wide range of applications, its extensions and alternatives.

In order to meet this ambitious aim with satisfaction, leading international scholars were invited to deliver plenary talks in such a way that all aspects from history to applications were covered as they relate to first-order logic. It was with no greater pride that the following list of distinguished philosophers, mathematicians, logicians, and computer scientists could be put together:

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 1–6

Hajnal Andréka (Hungary)

George Englebretsen (Canada)

William Ewald (USA)

Jeroen Groenendijk (The Netherlands)

Petr Hájek (Czech Republic)

Jaakko Hintikka (USA)

Wilfrid Hodges (Great Britain)

István Németi (Hungary)

Alan Robinson (USA)

Dana Scott (USA)

Valentin Shehtman (Russia)

Martin Stokhof (The Netherlands)

Equally important was to have a program committee capable of putting a feasible program together. Again a matching assembly was established consisting of

Johan van Benthem (The Netherlands)

J. Michael Dunn (USA)

Heinz-Dieter Ebbinghaus (Germany)

Dov M. Gabbay (Great Britain)

Gabriel Sandu (Finland)

together with the core group of organizers. The program committee was also responsible for assessing the large number of contributed papers submitted for presentation. The result of this careful selection process is reflected in these proceedings. Rather than attempting to categorize the papers and presenting them in potentially arbitrary and overlapping compartments, we decided to simply place them in alphabetical order after authors and let this listing in turn demonstrate the many theoretical aspects, considerations, discussions and practical applications first-order logic has generated over the past 75 years.

In the first paper 'Logical Analysis of Relativity Theories' by H. Andréka, J.X. Madarász, and I. Németi first-order logic is applied to analyze relativity theory. Among other interesting results is the presentation of a first-order logic axiomatization called *Spectrel*, which turns out to be 'a faithful, streamlined FOL axiomatization of (the kinematics of) special relativity theory. *Specrel* is intended to consist of simple, intuitively convincing, logically transparent, natural axioms.' Various fragments of the axiomatization are then scrutinized together with generalizations to the general theory of relativity.

'Safety Signatures for First-Order Languages and their Applications' by A. Avron develops a general framework for safely isolating the set of all *safe* formulas of various first-order languages. The significance of circumscriptions of this nature is then illuminated through examples drawn from a rich pool of applications in naive set theory with respect to the comprehension schema, in computability theory pertaining to recursive relations and in query languages for databases.

K. Brünnler and A. Guglielmi in 'A First Order System with Finite Choice of Premises' suggest a simple and elegant way of eliminating infinite choices in sequent systems of first-order logic which is of particular interest to automated deduction and logic programming. The elimination is accomplished by applying a particular calculus of structures allowing for derivations using inference rules 'deep' inside first-order formulae.

In 'Predicate Logic, Predicates, and Terms' G. Englebretsen outlines a Term Logic developed jointly with F. Sommers which is on par with the general inferential powers of first-order logic but simultaneously 'enjoys certain advantages in terms of simplicity and naturalness.' The outline of the term logic is accompanied by a becoming historical overview and motivation.

The history and chronological development of first-order logic before and around the publication of Hilbert and Ackermann's *Grundzüge der Theoretischen Logik* is discussed by W. Ewald in 'FOL75?'. It turns out that Hilbert's (and Bernays) meta-mathematical understanding of logic pre-dates both this publication and Hilbert's later research in proof theory. However his early understanding of first-order logic from 1918 was not quite the modern one later associated with him.

Petr Hájek begins by addressing the vagueness of propositions and from here on is lead to fuzzy logic in 'Fuzzy Logic and Arithmetical Hierarchy IV'. Fuzzy first-order logic (BL∀) is subsequently defined and a number of interesting arithmetical complexity properties are then either outlined or proved.

The expressibility of first-order logic is discussed by J. Hintikka in 'What is the True Algebra of First-Order Logic?' with particular emphasis on quantifier dependencies. Standard syntactical nesting of quantifiers is incapable of capturing many pertinent patterns of dependence between variables in the language. A remedy to this shortcoming is to introduce an independence operator into first-order logic which in turn implements Hintikka's independence-friendly (IF) logic. IF logic is subsequently situated in an algebraic setting.

In 'The Importance and Neglect of Conceptual Analysis: Hilbert-Ackermann iii.3' W. Hodges returns to the intellectual surroundings before and in the vicinity of the publication of *Grundzüge* with particular attention paid to the debate between Frege and Hilbert on conceptual analysis and logical inference. The twist and turns of this debate are closely followed and ends with an interesting outline of a program for mathematical work which both Hilbert, Frege and also Tarski could subscribe to.

Marcus Kracht continues by discussing substitution in 'Notes on Substitution in First-Order Logic'. Substitutions are used all the time and everywhere in first-order logic. Kracht rightfully asks whether the syntax of first-order logic actually naturally lends itself to a notion of replacement of formulae or terms in the first place. Through linguistic theory an interesting result is presented, which goes to show the apprehensions one may rightfully have towards substitution in first-order languages.

A new formal system of logic is presented in 'Logical Inquiries into a New Formal System with Plural Reference' by R. Lanzet and H. Ben-Yami. The idea is to base a formal system of logic on the semantics of natural language rather than on an artificial language like first-order. Furthermore, one may obtain a deductive power of this new system matching the deductive power of some versions of first-order logic.

The relation between first-order logic and the general theory of relativity is revisited in 'On Generalizing the Logic-Approach to Space-Time Towards General Relativity: First Steps' by J.X. Madarász, I. Németi, and C. Tőke. Using many-valued first-order logic and the axiomatization in *Spectrel* the authors outline two first steps of applying logic to the general theory of relativity in terms of (i) foundations for space-time theories, and (ii) logic-based conceptual analysis of theories of relativity.

'Constructive Predicate Logic and Constructive Modal Logic. Formal Duality versus Semantical Duality' by S.P. Odintsov and H. Wansing analyzes constructive modal and first-order logic together with duality properties of the modal operators. In particular, a translation of modal logic into constructive predicate logic is used to obtain natural examples of constructive modal logics with strong negation that fail to satisfy syntactic duality.

J.A. Robinson goes on to discuss mechanization, proof and formalized reasoning in 'Logic is not the Whole Story'. First-order logic is often considered to be capturing important features of actual reasoning in mathematical proofs notably 'objective validity or logical correctness'. With historical flashbacks Robinson provides an illuminating discussion as to whether another important feature of mathematical proofs, namely 'epistemological coherence' is reflected in the logical setting.

In 'First-Order Logic, Second-Order Logic, and Completeness' M. Rossberg focuses on second-order logic. While first-order logic is typically considered to being properly a *logic*, second-order logic is considered more dubious as incompleteness results are taken to demonstrate the intractability of the second-order consequence relation. The author provides arguments to the effect that although a completeness result of some kind is lacking for second-order logic this does not suffice for dismissing second-order logic as a logic.

First-order logic as a formal model of reasoning is revisited by M. Thielscher in 'Logic-Based Agents and the Frame Problem: A Case for Progression'. In Artificial Intelligence a common problem encountered is the Frame Problem. Using progression-based rather than regression-based logic programming systems, the superiority of the former progression-based system is discussed for dealing with the Frame Problem.

D.E. Willard introduces new variations of Gödel's second incompleteness theorem in 'A Version of the Second Incompleteness Theorem For Axiom Systems that Recognize Addition But Not Multiplication as a Total Function'. Such variations are interesting in that boundary-exceptions to the theorem have been uncovered, and the limits of the famous Gödel result in turn may be explored.

In the final paper 'First-Order Logic: (Philosophical) Pro and Contra' J. Wolenski takes first-order logic for a thorough philosophical treatment. Questions of purpose, scope, relation to other disciplines, the inherent nature of logic in the human mind, are dealt with. It is convincingly argued that firm and rigorous meta-logical results actually may shed illuminating light on the traditional philosophical questions about logic and its very nature.

The participants of the conference took an extraordinary interest in the celebration; both in recognition of the general multi-disciplinary importance of first-order logic, but apparently also because of the importance first-order logic has had in shaping their own ideas and subsequent intellectual accomplishments. Alongside with Gottlob Frege, Kurt Gödel, Leopold Löwenheim, Thoralf Skolem, Jacques Herbrand, Gerhad Gentzen, Alfred Tarski, Per Lindström and others, Leon Henkin has made path-breaking contributions to the development of first- and higher-order logic. Although Prof. Dr. Leon

Henkin was unable to attend for reasons explained by himself immediately below he eloquently expresses this sentiment in his reply to the invitation:

> Saturday, June 8 / 2002
>
> Dear Prof. Dr. Wansing,
>
> Thank you very much for your invitation to give a plenary talk at the conference FOL75. Please extend my thanks also to the other distinguished members of the organizing committee. The concept of a conference on the evolution of predicate logic during the first 75 years of Hilbert and Ackermann's *Grundzüge* is a wonderful one, and of course that evolution is entwined with my own development as a logician. Imagining a visit to Humboldt University in 2003 and participating in the conference excites me, and I would like to be able to accept. Unfortunately, I am 7 years older than the book we are admiring, and my own evolution is subject to biological forces that do not exist for predicate logic. Had I received your invitation at a time when I was 75, I surely would have accepted – if sureness is meaningful in a counterfactual conditional! – but in the last 6 years I have suffered serious declines in my vision, my hearing, and my memory, so that I must unwillingly decline. Please accept my best wishes for the wonderful conference you are starting to put together. I hope I'll be able to read its Proceedings.
>
> Leon Henkin
>
> Professor Emeritus

We dedicate this volume to Prof. Leon Henkin.

Logical analysis of relativity theories[1]

Andréka, H. Madarász, J. X. and Németi, I.
andreka@renyi.hu, madarasz@renyi.hu, nemeti@renyi.hu
Alfréd Rényi Institute of Mathematics, Budapest, H-1364 Hungary, Pf. 127

1 Introduction

In this paper we try to give a small sample illustrating the approach of Andréka et al. [2],[5],[21]–[24] to a logical analysis of relativity theory conducted purely in first-order logic, FOL. We stick with FOL for methodological reasons. Here we first concentrate on special relativity, but in [22],[2],[5],[24] steps are made in the direction of generalizing this FOL-approach towards general relativity. We discuss that direction in the second half of this introduction. In [5] we build up variants of relativity theory as "competing" axiom systems formalized in FOL. The reason for having several versions for the theory, i.e. several axiom systems, is that this way we can study the *consequences* of the various axioms, enabling us to find out which axiom is responsible for what interesting or "exotic" prediction of relativity theory. Among others, this enables us to refine the *conceptual analysis* of relativity in Friedman [12] and Rindler [28], or compare the Reichenbach-Grünbaum-Salmon approach to relativity (cf. [31] or [12]) with the standard one. Later we will refer to the just indicated "several competing axiom systems" feature of our theory as flexibility feature or *modularity feature*.

One of our FOL axiom systems will be called *Specrel*. We will see that *Specrel* is a faithful, streamlined FOL axiomatization of (the kinematics of)

[1]Research supported by the Hungarian National Foundation for Scientific Research grants No's T30314, T43242, T35192, as well as by COST grant No. 274.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 7–36

special relativity theory. *Specrel* is intended to consist of simple, intuitively convincing, logically transparent, natural axioms. Besides *Specrel* we study its fragments, its generalizations towards general relativity and other versions of relativity.

As explained in [5, §1], the present approach is (in some sense) more ambitious (as a relativity theory) than e.g. a formalization of, say, Minkowskian geometry in first-order logic would be, in various respects: (i) One respect is that *if* we identified Minkowskian geometry with special relativity, then this would yield an *uninterpreted* (in the physical sense) version of relativity, while the first-order theory which we develop in [5] contains "its own interpretation", too. (ii) It is not clear to us how the conceptual analysis[2] suggested e.g. in [12] could be squeezed into Minkowskian geometry. (iii) Our formalized relativity theory is undecidable, while the FOL-theory of Minkowskian geometry in [13] is decidable, pointing in the direction that in our theory one can talk about things which do not appear in pure Minkowskian geometry. Someone may argue that Minkowskian geometry is the heart of special relativity theory; but it is *only* the heart, and we would like to formalize the full theory and not only its heart. (iv) The observational/theoretical duality outlined in [12] motivates us to keep our vocabulary and axioms on the modular, observational side (while Minkowskian geometry remains more on the "monolithic", "theoretical" side).[3] (v) Besides building up observational relativity in FOL, we also formalize the "monolithic", theoretically oriented geometric theory of space-time in FOL in e.g. [21],[23]. Then we prove that these two FOL theories are FOL-definitionally equivalent. So the user of our FOL theory can switch between the observational and theoretical versions whenever he so wishes. (vi) We also work on generalizing gradually our FOL theory of special relativity in the direction of general relativity in e.g. [2],[22],[24]. For this gradual generalization we will rely on the modularity feature of our observational theory mentioned way above. This feature is tied to the theory's having many small building blocks each of which carries some intuitive and natural meaning and which blocks can be removed or added one-by-one like in a lego toy world yielding new, meaningful and coherent versions of the theory. Besides generalization towards general relativity, this

---

[2] Which axiom is responsible for what prediction, which axiom is intuitively more natural than the other, etc.

[3] We use the observational/theoretical distinction in relativity in the sense of [12] going back to Reichenbach (1920). Sometimes it is useful to think of this as bottom-up/top-down distinction.

modularity feature is used in answering the so-called *why-type questions* and for conceptual analysis. This degree of modularity does not seem to be easily available if one starts out with an axiomatization of Minkowskian geometry or some other "top-down" approach.

After having formalized relativity theory in first-order logic, one can use the well developed machinery of FOL for studying properties of the theory, e.g. *Specrel* (e.g. the number of non-elementarily equivalent models, or its relationships with Gödel's incompleteness theorems, independence issues, etc). The reasons why we find it important to stick with FOL as a framework throughout the logical analysis of relativity can be found e.g. in van Benthem [8] when read together with Sain [29]. These reasons are further explained in [5, Appendix], Väänänen [32], Woleński [34]. It is explained e.g. in Feferman [10] and in Ferreirós [11] why and how we can stay in the framework of FOL throughout all our developments, if we want to.

As already indicated, the present work intends to give samples from a broader ongoing project represented by e.g. [5],[2],[21],[22],[24],[1]. A general plan for this broader project goes as follows: First we build up (the kinematics of) special relativity theory in FOL obtaining the finitely axiomatized FOL theory *Specrel*. *Specrel* was mentioned already at the beginning of this introduction. We put emphasis on making the axioms of *Specrel* streamlined, transparent, and intuitively convincing. First, as usual, we establish adequateness of *Specrel* for special relativity (completeness theorem, independence of the axioms, etc).[4] Then we elaborate a conceptual analysis of special relativity, its variants, and its generalizations. This analysis is based on the FOL axiom system *Specrel*, on variants and fragments of *Specrel* and their generalizations. Among others, we analyse *Specrel* both from the logical point of view (model theory, proof theory, "reverse mathematics" etc) and from the physico-philosophical relativity theoretic point of view. Much of this is done in [5],[21],[2],[24]. As a natural continuation of all this, we also experiment with generalizing *Specrel* in the direction of general relativity.

The first two steps in this generalization are (I) and (II) below. (I) We extend *Specrel* to accomodate accelerated observers which, via Einstein's equivalence principle, enables us to study some features of gravity. E.g. the Twin Paradox and the Tower Paradox (gravity slows time down) become provable in the accelerated observers version *Acc*(*Specrel*) of *Specrel*, cf.

[4] In some sense, we consider this as "Step Zero".

e.g. [2]. (II) As a second step in this direction, we make *Acc*(*Specrel*) *local* where "local" is understood in the sense of general relativity. We do this via the so-called method of localization which can be applied basically to any version of *Specrel* and *Acc*(*Specrel*). The localized versions of these theories are built up also in FOL (we make special efforts to ensure this) for methodological reasons mentioned earlier. Since localization turns out to be such a general procedure, we can denote the so obtained theories as *Loc*(*Specrel*), *Loc*(*Acc*(*Specrel*)) etc. So, *Loc*(−) can be regarded as some kind of a general "operator" applicable to theories (which are variants of *Specrel*).

It is explained in the classic textbook [25, pp.163-5] on general relativity that by suitably combining accelerated observers and localization one can safely move towards general relativity by starting out from special relativity, cf. e.g. Box 6.1 on p.164 therein. This motivates our study of the FOL theory *Loc*(*Acc*(*Specrel*)) and its variants. The investigation of *Loc*(*Acc*(*Specrel*)) is analogous with that of *Specrel*, i.e. after introducing the theory and proving theorems about its basic properties comes a fine-scale conceptual analysis both from the logic point of view and from the relativity theoretic point of view. The operator *Loc*(−) and *Loc*(*Specrel*) in particular are discussed in the present volume in [22]. The ideas in [22] are easily combined with those in [2] on *Acc*(*Specrel*) in order to obtain a comprehensive understanding of *Loc*(*Acc*(*Specrel*)). More on *Loc*(−) and *Loc*(*Specrel*) is in [22],[24], while more on *Acc*(*Specrel*) is found in [2], and the works quoted therein.

The research project reported herein is part of a much broader background literature of logic-based approaches to space-time. E.g., axiomatizations of special relativity are abundant in the literature. To mention some: axiomatizations of special relativity have been studied in works of Robb, Reichenbach, Carathéodory, Alexandrov and his school, Suppes and his school, Szekeres, Ax, Friedman, Mundy, Goldblatt, Schutz, Walker. This is only a small sample. There are more works listed in the bibliographies of [2],[5],[21]. Latzer [19], Buseman [9] represent moves towards general relativity in a spirit similar to that of our [22],[2],[24].

2 The frame language

We introduce the first-order logic language, which we will use for formalizing (first special) relativity, with an eye open for the subsequent generalization of the theory. We want to talk about *motion of bodies*.[5] What is motion? It is changing location in time. Therefore we will talk about *bodies*, *time*, *space*, and about a *location-function* which tells us which body is where at a given time. We want to talk about *relativity* theories; therefore these location functions will depend on *observers*; different observers may see the same motion differently. (The location function determined by an observer m will be called the world-view function w_m of observer m.) We will treat observers as special bodies whose motion will be represented exactly the same way as that of the rest of the bodies. These observers are often called, in the literature, *reference frames*.[6]

It will be convenient for us to be flexible about the dimension of space: we will not only treat 3-dimensional space, but 1 or 2, or higher-dimensional spaces as well. We will treat time as a special dimension of *space-time*. n will denote the dimension of our space-time.[7] Thus, usually $n = 4$ (3 space-dimensions and 1 time-dimension), but we will consider also $n = 2, 3$ or $n > 4$. Our bodies will be idealized, pointlike.

The vocabulary of our language is the following: unary relations

B (bodies)

Obs (observers)

Ph (photons)

Q (quantities used for giving location and "measuring time");

an $n + 2$-ary relation, the *location*- or *world-view* relation

W (world-view relation, $W(m, b, t, s_1, \ldots, s_{n-1})$ intends to mean that according to observer (or reference-frame) m, the body b is present at time t and location $\langle s_1, \ldots, s_{n-1}\rangle$);

for dealing with quantities, we will have two binary functions, and a binary relation:

---

[5]In this paper we concentrate only on kinematics; the same kind of investigations can be carried out concerning mass, forces, energy etc. However, if a theorem can be proved without referring to these extra notions, we consider that a virtue.

[6]This difference is only a matter of linguistic convention.

[7]Recent generalizations of general relativity in the literature (e.g. M-theory) indicate that it might be useful to leave n as a variable.

$+, \ \cdot, \ \leq.$

In our theories, we will always assume the following:

- observers and photons are bodies,
- $W(m, b, t, s_1, \ldots, s_{n-1})$ implies that m is an observer, b is a body, and $t, s_1, \ldots, s_{n-1}$ are quantities,
- $\langle Q, +, \cdot, \leq\rangle$ is a Euclidean[8] linearly ordered field.

We found that the simplest way of treating these assumptions is to use a 2-sorted first-order language, where

B, Q are sorts or universes,

Obs, Ph are unary relations of sort B,

W is an $n+2$-ary relation of sort $B \times B \times Q \times Q \times \cdots \times Q$,

$+, \cdot$ and $\leq$ are operations and relation of sort Q.

Let
$$\mathsf{M} = \langle B^{\mathsf{M}}, Obs^{\mathsf{M}}, Ph^{\mathsf{M}}; Q^{\mathsf{M}}, +^{\mathsf{M}}, \cdot^{\mathsf{M}}, \leq^{\mathsf{M}}; W^{\mathsf{M}}\rangle$$
be a model of our two-sorted language. This means that B^{M} and Q^{M} are sets, they are called the *universes of sort B and Q* respectively, $Obs^{\mathsf{M}}, Ph^{\mathsf{M}} \subseteq B^{\mathsf{M}}$ etc. We will omit the superscripts $^{\mathsf{M}}$. We call M a *frame-model* if $\langle Q, +, \cdot, \leq\rangle$ is a Euclidean linearly ordered field and $W \subseteq Obs \times B \times Q \times \cdots \times Q$. $\models$ denotes the usual semantical consequence relation *induced by frame-models*, i.e. $Th \models \varphi$ means that for every *frame-model* M, if $\mathsf{M} \models Th$, then $\mathsf{M} \models \varphi$.

Next we introduce some terminology in connection with arbitrary frame-models $\mathsf{M} = \langle B, Obs, Ph; Q, +, \cdot, \leq; W\rangle$.

The essence, the "heart" of a frame-model is the world-view relation W. Since $W \subseteq Obs \times B \times {}^nQ$, for every observer $m \in Obs$ it induces a function $w_m : {}^nQ \to \{X : X \subseteq B\}$ as follows: for every $p \in {}^nQ$
$$w_m(p) := \{b \in B \ : \ W(m, b, p)\}.$$
Thus $w_m(p)$ is the set of bodies present at space-time location p for m. We

[8] An ordered field is called *Euclidean* if every positive element has a square root in it, i.e. if $(\forall x > 0)(\exists y) x = y \cdot y$ is valid in it.

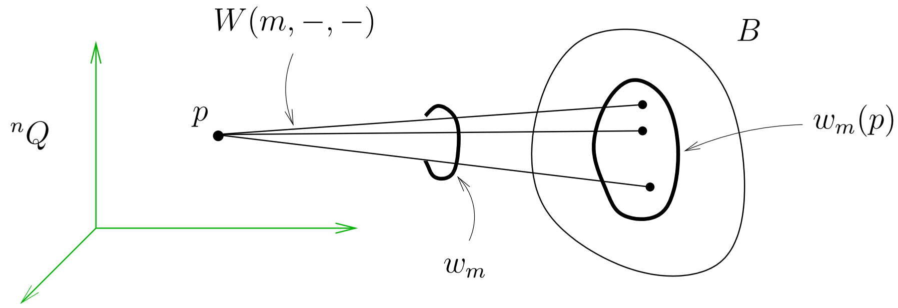


Figure 1: The world-view function w_m.

call a set of bodies an *event*, and w_m is called the *world-view function* of m, which to each space-time location p tells us what event observer m observes or "sees happening" at location p. "Seeing" has nothing to do with photons here, it really means "coordinatizing" only.

The *trace* or *life-line* of a body b according to an observer m is the set of space-time locations where m sees b, i.e.

$$tr_m(b) := \{p \in {}^nQ : W(m, b, p)\}.$$

The world-view function w_m can be recovered from the family of traces of all bodies (from $\langle tr_m(b) : b \in B\rangle$), and the world-view-relation W can be recovered from all the world-view functions (from $\langle w_m : m \in Obs\rangle$). Thus we can "represent" the function w_m by the *world-view* of m, which is just the indexed family $\langle tr_m(b) : b \in B\rangle$, and which, in turn, we represent by drawing the traces of bodies that we are interested in. See Figure 2.

Since $\mathsf{Q} = \langle Q, +, \cdot\rangle$ is a field, we can define n-dimensional straight lines as follows (these will be the life-lines of "inertial bodies"). We will use the vector-space structure of ${}^n\mathsf{Q}$, i.e. if $p, q \in {}^nQ$ and $\lambda \in Q$ then $p+q, p-q, \lambda p \in {}^nQ$ and $\overline{0}$ denotes the *origin*, i.e. $\overline{0} = \langle 0, \ldots, 0\rangle$, where 0 is the zero-element of the field. Let $\ell \subseteq {}^nQ$. We say that ℓ is a *straight line* iff there are $p, \alpha \in {}^nQ$ such that $\alpha \neq \overline{0}$ and

$$\ell = \{p + r \cdot \alpha \ : \ r \in Q\}.$$

Lines denotes the set of all straight lines (of dimension n). $\bar{t}$ denotes the *time axis*,

$$\bar{t} := \{\langle r, 0, \ldots, 0\rangle \ : \ r \in Q\}.$$

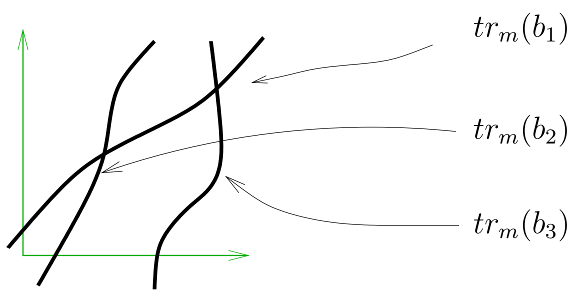


Figure 2: World-view of m.

$\bar{t}$ is a straight line. If $\ell \in Lines$, then $ang(\ell)$, defined below, represents the *angle*[9] between ℓ and $\bar{t}$ (where $\alpha = \langle \alpha_0, \ldots, \alpha_{n-1} \rangle$ is associated to ℓ as before):

$$ang(\ell) := \frac{\alpha_1^2 + \cdots + \alpha_{n-1}^2}{\alpha_0^2} \quad \text{if } \alpha_0 \neq 0, \text{ and}$$

$$ang(\ell) := \infty \quad \text{if } \alpha_0 = 0. \text{ Here } \infty \text{ is any element not in } Q.$$

$ang(\ell) = 1$ means intuitively that the angle between ℓ and $\bar{t}$ is 45^o. (See Figure 3.) Assume that $tr_m(k) = \ell$ is a straight line. Then $ang(\ell)$ represents the *velocity*[10] of k as seen by m:

$$v_m(k) := ang(tr_m(k)), \quad \text{if } tr_m(k) \in Lines.$$

E.g. $v_m(k) = 0$ means that $tr_m(k)$ is parallel with $\bar{t}$, i.e. k's location does not change with time, i.e. k is *at rest* w.r.t. m. The bigger $v_m(k)$ is, the bigger distance k travels in a unit time (as seen by m).

3 Basic axioms of special relativity

As already indicated, a *plurality* of "competing" axiom systems (or "relativity theories") is an essential feature of a logical analysis of relativity as developed in e.g. [5],[21],[2]. In this section we recall one of these axiom systems

[9] Actually, $ang(\ell)$ is the square of the tangent of the angle between ℓ and $\bar{t}$.

[10] Instead of "velocity", the precise expression would be "speed", since $v_m(k)$ is a scalar and not a vector.

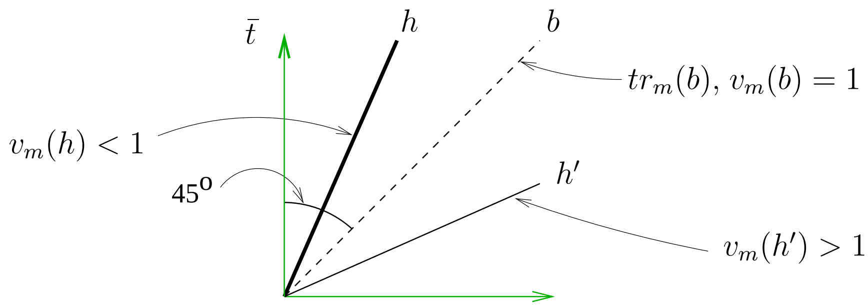


Figure 3: Velocities.

and will call it $Specrel_0$. It consists of five axioms. In the following axioms, m, k stand for arbitrary observers, h for an arbitrary body, ℓ for an arbitrary straight line (i.e. element of $Lines$), and ph for an arbitrary photon. We use the standard custom in logic that free variables should be understood as universally quantified, e.g. the axiom $tr_m(m) = \bar{t}$ means $(\forall m \in Obs) tr_m(m) = \bar{t}$.

Our first axiom says that the traces of observers and photons, as seen by any observer, are straight lines:

AxLine $tr_m(h) \in Lines$ for $h \in Obs \cup Ph$.

Since translating our intuitive statements to first-order formulas in the language of our frame-models (M's) will be straightforward, we will not give these translations,[11] we will only give the intuitive forms.

The second axiom says that any observer sees himself at rest in the origin:

AxSelf $tr_m(m) = \bar{t}$.

The third axiom says that we have the tools for thought-experiments: on any appropriate straight line we can assume there is a potential observer; and the same for photons:[12]

[11]They can be found in [2],[21].

[12]This axiom can be "tamed" by using modal logic, such that space-time does not get crowded with k's and ph's, cf. [5].

AxPot $ang(\ell) < 1 \quad \Rightarrow \quad (\exists k \in Obs)\ell = tr_m(k)$, and
$ang(\ell) = 1 \quad \Rightarrow \quad (\exists ph \in Ph)\ell = tr_m(ph)$.

The fourth axiom says that all observers "see" the same events (possibly at different space-time locations):[13]

AxEvents $Rng(w_m) = Rng(w_k)$.

The last axiom says that the velocity of a photon is 1, for each observer:

AxPh $v_m(ph) = 1 \quad$ (and $\quad tr_m(ph) \in Lines$).

Our choice for a "first possible" axiom system for special relativity is:

$Specrel_0 :=$ {**AxLine, AxSelf, AxPot, AxEvents, AxPh**}.

When we want to indicate explicitly the number of dimensions, we will write $Specrel_0(n)$ in place of $Specrel_0$. We note that **AxPh** together with the photon part of **AxPot** is the relativistic part of $Specrel_0$. (The rest are true in Newtonian Mechanics.)

Let $n > 2$. In this paper we show that $Specrel_0(n)$ is consistent, it is not independent, and it forbids faster than light observers but permits faster than light bodies.[14] We show that $Specrel_0$ generates an *undecidable* first-order theory but we can strengthen it so that it becomes decidable (moreover categorical); and also we can strengthen it so that it becomes hereditarily undecidable, further both of Gödel's incompleteness properties hold for this strengthened version. We will see that both kinds of extension of $Specrel_0$ are natural.

[13]This will have to be considerably weakened, when preparing for a generalization of our axiom systems like $Specrel_0$ towards general relativity, cf. [5],[2],[22]. For a function f, its *range* is $Rng(f) := \{y : \exists x(f(x) = y)\}$.

[14]The point in proving things like $Specrel_0 \models$ *no FTL observer* is in the small number of axioms and concepts needed. Actually in [22] we show that a much weaker version of $Specrel_0$ is enough for proving this conclusion. A more refined version of the theorem says that FTL observers "lose most of their meter rods", cf. [5].

4 Traveling with light, traveling faster than light

As a warm-up, we begin with a simple statement about our axiom system $Specrel_0$. When Einstein was a child, he once dreamed that he traveled together with a photon, and then he tried to imagine how the world could look like when one sees it while traveling with a photon. Our first proposition says that in models of $Specrel_0$, you can't see the world while traveling with a photon. (By "seeing" we mean "coordinatizing".)

Proposition 1. *$Specrel_0 \models tr_m(k) \neq tr_m(ph)$ for any $m, k \in Obs$ and $ph \in Ph$.*

Proof. Assume that $tr_m(k) = tr_m(ph)$ for some $m, k \in Obs, ph \in Ph$ in a model of $Specrel_0$. Then $tr_k(k) = \bar{t}$ and $v_k(ph) = 1$ by **AxSelf, AxPh**. Thus $tr_k(k) \neq tr_k(ph)$. Then k sees an event in which k is present but ph is not present (namely such is $w_k(p)$ for any $p \in tr_k(k) \setminus tr_k(ph)$). However, m does not see such an event by $tr_m(k) = tr_m(ph)$. This contradicts **AxEvents**, proving the proposition. See Figure 4. **QED**

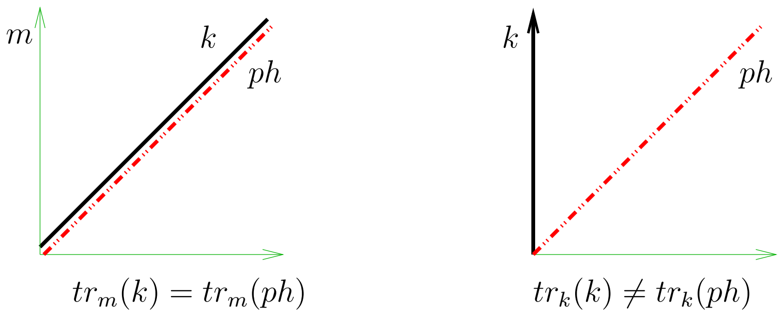


Figure 4: An observer cannot travel together with a photon.

Theorem 1. *Let $n > 2$.*

(i) $Specrel_0(n)$ is not independent, namely

$$\{\textbf{AxSelf, AxPot, AxEvents, AxPh}\} \models \textbf{AxLine}.$$

(ii) $Specrel_0(2)$ is independent, i.e. for any $\textbf{Ax} \in Specrel_0(2)$ we have

$$Specrel_0(2) \setminus \{\textbf{Ax}\} \not\models \textbf{Ax}.$$

Proof. For brevity, we will write $Specrel_0 - \mathbf{Ax}$ for $Specrel_0 \setminus \{\mathbf{Ax}\}$. It is not difficult to check that $Specrel_0 - \mathbf{Ax} \not\models \mathbf{Ax}$ for any $\mathbf{Ax} \in Specrel_0$, if $\mathbf{Ax} \neq \mathbf{AxLine}$. So we have to show that

$Specrel_0(n) - \mathbf{AxLine} \models \mathbf{AxLine}$ and

$Specrel_0(2) - \mathbf{AxLine} \not\models \mathbf{AxLine}$.

Assume that M is a model of $Specrel_0(n) - \mathbf{AxLine}$. Let $m, k \in Obs^{\mathsf{M}}$ and define

$$f_{mk} := \{\langle p, q\rangle \in {}^nQ \times {}^nQ \ : \ w_m(p) = w_k(q)\}.$$

Thus f_{mk} is a binary relation on space-time locations; two space-time locations are related when m and k see the same "events" at those points. We now show that

(*) f_{mk} is a bijective mapping of ${}^nQ$ onto ${}^nQ$, in any model of $\{$**AxPot, AxEvents**$\}$.

Let $p, q \in {}^nQ$ be distinct. Then there is a straight line ℓ with $ang(\ell) < 1$ separating them, i.e. $p \in \ell$ and $q \notin \ell$. By **AxPot**, ℓ is the trace of some observer h. Then $h \in w_m(p), h \notin w_m(q)$, showing that w_m is injective for any observer m. By **AxEvents** we have that both the domain and the range of f_{mk} is ${}^nQ$ (since $f_{mk}^{-1} = f_{km}$). These facts imply (*).

f_{mk} is called the *world-view transformation* between m and k: its intuitive meaning is that m thinks that k is "crazy" to the extent that his seeing is distorted by this function f_{mk} (whatever event m sees at space-time location p, k sees it at location $f_{mk}(p)$).

Now, **AxPh, AxPot** require that f_{mk} preserve *light-lines* (i.e. straight lines with angle 1). By a slight generalization of the celebrated Alexandrov-Zeeman theorem (that we will recall in a moment) then f_{mk} has to preserve all straight lines, in other words, it is a *collineation*. Then $tr_k(m) = f_{mk}(tr_m(m)) = f_{mk}(\bar{t})$ is a straight line by **AxSelf**. Thus **AxLine** holds.

To show $Specrel_0(2) - \mathbf{AxLine} \not\models \mathbf{AxLine}$ we construct a bijection $f : {}^2R \to {}^2R$, where R is the set of reals, which preserves light-lines, but which takes $\bar{t}$ onto a curve which is not a straight line. Here is the idea of the construction (see Figure 5):

Let t' be a "slightly bent" version of $\bar{t}$, and let f be any bijection between $\bar{t}$ and t'. We extend f to any point p not on $\bar{t}$ as follows: Let a and b be the two points where the two light-lines through p intersect $\bar{t}$, and let $f(p)$

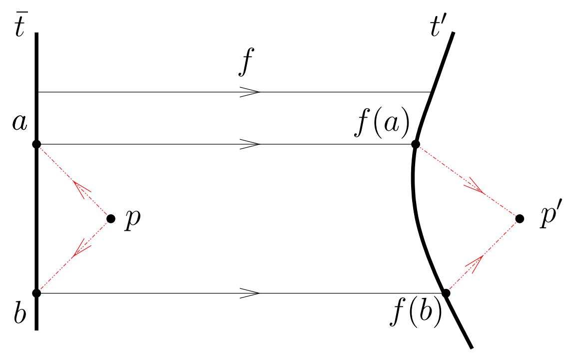


Figure 5: Illustration for the proof of Thm.1(ii). f preserves all light-lines but not all straight lines. Thus $\bar{t}$ cannot be defined from light-lines in ${}^2R$.

be the intersection point of the two corresponding light-lines going through $f(a)$ and $f(b)$. With some care this extension of f will be a bijection, and it preserves all light-lines by its construction. Now it is not difficult to construct a model of $Specrel_0(2)$ − **AxLine** where this f is one of the world-view transformations; and so in this model **AxLine** does not hold.

We now briefly recall the *Alexandrov-Zeeman theorem*. This theorem states that a permutation of ${}^4R$ which preserves light-lines is a collineation of a special form (namely a so-called Lorentz-transformation up to a dilation, a translation, and a field-automorphism-induced transformation[15]). An illuminating logical proof can be found in Appendix B of Goldblatt [13]. That proof can be generalized to any Euclidean field Q and $n > 2$ in place of R and 4. About the Alexandrov-Zeeman theorem see also [22] in this volume. We sketch the proof for $n = 3$. Let ℓ be any light-line. Let P be the set of those points through which no light-line intersecting ℓ goes through. Then it is not difficult to see that P is just the plane tangent to any light-cone[16] containing ℓ, see Figure 6. Now we can obtain all straight lines ℓ with $ang(\ell) > 1$ by intersecting such tangent planes; then we can define all planes using these newly obtained straight lines, and then we can obtain all the straight lines by intersecting again these new planes. Hence, any light-line preserving permutation is a collineation. We omit the proof of the rest, but for an idea of

[15] This latter will matter when R will be replaced with Q.

[16] A *light-cone* is the union of all light-lines going through a given point.

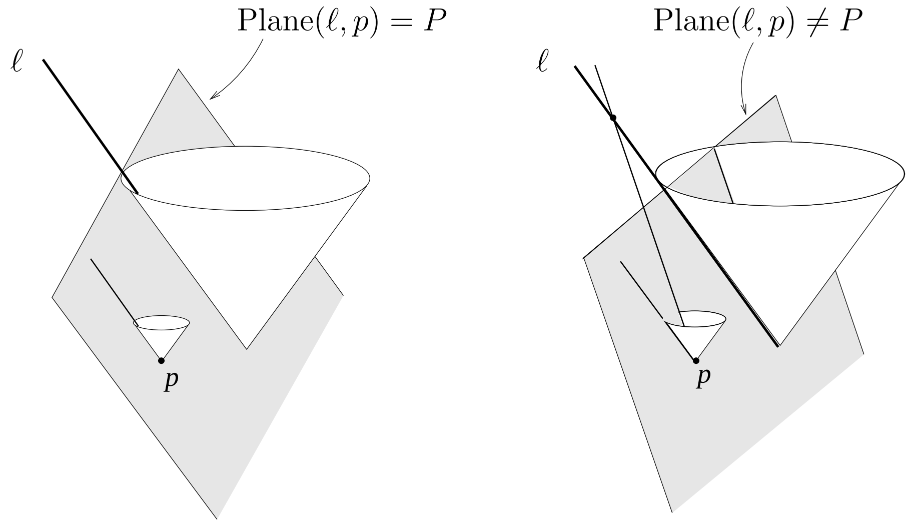


Figure 6: Illustration for the proof of the Alexandrov-Zeeman theorem. Definition of tangent planes: P is the set of points p through which no light-line intersecting ℓ goes. All straight lines can be defined from light-lines in ${}^3R$.

the proof see the proof of Thm.3 herein. **QED**

Let M be a frame-model, and k be an observer in it. We say that k is a *faster than light* (FTL) observer, if $v_m(k) > 1$ for some observer m. Below, *no FTL observer* abbreviates the sentence $(\forall m, k \in Obs) v_m(k) < 1$, i.e. that there is no FTL observer in the model.[17]

Theorem 2. *Let $n > 2$.*

(i) $Specrel_0(2) \not\models$ *no FTL observer.*

(ii) $Specrel_0(n) \models$ *no FTL observer.*

Proof. Since we want to stay visual, we give a proof for $n = 3$. We give a proof that is centered around the notion of Minkowski-orthogonality. Let

[17]There are well known common sense arguments, going back to Einstein, against FTL (cf. e.g. [27, p.11]). These involve "causality" among other undefined concepts. As e.g. Gödel pointed out, these arguments are *not* proofs in the logical sense. Our present Theorem 2 is of an essentially different character from this point of view (contrast e.g. (i) with (ii)).

ℓ, k be two straight lines. We say that ℓ is *Minkowski-orthogonal* (or shortly, M-orthogonal) to k if ℓ is orthogonal in the usual Euclidean sense to the reflection k' of k to the xy-plane. We say that ℓ is Minkowski-orthogonal to the plane P if it is Minkowski-orthogonal to at least two distinct straight lines lying in P, see Figure 7.

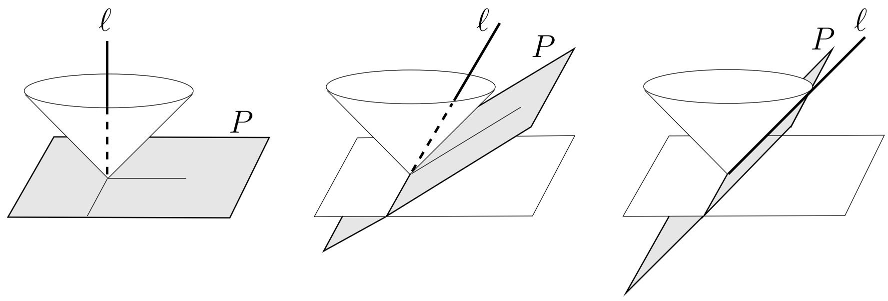


Figure 7: ℓ is Minkowski-orthogonal to P.

Minkowski-orthogonality is exhaustively investigated, e.g. fully axiomatized, in Goldblatt [13]. We will use here the following corollary of the generalized Alexandrov-Zeeman theorem:

(1) If a bijection of ${}^nQ$ preserves light-lines, then it preserves Minkowski-orthogonality.

We call a plane *space-like* if it contains no light-lines, and we call a straight line *time-like* if it is Minkowski-orthogonal to a space-like plane. It is not difficult to check (see Figure 8) that

(2) ℓ is time-like iff $ang(\ell) < 1$.

Clearly $\bar{t}$ is time-like, since it is M-orthogonal to the xy-plane which contains no light-line. Now we have seen in the proof of Theorem 1 that $f := f_{km}$ is a bijective collineation that preserves light-lines. Thus f takes the xy-plane to a space-like plane to which $f[\bar{t}]$ is M-orthogonal by (1), thus $f[\bar{t}]$ is time-like. By (2) then $ang(f[\bar{t}]) < 1$. But $f[\bar{t}] = f_{km}[tr_k(k)] = tr_m(k)$, thus $v_m(k) < 1$ in M.

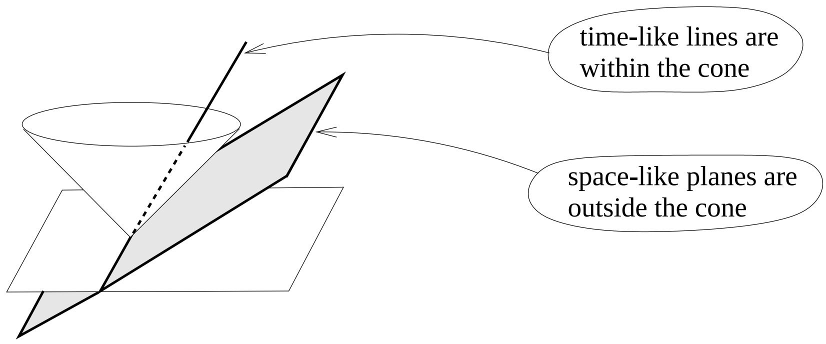


Figure 8: Time-like lines and space-like planes.

To show $Specrel_0(2) \not\models$ *no FTL observer*, we have to give a model of $Specrel_0(2)$ in which there are FTL-observers. Such models are given in [5], in section 2.4. **QED**

On pushing the limits of Theorem 2: The Alexandrov-Zeeman theorem is not true for functions f not defined everywhere in ${}^4R$. Therefore the above simple proof does not generalize to the local version $Loc(Specrel_0)$ of $Specrel_0$. In [22, Thm.3], "*no FTL observer*" is proved from a very weak axiom system, where the world-view transformations are only partial functions and where **AxPh** is substantially weakened. Theorems proving "*no FTL observer*" from weak axiom systems are also in [5] and in [21]. In the process of finding the "limits" of the "no FTL theorems", we also gave some intuitively appealing axiom systems (such is e.g. *Relphax* in [5, §3]) which do have models with faster than light observers. More about these FTL investigations is in e.g. [2]. We also note that in [22] a new Alexandrov-Zeeman style theorem is proved for the local version $Loc(Specrel_0)$ of $Specrel_0$.

Now we are going to introduce seven extra natural axioms that will make $Specrel_0$ categorical over any field. The theory $Specrel_0$ extended with these seven axioms (and with any decidable theory of fields) is decidable. We will see that if we leave out any one of six of these axioms, the theory will become undecidable, and such that it can be extended to a hereditarily undecidable theory where both Gödel's incompleteness theorems hold.

5 A principle of relativity

The *world-view transformation* f_{mk} between two observers m, k is defined as

$$f_{mk} := \{\langle p, q\rangle : w_m(p) = w_k(q) \text{ and } w_k(q) \neq \emptyset\} .$$

We already used f_{mk} in the proof of Theorem 1. From our previous axioms it follows that f_{mk} is a transformation of ${}^nQ$ (and not only an arbitrary binary relation) if m, k are observers.[18] Therefore we will use f_{mk} as a function. Then $f_{mk}(p)$ is the "place" where k sees the same event that m sees at p, i.e.

$$w_m(p) = w_k(f_{mk}(p)) .$$

When $p = \langle p_0, \ldots, p_{n-1}\rangle \in {}^nQ$, we will denote p_0 by p_t in order to emphasize that p_t is the "*time component*" of p. Let $p, q \in {}^nQ$. Then $p_t - q_t$ is the time passed between the events $w_m(p)$ and $w_m(q)$ as seen by m and $f_{mk}(p)_t - f_{mk}(q)_t$ is the time passed between the same two events as seen by k. Hence $\|(f_{mk}(p)_t - f_{mk}(q)_t)/(p_t - q_t)\|$ is the rate with which k's clock runs slow or fast as seen by m. Here, $\|a\|$ denotes the *absolute value* of a when $a \in Q$, i.e. $\|a\| \in \{a, -a\}$ and $\|a\| \geq 0$.

AxSym All observers see each other's clocks run slow to the same extent,

$$\|f_{mk}(p)_t - f_{mk}(q)_t\| = \|f_{km}(p)_t - f_{km}(q)_t\|, \text{ when } m, k \in Obs \text{ and } p, q \in \bar{t}.$$

AxSym states only that any two observers "see" each other's clocks "change" the same way. In principle, this allows the clocks run fast, be right, or run slow. In the Newtonian world **AxSym** is true because there each observer sees that the other's clocks are right. In models of $Specrel_0$, **AxSym** can be true only in the way that any observer sees that the clocks of any other observer not at rest wr.r.t. it *run slow.* Figure 12 in the proof of Thm.3 shows how it is possible in models of $Specrel_0$ that *both* observers "see" the clock of the other run slow.

On the choice of our symmetry axiom **AxSym**.: Under mild extra assumptions, $Specrel_0$ implies that **AxSym** is equivalent with an instance of

[18] This is a typical example of a property of special relativity which is relaxed in the process of localization (towards general relativity) in [22]. Namely, the axioms of the local theories in [22] will not imply that the function f_{mk} is everywhere defined in ${}^nQ$. This is an essential generalization towards general relativity.

Einstein's *special principle of relativity* SPR as it was formalized in [21, pp.87-89]. The principle SPR goes back to Galileo, intuitively it says that the "laws of nature" are the same for all inertial observers. A careful logic based analysis of SPR and its role in relativity is in [21, pp.84-91]. See also Friedman [12, p.153]. We note that, for $n > 2$, in models of $Specrel_0$, **AxSym** is equivalent to the potential axiom requiring that, in space, in the direction orthogonal (in the Euclidean sense) to the direction of the movement there is no relativistic distortion, i.e. there is no length-contraction. Other equivalent formalizations of **AxSym** can be found in [5, §3.7].

6 Axioms making $Specrel_0$ categorical

Here we introduce six more axioms that will make $Specrel_0$ categorical (over any given field). As in section 3, in the following m, k stand for observers, ℓ for a straight line, ph_i for photons; and free variables in the axioms should be understood as universally quantified.

The first two axioms deal with the direction of flow of time. We define for any two observers m, k

$$m \uparrow k \quad \text{iff} \quad (f_{km}(1_t) - f_{km}(\bar{0}))_t > 0.$$

Intuitively this means that time flows in the same direction for m and k, see Figure 9.

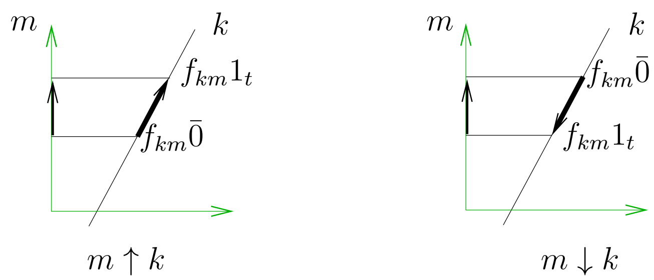


Figure 9: $m \uparrow k$ means that time flows in the same direction for m and k.

Our first axiom is a stronger version of part of **AxPot**, it says that every appropriate straight line is the life-line of an observer whose time flows "forwards".

AxPot$^+$ $ang(\ell) < 1 \quad \Rightarrow \quad (\exists k \in Obs)[\ell = tr_m(k)$ and $m \uparrow k]$.

The next axiom says that time flows in the same direction for any observers at rest in the origin.

Ax$\uparrow$ $tr_m(k) = \bar{t} \quad \Rightarrow \quad m \uparrow k$.

The next axiom says that every observer can "re-coordinatize" his world-view with a so-called Galilean transformation. To formalize the next axiom, first we single out special transformations, that we will call Galilean transformations. A mapping $f : {}^nQ \to {}^nQ$ is called a *Galilean transformation* if it preserves Euclidean distance and $f(1_t) - f(\bar{0}) = 1_t$ where $1_t = \langle 1, 0, 0, \dots \rangle$ and 1 denotes the unit element of the field Q. In other words, a Galilean transformation is a congruence transformation which is the identity map on $\bar{t}$, composed with a translation. See Figure 10. It is known that a Galilean transformation is a linear transformation composed with a translation, hence the next axiom is a first-order logic one.

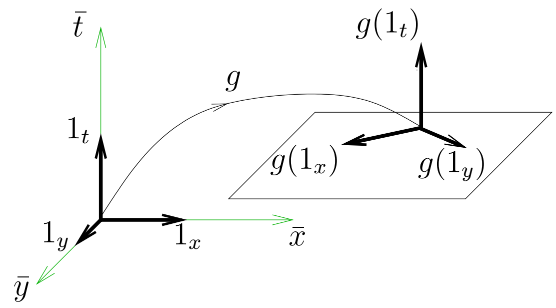


Figure 10: A Galilean transformation takes the unit vectors into pairwise orthogonal vectors of length 1, and does not change the direction of the time-unit vector.

AxGal $G(\bar{0}) \in \bar{t} \; \Rightarrow \; (\exists k \in Obs) f_{mk} = G,$ for every Galilean transformation G.

The next two axioms say, intuitively, that of each kind of observers and photons we have only one copy (or in other words, according to Leibniz's principle, if we cannot distinguish two observers or photons via some observable properties, then we treat them as equal).[19] In other words, these are so-called *extensionality axioms*. Id denotes the identity mapping.

AxExt$_1$ $f_{mk} = Id \quad \Rightarrow \quad m = k.$

AxExt$_2$ $tr_m(ph_1) = tr_m(ph_2) \quad \Rightarrow \quad ph_1 = ph_2.$

The last axiom says that every body is an observer or photon.

AxNobody $B = Obs \cup Ph.$

$Compl :=$ {**AxSym, AxPot**$^+$**, AxGal, AxExt**$_1$**, AxExt**$_2$**, AxNobody**}

$Specrel := Specrel_0 \cup$ {**AxSym**}

$Specrel^+ := Specrel \cup Compl \cup \{\mathrm{Ax}\uparrow\}$

In the terminology of e.g. Malament and Hogarth, $Specrel_0, Specrel$ and $Specrel^+$ correspond to *causal space-time* (or metric-free space-time), *space-time*, and *time-oriented space-time* respectively, cf. Hogarth [18]. $Specrel_0$ is also strongly connected to the "conformal structure of space-time". When we write "causal space-time", we have in mind the *symmetrized* version of the strict "causality relation" $\ll$. (Sometimes "metric-free space-time", "space-time", "time-oriented space-time" are used.)[20]

We did not include **Ax**$\uparrow$ into $Compl$ because, as we will see, its effects are different from those of the the elements of $Compl$.[21]

Theorem 3. *Let*[22] $n > 2$ *and let* $\mathsf{Q} = \langle Q, +, \cdot, \leq \rangle$ *be any Euclidean field.*

(i) There are exactly two models of $Specrel \cup Compl$ *with field-reduct* Q*, up to isomorphism.*

[19] We could have named these axioms after Occam, too.

[20] The terminology varies with different authors, but what we wanted to point out is that the levels of abstraction corresponding to $Specrel_0, Specrel$ and $Specrel^+$ seem to be generally distinguished levels of abstraction in the literature of relativity.

[21] Intuitively, **Ax**$\uparrow$ excludes only one model of two choices, while the rest exclude an infinite number of possibilities, cf. Thm.s 3-5. Sci.Am.(Sept. 2002, special issue, pp.30-31) discusses the justification of assuming **Ax**$\uparrow$ which turns out to be not as straightforward as one might think at first sight.

[22] We exclude the case $n = 2$ for simplicity only.

(ii) *There is a unique model of Specrel*$^+$ *with field-reduct* Q, *up to isomorphism.*

On the proof. We illustrate that in any model of *Specrel*, all the world-view transformations are so-called Poincaré-transformations (i.e. Lorentz-transformations composed with translations), and this is the most important part of the proof of Theorem 3.

Let m, k be observers in a model of *Specrel*, we will investigate the world-view transformation $f := f_{mk}$. We have already seen that $f : {}^nQ \to {}^nQ$ is a *bijection*. It is a *collineation* by the Alexandrov-Zeeman theorem in case $n > 2$, and by [22, Thm.2] in case $n = 2$. By **AxPh**, f takes light-lines onto light-lines, and this implies that f takes the unit vectors into vectors of the *same length* and *Minkowski-orthogonal* to each other. Figure 11 illustrates the idea of the proof of this part.

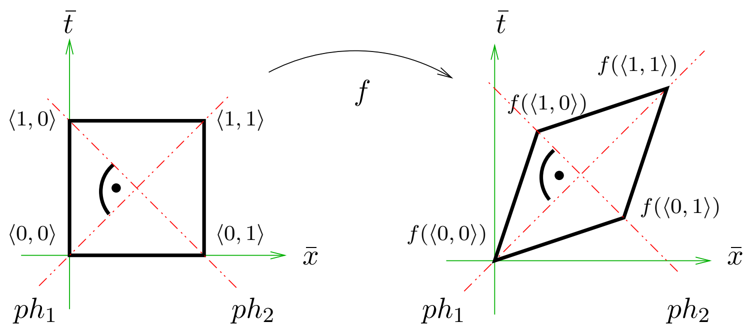


Figure 11: World-view transformations in models of $Specrel_0$ take the unit vectors to vectors Minkowski-orthogonal to each other and of the same length.

Finally, **AxSym** implies that the length of the unit vectors is fixed, as follows. We write out this part of the proof in more detail, because e.g. it shows how it is possible that both observers see each other's clocks run slow.

Let $1_t = \langle 1, 0, 0, \dots \rangle$, and let us see where $e := f_{km}(1_t)$ is on $tr_m(k)$. Let a, b and a' be as in Figure 12; i.e. they are the points on $tr_m(k)$ and on $\bar{t}$ such that the straight line connecting 1_t and a is parallel with $\bar{x}$, and the straight lines connecting 1_t and b and connecting a and a' are parallel with $f_{km}[\bar{x}]$. See Figure 12. If $e = a$, then m sees that k's clock shows 1 just when his clock shows 1, because 1_t and a are simultaneous for m. But k will see that m's clock shows $a' < 1$ when his clock shows 1, because for k, $e = a$ and a'

are simultaneous. So k will think that m's clocks run slow, but m will think that k's clocks are right. Analogously, m thinks that k's clocks are right (run slow or fast, respectively) iff $e = b$ ($> b$ or $< b$ respectively). And, k thinks that m's clocks are right (run slow or fast, respectively) iff $e = a$ ($< a$ or $> a$ respectively). Thus both think that the other's clocks run slow iff $b < e < a$. The rate of "slowness" is the same for them at a unique point in between a and b, because the change of rate is a continuous and strictly monotonic function (of the "number" $\|e\|$). Now, *Minkowski-distance* is defined so that the Minkowski-distance is 1 between $\overline{0}$ and this unique point (where the rates of "running slow" are the same for m and k). Figure 13 shows the points whose Minkowski-distance from $\overline{0}$ is 1, i.e. it shows Minkowski-circle with radius 1 and center $\overline{0}$.

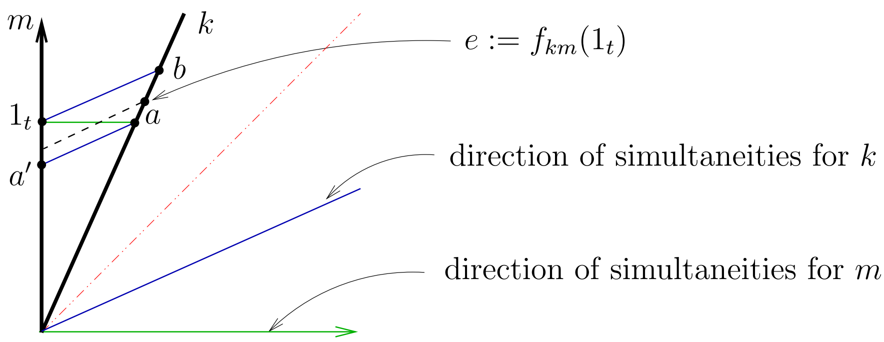


Figure 12: Both m and k think that the other's clocks run slow iff $f_{mk}(1_t)$ is in between a and b. The rates of "running slow" will be equal at a unique point.

It is known that any collineation is an affine transformation composed with a field-automorphism-induced transformation. Using that the above line of thought is valid for any $p \in \overline{t}$ in place of 1_t, one can show that the world-view transformations are actually *affine* transformations. Summing up: in models of *Specrel*, the world-view transformations take the unit vectors into pairwise Minkowski-orthogonal vectors of Minkowski-length 1. These kinds of affine transformations are called in the literature *Poincaré-transformations*. **QED**

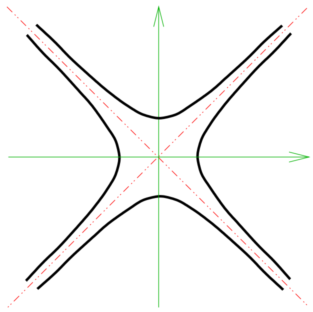


Figure 13: Minkowski-distance 1.

7 Decidability and Gödel incompleteness

We now turn to decidability questions. We start this by recalling the definition of real-closed fields and by recalling some facts from the literature.

An ordered field F is real-closed if it is Euclidean (i.e. every positive element has a square root), and if every polynomial of odd degree has zero as a value. This last requirement can be expressed with the infinite set $\{\phi_{2n+1} : n \in \omega\}$ of first-order formulas, where for every $n \in \omega$, ϕ_n denotes the following sentence

$$\forall x_0 \dots \forall x_n \exists y [x_n \neq 0 \rightarrow (x_0 + x_1 \cdot y + \cdots + x_n \cdot y^n = 0)].$$

By a *theory* we will understand an arbitrary set of first-order formulas (i.e. we will not assume that it is closed under semantical consequence). We call a theory *Th decidable* (or *undecidable* respectively) if the set of all first-order semantical consequences of *Th* is decidable (or undecidable respectively). We call *Th complete* if it implies either ϕ or $\neg\phi$ for each first-order formula ϕ (of its language). Propositions 2,3 below are known in the literature. Prop.2 is a corollary of Tarski's famous elimination of quantifiers for real-closed fields.

Proposition 2. *The theory of real-closed fields is decidable and complete.*

Proposition 3. *The theories of ordered fields and Euclidean fields are undecidable.* [23]

[23]Note that if a finitely (or more generally, recursively) axiomatizable theory is undecidable, then it is not complete.

Conjecture 1. *Any finitely axiomatizable consistent theory of ordered fields is undecidable.*

Corollary 1. *$Specrel_0$, Specrel and $Specrel^+$ are undecidable.*

Proof. This is a corollary of Prop.3, and the theorem that for any Euclidean field F there is a model of $Specrel^+$ with F as the field reduct (Theorem 3).: Let ϕ be any field-theoretic first-order formula written by using variables of our quantity sort. Then ϕ is valid in a frame-model M with field reduct F iff ϕ is valid in F. Thus ϕ is valid in the class of Euclidean fields iff ϕ is true in all models of $Specrel^+$. Since the first-order theory of the Euclidean fields is undecidable by Prop.3, the first-order consequences of $Specrel^+$ is undecidable, too. Since this is a finite theory, then any subset of it is undecidable, too. **QED**

The above suggests that if we want to obtain interesting decision-theoretic results, we have to concentrate on real-closed fields; or at least include a decidable theory of field-axioms into our theories. Let Φ denote the theory of real-closed fields.

Theorem 4. *Let $n > 2$.*

(i) $Specrel_0 \cup Compl \cup \Phi$ *is decidable.*

(ii) $Specrel_0 \cup Compl \cup \{\mathbf{Ax}\uparrow\} \cup \Phi$ *is decidable and complete.*

(iii) $Specrel_0 \cup (Compl \setminus \{\mathbf{Ax}\}) \cup \{\mathbf{Ax}\uparrow\} \cup \Phi$ *is undecidable, for any axiom* $\mathbf{Ax} \in Compl$.

Proof. We show that (i) and (ii) are corollaries of Theorem 3, we sketch the proof of (ii). Let M and M' be models of $Specrel_0 \cup Compl \cup \{\mathbf{Ax}\uparrow\} \cup \Phi$. We cannot apply Theorem 3 yet, because the field-reducts F and F' of M and M' respectively may not be the same. But they are elementarily equivalent, because Φ is complete, so by the Keisler-Shelah isomorphic ultrapowers theorem they have isomorphic ultrapowers, say F_1 and F'_1. Let M_1 and M'_1 be the ultrapowers of M and M' respectively, taken by the same ultrafilter. Then the field-reducts of these are F_1 and F'_1 respectively. Now we can apply Theorem 3 to M_1 and M'_1 because F_1 and F'_1 are isomorphic, getting that M_1 and M'_1 are isomorphic, so elementarily equivalent. But then M and M' are elementarily equivalent, too, since the former two models are ultrapowers of these. This finishes the proof of (ii). (iii) is a corollary of the next theorem; we included it here because it nicely contrasts (i) and (ii). **QED**

We now turn to the analog of Gödel's first incompleteness theorem.

Theorem 5. *Let $n > 0$ and let* **Ax** *be any member of Compl. There is a formula ν (in our frame-language) such that*

(i) ν is consistent with $Specrel_0 \cup (Compl \setminus \{\mathbf{Ax}\}) \cup \{\mathbf{Ax}\!\uparrow\} \cup \Phi$

and for any theory Th consistent with ν

(ii) Th is hereditarily undecidable in the sense that no consistent extension of Th is decidable.

(iii) The conclusion of Gödel's first incompleteness theorem applies to the theory Th, i.e. no consistent recursively enumerable extension of Th is complete; moreover there is an algorithm that to each consistent, recursively enumerable extension Th' of Th gives us a formula ϕ such that $Th' \not\models \phi$ and $Th' \not\models \neg\phi$.

Proof. The idea of the proof is to show that absence of any member of *Compl* allows us to interpret Robinson's Arithmetic into our theory. We sketch this for the case $\mathbf{Ax} = \mathbf{AxNobody}$. We will see that in this case ν will be quite natural: it will state the existence of a periodically moving body. Consider the following formulas (with free variables m, b and t):

$$
\begin{aligned}
I(t) \ &:= \ I(m,b,t) \ := \ W(m,b,t,\overline{0}), \quad \text{and} \\
\nu \ &:= \ I(0) \ \wedge \ (\forall t, s) \\
&\quad ([t < 1 \ \wedge \ t \neq 0] \to \neg I(t) \quad \wedge \\
&\quad t \geq 0 \to [I(t) \leftrightarrow I(t+1)] \quad \wedge \\
&\quad [I(t) \wedge I(s)] \to [I(t+s) \wedge I(t \cdot s)]).
\end{aligned}
$$

Add, for a moment, m and b as constants to our language. Then t remains the only free variable of I which then specifies a subset of the field-reduct in any frame-model: the set of time-points where the observer m sees the body b at the origin. Now the formula ν requires that this subset behaves like the set of integers: it is a discrete periodic subset containing $0, 1$ and closed under $+, \cdot$. Since the field-reduct of a frame-model is a field, then Robinson's arithmetic will be true in the field-reduct restricted to the subset defined by I. In other words, I is an *interpretation of Robinson's Arithmetic in* $Th \cup \{\nu\}$, whenever ν is consistent with Th. For definition of Robinson's Arithmetic and (semantical) interpretation see e.g. Monk [26, Def.14.17, Def.11.43]. Thus, Robinson's Arithmetic can be interpreted in $Th \cup \{\nu\}$. Then $Th \cup \{\nu\}$ is inseparable (which is a strong version of undecidability) by Thm.16.1 and Prop.15.6 in [26]; and thus (ii) and (iii) of our Theorem hold by Monk [26, Thm.s 15.9 and 15.8]. Finally, if we omit the constants m, b, then semantical consequence does not change, so (ii) and (iii) will hold for the original language (set of formulas not containing the constants m or b), too (in (iii) a further little argument is needed).

To show (i), we have to construct a model of $Specrel_0 \cup \{\mathbf{Ax}\uparrow\} \cup \Phi \cup \{\nu\} \cup (Compl \setminus \{\mathbf{AxNobody}\})$. This is not difficult as ν basically states the existence of a periodically moving body; see Figure 14.

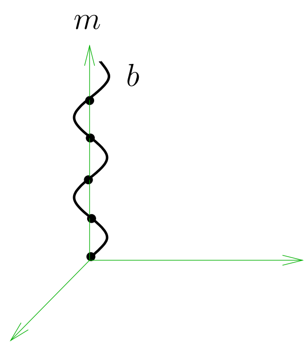


Figure 14: b is a periodically moving body in m's world-view.

Take a "standard" model with minimum set of observers and photons; and add one periodically moving body. We omit the details of the definition of this model.

The proofs for the other cases are analogous; we only give different interpretations of Robinson's arithmetic. This means that we give a different formula I, but ν will be the same (speaking about I), and then we only have to show that $Th \cup \{\nu\}$ is consistent, where Th is the theory in (i). To give a flavor, we give this new interpretation I for the case when $\mathbf{Ax} = \mathbf{AxPot}^+$.

$$I(m,t) := (\forall \ell)[ang(\ell) = \tfrac{1}{t} \quad \Rightarrow \quad (\exists k)(tr_m(k) = \ell \wedge m \uparrow k)] \text{ or } t = 0, 1.$$

This finishes the proofidea of Theorem 5. **QED**

A theorem analogous to Theorem 5 but concerning *Gödel's second incompleteness theorem* can also be stated and proved with analogous methods. For details see [4].

For current research directions in logic started by Gödel's incompleteness theorems we refer to Hájek and Pudlák [14], Willard [33], the latter in the present volume. The connections between the "observations oriented" and the "theoretically oriented" approaches to relativity were studied in [21] where the logical theories of definability and identifiability are used and further elaborated in the spirit of works of Hodges (cf. [17]) and Hintikka [15].[24] Actually, these logic based relativistic investigations induced new research in definability and identifiability theory. In later work continuing [2],[22] we plan to look into the logical structure of general relativistic space-times permitting *closed time-like loops* (which can be regarded as causing a kind of self-reference[25]). In Scheffler [30, p.179], and in Lewis [20, pp.67-80, pp.212-3] it is pointed out that these causal loops do not imply logical contradictions or even logical paradoxes. They simply have more complex logical structures than "linear causation". We plan to extend the mathematical logic based approach to further analyzing these and related possibilities thoroughly and carefully.

Acknowledgements. We would like to express our deep gratitude to the organizers and to the participants of the FOL75 conference for useful discussions, suggestions, encouragement and for a very fruitful, creative, and supportive atmosphere.

8 Bibliography

[1] Andai, A. Andréka, H. Madarász, J. X. and Németi, I. Logical approach to relativity theory. Mathematical Physics Meeting, Sept. 2001. (In Hungarian.) www.kfki.hu/~elftterm/matfiz01.html

[24] In [21, Thm.2.2.23, p.168] we use and extend Tarski's elimination of quantifiers for real-closed fields [16], for analyzing our theories of relativity.

[25] like the one in Gödel's incompleteness proof, Tarski's proof of undefinability of truth, or Barwise and Etchemendy's book on the "Liar"

[2] Andréka, H. Madarász, J. X. and Németi, I. Logical axiomatizations of space-time. Samples from the literature. In: Non-Euclidean Geometries (János Bolyai Memorial Volume). Kluwer Academic Publishers, to appear.

[3] Andréka, H. Madarász, J. X. and Németi, I. Defining new universes in many-sorted logic. Preprint, Rényi Institute of Mathematics, Budapest, 2001.

[4] Andréka, H. Madarász, J. X. and Németi, I. Decidability, undecidability, and Gödel incompleteness in relativity theories. Manuscript, Budapest, 1999.

[5] Andréka, H. Madarász, J. X. Németi, I. Sain, I. and Sági, G. Analyzing the logical structure of relativity theory via model theoretic logic. Lecture Notes for A'dam course (spring 1998). Updated version July 5, 2002, 1312pp. http://www.math-inst.hu/pub/algebraic-logic/Contents.html.

[6] van Benthem, J. The logic of time. Reidel, Dordrecht, 1983.

[7] van Benthem, J. Semantics of time. In: Time, Mind, and Behaviour (eds: Michon, J. A. and Jackson, J. L.), Springer-Verlag Berlin 1985. pp.266-278.

[8] van Benthem, J. Content versus wrapping: an essay in semantic complexity. In: Arrow Logic and Multimodal Logic (eds: Marx, M. Pólos, L. and Masuch, M.) CSLI Publications, Stanford, 1996. pp.203-219.

[9] Buseman, H. Time-like spaces. Dissertationes Math. (Rozprawy Math.), 53, 1967.

[10] Feferman, S. Theories of finite type related to mathematical practice. In J. Barwise, editor, Handbook of Mathematical Logic, North-Holland, 1977. pp.913-971.

[11] Ferreirós, J. The road to modern logic - an interpretation. The Bulletin of Symbolic Logic 7,4 (2001), 441-484.

[12] Friedman, M. Foundations of space-time theories. Relativistic physics and philosophy of science. Princeton University Press, 1983.

[13] Goldblatt, R. Orthogonality and space-time geometry. (With 126 illustrations.) Springer-Verlag, 1987.

[14] Hájek, P. and Pudlák, P. Mathematics of first order arithmetic. Springer-Verlag, Berlin, 1993.

[15] Hintikka, J. Towards a general theory of identifiability. In: Definitions and definability: philosophical perspectives. Kluwer Academic Publishers, 1991. pp.161-183.

[16] Hodges, W. Model theory. Cambridge University Press, 1993.

[17] Hodges, W. Hodkinson, I. and Macpherson, H. Omega-categoricity, relative categoricity and coordinatisation. Annals of Pure and Applied Logic 46,4 (1990), 169-199.

[18] Hogarth, M. Conventionality of simultaneity: Malament's result revisited. Preprint, Cambridge University, 2003.

[19] Latzer, R. W. Nondirected light signals and the structure of time. Synthese 24 (1972), 236-280.

[20] Lewis, D. The paradoxes of time travel. In: David Lewis Philosophical Papers Vol. II, Oxford University Press, 1986. pp.67-80.

[21] Madarász, J. X. Logic and Relativity (in the light of definability theory). PhD Thesis, Eötvös Loránd University, Budapest, 2002. http://www.math-inst.hu/pub/algebraic-logic/Contents.html.

[22] Madarász, J. X. Németi, I. and Tőke, Cs. On generalizing the logic-approach to space-time towards general relativity: first steps. This volume.

[23] Madarász, J. X. and Tőke, Cs. Duality between observational and theoretical concepts in relativity theory analyzed by FOL and definability theory. In: Proceedings of the Kalmár Workshop on Logic and Computer Science, July 2003. pp.133-142.

[24] Madarász, J. X. and Tőke, Cs. Partial relativity theories. Research report, Alfréd Rényi Institute of Mathematics, Budapest, July 2003.

[25] Misner, C. W. Thorne, K. S. and Wheeler, J. A. Gravitation. W. H. Freeman, San Francisco, 1973.

[26] Monk, J. D. Mathematical Logic. Springer-Verlag, Berlin, 1976.

[27] Rakić, N. Common sense time and special relativity. Dissertation, University of Amsterdam, 1997. ILLC Dissertation series 97-02.

[28] Rindler, W. Essential relativity. Special, general and cosmological. Springer-Verlag, Berlin, 1969, 1977.

[29] Sain, I. There are general rules for specifying semantics: observations on abstract model theory. CL&CL 13 (1990), 195-250.

[30] Scheffler, U. Ereignis und Zeit. Ontologische Grundlagen der Kausalrelationen. Logos Verlag Berlin, 2001.

[31] Szabó, L. E. The problem of open future, determinism in the light of relativity and quantum theory. Manuscript, Dept. of Theoretical Physics, Eötvös Loránd University, Budapest 1999. (Also: Typotex Kiadó, Budapest, 2002. 256pp. (In Hungarian.))

[32] Väänänen, J. Second-order logic and foundations of mathematics. The Bulletin of Symbolic Logic 7,4 (2001), 504-520.

[33] Willard, D. A version of the second incompleteness theorem for axiom systems that recognize addition but not multiplication as total functions. This volume.

[34] Woleński, J. First-order logic: (philosophical) pro and contra. This volume.

Safety Signatures for First-order Languages and Their Applications

Arnon Avron
aa@math.tau.ac.il
School of Computer Science
Tel Aviv University
Tel Aviv 69978, Israel

1 Introduction

In several areas of Mathematical Logic and Computer Science one would ideally like to use the set $Form(L)$ of all formulas of some first-order language L for some goal, but this cannot be done safely. In such a case it is necessary to select a subset of $Form(L)$ that can safely be used. Three main examples of this phenomenon are:

- The main principle of naive set theory is the *comprehension* schema:

$$\exists Z(\forall x.x \in Z \Leftrightarrow A)$$

  where A is a formula in which Z is not free (but may contain other parameters). Ideally, every formula A should be used in this schema. Unfortunately, it is well known that this would lead to paradoxes. What the various axiomatic set theories do is to replace the general comprehension schema by "safer" versions. Thus most of the axioms of ZF, the most famous axiomatic set theory, are just particular instances of the comprehension schema. Historically, the guiding line behind the choice of these instances has been the "limitation of size doctrine" ([8, 10]).

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 37–58

However, the criterion provided by this doctrine is not constructive, so ZF uses some constructive substitutes to select formulas which seem to meet it. These principles are usually explained and justified ([18]) on *semantic* ground, using certain general ontological assumptions. (some of which, like the "cofinality principle", may be debatable).

- A main goal of computability theory is to characterize the *decidable* relations. Now the most straightforward method of defining relations is by using formulas of an appropriate formal language L (like the language of Peano Arithmetics PA in the case of arithmetical relations). However, usually not every formula of L defines a decidable relation. Hence a major problem here is: what are the "safe" formulas which do? A strongly related problem of crucial importance for proof theory and the foundations of Mathematics (especially Gödel theorems) is: what formulas of L binumerate relations within a given theory T? [1] Again it is well known that in the case of PA no constructive general solution can be given for either problem. Therefore some constructively defined classes of "safe" formulas, broad enough for the various applications, have been selected in its language. Two major examples are the class of primitive recursive (p.r.) formulas ([9, 14]) and the class of bounded formulas ([21]).

- A query language for a database ([24, 2]) is an ordinary first-order language with equality, the signature of which includes predicate symbols for the database relations. Ideally, every formula ψ of a query language can serve as a query. If ψ is closed then the answer to the query is either "yes" or "no". If ψ has free variables then the answer to ψ is the set of tuples which satisfy it in the intended structure. However, an answer to a query should be finite and computable, even if the intended domain is infinite. Hence only "safe" formulas, the answers to which always have these properties, should be used as queries. Unfortunately, it is again undecidable which formulas are "safe". Therefore all commercial query languages (like SQL) allow to use as queries only formulas from some syntactically defined class of safe formulas.

In all these examples the same pattern repeats: a certain undecidable class of f.o. formulas, originally characterized by some semantic criterion, is

[1] If T is r.e. then such a relation is necessarily decidable.

singled out for some fundamental application. Then an effective, syntactically defined subclass that can serve as a sufficient substitute is found. In what follows we show that despite the different purposes and intuitions, the principles which have been used in all these areas in order to secure safety are similar (although they have independently been developed), and are directly based on the role of the first-order logical constants. By merging them we will be able to develop a unified, purely logical framework for dealing with "safety". The key feature of this framework is the use of a generalized concept of a f.o. signature. The idea is that a generalized signature for a language can contain more than just the arity of the possible interpretations of the primitive symbols of the language. It can contain e.g. also information about the size and/or the computability of their *intended* interpretations (reducing by this the class of allowed models).

Three concrete applications of our framework are:

- In set theory it provides a new, concise presentation (and in our opinion, a new understanding) of ZF. This presentation is based on purely *syntactic* criteria concerning the role the f.o. connectives and quantifiers have in defining legitimate new sets.
- In Computability Theory it provides a general framework for analyzing relative computability of both extensional and intensional relations and functions, on arbitrary (or at least countable) f.o. structures.
- In database theory it provides a simple syntactical notion of safety, which allows to use properties of relations and functions which do not belong to the database scheme. This notion is adequate not only for conventional databases, but also for databases in which there is only a partial access to some of the relations (like in the world wide web).

2 The General Framework

In the examples above *two* different factors were involved in questions of "safety": *size* (of the class of tuples which satisfy a given formula) and *computability* (of this class). Now of the three example above only safety in databases is connected with both. It is reasonable therefore to take database theory as our starting point. Another reason for this choice is that many explicit proposals of decidable, syntactically defined classes of safe formulas have been made in this theory. Examples are: "range separable formulas" ([5]), "range restricted formulas" ([16]), "evaluable formulas" ([6]), "allowed formulas" ([23]), and "range safe formulas" ([2]) [2]. The simplest among them (and the closer to what has actually been implemented) is perhaps the following class $SS(D)$ ("syntactically safe" formulas for a database scheme D) from [24] (originally designed for languages with no function symbols) [3]:

1. $p_i(t_1, \ldots, t_{n_i}) \in SS(D)$ in case p_i (of arity n_i) is in D, and each t_i is either a variable or a constant.
2. $x = c$ and $c = x$ are in $SS(D)$ (where x is a variable and c is a constant).
3. $A \vee B \in SS(D)$ if $A \in SS(D)$, $B \in SS(D)$, and they have the same free variables.
4. $\exists x A \in SS(D)$ if $A \in SS(D)$.
5. If $A = A_1 \wedge A_2 \wedge \ldots \wedge A_k$ then $A \in SS(D)$ if both of the following conditions are met:
   (a) For each $1 \leq i \leq k$, either A_i is atomic, or A_i is in $\mathcal{SS}(\mathcal{D})$, or A_i is a negation of a formula of either type.
   (b) Every free variable x of A is limited in A. This means that there exists $1 \leq i \leq k$ s.t. x is free in A_i, and either $A_i \in SS(D)$, or $A_i \in \{x = y, y = x\}$, where y is already limited in A.

There is one clause in this definition which is somewhat strange: the last one, which treats conjunction. The reason why this clause does not simply tell us (like in the case of disjunction) when a conjunction of *two* formulas is in $SS(D)$, is the desire to take into account the fact that once the value of y (say) is known, the formula $x = y$ becomes safe. One of the crucial observations on which our framework is based is that in order to find

[2]In our opinion there is a mistake in the definition of the last one. According to this definition a formula like $x = c \wedge (\neg \exists y (y \neq x))$ is range safe, although it is clearly not domain independent (despite a theorem to the converse which is proved in [2]). We believe that the source of the problem is a mistake in the way negation is handled there, and that it should be corrected along the lines this is done below.

[3]What we present below is both a generalization and a simplification of Ullman's original definition. It should be noted that Ullman's main concern is the stronger property of domain-independence that we discuss in subsection 3.2.

a common generalization of the various notions described above one should indeed consider *partial* safety. In other words: safety should be viewed as a *relation* between formulas and (finite) sets of variables rather than as a *property* of formulas [4]. Since two different issues are involved here (size and computability), this observation leads to the following two generalizations of $SS(D)$ (where $Fv(E)$ denotes the set of free variables of E):

Definition 1. *A relation $\succ$ between formulas A of a first-order language L in a signature σ and subsets of $Fv(A)$ is a* size-safety (s-safety) relation *if it satisfies the following conditions:*

(1) *$A \succ \emptyset$ for all A.*

(2) *If $x \notin Fv(t)$ then $x = t \succ \{x\}$ and $t = x \succ \{x\}$.*

(3) *If $A \succ X$ and $B \succ X$ then $A \vee B \succ X$.*

(4) *If $y \notin X$ and $A \succ X \cup \{y\}$ then $\exists y A \succ X$.*

(5) *If $A \succ X$, $B \succ Y$, and $X \cap Fv(B) = \emptyset$ or $Y \cap Fv(A) = \emptyset$, then $A \wedge B \succ X \cup Y$.*

Definition 2. *A* c-safety relation *between formulas of a language L and finite sets of variables is defined like in Definition 1, except that condition* (1) *is replaced by the following weaker conditions:*

(1a) *$p(x_1, \ldots, x_n) \succ \emptyset$ in case p is a primitive n-ary predicate symbol of σ.*

(1b) *If $A \succ \emptyset$ then $\neg A \succ \emptyset$.*

Our standard interpretation of s-safety is that $A(x_1, \ldots, x_n, y_1, \ldots, y_k)$ is s-safe w.r.t. $\{x_1, \ldots, x_n\}$ in a given structure S, iff either $n = 0$, or for any assignment $c_1 \ldots, c_k$ of values from S for $y_1, \ldots, y_k$, the set of tuples $\langle d_1, \ldots, d_n \rangle$, which together with $c_1 \ldots, c_k$ satisfy A in S, is finite. It is easy to prove that this interpretation indeed defines an s-safety relation (see section 4). To get an intuition concerning definition 2, think of $A(x_1, \ldots, x_n, y_1, \ldots, y_k)$ as a query with parameters $y_1, \ldots, y_k$, and interpret "$A(x_1, \ldots, x_n, y_1, \ldots, y_k) \succ \{x_1, \ldots, x_n\}$" as: "The answer to the query A

[4]This may be compared with Tarski's definition of the validity *property* of formulas in structures via the satisfaction *relation* between formulas and assignments in structures.

is finite and effectively computable for any values of the parameters". *Intuitively* (see again section 4), this defines a c-safety relation, provided that the interpretations of the primitive function symbols of σ are all effectively computable, and the interpretations of the primitive predicate symbols of σ are all effectively decidable (This cannot be rigorously proved, though, since we do not have a precise definition of an "effectively computable answer to a query". We shall return to this point in section 4).

Note 1. For the present framework it is preferable to take $\wedge, \vee, \neg$ and $\exists$ as primitive, and $\rightarrow$ and $\forall$ as defined in terms of them. Moreover: we take $\neg(A \rightarrow B)$ as an abbreviation for $A \wedge \neg B$, and $\forall x_1 \ldots x_k A$ as an abbreviation for $\neg \exists x_1 \ldots x_k \neg A$. This entails the following important property of "bounded quantification": *If $\succ$ is a c-safety relation, $A \succ \{x_1, \ldots, x_n\}$, and $B \succ \emptyset$, then $\exists x_1 \ldots x_n.A \wedge B \succ \emptyset$ and $\forall x_1 \ldots x_n.A \rightarrow B \succ \emptyset$.*

Note 2. In all examples we know, whenever a safety relation $\succ$ is defined by some semantic property, it obeys the following principle: If $A \succ X$, B is logically equivalent to A, and $Fv(A) = Fv(B)$, then $B \succ X$. s-safety is usually closed under the even stronger principle: If $A \succ X$ where $X = \{x_1 \ldots x_k\}$, $\exists y_1 \ldots y_n \forall x_1 \ldots x_k (A \leftrightarrow B)$ is logically valid, and $\{y_1 \ldots y_n\} \cap Fv(B) = \emptyset$, then $B \succ X$. The reason we have not included these principles in the definitions above is that we want to be able to define *decidable* safety relations that can serve in applications as good substitutes for the undecidable, semantically defined ones. Still, for convenience one may incorporate into the definitions useful special cases of these properties, like standard boolean identities, and the following facts concerning substitutions (which follow from the equivalence between $A(t/y)$ and $\exists z \exists y(z = t \wedge y = z \wedge A)$, where $z \notin Fv(t) \cup Fv(A)$):

- If $y \notin X$, $A \succ X \cup \{y\}$, $Y \subseteq Fv(t)$, $Y \cap Fv(A) \subseteq \{y\}$, and $z = t \succ Y$ for $z \notin Fv(t) \cup Fv(A)$, then $A(t/y)$ is equivalent to some B s.t. $B \succ X \cup Y$.
- If $y \notin X$, $A \succ X$, and $X \cap Fv(t) = \emptyset$, then $A(t/y)$ is equivalent to some B such that $B \succ X$.

The straightforward way of defining a reasonable syntactical substitute for a given semantical safety relation is to use definitions 1 or 2 as a basis for an inductive definition. In most cases this amounts to specifying what atomic formulas (other than those of the form $x = t$ or $t = x$) are taken as safe w.r.t. what variables. For this it is usually best to use the following generalization of the notion of a signature for a language (see the introduction for the motivation):

Definition 3. *A* safety-signature *is a pair* (σ, F)*, where* σ *is an ordinary first-order signature and* F *is a function which assigns to every n-ary symbol* s *from* σ *(other than equality) a nonempty subset of* $\mathcal{P}(\{1,\ldots,n\})$*, so that if* $I \in F(s)$ *and* $J \subset X$ *then* $J \in F(s)$*.*

Definition 4. *Let* (σ, F) *be a safety-signature.* $\succ_{(\sigma,F)}$ $(\succ^s_{(\sigma,F)})$ *is the (inductively defined) minimal c-safety (s-safety) relation* $\succ$ *(in the first order language induced by* σ*) which satisfies the following conditions:*

1. *If* p *is an n-ary predicate symbol of* σ*;* $x_1,\ldots,x_n$ *are* n *distinct variables, and* $\{i_1,\ldots,i_k\}$ *is in* $F(p)$*, then* $p(x_1,\ldots,x_n) \succ \{x_{i_1},\ldots,x_{i_k}\}$*.*
2. *If* f *is an n-ary function symbol of* σ*;* $y, x_1,\ldots,x_n$ *are* $n{+}1$ *distinct variables, and* $\{i_1,\ldots,i_k\} \in F(f)$*, then* $y = f(x_1,\ldots,x_n) \succ \{x_{i_1},\ldots,x_{i_k}\}$*.*

Proposition 1. *Both* $\succ_{(\sigma,F)}$ *and* $\succ^s_{(\sigma,F)}$ *satisfy the following conditions:*

1. *If* $A \succ X$ *then* $X \subseteq Fv(A)$*.*
2. *If* $A \succ X$ *and* $Z \subseteq X$*, then* $A \succ Z$*.*

In the coming sections we shall see several applications of these notions.

3 Safety in Databases

From a logical point of view, a database of scheme $D = \{p_1,\ldots,p_n\}$ is just a given set of *finite* interpretations of $p_1,\ldots,p_n$. As noted in the introduction, a corresponding query language is an ordinary first-order language with equality, the signature of which contains D. A query is "safe" if its answer is finite and computable for all interpretations in which $p_1,\ldots,p_n$ are finite (and given), while the interpretations of all other predicate symbols are decidable, and function symbols (if any) are interpreted by computable functions. Our framework leads in this case to the following syntactical counterpart:

Definition 5. *Let* D *be a subset of* σ *such that each* $q \in D$ *is a predicate symbol of arity* k_q*.*

1. *The safety signature* (σ, F_D) *corresponding to* σ *and* D *is defined by:*

$$F_D(q) = \begin{cases} \mathcal{P}(\{1,\ldots,k_q\}) & \textit{if } q \in D \\ \{\emptyset\} & \textit{otherwise} \end{cases}$$

2. *A formula A is called (σ, F_D)-safe if $A \succ_{(\sigma,F_D)} Fv(A)$.*

It is easy to show inductively (using the intuitive meaning of c-safety given in the previous section) that each (σ, F_D)-safe formula can safely be used as a query for any database of scheme D, and that in a language without function symbols every formula in Ullman's $SS(D)$ is logically equivalent to some (σ, F_D)-safe formula[5]. It is important to note that our notion can in fact be used even if function symbols *are* allowed (provided that their intended interpretations are computable). Moreover: it is very easy to extend it *in a natural way* in order to be able to take into account safety properties that other functions and relations (not in the database scheme) might have in the intended domain(s). Suppose for example that "$<$" is in the language, and that its intended interpretation is the usual order relation of $\mathcal{N}$, or the substring relation on strings. In such a case the set $\{x \mid x < a\}$ is finite and computable for every a in the intended domain. This fact can be exploited by taking $F_D(<) = \{\emptyset, \{1\}\}$. An example of a query that will become then (σ, F_D)-safe is $\exists x \exists y (p_1(x, y) \land z < x + y)$.

3.1 An Application: Querying the Web

An interesting application of c-safety has implicitly been made in [15]. There the web is modeled as an ordinary database augmented with three more special relations: [6] $N(id, title, \ldots)$, $L(source, destination, \ldots)$, $C(node, value)$. The intuitive interpretations of these relations are the following:

- The relation N contains the Web objects which are identified by a Uniform Resource Locator (URL). *id* represents the URL and is a key.
- The relation L holds between two nodes *source* and *destination* if there is a hypertext link from the first to the second.
- The meaning of the relation C is that the string which is represented by its second argument occurs within the body of the document in the URL which is represented by its first argument.

The question investigated in [15] is: what queries should be taken as safe, if we assume that what is practically possible in the case of N and L is to

---

[5] If we strengthen $\succ_{(\sigma,F_D)}$ as suggested in Note 2, then $SS(D)$ becomes a proper subset of the set of (σ, F_D)-safe formulas.

[6] For simplicity, we ignore other special relations which are used there.

list all their tuples which *correspond to a given first argument*, while C is only assumed to be decidable. A special "Safe Web Calculus" based on these assumptions is then introduced. It is not difficult to see that the notion of safety which is defined by this calculus is in fact equivalent to (σ_{web}, F_{web})-c-safety in our sense, where $\{L, N, C\} \subseteq \sigma_{web}$, and F is defined like in ordinary databases, except that $F(L) = \mathcal{P}(\{2, \ldots, m\})$ (where m is the arity of L), $F(N) = \mathcal{P}(\{2, \ldots, k\})$ (where k is the arity of N), and $F(C) = \{\emptyset\}$.

3.2 Domain-independence

Another property of queries to databases which is considered to be crucial ([12, 24, 2]) is *domain-independence* (d.i.). Its definition (in the case of ordinary databases) is the following:

Definition 6. [7] *Let σ be a signature which includes $\overrightarrow{P} = \{P_1, \ldots, P_n\}$, and optionally constants and other predicate symbols (but no function symbols). A query in σ is called* $\overrightarrow{P}-$d.i. ($\overrightarrow{P}-$domain-independent) *if it has the same answer in S_1 and S_2, whenever S_1 is a substructure of S_2, and the interpretations of $\{P_1, \ldots, P_n\}$ in S_1 and S_2 are identical.*

Our next goal is to show that domain-independence can also successfully be handled within our framework (using safety-signatures as our main tool). For this we generalize first the ordinary notion of an extension of a structure (for a signature σ) to structures for safety-signatures:

Definition 7. *Let (σ, F) be a safety-signature with no function symbols (other than constants). Let S_1 and S_2 be two structures for σ s.t. $S_1 \subseteq S_2$. S_2 is called a* $(\sigma, F)-$extension *of S_1 if the following condition is satisfied: If $p \in \sigma$ is of arity n, $I \in F(p)$, and $a_1, \ldots, a_n$ are elements of S_2 such that $a_i \in S_1$ in case $i \notin I$, then $S_2 \models p(a_1, \ldots, a_n)$ iff $a_i \in S_1$ for* all *i and $S_1 \models p(a_1, \ldots, a_n)$.*

Note 3. Since $\emptyset \in F(P)$ for all $P \in \sigma$, S_1 is a substructure of S_2 whenever S_2 is a $(\sigma, F)-$extension of S_1.

Examples:

[7] This is a slight generalization of the definition in [Su98], which in turn is a generalization of the usual one ([Ki88,Ul88]). The latter applies only to free Herbrand structures which are generated by adding to σ some new set of constants.

1. Let $\sigma = \{=, <\}$, and let $F(<) = \{\emptyset, \{1\}\}$. Obviously, a structure for σ is a $(\sigma, F)-$extension of its initial segments (and only of them).

2. Let $\sigma_{AS} = \{=, \in\}$ and let $F_{AS}(\in) = \{\emptyset, \{1\}\}$. In this case the "universe" is a $(\sigma_{AS}, F_{AS})-$ extension of what are known in Set Theory as "transitive sets".

Definition 8. *Let (σ, F) be as in Definition 7. A formula A of σ is called* $(\sigma, F)-$d.i. w.r.t. *X ($A \succ^{di}_{(\sigma,F)} X$) if whenever S_2 is a $(\sigma, F)-$extension of S_1, and A^* resulted from A by substituting values from S_1 for the free variables of A that are not in X, then the sets of tuples which satisfy A^* in S_1 and in S_2 are identical.* [8] *A formula A of σ is called* $(\sigma, F)-$d.i. *if $A \succ^{di}_{(\sigma,F)} Fv(A)$.*

It can easily be proved that $\succ^{di}_{(\sigma,F)}$ is a c-safety relation. It follows that if $A \succ_{(\sigma,F)} X$ (Definition 4) then $A \succ^{di}_{(\sigma,F)} X$. In particular: if $A \succ_{(\sigma,F)} Fv(A)$ then A is $(\sigma, F)-$d.i. It is also obvious that if D is a database scheme in a signature σ, then a formula A is $(\sigma, F_D)-$d.i. iff it is d.i. for D in the usual sense (of Definition 6). Since already in this case the notion of d.i. is undecidable ([7]), the class of $(\sigma, F_D)-$safe formulas is again a good syntactical substitute.

Note 4. Despite the close connection, safety of queries (in the sense of being "effectively finite") and domain independence of them are in general independent notions. Thus every logically valid sentence (or a logical contradiction) is d.i. w.r.t. $\emptyset$, but not necessarily safe w.r.t. $\emptyset$. On the other hand $\forall x.x = c$ is (effectively) safe w.r.t. $\emptyset$ (it is true if the domain is a singleton, false otherwise), but not d.i. w.r.t. $\emptyset$ (for precisely the same reason).

4 Safety in Computability Theory and in Metamathematics

We have followed up to now the intuition that a query is safe iff its answer is finite and computable, but we have not defined what "computable" means. The intuitive notion we have in mind is *not* identical to that investigated in Classical Computability Theory (CCT). CCT provides answers to the questions "what *extensional* relations are decidable?" and "what *extensional*

[8] A^* is a formula only in a generalized sense, but the intention should be clear.

functions are computable?". Queries define however *intensional* relations, and CCT provides only necessary conditions (like finiteness and decidability) for the computability of such relations and functions. Thus every extensional finite relation is "computable" according to CCT, but in reality we might not know how to actually compute an intensional relation even if its extension is finite. However, it is precisely the question of computing intensional relations and functions that we encounter in practice. A particularly delicate question in this context is what is the interpretation of "The answer to the query φ is computable" in case φ is a sentence (i.e.: a query with a "yes" or "no" answer). This is a question to which CCT provides no clue, but is important for database theory, and is also the main point of difference between s-safety and c-safety. It should be noted that this question is crucial for constructive computability theory too. Thus Bridges explicitly gives in [4] the following example of a function f from $\mathcal{N}$ to $\mathcal{N}$, which is "computable" according to CCT, but not constructively so: For all n, $f(n)$ is 1 if Goldbach conjecture is true, 0 otherwise.

I am not aware of any precise definition in the literature (or an analogue of Church's thesis) for the concepts of "computable intensional function" and "computable intensional relation". What can therefore be done at present is to provide obvious properties of these notions and use them for developing a useful corresponding general theory.[9] Where should we start? Well, the usual approach to CCT is to characterize first the class of computable functions, and then to define the class of decidable relations as those relations whose characteristic functions are computable. However, in modern mathematics functions are defined as a special type of relations, and so it seems more reasonable from its point of view to go the other way around. This is certainly more natural when a theory of intensional computability is sought, since even intensional relations and functions should be defined in some formal language — and what first-order languages (which are the most natural languages to use for this purpose) directly define are *relations.* [10] Now for *intensional* relations the general framework suggested here does provide a general relative computability theory (though we do not claim it to be the ultimate one). In fact it provides general sufficient criteria for a parametric formula in some first-order language L to define a (finite) computable (intensional) relation

[9] A similar procedure is suggested in [17] as a possible approach for trying to *prove* Church's thesis in the extensional case.

[10] This has indeed been the approach of [21], where the class of "bounded formulas" is used for defining the basic notions of CCT. We will return to this class below.

in all structures for L in which the interpretations of the primitive predicates and functions of L have certain computational properties:

Definition 9. *Let L be a first-order language with equality, and let S be a structure for L.*

1. *A formula A of L which is not a sentence is $S-$*effective *if the number of tuples which satisfy it in S (for some order of its free variables) is finite, and they can effectively be listed. A sentence is $S-$*effective *if its truth-value in S can effectively be computed.*

2. *A formula A of L is $S-$*effective *w.r.t a finite subset X of its free variables if for any substitution of concrete (syntactic names for) elements from S for the free variables of A which are not in X, we get an $S-$effective formula (in the extended language $L(S)$ of S ([17])).*

Note 5. It should again be emphasized that this definition assumes the intuitive notion of "effective computability" which is left here *undefined* (and probably cannot be defined!). Note also that $S-$effectiveness of a formula A w.r.t. $\emptyset$ means that the relation on S which A defines is effectively *decidable.*

Definition 10. *Let (σ, F) be a safety-signature. A structure S for σ is* appropriate *for (σ, F) if it satisfies the following two conditions:*

- *If p is an n-ary predicate symbol of σ; $x_1, \ldots, x_n$ are n distinct variables, and $\{i_1, \ldots, i_k\} \in F(p)$, then $p(x_1, \ldots, x_n)$ is $S-$effective w.r.t. $\{x_{i_1}, \ldots, x_{i_k}\}$.*

- *If f is an n-ary function symbol of σ; $y, x_1, \ldots, x_n$ are $n+1$ distinct variables, and $\{i_1, \ldots, i_k\} \in F(f)$, then $y = f(x_1, \ldots, x_n)$ is $S-$effective w.r.t. $\{y\}$ and w.r.t. $\{x_{i_1}, \ldots, x_{i_k}\}$.*

Theorem 1. *If S is appropriate for (σ, F), and $C \succ_{(\sigma,F)} X$, then the formula C is $S-$effective w.r.t. X.*

Intuitive Proof: By induction on the structure of C. If C is of the form $x = t$ or $t = x$ then the claim is proved by induction on the structure of t (using the assumption that every formula of the form $y = f(x_1 \ldots, x_n)$ is $S-$effective w.r.t. $\{y\}$). The other safety conditions concerning atomic formulas directly follow from the fact that S is appropriate for (σ, F). The induction step splits into four cases. We do here the case where $C = A \wedge B$,

$A \succ X$, $B \succ Y$, and $Y \cap Fv(A) = \emptyset$. To simplify notation, assume that $Fv(A) = \{x, z\}$, $Fv(B) = Fv(C) = \{x, y, z\}$, $X = \{x\}$, $Y = \{y\}$. Let c be an element of S. To compute $\{< x, y >\in S^2 \mid C(c/z)\}$, compute first the set $Z(c) = \{x \in S \mid A(c/z)\}$ ($Z(c)$ is finite and effectively computable by our assumptions on A). Then for each $d \in Z(c)$ compute the set $W(c, d) = \{y \in S \mid B(d/x, c/z)\}$ ($W(c, d)$ is finite and effectively computable by our assumptions on B). The set $\{< x, y >\in S^2 \mid C(c/z)\}$ is the finite union of the sets $\{d\} \times W(c, d)$ ($d \in Z(c)$).

We present now two famous applications of Theorem 1 from the literature on Metamathematics. This is another area in which one needs (especially for the proof of Gödel's second incompleteness theorem) a class of effective *intensional* relations, defined by formulas in some particular f.o. languages.

Bounded safety: Let $\sigma_b = \{=, 0, S, +, \times, <\}$. Define $F_b(<) = \{\emptyset, \{1\}\}$ and $F(f) = \{\emptyset\}$ for any function symbol f. Let $\succ_b = \succ_{(\sigma_b, F_b)}$.

Primitive recursive safety: Let $\sigma_{PA} = \{=, 0, S, +, \times\}$, and let σ_{PR} be σ_{PA} augmented with function symbols for every primitive recursive (p.r.) function. Define $F_{PR}(S) = \mathcal{P}(\{1\})$, $F_{PR}(+) = \mathcal{P}(\{1, 2\})$, and $F_{PR}(f) = \{\emptyset\}$ otherwise. Let $\succ_{PR} = \succ_{(\sigma_{PR}, F_{PR})}$.

Definition 11. *A formula A is* $\succ_b$-effective *if* $A \succ_b \emptyset$. *A formula A is* $\succ_{PR}$-effective *if* $A \succ_{PR} \emptyset$.

It is easy to see that the structure $\mathcal{N}$ of the natural numbers (with the standard interpretations of the symbols in σ_b and σ_{PR}) is appropriate for (σ_b, F_b) and (σ_{PR}, F_{PR}). Thus $\{1\} \in F_b(<)$ means that $x < y \succ_b \{x\}$ (where x and y are two different variables). This in turn means that given any $n \in \mathcal{N}$, there is only a finite number of k's such that $k < n$, and they can effectively be listed. Similarly, the fact that $\{1, 2\} \in F_{PR}(+)$ means that $y = x_1 + x_2 \succ_{PR} \{x_1, x_2\}$. This in turn means that given any $n \in \mathcal{N}$, there is only a finite number of pairs $\langle k_1, k_2\rangle$ such that $n = k_1 + k_2$, and they can effectively be listed. Both claims are obvious. It follows from Theorem 1 that if A is $\succ_b$-*effective* or $\succ_{PR}$-effective then A defines a decidable relation on the structure $\mathcal{N}$.

By letting $x < y$ abbreviate in σ_{PA} $\exists w \exists x \exists z . y = x + w \wedge w = S(z)$, we get that $x < y \succ_{PR} \{x\}$ as well. Hence $\succ_b$ is contained in $\succ_{PR}$. Now Note 1 implies that the set of $\succ_b$-effective formulas is closed under bounded quantification (the same applies to the set of $\succ_{PR}$-effective formulas). It follows that this class is an extension of the class of bounded formulas ([21]). It is not difficult to see, in fact, that the two classes define the same class of (extensional) relations on $\mathcal{N}$. Similarly, the class of $\succ_{PR}$-effective formulas [11] is equivalent to Feferman's class of primitive recursive formulas ([9, 14]), which is of crucial importance in Metamathematics. Actually, this importance is easily seen to be due to the following connections between $\succ_{PR}$ and provability:

Theorem 2. *Let Q^* and PA^* be the extensions in σ_{PR} of Q and PA (respectively) with the defining axioms of every p.r. function*[12]*. Let A be a formula in σ_{PR} such that $A \succ_{PR} \{x_1, \ldots, x_k\}$.*

1. *Assume that A' is a closed substitution instance of A. If A' is true (in $\mathcal{N}$) then $\vdash_{Q^*} A'$, while if it false then $\vdash_{Q^*} \neg A'$.*
2. *Let $F_v(A) - X = \{y_1, \ldots, y_n\}$. If $k > 0$ then there exists a p.r. function f_A s.t. $\vdash_{Q^*} A \to \big(x_1 < f_A(y_1, \ldots, y_n) \wedge \ldots \wedge x_k < f_A(y_1, \ldots, y_n)\big)$.* [13]
3. *$A \to Pr_{PA^*}(\ulcorner A(\dot{x}_1, \ldots, \dot{x}_n, \dot{y}_1, \ldots, \dot{y}_n)\urcorner)$ is provable in PA^*.*[14]

Proof: The proof of (3) is similar to the usual proofs of such results in the literature, using the fact that by (2), if $A \succ_{PR} \{x\}$ then the existential quantification of A on x can be replaced by bounded quantification on x. The proofs of (1) and (2) are done simultaneously, using an induction on the construction of $\succ_{PR}$. We do here the case $A = \exists y B$ as an example. To simplify notation we assume that $F_v(B) = \{x, y, z\}$, $B \succ_{PR} \{x, y\}$ (and so indeed $A \succ_{PR} \{x\}$). By induction hypothesis there is a p.r. function f_B s.t.:

$$\text{(I)} \qquad \vdash_{Q^*} B \to x < f_B(z)$$
$$\text{(II)} \qquad \vdash_{Q^*} B \to y < f_B(z)$$

[11] More accurately: the class of formulas which result from the $\succ_{PR}$-effective ones by substituting everywhere $\delta_f(x_1, \ldots, x_n, y)$ for $y = f(x_1, \ldots, x_n)$, where $\delta_f(x_1, \ldots, x_n, y)$ is the standard formula in σ_{PA} which binumerates f in PA ([14]).

[12] It is well known ([14]) that Q^* and PA^* are conservative extensions of Q and PA.

[13] For convenience, we use here the same symbol for f_A in both σ_{PR} and in our meta-language.

[14] The notation here follows that of [20].

It immediately follows from (I) that $\vdash_{Q^*} A \to x < f_B(z)$, proving (2) for A (take $f_A = f_B$). To prove (1), assume that A' is a closed instance of A. Then there are numbers n and k such that A' is equivalent in Q^* to $\exists y B(\overline{n}, y, \overline{k})$. If A' is true in $\mathcal{N}$ there is a number m such that $B(\overline{n}, \overline{m}, \overline{k})$ is true, and so (by induction hypothesis) provable in Q^*. This entails that A' is provable in Q^*. If, on the other hand, A' is false then $B(\overline{n}, \overline{i}, \overline{k})$ is false for every $i < f_B(k)$. Hence, by induction hypothesis, we have for every $i < f_B(k)$:

$$\text{(III)} \quad \vdash_{Q^*} \neg B(\overline{n}, \overline{i}, \overline{k})$$

On the other hand (II) above implies that

$$\text{(IV)} \quad \vdash_{Q^*} B(\overline{n}, y, \overline{k}) \to y < \overline{f_B(k)}.$$

Together, (III) and (IV) imply that $\vdash_{Q^*} \neg\exists y B(\overline{n}, y, \overline{k})$, and so $\vdash_{Q^*} \neg A'$.

Note 6. Let $\sigma_p = \{=, 1, S, +, \times\}$. Define $F_p(S) = P(\{1\})$, $F_p(+) = F_p(\times) = P(\{1, 2\})$. Let $\succ_p = \succ_{(\sigma_p, F_p)}$. It is possible to show that the class of $\succ_p$-effective relations is exactly the class of arithmetical relations that can be decided in polynomial time (some other complexity classes can similarly be characterized).

The class of $\succ_b$-effective relations is a proper subclass of the class of $\succ_{PR}$-effective relations. Our general definition allows us, accordingly, to capture different notions of "effectiveness". None of them can exactly capture the intuitive notion of "constructive effectiveness" (by a diagonalization argument). What *does* seem to be robust is the notion of *semi-decidable* relations:

Definition 12. *A formula is called* $\succ_b$-r.e. *if it is of the form* $\exists x A$*, where* A *is a* $\succ_b$*-effective formula. A relation is called* $\succ_b$-r.e. *if it is defined by a* $\succ_b$-r.e. *formula. The classes of* $\succ_p$*-r.e. and* $\succ_{PR}$*-r.e. formulas and relations are defined similarly.*

Proposition 2. *The classes of* $\succ_b$*-r.e. relations,* $\succ_p$*-r.e. relations, and* $\succ_{PR}$*-r.e. relations are all identical (to the usual class of r.e. relations).*

Note 7. From the last proposition it is clear that a possible formulation of Church's Thesis is that a relation R on $\mathcal{N}$ is semi-decidable iff it is definable by a formula of the form $\exists x A$, where A is either a $\succ_b$-effective or $\succ_{PR}$-effective (R is of course decidable iff both R and its complement are semi-decidable). It follows that CCT for extensional relations does not really involve principles that go beyond those that are suggested in our framework for c-safety of f.o. formulas and for intensional computability.

Note 8. Unlike in databases, the interest in CCT and in Metamathematics has been in safety of a formula A w.r.t. $\emptyset$ (rather than w.r.t. $Fv(A)$).

5 Safety in Set Theory

As we have noted in the introduction, what the various axiomatic set theories actually provide are syntactic criteria for classes of formulas which may be assumed to be "safe" for applying the naive comprehension schema. This is evident, e.g., in the case of Quine's NF, in which the notion of a "stratified formula" is used. However, we show in this section that it is true also in the case of ZF, the most famous (and universally accepted) among the axiomatic set theories. Most of the axioms of ZF are indeed just particular instances of the comprehension schema. As noted in the introduction, the guiding line behind the choice of these instances has been the semantic "limitation of size doctrine" ([8, 10]). According to this criterion, only collections which are not "too big" can be accepted as sets. Here "not too big" is an intuitive notion (which encompasses quite large infinite sets). With this intuitive notion in mind, a formula A of set theory may be called "size-safe" ("s-safe") w.r.t x, if $\{x \mid A\}$ determines a collection which is "not too big". The comprehension axioms of ZF lists all the cases which are universally recognized to be "s-safe" in this sense. Now in databases "size-safe" means "finite". In set theory it means something completely different (like "not equipotent with the collection of all sets"). We show now that the principles which have been used in these two disciplines in order to secure limitation of size are nevertheless the same, although they have been developed independently. This, we believe, provides strong support to the claim (recently made, e.g., by H. Friedman) that with the exception of the infinity axiom, all the other comprehension axioms of ZF are obtained by an extrapolation from the finite case to the general one. It also leads to new presentations of ZF which are based on purely *syntactic* considerations — in contrast to the usual semantical justifications (as presented, e.g., in [18]).

To achieve these goals we should use of course an appropriate s-safety relation rather than a c-safety relation (computability is not an issue here!). To be able to present one, we need (because of the Powerset axiom) to assume that $=$, $\in$ and $\subseteq$ are all primitive symbols of ZF [15]. Finally, in order to get a real insight into the nature of ZF, we follow its presentation in [18]

[15] Hence the usual definition of $\subseteq$ in terms of $\in$ should be taken as one of the axioms.

(where the axioms of ZF are explained and an attempt is made to justify them on a semantic ground). Practically this means that we use for ZF a dynamic language which has the means to introduce new symbols for definable functions. In other words: once $\exists! y A(x_1, \ldots, x_n, y)$ is proved, it is possible to introduce a new function symbol F_A together with the axiom $\forall x_1, \ldots, x_n (A(x_1, \ldots, x_n, F_A(x_1, \ldots, x_n)))$ (see [17], section 4.6). Officially we assume that the language includes all these function symbols from the start, and that every instance of the following schema is an axiom:

$$\forall x_1, \ldots, x_n (\exists! y A(x_1, \ldots, x_n, y) \to A(x_1, \ldots, x_n, F_A(x_1, \ldots, x_n)))$$

By this we obtain a conservative extension of ZF which we denote by ZF^f. We next introduce a corresponding safety-signature:

Definition 13. *Let* $\sigma_{ZF} = \{=, \in, \subseteq\}$*, and let* σ_{ZF^f} *be* σ_{ZF} *augmented with all the function symbols of* ZF^f*. Define* $F_{ZF}(\in) = F_{ZF}(\subseteq) = \{\emptyset, \{1\}\}$*. Extend* F_{ZF} *to* σ_{ZF^f} *by letting* $F_{ZF^f}(g) = \{\emptyset\}$ *for every function symbol* g*.*

In the rest of this section "safety" will mean $(\sigma_{ZF^f}, F_{ZF^f})$-s-safety. For the reader convenience, we recall that this relation is defined here as follows:

(A) Every formula is safe w.r.t $\emptyset$.

(B) If $x \notin Fv(t)$ then $x = t$, $t = x$, $x \in t$, and $x \subseteq t$ are safe w.r.t $\{x\}$.

(C) If A and B are both safe w.r.t. X, then so is $A \vee B$.

(D) If A is safe w.r.t. X, B is safe w.r.t. Y, and $X \cap Fv(B) = \emptyset$ or $Y \cap Fv(A) = \emptyset$, then $A \wedge B$ is safe w.r.t. $X \cup Y$.

(E) If $y \notin X$ and A is safe w.r.t. $X \cup \{y\}$, then $\exists y A$ is safe w.r.t. X.

Theorem 3. *The standard comprehension axioms of* ZF^f *(Pairing, Powerset, Union, Separation, and Replacement) can be replaced by the following single* safe comprehension schema (SCn^f):

$$\exists Z \forall x. x \in Z \Leftrightarrow A, \text{ where } A \text{ is safe w.r.t. } \{x\}, \text{ and } Z \notin Fv(A).$$

Proof: The comprehension axioms of ZF^f are all instances of SCn^f:

Pairing: $\exists Z \forall x. x \in Z \Leftrightarrow (x = y \vee x = z)$

The formula used here is safe w.r.t. $\{x\}$ by (B) and (C).

Powerset: $\exists Z \forall x . x \in Z \Leftrightarrow (x \subseteq y)$

The formula used here is safe w.r.t. $\{x\}$ by (B). [16]

Union: $\exists Z \forall x . x \in Z \Leftrightarrow (\exists v . x \in v \wedge v \in y)$

The formula used here is safe w.r.t. $\{x\}$ by (B), (D), and (E).

Separation: $\exists Z \forall x . x \in Z \Leftrightarrow (x \in y \wedge B)$

The formulas used in this schema are safe w.r.t. $\{x\}$ by (A), (B), (D).

Replacement: $\exists Z \forall x . x \in Z \Leftrightarrow (\exists v . v \in y \wedge x = t)$

Here x, y, v are 3 distinct variables, and t is a term in which x does not occur free. The formulas used in this scheme are safe w.r.t. $\{x\}$ by (B), (D), and (E).

For the converse, let C be a formula which is safe w.r.t. $\{x_1, \ldots, x_n\}$ (where $\{x_1, \ldots, x_n\}$ are all free in C). Define $Set_{x_1,\ldots,x_n} C$ to be $\neg C \vee C$ in case $n = 0$, and $\exists Z(\forall x_1 \ldots \forall x_n (< x_1, \ldots, x_n > \in Z \Leftrightarrow C))$ in case $n > 0$. We show by induction on the structure of C that $Set_{x_1,\ldots,x_n} C$ is a theorem of ZF^f (the principle we want to show is obtained from this result as the particular case in which $n = 1$). Most of the cases are straightforward. We again do here the case of conjunction (which is the most complicated) as an example. To simplify notation, assume again that $Fv(A) = \{x, z\}$, $Fv(B) = \{x, y, z\}$, A is safe w.r.t. $\{x\}$, B is safe w.r.t. $\{y\}$ (and so $A \wedge B$ is safe w.r.t. $\{x, y\}$). By induction hypothesis, $\vdash_{ZF^f} Set_x A$, and $\vdash_{ZF^f} Set_y B$. We show that $\vdash_{ZF^f} Set_{x,y}(A \wedge B)$. Now the assumptions imply that there are sets Z(z) and W(x,z) such that:

$$\vdash_{ZF^f} x \in Z(z) \Leftrightarrow A \qquad \vdash_{ZF^f} y \in W(x, z) \Leftrightarrow B$$

It follows that $\{< x, y >\mid A \wedge B\} = \bigcup_{x \in Z(z)} \{< x, y >\mid y \in W(x, z)\}$. Hence $Set_{x,y}(A \wedge B)$ follows from the axiom of replacement and the axiom of union.

[16] Note that the validity of the Powerset axiom is enforced here by taking $\subseteq$ as primitive, and letting $\{1\} \in F_{ZF}(\subseteq)$.

Example: The existence of the Cartesian product of two sets, U and V, is due to the safety w.r.t. $\{x\}$ of $\exists a \exists b . a \in U \wedge b \in V \wedge x = < a, b >$ (One should here justify first the use of the term $< a, b >$. This is easy.

Already Theorem 3 suffices for supporting the claim that the construction principles behind ZF are nothing more than standard syntactical principles concerning the first-order logical constants which are normally used to secure *finiteness*[17]. However, if one insists on using just standard first-order formulas of the signature σ_{ZF}, then replacement causes a problem. The reason is that unlike the other comprehension axioms of ZF, its official formulation in ZF has the form of a conditional. A possible solution to this problem is to translate into the language of ZF the conditions which define safety, and take these translations as our axioms. For this we assume first that a binary function symbol $<,>$ for forming ordered pairs is added to σ_{ZF}, together with an axiom which corresponds to its usual definition.[18]

Theorem 4. *The various comprehension axioms of ZF can be replaced (in the language with $<,>$) by the following axioms:*

(A) $Set_{x_1,\ldots,x_n} A \Rightarrow Set_{z_1,\ldots,z_n} A$ *where* $z_1, \ldots, z_n$ *is a permutation of* $x_1, \ldots, x_n$.

(B1) $set_x x = y$

(B2) $set_x x \subseteq y$

(C) $(Set_{x_1,\ldots,x_n} A \wedge Set_{x_1,\ldots,x_n} B) \Rightarrow Set_{x_1,\ldots,x_n} A \vee B$

(D) $(Set_{x_1,\ldots,x_n} A \wedge (\forall x_1 \ldots \forall x_n Set_{y_1,\ldots,y_m} B)) \Rightarrow Set_{x_1,\ldots,x_n,y_1,\ldots,y_m} A \wedge B$ *in case* $\{y_1, \ldots, y_m\} \cap Fv(A) = \emptyset$.

(E) $Set_{x_1,\ldots,x_n,y} A \Rightarrow Set_{x_1,\ldots,x_n} \exists y A$

Proof: We shall show here how to prove replacement from the new set of axioms, leaving the rest for the reader. For this it is convenient to use the version of replacement given in [17]. In the present notation, this version can be formulated as follows:

$$\forall y Set_x A \Rightarrow Set_x \exists y . y \in w \wedge A$$

[17] This basing of the axioms of ZF on a syntactically defined notion of "smallness" is similar in spirit to recent works on category-theoretic models of ZF (see [11, 19]).

[18] Alternatively, one may add a function symbol for forming unordered pairs.

So assume $\forall y Set_x A$. Since $Set_y y \in w$ is logically valid, this assumption implies $(Set_y y \in w) \wedge (\forall y Set_x A)$. By axiom (D) we can infer therefore $Set_{x,y} y \in w \wedge A$. From this $Set_x \exists y.y \in w \wedge A$ follows by axiom (E).

Note 9. Still another approach, in which an extra case is added to the notion of safety in ZF (and the explicit use of the abstraction operation is allowed) is outlined in [1]. In that paper also the infinity axiom is presented as a particular case of the safe comprehension schema, but for this one needs to use an extension of first-order logic with a transitive closure operation.

Note 10. We have seen that in database theory the interest is in safety of a formula w.r.t. to its whole set of free variables. Then we saw that in computability theory and in metamathematics the interest is in safety of a formula w.r.t. the empty set of variables. Now we see that in set theory, in contrast, the main interest is in safety of a formula w.r.t. exactly *one* of its free variables! These differences might be the reason why the connection between the three cases has been hidden for so long.

5.1 Absoluteness

It is interesting to note that also an analogue of the concept of domain-independence from database theory (see subsection 3.2) has independently been introduced in the literature on set theory. This is the notion of *absoluteness*, which is crucial for independence proofs (see, e.g., [13]). Indeed, it is easy to see that a formula is $(\sigma_{AS}, F_{AS})-$d.i. with respect to $\emptyset$ (see the second example after Definition 7) iff it is absolute according to the literature on set theory. Again we see here an interesting difference between what has been taken to be important in database theory and in set theory: while in database theory the main interest is in d.i. of a formula w.r.t. its set of free variables, in set theory the interest has been in d.i. of a formula w.r.t. $\emptyset$!.

6 Bibliography

[1] A. Avron, *Transitive Closure and the mechanization of Mathematics*, in F. Kamareddine (ed.), **Thirty Five Years of Automating Mathematics**, Kluwer Academic Publishers, 2003, 149-171.

[2] S. Abiteboul, R. Hull and V. Vianu, **Foundations of Databases**, Addison-Wesley, 1995.

[3] J. Barwise, **Admissible Sets and Structures**, Perspectives in Mathematical Logic, Springer-Verlag, 1975.

[4] D. S. Bridges, *Constructive Truth in Practice*, in H. G. Dales and G. Oliveri (eds.), **Truth in Mathematics**, Calendron Press, 1998, 53-69.

[5] E.F. Codd, *Relational Completeness of Database Sublanguages*, in R. Rustin (ed.), **Database Systems**, Prentice-Hall (1972), pp. 65-98.

[6] R. Demolombe, *Syntactical characterization of a subset of domain independent formulas*. Technical report, ONERA-CERT, Toulouse (1982).

[7] R.A. Di Paola, *The recursive unsolvability of the decision problem for the class of definite formulas*. J. ACM 16, 324-327, 1969.

[8] A. Fraenkel, Y. Bar-Hillel, and A. Levy, **Foundations of Set Theory**, North-Holland, Amsterdam, 1973.

[9] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fundamenta Mathematica 49 (1960) 35-92.

[10] M. Hallett **Cantorian Set Theory and Limitation of Size**, Clarendon Press, 1984.

[11] A. Joyal, and I. Moerdijk, **Algebraic Set Theory**, Cambridge University Press, 1995.

[12] M. Kiffer, *On Safety Domain independence and capturability of database queries*, Proc. International Conference on database and knowledge bases, 405-414, Jerusalem 1988.

[13] K. Kunen, **Set Theory: an Introduction to Independence Proofs**, North-Holland, Amsterdam, 1980.

[14] P. Lindström, **Aspects of Incompleteness**, Lectures Notes in Logic 10, Springer, 1997.

[15] A. O. Mendelzon, and T. Milo, *Formal Models of Web Queries*, Proceedings of the Sixteenth ACM Symposium on Principles of Database Systems, 134-143, 1997.

[16] J.M. Nicolas, *Logic for Improving Integrity Checking in Relational Data Bases*, Acta Informatica, Vol. 18 (1982), pp. 227-253.

[17] J. R. Shoenfield, **Mathematical Logic**, Addison-Wesley, 1967.

[18] J. R. Shoenfield, *Axioms of Set Theory*, in: **Handbook Of Mathematical Logic** (J. Barwise, Ed.), North-Holland, Amsterdam, 1977.

[19] A. Simpson, *Axioms for the Category of Classes*, Proceedings of the Fourteenth IEEE Symposium on Logic in Computer Science, 77-85.

[20] C. Smorynski, *The Incompleteness Theorems*, in **Handbook of Mathematical Logic** (J. Barwise, Ed.), North-Holland, Amsterdam, 1977.

[21] R. M. Smullyan, **The Incompleteness Theorems**, Oxford University Press, 1992.

[22] D. Suciu, *Domain-independent queries on databases with external functions*, Theoretical Computer Science 190, 279-315, 1998.

[23] R.W. Topor, *Domain independence formulas and databases*. Theoretical Computer Science 52 (1987) 281-306.

[24] J.D. Ullman, **Principles of database and knowledge-base systems**, Computer Science Press, 1988

A First Order System with Finite Choice of Premises

Kai Brünnler
kai.bruennler@gmx.net
Technische Universität Dresden, Fakultät Informatik, 01062 Dresden, Germany
and
Institut für Informatik und angewandte Mathematik, Fachgruppe für theoretische Informatik und Logik, Neubrückstr. 10, CH–3012 Bern, Switzerland
Alessio Guglielmi
alessio.guglielmi@inf.tu-dresden.de
Technische Universität Dresden, Fakultät Informatik, 01062 Dresden, Germany

1 Introduction

We call *finitely generating* an inference rule in a sequent system [7] if, given its conclusion, there is only a finite set of premises to choose from. This property is desirable from the viewpoint of proof search, since it implies that the search tree is finitely branching. It is also desirable for showing consistency, since the biggest obstacle to showing consistency is the cut rule, which is not finitely generating.

Much effort has been devoted to eliminating the cut rule in various systems: theorems of cut elimination are at the core of proof theory. In addition to the cut, there is another source of infinite choice in the bottom-up construction of a first order proof, namely the choice in instantiating an existentially quantified variable. Research grounded in Herbrand's theorem [11] deals with this aspect and is at the core of automated deduction and logic programming.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 59–74

Implementations of sequent systems for first order predicate logic are usually based on rules that are finitely generating. Such rules can be obtained for example as follows: first prove cut elimination to get rid of the cut rule, then use unification instead of blindly guessing instantiations.

This paper shows how one can eliminate all sources of infinite choice in a system of first order classical logic in a much simpler way, in particular without the use of cut elimination. To do so we use the *calculus of structures* [9], a formalism based on *deep inference*, which is the possibility of applying inference rules deep inside formulae. Systems in the calculus of structures offer the same proof theoretical properties as systems in the sequent calculus, in particular it is possible to prove cut elimination and many other normalisation results [9, 3, 1, 2]. The point here is that it is possible to obtain finitely generating systems without having to use these complex methods.

The main idea we exploit is that there are actually two sources of infinite choice in the cut rule: an infinite choice of atoms and an infinite choice in how these atoms can be combined for making formulae. Deep inference allows us to separate these two kinds of infinite choice. First we reduce the cut rule to atomic form, which is immediate thanks to deep inference. Then we eliminate only those cuts which (seen bottom-up) introduce atoms that do not occur in the conclusion. This is much simpler than a full-blown cut elimination. The instances of cut that we retain introduce atoms that already occur in the conclusion, so they are finitely generating.

Just like the cut rule, the rule for existential quantification also has two sources of infinite choice: an infinite choice of function symbols and an infinite choice in how these function symbols can be combined for making terms. We eliminate them by using the same technique that we used for the cut: first we reduce the instantiation rule to a form which instantiates only with one function symbol at a time. Then we eliminate those instances which introduce function symbols that do not occur in the conclusion. We are left with a finitely generating instantiation rule.

In the sequent calculus, which restricts the application of inference rules to the main connective of a formula, it is impossible to eliminate infinite choice in such a simple manner. First, the cut rule can not immediately be reduced to atomic form: one has to use cut elimination for that. Second, the rule for instantiating existentially quantified variables can not be restricted to instantiating with only one function symbol. The adoption of deep inference instead allows this.

When proving in the system we obtain, the only infinity that remains is in the unboundedness of the proofs themselves, every other aspect in proof construction is finite: at any given step, there are finitely many inferences possible, and each inference rule can only be applied in a finite number of different ways. Also, the consistency of the system is easily shown.

The point we make in this paper is not so much the existence of the finitely generating system that we show, but the simplicity of the techniques that are used to obtain it, which are purely syntactic and much less complex than cut elimination.

The notion of a finitely generating inference rule is closely related to that of an *analytic* rule, cf. Smullyan [12]. An analytic rule is one that obeys the subformula property. We tend to think of the notion of being finitely generating as a more general, weaker subformula property: there are interesting rules that are finitely generating but do not obey the subformula property, for example in system GS4ip, cf. Dyckhoff [6]. However, not all analytic rules are finitely generating, as witnessed by the existential-right rule. This is due to the fact that analyticity is defined with respect to the notion of Gentzen subformula (where instances of subformulae count as subformulae), rather than the literal notion of subformula.

In previous work, Brünnler and Tiu proved that classical logic can be presented in the calculus of structures in such a way that applying a rule only requires a bounded effort [1, 3]. This paper improves on that result by bounding choice.

In Section 2 we introduce first order logic in the calculus of structures and in Section 3 we show how to eliminate infinitely generating rules and we show the consistency argument.

2 First Order Logic in the Calculus of Structures

Variables are denoted by x and y. *Terms*, denoted by τ, are defined as usual in first-order predicate logic. *Atoms*, denoted by a, b, etc., are expressions of the form $p(\tau_1, \ldots, \tau_n)$, where p is a *predicate symbol* of *arity* n and $\tau_1, \ldots, \tau_n$ are terms. The negation of an atom is again an atom.

The *structures* of the language KSq are generated by

$$S ::= \mathsf{f} \mid \mathsf{t} \mid a \mid [\underbrace{S, \ldots, S}_{>0}] \mid (\underbrace{S, \ldots, S}_{>0}) \mid \exists x S \mid \forall x S \mid \bar{S} \quad .$$

where t and f are the units *true* and *false*, $[S_1, \ldots, S_h]$ is a *disjunction*, $(S_1, \ldots, S_h)$ is a *conjunction*, $\exists$ is the *existential quantifier* and $\forall$ is the

Associativity

$$[\vec{R},[\vec{T}],\vec{U}] = [\vec{R},\vec{T},\vec{U}]$$
$$(\vec{R},(\vec{T}),\vec{U}) = (\vec{R},\vec{T},\vec{U})$$

Commutativity

$$[R,T] = [T,R]$$
$$(R,T) = (T,R)$$

Units

$$(\mathsf{f},\mathsf{f}) = \mathsf{f} \qquad [\mathsf{f},R] = R$$
$$[\mathsf{t},\mathsf{t}] = \mathsf{t} \qquad (\mathsf{t},R) = R$$

Negation

$$\bar{\mathsf{f}} = \mathsf{t}$$
$$\bar{\mathsf{t}} = \mathsf{f}$$
$$\overline{[R,T]} = (\bar{R},\bar{T})$$
$$\overline{(R,T)} = [\bar{R},\bar{T}]$$
$$\overline{\exists x R} = \forall x \bar{R}$$
$$\overline{\forall x R} = \exists x \bar{R}$$
$$\bar{\bar{R}} = R$$

Context Closure

$$\text{if} \quad R = T \quad \text{then} \quad \begin{array}{c} S\{R\} = S\{T\} \\ \bar{R} = \bar{T} \end{array}$$

Vacuous Quantifier

if y is not free in R then $\forall y R = \exists y R = R$

Variable Renaming

$$\text{if } y \text{ is not free in } R \quad \text{then} \quad \begin{array}{c} \forall x R = \forall y R[x/y] \\ \exists x R = \exists y R[x/y] \end{array}$$

Figure 1: Syntactic equivalence of structures

universal quantifier. $\bar{S}$ is the *negation* of the structure S. The units are not atoms. Structures are denoted by S, R, T, U and V. *Structure contexts*, denoted by $S\{\ \}$, are structures with one occurrence of $\{\ \}$, the *empty context* or *hole*, that does not appear in the scope of a negation. $S\{R\}$ denotes the structure obtained by filling the hole in $S\{\ \}$ with R. We drop the curly braces when they are redundant: for example, $S[R,T]$ stands for $S\{[R,T]\}$. Structures are *equivalent* modulo the smallest equivalence relation induced by the axioms shown in Fig. 1, where $\vec{R}$ and $\vec{T}$ are finite, non-empty sequences of structures. In general we do not distinguish between equivalent structures.

An *inference rule* is a scheme of the kind $\rho\,\dfrac{S\{T\}}{S\{R\}}$, where ρ is the *name* of the rule, $S\{T\}$ is its *premise* and $S\{R\}$ is its *conclusion*. A (*formal*) *system* $\mathcal{S}$ is a set of inference rules. The *dual* of a rule is obtained by exchanging premise and conclusion and replacing each connective by its De Morgan dual.

A *derivation* Δ is a finite chain of instances of inference rules:

$$\begin{array}{r@{\,}c} \pi' & \dfrac{T}{V} \\ \pi & \overline{} \\ & \vdots \\ \rho' & \overline{} \\ & U \\ \rho & \overline{R} \end{array} \quad .$$

A derivation can consist of just one structure. The topmost structure in a derivation is called the *premise* of the derivation, and the structure at the bottom is called its *conclusion*. A derivation Δ whose premise is T, whose conclusion is R, and whose inference rules are in $\mathcal{S}$ will be indicated with $\begin{array}{c}T\\ \Delta\Vert\mathcal{S}\\ R\end{array}$. A *proof* Π in the calculus of structures is a derivation whose premise is the unit true. It will be denoted by $\begin{array}{c}\Pi\,\overline{\Vert}\,\mathcal{S}\\ R\end{array}$. A rule ρ is *derivable* for a system $\mathcal{S}$ if for every instance of $\rho\,\dfrac{T}{R}$ there is a derivation $\begin{array}{c}T\\ \Vert\mathcal{S}\\ R\end{array}$. A rule ρ is *admissible* for a system $\mathcal{S}$ if for every proof $\begin{array}{c}\overline{\Vert}\,\mathcal{S}\cup\{\rho\}\\ S\end{array}$ there is a proof $\begin{array}{c}\overline{\Vert}\,\mathcal{S}\\ S\end{array}$.

Besides deep inference, the calculus of structures employs a notion of top-down symmetry for derivations. Symmetry makes possible to reduce the cut rule to its atomic form without performing cut elimination: this would be impossible by solely adopting deep inference. The *dual* of a derivation is obtained by flipping it upside-down, negating each structure in it, and replacing each rule by its dual.

System SKSgq, shown in Fig. 2, has been introduced and shown to be sound and complete for classical predicate logic in [1]. The first S stands for "symmetric" or "self-dual", meaning that for each rule, its dual (or contrapositive) is also in the system. The K stands for "klassisch" as in Gentzen's LK and the second S says that it is a system in the calculus of structures.

$$\mathsf{i}{\downarrow}\,\frac{S\{\mathsf{t}\}}{S[R,\bar{R}]} \qquad \mathsf{i}{\uparrow}\,\frac{S(R,\bar{R})}{S\{\mathsf{f}\}}$$

$$\mathsf{s}\,\frac{S([R,U],T)}{S[(R,T),U]}$$

$$\mathsf{u}{\downarrow}\,\frac{S\{\forall x[R,T]\}}{S[\forall xR,\exists xT]} \qquad \mathsf{u}{\uparrow}\,\frac{S(\exists xR,\forall xT)}{S\{\exists x(R,T)\}}$$

$$\mathsf{w}{\downarrow}\,\frac{S\{\mathsf{f}\}}{S\{R\}} \qquad \mathsf{w}{\uparrow}\,\frac{S\{R\}}{S\{\mathsf{t}\}}$$

$$\mathsf{c}{\downarrow}\,\frac{S[R,R]}{S\{R\}} \qquad \mathsf{c}{\uparrow}\,\frac{S\{R\}}{S(R,R)}$$

$$\mathsf{n}{\downarrow}\,\frac{S\{R[x/\tau]\}}{S\{\exists xR\}} \qquad \mathsf{n}{\uparrow}\,\frac{S\{\forall xR\}}{S\{R[x/\tau]\}}$$

Figure 2: System SKSgq

The g is for “general” (as opposed to atomic) contraction. The q denotes (first-order) quantifiers.

The first and last column show the rules that deal with quantifiers, in the middle there are the rules for the propositional fragment. The propositional rules $\mathsf{i}{\downarrow}, \mathsf{s}, \mathsf{w}{\downarrow}$ and $\mathsf{c}{\downarrow}$ are called respectively *identity*, *switch*, *weakening* and *contraction*. The rule $\mathsf{u}{\downarrow}$ is called *universal*, because it roughly corresponds to the $\mathsf{R}\forall$ rule in sequent systems, while $\mathsf{n}{\downarrow}$ is called *instantiation*, because it corresponds to $\mathsf{R}\exists$.

In the sequent calculus, going up, the $\mathsf{R}\forall$ rule removes a universal quantifier from a formula to allow other rules to access this formula. In system SKSgq, inference rules apply deep inside formulae, so there is no need to remove the quantifier. Note that the premise of the u↓ rule implies its conclusion, which is not true for the R∀ rule of the sequent calculus. In all rules of SKSgq the premise implies the conclusion.

As usual, the substitution operation in the rules n↓ and n↑ requires τ to be free for x in R: quantifiers in R do not capture variables in τ. The term τ is not required to be free for x in $S\{R\}$: quantifiers in S may capture variables in τ.

The dual of rule carries the same name prefixed with a "co-", so e.g. w↑ is called *co-weakening*. The rule s is self-dual. The rule i↑ is special, it is called *cut*. Rules with a name that contains an arrow pointing downward are called *down-rules* and their duals are called *up-rules*. The system enjoys cut elimination: all up-rules are admissible, as has been shown in [1].

Sequent calculus derivations easily correspond to derivations in system SKSgq. For instance, the cut of sequent systems in Gentzen-Schütte form [16]:

$$\mathsf{Cut}\,\frac{\vdash \Phi, A \quad \vdash \Psi, \bar{A}}{\vdash \Phi, \Psi} \qquad \text{corresponds to} \qquad \mathsf{i}\!\uparrow \frac{\mathsf{s}\,\dfrac{\mathsf{s}\,\dfrac{([\Phi, A], [\Psi, \bar{A}])}{[\Phi, (A, [\Psi, \bar{A}])]}}{[\Phi, \Psi, (A, \bar{A})]}}{[\Phi, \Psi]} \quad .$$

3 Eliminating Infinite Choice in Inference Rules

There are several rules with infinite choice in system SKSgq: the co-weakening, the cut, the instantiation and the co-instantiation rule. The equivalence on structures is infinitely generating as well: equivalence classes are infinite. In the following we will see for each of these rules and the equivalence how to to replace them by finitely generating rules without affecting provability.

3.1 The Co-weakening Rule

The rule w↑ is clearly infinitely generating, since there is an infinite choice of atoms, but it can immediately be eliminated by using a cut and an instance of w↓ as follows:

$$\mathsf{w}{\uparrow}\,\frac{S\{a\}}{S\{\mathsf{t}\}} \quad \leadsto \quad \mathsf{s}\,\cfrac{S\{R\}}{\mathsf{s}\,\cfrac{S(R,[\mathsf{t},\mathsf{f}])}{\mathsf{w}{\downarrow}\,\cfrac{S[\mathsf{t},(R,\mathsf{f})]}{\mathsf{i}{\uparrow}\,\cfrac{S[\mathsf{t},(R,\bar{R})]}{=\,\cfrac{S[\mathsf{t},\mathsf{f}]}{S\{\mathsf{t}\}}}}}} \quad .$$

3.2 The Cut Rule

The cut is the most prominent infinitely generating rule. The first source of infinite choice we will remove is the arbitrary size of the cut formula. To that end, consider the *atomic cut* rule:

$$\mathsf{ai}{\uparrow}\,\frac{S(a,\bar{a})}{S\{\mathsf{f}\}}$$

The following theorem, also proved in [1], allows us to restrict ourselves to atomic cuts.

Theorem 1. *The rule* $\mathsf{i}{\uparrow}$ *is derivable for* $\{\mathsf{ai}{\uparrow},\mathsf{s},\mathsf{u}{\uparrow}\}$.

Proof. By an easy structural induction on the structure that is cut. A cut introducing the structure (R,T) together with its dual structure $[\bar{R},\bar{T}]$ is replaced by two cuts on smaller structures:

$$\mathsf{i}{\uparrow}\,\frac{S(R,T,[\bar{R},\bar{T}])}{S\{\mathsf{f}\}} \quad \leadsto \quad \mathsf{s}\,\cfrac{S(R,T,[\bar{R},\bar{T}])}{\mathsf{s}\,\cfrac{S(R,[\bar{R},(T,\bar{T})])}{\mathsf{i}{\uparrow}\,\cfrac{S[(R,\bar{R}),(T,\bar{T})]}{\mathsf{i}{\uparrow}\,\cfrac{S(R,\bar{R})}{S\{\mathsf{f}\}}}}} \quad .$$

A cut introducing the structure $\forall xR$ together with its dual structure $\exists x\bar{R}$ is replaced by a cut inside an existential quantifier followed by an instance of $\mathsf{u}{\uparrow}$:

$$\text{i}\!\uparrow \frac{S(\forall x R, \exists x \bar{R})}{S\{\mathsf{f}\}} \quad \rightsquigarrow \quad = \frac{\text{i}\!\uparrow \dfrac{\text{u}\!\uparrow \dfrac{S(\forall x R, \exists x \bar{R})}{S\{\exists x (R, \bar{R})\}}}{S\{\exists x \mathsf{f}\}}}{S\{\mathsf{f}\}} \quad .$$

These reductions can be repeated until all cuts are atomic. □

The rule ai↑ still is infinitely generating, since there is an infinite choice of atoms. Let us take a closer look at the atoms:

$$\text{ai}\!\uparrow \frac{S(p(\tau_1, \dots, \tau_n), \overline{p(\tau_1, \dots, \tau_n)})}{S\{\mathsf{f}\}} \quad .$$

There are both an infinite choice of predicate symbols p and an infinite choice of terms for each argument of p. Let $\vec{\tau}$ denote $\tau_1, \dots, \tau_n$ and $\vec{x}$ denote $x_1, \dots, x_n$. Since cuts can be applied inside existential quantifiers, we can delegate the choice of terms to a sequence of n↓ instances:

$$\text{ai}\!\uparrow \frac{S(p(\vec{\tau}), \overline{p(\vec{\tau})})}{S\{\mathsf{f}\}} \quad \rightsquigarrow \quad = \frac{\text{ai}\!\uparrow \dfrac{\text{n}\!\downarrow^n \dfrac{S(p(\vec{\tau}), \overline{p(\vec{\tau})})}{S\{\exists \vec{x}(p(\vec{x}), \overline{p(\vec{x})})\}}}{S\{\exists \vec{x} \mathsf{f}\}}}{S\{\mathsf{f}\}} \quad .$$

The remaining cuts are restricted in that they do not introduce arbitray terms but just existential variables. Let us call this restricted form vai↑:

$$\text{vai}\!\uparrow \frac{S(p(\vec{x}), \overline{p(\vec{x})})}{S\{\mathsf{f}\}} \quad .$$

The only infinite choice that remains is the one of the predicate symbol p. To remove it, consider the rule *finitely generating atomic cut*

$$\text{fai}\!\uparrow \frac{S(p(\vec{x}), \overline{p(\vec{x})})}{S\{\mathsf{f}\}} \quad \text{where } p \text{ appears in the conclusion.}$$

This rule is finitely generating, and we will show that we can easily transform a proof into one where the only cuts that appear are fai↑ instances.

Take a proof in the system we obtained so far, that is SKSgq without w↑, and with vai↑ instead of i↑. Individuate the bottommost instance of vai↓ that violates the proviso of fai↑:

$$\mathsf{vai}\uparrow \frac{\overset{\displaystyle\mathbb{T}}{\displaystyle S(p(\vec{x}), \overline{p(\vec{x})})}}{S\{\mathsf{f}\}} \quad ,$$

where p does not appear in $S\{\mathsf{f}\}$. We can then replace all instances of $p(\vec{x})$ and $\overline{p(\vec{x})}$ in the proof above the cut with t and f, respectively, to obtain a proof of $S\{\mathsf{f}\}$. It is easy to check that all rule instances stay valid or become trivial; the cut

$$\mathsf{vai}\uparrow \frac{S(\mathsf{t}, \mathsf{f})}{S\{\mathsf{f}\}} \quad ,$$

can just be removed, since $(\mathsf{t}, \mathsf{f}) = \mathsf{f}$.

Please notice that if p appeared in $S\{\mathsf{f}\}$, then this would not work, because it could destroy the rule instance below $S\{\mathsf{f}\}$.

Proceeding inductively upwards, we remove all infinitely generating atomic cuts.

3.3 The Instantiation Rule

The same techniques also work for instantiation. Consider these two restricted versions of n↓:

$$\mathsf{n}\downarrow_1 \frac{S\{R[x/f(\vec{x})]\}}{S\{\exists x R\}} \qquad \text{and} \qquad \mathsf{n}\downarrow_2 \frac{S\{R[x/y]\}}{S\{\exists x R\}} \quad .$$

An instance of n↓ that is not an instance of $\mathsf{n}\downarrow_2$ can inductively be replaced by instances of $\mathsf{n}\downarrow_1$ (chose variables for $\vec{x}$ that are not free in R):

$$\mathsf{n}\downarrow \frac{S\{R[x/f(\vec{\tau})]\}}{S\{\exists x R\}} \quad \leadsto \quad = \frac{\mathsf{n}\downarrow_1 \dfrac{\mathsf{n}\downarrow^n \dfrac{S\{R[x/f(\vec{\tau})]\}}{S\{\exists \vec{x} R[x/f(\vec{x})]\}}}{S\{\exists \vec{x} \exists x R\}}}{S\{\exists x R\}} \quad .$$

This process can be repeated until all instances of $\mathsf{n}\downarrow$ are either instances of $\mathsf{n}\downarrow_1$ or $\mathsf{n}\downarrow_2$.

Now consider the following rules, which are finitely generating

$$\mathsf{fn}\downarrow_1 \frac{S\{R[x/f(\vec{x})]\}}{S\{\exists xR\}} \qquad \text{and} \qquad \mathsf{fn}\downarrow_2 \frac{S\{R[x/y]\}}{S\{\exists xR\}} \qquad .$$

where $\mathsf{fn}\downarrow_1$ carries the proviso that the function symbol f either occurs in the conclusion or is a fixed constant c, and $\mathsf{fn}\downarrow_2$ carries the proviso that the variable y appears in the conclusion (no matter whether free or bound or in a vacuous quantifier).

Infinitely generating instances of $\mathsf{n}\downarrow_1$ and $\mathsf{n}\downarrow_2$, i.e. those that are not instances of $\mathsf{fn}\downarrow_1$ and $\mathsf{fn}\downarrow_2$, respectively, are easily replaced by instances of finitely generating rules similarly to how the infinitely generating cut was eliminated. Take the constant symbol c that is fixed in the proviso of $\mathsf{fn}\downarrow_1$, and throughout the proof above an infinitely generating instance of $\mathsf{n}\downarrow_1$, replace all terms that are instances of $f(\vec{x})$ by c. For $\mathsf{n}\downarrow_2$ we do the same to all occurences of y, turning it into an instance of $\mathsf{fn}\downarrow_1$.

3.4 The Equivalence and the Co-instantiation Rule

The equivalence can be broken up into several rules, for associativity, commutativity, and so on. Those rules are clearly finitely generating except for variable renaming and vacuous quantifier, which, technically speaking, have an infinite choice of names for bound variables. The same goes for the co-instantiation rule. Of course these rules can be made finitely generating since the choice of names of bound variables does not matter. There is nothing deep in it: the only reason for us to tediously show this obvious fact is to avoid giving the impression that we hide infinity under the carpet of the equivalence. The need for the argument below just comes from a syntax which has infinitely many different objects for essentially the same thing, e.g. $\forall xp(x), \forall yp(y)$ and $\forall y \forall xp(x)$ .... If you are not concerned about this 'infinite' choice of names of bound variables, then please feel invited to skip ahead to the next section.

Consider the following rules for variable renaming and vacuous quantifier, they all carry the proviso that x is not free in R:

$$\alpha\downarrow \frac{S\{\forall x R[y/x]\}}{S\{\forall y R\}} \qquad \alpha\uparrow \frac{S\{\exists x R[y/x]\}}{S\{\exists y R\}}$$

$$\mathsf{v}\downarrow \frac{S\{\exists x R\}}{S\{R\}} \qquad \mathsf{v}\uparrow \frac{S\{R\}}{S\{\forall x R\}}$$

Let us now consider proofs in the system that is obtained from SKSq by restricting the equivalence rule to not include vacuous quantifier and variable renaming and by adding the above rules. This system is strongly equivalent to SKSq as can easily be checked.

The rule $\mathsf{v}\uparrow$ is clearly finitely generating. Let us see how to replace the rule $\mathsf{v}\downarrow$ by finitely generating rules, the same technique also works for the rules $\alpha\uparrow$ and $\alpha\downarrow$. Consider the finitely generating rule $\mathsf{fv}\downarrow_1$, which is $\mathsf{v}\downarrow$ with the added proviso that x occurs in the conclusion (no matter whether bound or free or in a vacuous quantifier) and the infinitely generating rule $\mathsf{v}\downarrow'$ which is $\mathsf{v}\downarrow$ with the proviso that x does not occur in the conclusion.

Fix a total order on variables. Let $\mathsf{fv}\downarrow_2$ be $\mathsf{v}\downarrow$ with the proviso that x is the lowest variable in the order that does not occur in the conclusion. This rule is clearly finitely generating: there is no choice.

Each instance of $\mathsf{v}\downarrow$ is either an instance of $\mathsf{fv}\downarrow_1$ or of $\mathsf{v}\downarrow'$. In a given proof, all instances of $\mathsf{v}\downarrow'$ can be replaced by instances of $\mathsf{fv}\downarrow_2$ as follows, as we we will see now. Starting from the conclusion, going up in the proof, identify the first infinitely generating vacuous quantifier rule:

$$\begin{array}{c} \top \\ \Vert \\ \mathsf{v}\downarrow' \dfrac{S\{\exists x R\}}{S\{R\}} \quad x \text{ does not occur in } S\{R\} \\ \Vert \\ T \end{array} \quad ,$$

where x is not the lowest in our fixed order that does not occur in the conclusion. Let y be the lowest variable that does not occur in the conclusion. Now, throughout the proof above, do the following:

1. Choose a variable z that does not occur in the proof. Replace y by z.

2. Replace x by y.

By definition neither x nor y occur in the conclusion, so the conclusion is not broken. All the replacements respect that variable occurrences with different names stay different and variable occurrences with the same names stay the same. So the proof above stays intact. Replace the $\mathsf{v}{\downarrow}'$ instance by a $\mathsf{fv}{\downarrow}_2$ instance and proceed inductively upwards.

4 A Finitely Generating System for Predicate Logic

We now define the finitely generating system FKSgq to be

$$(\mathsf{SKSgq} \setminus \{\mathsf{i}{\uparrow}, \mathsf{w}{\uparrow}, \mathsf{n}{\downarrow}\}) \cup \{\mathsf{fai}{\uparrow}, \mathsf{fn}{\downarrow}_1, \mathsf{fn}{\downarrow}_2\} \quad ,$$

and, for what we showed in the previous section, state

Lemma 1. *Each structure is provable in system* SKSgq *if and only it is provable in system* FKSgq.

To put the finitely generating system at work, we use it to show consistency of system SKSgq. Of course, for this purpose it suffices to have finitely generating cut. Having infinite choice in instantiation would not affect the following argument.

Lemma 2. *The unit* f *is not provable in system* FKSgq.

Proof. No atoms, but only f, t and vacuous quantifiers can appear in such a proof. It is easy to show that f is not equivalent to t. Then we show that no rule can have a premise equivalent to t and a conclusion equivalent to f. This is simply done by inspection of all the rules in FKSgq. □

From the two lemmas above we immediately get consistency:

Theorem 2. *The unit* f *is not provable in system* SKSgq.

Now we can make use of symmetry by flipping derivations:

Theorem 3. *For all structures R, if there is a proof of R then there is no proof of $\bar{R}$.*

Proof. We assume that we have both proofs:

$$\overset{\top}{\underset{R}{\|}} \quad \text{and} \quad \overset{\top}{\underset{\bar{R}}{\|}} \quad .$$

Dualise the proof of R to get

$$\begin{array}{c}\bar{R}\\ \Vert \\ \mathsf{f}\end{array} \quad ,$$

and compose this derivation with the proof of $\bar{R}$ to get a proof of f, which is in contradiction to Theorem 2:

$$\begin{array}{c}\overline{\Vert}\\ \bar{R}\\ \Vert \\ \mathsf{f}\end{array} \quad .$$

□

5 Conclusion

In this paper we showed simple proof theoretical techniques for making a system of first order classical logic finitely generating. We believe that these considerations help make clear that finite choice and cut elimination, or other normalisation techniques, are conceptually independent.

Some of the techniques we used, for example the replacement of an atom and its dual by t and f, are folklore. However, in order to produce a finitely generating system they have to be combined with the reduction of the cut rule to its atomic form. This crucial ingredient is provided by deep inference and top-down symmetry, which are not available in the sequent calculus.

In the calculus of structures, there are presentations of various modal logics [13], linear logic [14, 15] and various extensions of it [9, 10, 4] and noncommutative logics [5]. All these systems are similar to system SKSgq in that they include rules which follow a scheme [8], which ensures atomicity of cut and identity. The techniques shown here thus also work for these logics.

6 Bibliography

[1] Kai Brünnler. Locality for classical logic. Technical Report WV-02-15, Dresden University of Technology, 2002. Available at http://www.wv.inf.tu-dresden.de/~kai/LocalityClassical.pdf.

[2] Kai Brünnler. Atomic cut elimination for classical logic. In M. Baaz and J. A. Makowsky, editors, *CSL 2003*, volume 2803 of *Lecture Notes in Computer Science*, pages 86–97. Springer-Verlag, 2003.

[3] Kai Brünnler and Alwen Fernanto Tiu. A local system for classical logic. In R. Nieuwenhuis and A. Voronkov, editors, *LPAR 2001*, volume 2250 of *Lecture Notes in Artificial Intelligence*, pages 347–361. Springer-Verlag, 2001.

[4] Paola Bruscoli. A purely logical account of sequentiality in proof search. In Peter J. Stuckey, editor, *Logic Programming, 18th International Conference*, volume 2401 of *Lecture Notes in Artificial Intelligence*, pages 302–316. Springer-Verlag, 2002.

[5] Pietro Di Gianantonio. Structures in cyclic linear logic. Technical report, Università di Udine, 2003.

[6] Roy Dyckhoff. Contraction-free sequent calculi for intuitionistic logic. *The Journal of Symbolic Logic*, (57):795–807, 1992.

[7] Gerhard Gentzen. Investigations into logical deduction. In M. E. Szabo, editor, *The Collected Papers of Gerhard Gentzen*, pages 68–131. North-Holland Publishing Co., Amsterdam, 1969.

[8] Alessio Guglielmi. Recipe. Manuscript. http://www.wv.inf.tu-dresden.de/˜guglielm/Research/Notes/AG2.pdf, 2002.

[9] Alessio Guglielmi. A system of interaction and structure. Technical Report WV-02-10, Technische Universität Dresden, 2002. Available at http://www.wv.inf.tu-dresden.de/˜guglielm/Research/Gug/Gug.pdf.

[10] Alessio Guglielmi and Lutz Straßburger. A non-commutative extension of MELL. In Matthias Baaz and Andrei Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning, LPAR 2002*, volume 2514 of *LNAI*, pages 231–246. Springer-Verlag, 2002.

[11] J. Herbrand. *Recherches sur la théorie de la démonstration*. PhD thesis, Université de Paris, 1930.

[12] Raymond M. Smullyan. Analytic cut. *The Journal of Symbolic Logic*, 33:560–564, 1968.

[13] Charles Stewart and Phiniki Stouppa. A systematic proof theory for several modal logics. Technical Report WV-03-08, Technische Universität Dresden, 2003. Submitted.

[14] Lutz Straßburger. A local system for linear logic. In Matthias Baaz and Andrei Voronkov, editors, *Logic for Programming, Artificial Intelligence, and Reasoning, LPAR 2002*, volume 2514 of *LNAI*, pages 388–402. Springer-Verlag, 2002.

[15] Lutz Straßburger. MELL in the Calculus of Structures. *Theoretical Computer Science*, 309(1–3):213–285, 2003.

[16] Anne Sjerp Troelstra and Helmut Schwichtenberg. *Basic Proof Theory.* Cambridge University Press, 1996.

Predicate Logic, Predicates, and Terms

George Englebretsen
genglebr@ubishops.ca

My primary aim here is to introduce in a very preliminary way a system of formal logic that has been built by Fred Sommers and myself over the past few years. This Term Logic matches the inferential power of the standard first-order Predicate Logic, but enjoys certain advantages in terms of simplicity and naturalness. What I hope this can offer is some insight into ideas concerning formal logic that are extremely old but not often encountered today. I may rightly be accused of atavism for touting such antiques, but perhaps the contrast between these ideas and more contemporary ones will be of interest. So, some of my remarks will concern some central logical concepts (especially the concept of predication), while others will be a bit historical.

The very title of this conference, "75 Years of Predicate Logic," and its recognition of the anniversary of Hilbert and Ackermann's great work, inevitably puts one in a historical frame of mind. I shall begin with a few historical remarks.

1 Historical Remarks

A century and a half ago De Morgan wrote:

> We know that mathematicians care no more for logic than logicians for mathematics. The two eyes of exact science are mathematics and logic: the mathematical sect puts out the logical eye, the logical sect puts out the mathematical eye, each believing that it can see better with one eye than with two.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 75–88

That was a long time ago. Before Boole no one, with the obvious exception of Leibniz, thought much about the connection – if any – between logic and mathematics. But by the time of Boole and De Morgan the algebraic logicians were contending that logic was a branch of mathematics (indeed, a branch of algebra). Soon after, the logicists were turning the tables, insisting that mathematics (or much of it), if not a branch of logic, was at least founded on logic. Since 1931 logicism has lost much of its steam. But long before that the new logic (Predicate Logic) had supplanted the old logic in all of its old guises – including algebraic. Today, it seems to me, logicians and mathematicians don't spend much time poking out each other's eyes. But the exact nature of the relation between logic and mathematics is still not clear. De Morgan had called them "the two eyes of exact science." And that sounds pretty much correct. Perhaps they are just the two branches of a more general discipline – the science of reckoning.

Back in the '60s, when I began teaching logic, students had been more thoroughly instructed in Grammar [it seems to me that Rhetoric holds the premier position in the trivium today]. They knew that the natural complement of the term 'predicate' was 'subject', and would sometimes ask the whereabouts of the Subject in Predicate Logic. Of course there are expressions that play the grammatical role of subject in Predicate Logic – individual constants and variables. But so-called Subject-Predicate Logic, the logic of the Schoolmen, of Leibniz, of the algebraic logicians from Boole to Schröder and Peirce, had been thoroughly replaced by a logic giving pride of place to the predicate alone. Frege had written: "I believe that the replacement of the concepts of *subject* and *predicate* by *argument* and *function*, respectively, will stand the test of time." The smart money, it turned out, was on Frege. Russell impeached Leibniz's entire philosophy because he took it to rest on Leibniz's logic – a logic that still recognized subjects and predicates, a Subject-Predicate Logic. How was it that the old Subject-Predicate Logic gave way to the new Predicate Logic? And Why?

Well, the answer to the second question is easy. The new logic could do far more, and usually do it far more efficiently, than the old. The old logic was bad at offering any insight into the logic of inferences involving singular terms, compound propositions, or relationals. The latter require an adequate representation of propositions of so-called multiple generality, and, as Fregeans like Geach and Dummett are fond of pointing out, the best the old logic could do here was the cumbersome Scholastic semantic theory of supposition. So, on these grounds, the victory of the new over the old was easy. The answer to the first question is a longer, more complex story. Its historical outlines are familiar, and have been recounted in detail elsewhere, so I need only briefly highlight them here – the rest of the answer will come later.

Philosophers (not to mention mathematicians, linguists, cognitive psychologists, and others) have been interested in formal logic for a variety of reasons. Attempts to articulate a formal logic can throw light on how we do, or should, reason, might reveal the foundations of mathematics (or at the very least can illuminate basic mathematical concepts and procedures), could (once fitted with a practicable algorithm) ease the burden of logical reckoning, and so forth. Aristotle, who started it all, laid out the principles governing correct argumentation (from one or two categorical premises to a categorical conclusion) – syllogistic – as a tool to be used for the teaching and doing of theoretical sciences (physics, mathematics, metaphysics, for him). He said a little, but not much, about other forms of argumentation and other kinds of propositions. The Stoics, though soon forgotten, took on the task of formalizing the rules governing arguments involving compound propositions – truth-functional logic as it is usually called today. Fusing, and sometimes confusing, Aristotelian and Stoic insights, the Scholastic logicians codified formal logic and added a number of their own insights concerning both logical syntax and semantics. In the 17^{th} century Leibniz made a number of attempts to use the tools of mathematics to formalize a symbolic algorithm for a logic that was essentially syllogistic. Two centuries later most of his ideas were rediscovered by the algebraic logicians and by Frege. It is the logic of Frege, filtered through Russell and Whitehead, Hilbert, Gentzen, Gödel, Quine, and others that we teach today.

2 Formatives

Logicians recognize a fundamental distinction between expressions that have some "material" sense and those that merely determine the form of more complex expressions in which they occur. For the Scholastics this was the categoremata /syncategoremata distinction – the constant/variable distinction today. However, defining or characterizing logical constants (or formatives, as I shall call them) so as to draw this distinction sharply and objectively has posed a challenge for post-Fregeans. Logicians such as Russell, Tarski, Quine, and many others have conceded that the best one can do is simply enumerate those expressions that one recognizes as determining logical form. The task of defining or characterizing formatives is simply abandoned. Thus Quine wrote that "a morpheme is a particle or a lexical element according as there are fewer or many expressions in its grammatical category" ([6], pp. 18–19). Some logicians, on the other hand, have claimed that a definition (of sorts) can be given by Gentzen-like rules for the introduction or elimination of formatives in proofs. The idea is that one begins with a simple formal language with no formatives (perhaps consisting of just Tractarian atomic formulae) and then formulates rules of proof that introduce or eliminate each formative. But there remains a question about whether or not this process actually reveals anything about the nature of the formatives rather than about the nature of proof.

If you believe, as I do, that logic is the science of how we ideally reason and express our reasoning in natural language, then whatever one might offer as a characterization of formatives must reflect how their natural language analogues actually work when properly used. A number of traditional logicians had a good idea of this. Hobbes thought reasoning should be a kind of calculating in which ideas are added or subtracted; Leibniz thought all formatives could be defined in terms of copulae ('is'/'isn't') and could be symbolized by plus and minus; De Morgan, too, held that all formatives amount to algebraic operators of addition and subtraction. Many years after Boole's death his widow wrote of his aim in formulating his logic: "...to express ordinary statements about facts in some sort of arithmetical or algebraic notation so as to be able to work out the logical consequences of premises with the same ease as we work sums" (quoted in [3], 59). The idea that formatives share some unique characteristic and the idea that one can thus build a symbolic algorithm in which logical reckoning becomes a matter of algebraic addition or subtraction have been carried out to the fullest extent by the American logician Fred Sommers. His version of formal logic is Term Logic. It takes seriously the notion that all logical formatives are oppositional in nature. And, just as in arithmetic or algebra, there is both a unary and a binary version of each.

3 Logical Copulation

To over-simplify, logic began twice – first with Aristotle and then with Frege. Traditional logic was a Term Logic; modern logic is Predicate Logic. When

Frege began his logical researches, seeking a way to place arithmetic on the firmer foundation of logic, he looked at the logic of Boole and the algebraists. But he soon concluded that such a logic was flawed and limited in a number of ways. It was plagued by what he termed "psychologism," it was too closely tied to natural language with all its ambiguity and untidiness, it mistakenly took terms to be logically prior to sentences, it made use of mathematical concepts and notations, rendering it inappropriate for use as a foundation for mathematics, and, of course, it was burdened by the old-fashioned notion that sentences should be analyzed in terms of subjects and predicates. Frege offered an account of the logical syntax of propositions, an account of why a proposition is a logical unit rather than just a string of terms, that was far different from the logical syntax of traditional term logic.

In the *Sophist* Plato claimed that propositions result from combining nouns (*onoma*) with verbs (*rhema*). In the simplest case a noun is combined with an intransitive verb (e.g., 'Men reason'). His logical insight was gleaned from Greek grammar, and his intention was to show that in a true proposition the combining of a noun and a verb reflects the mixing of Forms (the form Man mixes with the form Rational). Both the ontological mixing and the logical combining are natural; nothing further is required. At least for the simplest cases, this *binary* theory of logical syntax is the one Frege adopted many centuries later. But where Plato looked to grammar for logical guidance, finding *onoma* and *rhema*, Frege looked to mathematics, finding *arguments* and *functions*.

Early on Aristotle seems to have accepted his teacher's binary theory of logical syntax, with its implicit explanation of propositional unity. However, once on the road to formulating the syllogistic in *Prior Analytics* he was forced to adopt a theory that would allow terms to play different logical roles in different propositions (because in any valid syllogism at least one term occurs once as a subject-term and once as a predicate-term). Grammar was of little use here. Aristotle had to look deeper. What he discovered was that any proposition fit for use in a syllogism (i.e., any categorical) had to consist of a pair of terms whose logical roles were independent of their grammatical roles. The logician could simply view them as *terms* (*horos*). But this left the question of propositional unity unsolved. The binary theory holds that propositional unity is the result of combining a pair of expressions (terms) that are just grammatically fit for one another (axe heads and axe handles forming axes). But how do two terms whose grammatical features are inert form a unified proposition? Aristotle's answer was the *logical copula*. He generally formulated propositions with a pair of terms flanking an expression that was itself not a term – it was a formative. Its job was to mediate between the two terms, tie them together to form a unified proposition. There were four such formatives: 'belongs to some', 'belongs to no', 'belongs to every', and 'does not belong to some'. Aristotle's formatives acted as the glue that binds pairs of terms to form a logical unit. His theory of logical syntax is thus *ternary*, analyzing a proposition as a pair of terms and a logical copula. The ternary theory made syllogistic possible.

The Scholastic logicians had a field day in the land of syllogistic, amending and emending at will while preserving Aristotle's core insights. The most obvious change they made was this: Aristotle's propositions consisted of a pair of terms at the two ends (literally *termini*) of the proposition with a copula in the middle. The Scholastics recognized first that this was not only awkward in both Greek and Latin (and English, German, French, etc.), but that the copula was providing the proposition with two features simultaneously – quantity and quality. Their solution was to

1. split the copula into two parts (e.g., 'wisdom belongs to / some men'),

2. reorder the results ('some men / wisdom belongs to'),

3. reorder the second part ('some men / belongs to wisdom'),

4. replace the second half of the now split copula with a grammatically appropriate version of the verb 'to be', which Abelard had first called a 'logical copula' anyway ('some men / are wisdom'), and then

5. put the terms in the appropriate cases ('some men are wise').

On this analysis the two fragments of the logical copula are the quantifier and the qualifier, each attached to its own term. The first became the subject-term and the second became the predicate-term. The quantifier plus subject-term became the Subject; the qualifier plus predicate-term became the Predicate. This *quaternary* theory of logical syntax is the heart of so-called Subject-Predicate Logic. It is important to recognize however that the quaternary analysis is just a version of the ternary one. Even though the logical copula of the ternary theory has been *split*, its two fragments still constitute a single copula.

4 Frege's Theory of Logical Syntax

Frege saw a sharp and unbridgeable chasm between function expressions and arguments. The lowest level function expresssions are *predicates*; arguments are *names* (singular terms, including proper names, personal pronouns, and definite descriptions). The *Bedeutung* (reference or even supposition perhaps) of a predicate is a *concept*; the *Bedeutung* of a name is an object. The asymmetry here is complete and absolute. No concept is ever an object; no object is ever a concept (not even the concept of a horse). And likewise for their linguistic counterparts. In the logically simplest cases of propositions, so-called atomic propositions, only predicates can occur in predicate positions, only names can occur in subject positions. Geach has said it quite clearly:

> It is logically impossible for a term to shift about between subject and predicate position without undergoing a change of sense as well as a change of role. Only a name can be a logical subject; and a name cannot retain the role of name if it becomes a logical predicate. ([2], p. 48)

But just how does one know which term in such a proposition is the predicate and which the subject? Frege's answer was that predicates are incomplete, unsaturated, while names are complete, saturated. Predicates contain (one or more) gaps; names don't. And this is how he can account for the unity of a proposition. Names just naturally happen to be fit to fill the gaps in predicates. Round pegs in round holes. No glue, no logical copula is required. A new version of Plato's binary syntax.

Geach spoke of both subjects and predicates in characterizing Frege's new theory of logical syntax, so why is that logic a Predicate Logic rather than a Subject-Predicate Logic? Predicate Logic must distinguish between propositions consisting of a single predicate (having one or more gaps) and names filling each gap – atomic propositions. If logic were to limit itself to just dealing with such propositions it might be acceptable to call it a Subject(s)-Predicate Logic. But an adequate logic must be able to handle more complex types of propositions (compound propositions and multiply quantified propositions, especially). The foundation of Predicate Logic is the *propositional calculus*. This basic but elementary part of the logic accounts for compound propositions (conjunctions, conditionals, etc.) whose constituent sub-propositions need not be analyzed into predicates and subjects. The rest of standard Predicate Logic consists of the *predicate calculus (with identity)*. Here quantifiers are introduced as higher-order function expression operating on entire propositions that normally involve both predicates and names – including pronouns (individual variables). The quantifiers *bind* those variables. The propositional calculus is basic here because the function expressions that connect sub-propositions (*propositional connectives*) are to be found in the propositions to which quantifiers are applied. Propositions of identity are treated as special since there are special restrictions on both the syntax and semantics of the identity predicate.

5 Pluses and Minuses

Predicate Logic is a very powerful system for analyzing logical reckoning. It can do all the things that traditional logic could not do or only do poorly. In particular, it can account for inferences involving singular terms, relationals, and compound propositions. It provides relatively simple tests for deciding validity and clear accounts of semantic modelling and of proof. Moreover, it is equipped with a symbolic algorithm for carrying out much of this work mechanically – the goal glimpsed by Leibniz and Boole. But in spite of these advantages it fairs poorly with traditional logic in terms of simplicity and naturalness. Traditional logic was relatively powerless to account for various kinds of inferences (especially those involving relationals), but it was both more natural and simpler than Predicate Logic. It was more natural in the sense that its account of the logical syntax of propositions was close to that of natural language. It was simpler in that it required a much smaller number of kinds of formatives and fewer kinds of rules of inference.

The challenge for any term logician is to build a logic that enjoys the power of Predicate Logic as well as the simplicity and naturalness of traditional logic. A Term Logic, such as the one devised by Sommers, aims to meet that challenge. I want to sketch out very briefly how one might begin to go about building such a system. We begin with two ideas: Aristotle's idea that propositions can be viewed as pairs of copulated terms, and Leibniz, De Morgan, and Boole's idea that logical formatives can be seen as signs of opposition. We begin with a lexicon. Frege's lexicon consisted of two parts: formatives and non-formatives. The formatives consisted of propositional connectives, quantifiers, and a sign for identity. The non-formatives consisted of predicates and names. Initially our lexicon will consist of just the two formatives, plus and minus, and our non-formatives – terms. The plus here is binary; like addition, it comes between pairs of terms to form a more complex term (called a *dyad*). It shares the formal features of addition as well: it is symmetric, and associative. The minus is unary, like the negative in arithmetic. Prefixing a minus to a term (whether simple or complex) yields a new term. Using the plus to symbolize Aristotle's 'belongs to some' and the minus to form negated terms, all the four standard categoricals can be formulated in this new language. But this language has little expressive power. As a first step to increasing its power we define two new formatives in terms of our initial two. We can define a unary plus as the negation of a negative and a binary minus as the negation of a dyad whose first term is negated. The binary minus is reflexive and transitive. The systematic ambiguity of our pluses and minuses matches their systematic ambiguity in arithmetic. Still the expressive power of the language is limited. To increase its power further we take advantage of the Scholastics' idea that logical copulae can be split. Mimicking their procedure, we introduce split versions of our binary formatives, yielding in each case a pair consisting of a quantifier and a qualifier. Now any well-formed dyad must consist of either a pair of terms flanking an unsplit binary formative or a quantified term and a qualified term (the quantifier and qualifier constituting a split copula).

For this language to match the expressive power of Predicate Logic it must be able to offer a systematic logical formulation for singular, relational, and compound propositions. And it can. Keep in mind that the non-formative lexicon is not divided into general term predicates and singular term names. The singular/general distinction is semantic and is ignored by this syntactical analysis. Singular terms are treated on a par with any other term. However, since we happen to know that a singular term denotes just one thing, we can take advantage of an idea first encountered by Leibniz and then exploited by Sommers. We can treat singular subject-terms has having, implicitly, "wild" quantity, being indifferently either universal or particular. That's why we can derive 'Some senator is a philosopher' *syllogistically* from 'Cicero is a senator' and 'Cicero is a philosopher'. Perhaps more importantly, the wild quantity of singular subjects relieves us of the need for a special "identity theory." Propositions such as 'Cicero is Tully' can be treated as straightforward categoricals with a tacit wild quantity. The formal features of our binary plus and our binary minus (both split in this case) guarantee reflexivity, symmetry, and transitivity. The Laws of Identity, that govern inferences involving the special identity *relation* are unnecessary.

The Port Royal logicians came close to showing how relationals could be accommodated by a term logic. The problem with relational propositions for Subject-Predicate Logic is that they have more than one subject, so that the proposition cannot be parsed into two parts. What Arnauld and the other Port Royalists recognized was that while any complex term is a dyad (a pair of copulated terms, or as they said, *predicated* terms), any dyad is itself a term that can be copulated, predicationally tied, to any other term. A favorite example was 'Invisible God created the visible world'. There are five simple terms here, but there are four dyads as well. 'Invisible' is predicated of 'God', 'visible' is predicated of 'world', 'created' is predicated of the complex term, the dyad, 'invisible world', and that more complex term is itself predicated of the dyad 'invisible God' to yield the entire proposition. The proposition is still a dyad. Relationals demand dyads nested in dyads. Our system exploits this idea, adding to it the recognition that in natural languages such as English or German the logical copula of a dyad using a simple relational term – a transitive verb – is usually unsplit. For example, in 'Every boy is kissing some girl' the two terms 'kissing' and 'girl' are connected by an unsplit copula (indicated in English by 'some' and symbolized by our unsplit binary plus), while the resulting dyad and 'boy' are connected by a split copula ('every ... is ... ', symbolized by our split binary minus).

The Stoics, and much later Frege, saw the importance of accounting for the logic of compound propositions whose ultimate sub-propositions are not analyzed. Traditionally, while many term logicians simply tried to ignore unanalyzed propositions, there were some who believed these could be incorporated into a logic of terms (e.g., Aristotle) or reduced to it (Leibniz). Others believed that the logic of terms and the logic of compound propositions were simply isomorphic (Peirce). Leibniz expressed clearly his desire to do this.

> If, as I hope, I can conceive all propositions as terms, and hypotheticals as categoricals, and if I can treat all propositions universally, this promises a wonderful ease in my symbolism and analysis of concepts, and will be a discovery of the greatest importance. ([4], p. 66)

He was right about what had to be done (and about its importance for the project of building a powerful term logic). First, entire propositions must be taken as terms, then compound propositions (so-called hypotheticals) must be construed as categoricals. Each of these tasks is easily done by our theory of terms and oppositional formatives. As I've said, propositions are simply dyads (pairs of copulated terms) and dyads are themselves terms. Moreover, recognizing that conjunctive propositions share the same formal features as particularly quantified categoricals, and that conditionals share the same formal features as universals, allows us to use the logical copulae already in hand (symbolized by either split or unsplit versions of our two binary formatives) to analyze compound propositions on all logical fours with categoricals. As it happens, Sommers has shown that propositions are special terms in that they are semantically singular. This makes their logic a special branch of term logic. Just how this is so is a story for another day, however.

Also a story for another day is how logical reckoning (assessing arguments for validity, assessing proposition sets for consistency, proving, etc.) turn out to be essentially matters of algebraic addition/subtraction and the application of rules such as the *dictum de omni* – just as Leibniz and Boole had invisioned.

6 Predicate-Functor Algebra

In the formal language of Predicate Logic bound variables, the formal analogues of natural language pronouns, are ubiquitous. Natural language propositions that contain quantifier expressions but no pronouns are uniformly symbolized with bound individual variables. There is more than just logic at stake in this contrast. Quine held that pronouns/variables "carry the burden of reference," so that one could reveal any speaker's ontological commitment by translating his or her propositions into a formal language that brings such reference carriers to the surface.

At various points in his career Quine produced studies intended to show just what role bound variables play in Predicate Logic. Ironically, the logic that resulted was a Term Logic. Quine's Predicate Functor Logic was meant to reveal just what the roles of predicates and bound variables were in the classic Predicate Logic. What Quine recognized was that a relative clause (e.g., 'which is F', 'that Gs', 'who is H') can be construed as an abstraction from a complex sentence, forming a complex general term. For example, from the sentence 'Some philosopher formulated the principles of reasoning' we can form 'Someone who is a philosopher formulated the principles of reasoning', where the relative clause is a term that captures the content of the predicate in the sentence. Taking the expression 'such that' as the relative pronoun *par excellence*, we then get 'Some x is such that x is a philosopher and x formulated the principles of reasoning'. Bound variable are seen as essentially devices for constructing complex terms. The "basic" use of the bound variable, then, is to isolate from a complex sentence that makes reference to an individual a term (viz., noun or adjective) that can be predicated of that individual without loss of information. For the Quine of Predicate Functor Logic this use of variables is more basic than their use in quantification. Variables can be eliminated and their work carried out by functors applied to the predicates that remain. Applying the various functors (in most versions there are four or five along with what amounts to a modified Sheffer stroke operating on pairs of terms; the existential quantifier becomes Quines *cropping* functor) to predicates permits the permutation, iteration, and finally elimination of variables. The result is a formal language consisting of nothing other than terms and functors. This Predicate Funtor Logic can serve all the purposes of the Predicate Logic. The important point to notice, of course, is that a logic of terms and functors is nothing more than a version of Term Logic. It even includes the recognition of wild quantity for singulars ([5], p. 97). Consequently, Predicate Logic is, if Quine is correct, at heart a Predicate Functor Logic, that is, a Term Logic.

Quine was well aware that most contemporary logicians are uncomfortable accepting a formal language in which predicates are allowed to stand free of their arguments. They do this from a natural disinclination to allow predicates to be names of any sort. They fail to see the proper role of 'such that' phrases as devices for forming terms. Thus few contemporary logicians are prepared to look for *terms* in predications.

7 Concluding Remarks

Having outlined Sommers' Term Logic and then intimated that Quine's Predicate Functor Logic reveals the terminist heart of Predicate Logic, one still must be cautious in drawing too many rosy conclusions. There are fundamental differences between the way Term Logic views the logic of natural language and the way Predicate Logic views it. Differences in how natural language predication is conceived account for most of this. The Predicate logician must assume that the saturated/unsaturated (complete/incomplete) distinction is not up for either analysis or debate. It's settled when the lexicon is given. But one can still ask, even at this late date, How is the distinction to be drawn? Is it semantic (or, perhaps, onto-semantic)? For it certainly is *not* syntactic. And even if it is a semantic distinction, one can ask, à la Ramsey, which term in an atomic proposition is supposed to be incomplete ('Socrates is . . . ' or '. . . is wise')? This is important. In Predicate Logic there is no tie binding together the subject-term(s) and the predicate-term. In the absence of such a logical copula the gap(s) account takes on great importance. For the features of symmetry and transitivity that play so largely in both traditional and newer Term Logics, and which are borne of the copulae, have to be found elsewhere in a logic that achieves predication without copulation.

Appendix

I'd like to append here a few brief remarks regarding the topic of logical adequacy. It is hardly and exaggeration to say that modern Predicate Logic was developed as part of the search for *mathematical* adequacy. The implicit claim was that this is how reason should work if one is to produce (eventually) mathematics. Consequently, the standard Predicate Logic is able to reveal very little about how we actually reason in everyday life (and so much the better for logic, thought Frege). By contrast, Sommers' Term Logic was developed as part of the search for *cognitive* adequacy. The implicit claim here was that this is how reason should work if one is to reason correctly. Moreover, recognizing that we actually do reason in everyday life in mostly correct ways, this logic can be used as a model of everyday logical reckoning. Certain consequences of Sommers' claims account for some obvious differences between Predicate Logic and Term Logic. The latter requires a notion of logical syntax that is similar to the syntax of natural language. The former has no such requirement. Thus, while formalization into Predicate Logic is a matter of translation, formalization into Term Logic is little more than transcription. As well, since Term Logic aims to be cognitively adequate, it must have simple, perspicuous rules of logical reckoning that are easily and quickly applicable since most everyday reasoning is relatively fast and easy. Finally, modern Predicate Logic is unable to provide any principled account of the general nature of formatives. Sommers, however, empirically discovered that natural language formatives are always signs of opposition.

8 Bibliography

[1] G. Englebretsen. *Something to Reckon With.* University of Ottawa Press, Ottawa, 1996.

[2] P. T. Geach. *Logic Matters.* University of California Press, Berkeley, 1980.

[3] L. M. Laita. Boolean algebra and its extra-logical sources: The testimony of Mary Everest Boole. *History and Philosophy of Logic*, 1:57–60, 1980.

[4] G. W. Leibniz. General inquiries about the analysis of concepts and of truths. In G. H. R. Parkinson, editor, *Leibniz: Logical Papers.* Clarendon Press, Oxford, 1966. [1686].

[5] W. V. Quine. *The Roots of Reference.* Open Court, La Salle, IL, 1974.

[6] W. V. Quine. Grammar, truth and logic. In S. Kanager and S. Öhmar, editors, *Philosophy and Grammar.* D. Reidel, Dordrecht, 1980.

[7] F. Sommers. *The Logic of Natural Language.* Clarendon Press, Oxford, 1982.

[8] F. Sommers and G. Englebretsen. *An Invitation to Formal Reasoning.* Ashgate Publishing, Aldershot, 2000.

FOL 75?[1]

William Ewald
wewald@law.upenn.edu
University of Pennsylvania

0

First-order logic is commonly said to have emerged to light seventy-five years ago with the publication of *Hilbert & Ackermann* [5] in 1928. The argument for this claim is strong, and, in its main outlines, indisputable. The earlier logical systems of Frege and of Whitehead-Russell employed higher-order quantifiers and variables throughout. Although (with some effort) one can retrospectively extract from their systems a first-order fragment, such an enterprise is plainly anachronistic: they themselves display no inclination to extract such a subsystem, nor do they display any sense that it would be significant either mathematically or philosophically. In *Hilbert & Ackermann*, however, first-order logic is explicitly singled out for study, receiving a forty-page chapter to itself (almost a third of the entire monograph). Metalogical questions are posed about this system, most significantly the question of completeness and the decision problem. In this sense, the publication of

[1]This note on Hilbert and first-order logic treats a topic more directly relevant to the theme of the "FOL 75" conference than the talk actually delivered in Berlin, which will appear elsewhere. §1 below draws heavily from the introductory notes for the Hilbert lectures of 1917–18, which were written together with Wilfried Sieg, and which will appear in the forthcoming volume of the Hilbert Edition dealing with logic and arithmetic, 1917–1934; the Edition, in six volumes, is published by Springer Verlag. A more detailed examination of the entire series of lecture notes from the period 1917–1922 is to be found in [11]

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 89–105

Hilbert & Ackermann was a major turning point, not just for first-order logic, but for logical studies generally. Nevertheless, a couple of caveats are in order. The first is a point of chronology. It is clear from Hilbert's unpublished lecture notes that already by the spring of 1918 at the latest he was in possession of substantially the entire conceptual apparatus of *Hilbert & Ackermann*, a full decade before the book was formally published; it is moreover clear that Ackermann, although an important logician in his own right, and although he is named as Hilbert's co-author, had almost nothing to do with the mathematical content of the book: his role was closer to that of a textual editor. (These lecture notes are soon to be published in the volume of the Hilbert Edition covering his logical writings from 1917 to the early 1930s.) The chronological point is important. It means that Hilbert (and Bernays, and the Hilbert School in Göttingen) were in possession of the modern, metamathematical conception of logic well before the start of Hilbert's research in proof theory, and not as its capstone. (Hilbert did not begin his investigations into proof theory, in the strict sense of the term, until the early 1920s.) Secondly, although Hilbert had already by 1918 isolated first-order logic, his understanding of that system, both in the lectures and in the published book, is not quite the modern understanding. The central questions had been posed, and the modern understanding lay just over the horizon: but it had not yet arrived. Here I wish to expand on these two points.

1

Let us begin by canvassing what is known about the background of the 1917-18 lectures on *Prinzipien der Mathematik*. It is usual to divide Hilbert's work in foundations into two distinct phases. The first phase lasts from roughly 1899 to about 1904, during which time he was mostly occupied with the axiomatics of geometry and the consistency of arithmetic. The second phase is taken to begin primarily as a response to Brouwer, Weyl and the paradoxes roughly in 1922 with the publication of 'Neubegründung der Mathematik;' this phase is mostly occupied with proof theory and the quest for a finitist consistency proof for arithmetic. During this period Hilbert is held to have adopted a 'formalist' philosophy of mathematics; this second phase culminates in two co-authored books: *Hilbert-Ackermann 1928* [5], and *Hilbert-Bernays 1934* [6] and *1939* [7].

Hilbert's research publications on foundational matters do indeed cease with his Heidelberg talk, 'Über die Grundlagen der Logik und der Arithmetik' (which was delivered in August of 1904 and published in 1905). But the lecture notes tell a more complex story about the development of his thought. Hilbert continued to lecture on foundational matters throughout the 'fallow' period from 1904 to 1922; he delivered roughly one lecture course every other year, and in this way kept abreast of the subject. The standard account is correct, however, that his research can be divided into two phases. But the break occurs during the summer of 1917, and not in the early 1920s. In the spring of 1917 he delivered a series of lectures on set theory which contain only the faintest hints of the new approach; then, in the fall, he embarks on a series of lectures which launch the modern subject of mathematical logic. (There is by the way no indication that the changes came about as a reaction to Brouwer, Weyl, or the paradoxes.)

The spring set theory lectures are for the most part an elegant and polished presentation of familiar results. But near the end of the lectures (which finished around August 15) Hilbert remarked without further comment that 'Next semester I hope to be able to go more deeply into the foundations of logic.' During the short summer vacation Hilbert, as was his custom, traveled to Switzerland; there, on September 11, he delivered his lecture 'Axiomatisches Denken' to the Swiss Mathematical Society in Zürich. On this trip he invited Paul Bernays, a promising young mathematician who had studied in Göttingen and who had strong philosophical interests, to return to Göttingen as his assistant in foundations of mathematics. These two events signal the new turn in his foundational research. A new approach to foundational issues was to evolve over the next six years and to be presented in a remarkable series of lecture courses, most of them written up by Bernays. This period saw the development of axiomatic investigations of logic and arithmetic, the birth of proof theory, and the beginnings of work on the *Entscheidungsproblem*.

It is important to observe that (as the very title of the Zürich talk indicates) the new approach was firmly grounded in Hilbert's earlier work in axiomatics. He had long viewed the axiomatic method as holding the key to a systematic organization of any developed mathematical subject. He also saw it as providing the basis for metamathematical investigations of independence and completeness. In particular, the problem of *consistency* had been of central importance ever since he turned his attention to the foundations of analysis at the end of the last decade of the 19^{th} century. Hilbert stressed (for example in his Paris address of 1900) that a rigorous proof of consistency was necessary to underwrite the legitimacy of any axiom system and to establish the existence of its range of objects. This issue occupied him intermittently until the Heidelberg talk of 1904, in which he proposed a simultaneous development of logic and arithmetic, and sketched a 'direct' syntactic consistency proof for a fragment of arithmetic. His approach was severely criticized by Poincaré [10], and in the intervening years he had nothing further to say on the matter. At any rate, the lectures given during the period from 1905 to 1917 do not break new ground, and in particular do not advance the 'proof theoretic' approach of the Heidelberg talk. (Nor do they embrace a philosophy of 'formalism.' Indeed, the notes for his *Mengenlehre* course in the spring of 1917 and the talk *Axiomatisches Denken* reveal an inclination at least to consider seriously the claims of logicism; though whether Hilbert was ever in any straightforward sense a logicist is more questionable.)

This interest in logicism was related to his more particular interest in the work of Whitehead and Russell, which seems to have begun around 1913. Before the outbreak of the war, Hilbert had planned to invite Russell to Göttingen, and he discussed type theory briefly in his lectures of 1914-15. (The exchanges of correspondence between Hilbert and Russell are documented and examined in [11]; further details about Hilbert's student Heinrich Behmann, who seems to have been the principal conduit of Russell's ideas to Göttingen, are provided in [8].) Between 1914 and 1917 several members of the Hilbert school gave lectures on logic and the foundations of mathematics to the Colloquium of the Göttingen Mathematical Society. In particular, the meetings during the month of July 1917 were entirely devoted to foundational questions. On 3 and 10 July Behmann lectured on 'The Russell-Whitehead Theory and the Foundations of Arithmetic.' On 17 and 24 July Felix Bernstein lectured on 'The History of Set Theory', and on 31 July Hilbert himself reported on his set theory lecture course. The Summer Semester ended on 15 August; Hilbert's remarks to his class about hoping 'to be able to explore a foundation for logic more deeply next semester' thus fall squarely within this period, and may possibly have been prompted by Behmann's report on the *Principia*.

There are certainly signs that the *Principia*, the chief technical advance since he had last intensively worked on foundations of mathematics, loomed large in Hilbert's thought during the summer of 1917, and in particular when he delivered his lecture *Axiomatisches Denken*. After discussing the axiomatic method and the reduction of mathematics to set theory, he succinctly sets forth the issues that were to be at the heart of the upcoming lectures, beginning with a statement of the logicist *credo*:

> Since the examination of consistency is a task that cannot be avoided, it appears necessary to axiomatize logic itself and to prove that number theory and set theory are only parts of logic.
>
> This method was prepared long ago (not least by Frege's profound investigations); it has been most successfully explained by the acute mathematician and logician Russell. One could regard the completion of this magnificent Russellian enterprise of the *axiomatization of logic* as the crowning achievement of the work of axiomatization as a whole.
>
> But this completion will require further work. When we consider the matter more closely we soon recognize that the question of the consistency for integers and for sets is not one that stands alone, but that it belongs to a vast domain of difficult epistemological questions which have a specifically mathematical tint: for example (to characterize this domain of questions briefly) the problem of the *solvability in principle of every mathematical question*, the problem of the subsequent *checkability* of the results of a mathematical investigation, the question of a *criterion of simplicity* for mathematical proofs, the question of the relationship between *content and formalism* in mathematics and logic, and finally the problem of the *decidability* of a mathematical question in a finite number of operations. [4, §§39-41]

In retrospect, these brief remarks can be seen as Hilbert's first public announcement of the material he was to begin teaching three weeks later. Hilbert's core insight was the realization that the techniques of axiomatics that he had first developed in the work on geometry in the late 1890s, culminating in the *Grundlagen der Geometrie*, could be extended to the logic of *Principia* and that the latter could provide a foundation for all of mathematics. The detailed pursuit of that goal required the presentation of a formal language (for capturing the logical form of informal statements), the use of a formal calculus (for representing the structure of logical arguments), and the formulation of 'logical' principles (for defining mathematical objects). This project was to be executed with remarkable focus in the lectures of the winter semester 1917-18, which began on October 1.

During this semester Hilbert carried a heavy teaching load. He lectured on Mondays, 9-11, on 'The Theory of the Electron,' and then again from 4-6 (with Emmy Noether) on 'Mathematical Principles.' (No record of the contents of those classes with Noether appears to have survived.) On Wednesdays, 4-6, he lectured with Peter Debye on 'The Structure of Matter.' The logic lectures came last, on Thursdays, 9-11. He was intensely involved during this time in research on the foundations of mathematical physics; the rigors of wartime had imposed additional burdens. The speed with which his logical ideas emerge in the lectures is therefore quite remarkable.

Hilbert begins by announcing his intention to talk about the axiomatic method in geometry, arithmetic, and mechanics; and the first 62 pages of Bernays's typescript – 'Part A,' fully one-quarter of the total – are devoted to a recapitulation of Hilbert's ideas on axiomatics. Most of this material goes back to the period of his foundational investigation of geometry, and was familiar territory. There is certainly a thematic appropriateness to his beginning with the axioms of geometry, for his new material would in effect demonstrate how the axiomatic techniques he had developed in geometry, and that had been the *leitmotif* of his intervening work in physics, could now be extended to encompass logic. The emphasis throughout Part A is not on the negative goal of avoiding paradox, but squarely on the programmatic gains for mathematics from axiomatization, and from what in *Axiomatisches Denken* he had called the '*Tieferlegung der Fundamente*.' Hilbert stresses that the axiomatic method yields deep insights into the structure and interrelationships between the theorems of geometry, and he spends several lectures illustrating the point, exploring in particular the role played by continuity assumptions in basic geometry. There is no mention of any 'crisis in the foundations of mathematics' as a motivating factor for his axiomatic investigations; the disagreements with Weyl and Brouwer still lie a couple of years in the future. At the end of the discussion of geometry Hilbert suddenly announces that he has said enough to explain the fundamental ideas of axiomatics. He will not, as originally planned, treat the principles of mechanics, but will instead direct his attention immediately to the logical foundations of mathematics.

What follows in Part B of the typescript is an incisive and carefully organized development of the very core of modern mathematical logic. Systemic and meta-systemic issues are clearly distinguished; metamathematical questions about consistency and completeness are crisply formulated. Each step forward is carefully explained and carefully motivated.

Part B of the notes is divided into five chapters:

1. The Propositional Calculus;

2. The Predicate Calculus and the Class Calculus;

3. Transition to the Function Calculus;

4. Systematic Presentation of the Function Calculus; and,

5. The Extended Function Calculus.

Here the predicate calculus is just monadic logic; the class calculus is its semantic, Boolean interpretation in terms of sets. The function calculus (which was renamed, following a suggestion by Hilbert, the 'predicate calculus' in the second edition of [5]) is in essence many-sorted first-order logic with variables for sentences as well as for functions (i.e. relations); the extended function calculus is the corresponding second-order system, but with quantifiers allowed to range, not just over set-theoretical objects, but also over propositions. Each transition from one section to the next is systematically argued for, usually by pointing out that the existing calculus, in some way or other, is in need of augmentation or 'completion.'

Broadly speaking, the chief accomplishment of these lectures consists in (1) the presentation of a series of precisely formulated axiomatic calculi, (2) the formulation and investigation of metalogical questions such as decidability, independence, consistency, and completeness, and finally (3) the development of arithmetic (including analysis). The most distinctive contribution lies in the metalogical investigations; but the axiomatic calculi in the tradition of Frege, Peano, Schröder, and Russell are more sharply presented than by Hilbert's predecessors, and the sharper formulation is essential to the metamathematical advances.

For example (and in contrast to *1 of *Principia*) Hilbert clearly distinguishes the axioms from the rules of inference; and where the *Principia* had employed the 'primitive proposition' *1.1: 'Anything implied by a true elementary proposition is true,' Hilbert provides both an explicit substitution rule and *modus ponens*. This distinction shows a grasp of the purely syntactic character of derivations, and is fundamental for the later metamathematical results, because it made possible rigorous proofs by induction on derivations. (Of course, that grasp goes back to his Heidelberg talk of 1904.) Likewise, the quantificational axioms and rules of inference he presents in the lectures are new, and considerably more perspicuous than the treatment in *Principia*.

The most innovative part of the lectures is the investigation of the various metalogical questions that Hilbert had adumbrated in Zürich. But there are some gaps and oddities of presentation that most likely reflect the speed with which the new ideas were being developed. For example, he had long been in possession of a decidability proof for propositional logic; it is odd that he does not mention the decision problem in the present lectures, although he states and proves the normal form theorem for the propositional calculus, and thus had in hand the tools for a decidability proof. Likewise, he briefly treats some independence results in the propositional calculus (using techniques that go back to his lectures in 1905), but does not work out the details. He devotes considerably more attention to the problem of consistency, providing first a consistency proof for the twelve axioms of his algebraic version of propositional logic. He furnishes a two-element model: all atomic propositions are interpreted as 0 or 1; 'or' is interpreted as the minimum, 'and' as the maximum, and 'not' as 1-X. He observes that all twelve axioms are satisfied in this model, and are therefore consistent. Later in the lectures he provides a proof of the consistency of the function calculus. He divides the proof into two parts and considers first the propositional sub-system of the function calculus. He has here to deal with a stock of formulas generated by axioms and rules of inference; in order to establish the conclusion that the stock of formulas thus generated cannot contain a formula and its negation, he argues by induction on derivations. Hilbert uses the same two-valued model as before and shows that the five propositional axioms all have the value 0, and that the two rules of inference preserve this property. He then extends this interpretation to the quantificational part of the function calculus, showing likewise that the derivable formulas all have value 0; since their negations have value 1, not every formula is derivable, and the system is consistent. In a footnote immediately after this proof he observes: 'One should not overestimate the significance of this result. We do not yet have any guarantee that with the symbolic introduction of contentually unobjectionable presuppositions the system of provable formulas remains free of contradiction.' So consistency proofs for expanded axiom systems, in particular for those requiring infinite models, will still have to be given.

Hilbert's treatment of completeness in the lectures is roundabout, and it is natural to conjecture that he was still working his way to an appropriate formulation. Indeed, the lecture notes appear to employ the word

'Vollständigkeit' in four distinct senses.

Early in the lectures, Hilbert appears to speak of 'completeness' in an informal, quasi-empirical sense, either meaning (1) 'capturing all traditional logical inferences' or meaning (2) 'being adequate to the analysis of a particular subject matter.' It is not until quite late in the typescript that he states a formal criterion of completeness and provides a completeness proof for the propositional fragment of the function calculus. The discussion of propositional logic lay several weeks in the past; but in those earlier discussions there is no hint of the new, formal sense of completeness. The proof establishes what today is called (3) the *Post completeness* (or *syntactic completeness*) of propositional logic: the axiom system is said to be complete in this sense, if the addition of a previously unprovable formula to the axioms always results in an inconsistent system. (The similarity of this criterion to Hilbert's axiom of completeness in geometry should be evident: both criteria intuitively say that it is impossible to add any further elements to the system.) Hilbert's proof first establishes a lemma: a formula ('*Ausdruck*') is a provable propositional formula ('*logische Aussagen-Formel*') if and only if it is the sum of simple products, each of which contains a sentence-letter and its negation. Using this lemma, Hilbert is then able to establish easily the syntactic completeness of his axioms. But in the course of establishing the lemma he also proves a semantic result: every provable propositional formula is identically 0, thus establishing the system's soundness (and consistency). The converse of soundness is proved in a footnote; this establishes, from our perspective, (4) the *semantic completeness* of the system as well.

Hilbert in these lectures does not explicitly define a semantic notion of completeness. It is only a few months later, that Bernays provides this notion and a direct formulation of the semantic completeness theorem in his *Habilitation*. Hilbert appears to treat the result, formulated in the footnote, as not especially significant. He outlines an argument that the function calculus is not Post-complete, but notes that a rigorous proof remains to be found; the proof, along the lines sketched by Hilbert, was provided by Ackermann in [*Hilbert & Ackermann*, pp. 66-68]. Hilbert in the 1917-18 lectures does not even *state* the problem of the semantic completeness for the function calculus. That problem seems to have been first explicitly formulated as an open problem in [*Hilbert & Ackermann*, p. 28], and then of course solved by Gödel in his dissertation. – Many of the loose ends of the 1917-18 lectures were tidied up in the summer of 1918 in the unpublished *Habilitation* thesis of Paul Bernays, which was submitted in July, 1918, and provided a metalogical analysis of the propositional logic of *Principia Mathematica*. Although Bernays had had no experience of logical research before coming to Göttingen as Hilbert's assistant, the thesis contains several important advances. He provides a sharper formulation than Hilbert had done of the semantic completeness theorem ('Every valid formula is provable and conversely'); a careful, model-theoretic investigation of the independence and dependence of various groups of axioms; and an investigation of ways in which axioms can be replaced by rules of inference.

The similarity of the topics covered in the 1917-18 lectures to [5] is clear enough. It has often been assumed that *Hilbert & Ackermann* represents the culmination of years of collective research into logic by the members of the Hilbert school. But in fact virtually the whole of *Hilbert & Ackermann*, from §10 of Chapter One onwards, is taken, usually verbatim, from Part B of Bernays's 1917-18 typescript. (§§1-9 of Chapter One are similarly taken from Bernays's typescript of the lectures from the Winter Semester 1920.) The most important divergences between the two texts are the following.

Chapter One, §§12-13 (of the first edition of *Hilbert & Ackermann*): The discussion of the completeness and consistency of the propositional calculus has been repositioned, sharpened, and expanded, showing the influence of Bernays's [1] (which is cited), and distinguishing two formal conceptions of completeness (now known as semantic and Post completeness). Curiously, the formulation of semantic completeness is less perspicuous than in Bernays's uncited *Habilitation*. In the 1917-18 lectures this material had come in the middle of the discussion of the function calculus, and was apparently presented when Hilbert first formulated the argument; in the book it has been moved forward to a point where it more naturally belongs.

Chapter Two, §2: The blending of the propositional calculus with the predicate calculus has been shortened and simplified.

Chapter Three, §5: The axiom system for the function calculus, and in particular the axioms and inference rules for the quantifiers, have been streamlined; credit for this improvement is given to Bernays.

Chapter Three, §9: The lecture notes announce and sketch on p. 156 a proof that the function calculus is not Post complete, but remark parenthetically that 'to be sure, a strict formal proof that this formula cannot be derived from the axioms remains to be found.' Ackermann supplies the missing details (and takes credit in a footnote). The problem of the semantic completeness of the function calculus is in addition now explicitly stated as an open problem; cf. Section 4.3.

Chapter Three, §§11-12: Two new sections, on the *Entscheidungsproblem* and on special cases of the Löwenheim-Skolem results, report on work done after the lecture notes and published elsewhere by Ackermann, Behmann, Bernays, Löwenheim, Schönfinkel, and Skolem.

Chapter Four, §§5, 8, and especially 9: The discussion of the theory of types and the axiom of reducibility has become sharper and more focused. It should be noted that the critical attitude was already hinted at in the 1917-18 lectures, and explicitly formulated in the 1920 lectures.

It will be seen from this list that Ackermann's only new mathematical contribution to *Hilbert & Ackermann* consists in supplying the details for the proof of the Post incompleteness of the function calculus; otherwise his role seems to have been more that of textual editor than of co-author.

But to return to the lectures. At the end of the discussion of first-order logic (p. 188 of the typescript), Hilbert makes the important remark that this system is adequate for the formalization of the inferences found in ordinary mathematics:

> The basic discussion of the logical calculus could cease here if we had no other end in view for this calculus than the formalization of logical inference. But we cannot be satisfied with this application of symbolic logic. Not only do we want to be able to develop individual theories from their principles in a purely formal way, but we also want to investigate the foundations of the mathematical theories themselves and examine how they are related to logic and how far they can be built up from purely logical operations and concept formations; and for this purpose the logical calculus is to serve us as a tool.

Hilbert accordingly turns to an examination of higher-order logic, pointing out that the widening of the function calculus enables one to formalize the axiom of complete induction, and to give the usual second-order definition of identity; more importantly, it permits also a Frege-style definition of the natural numbers as properties of predicates. Remarking that it is more natural for a mathematician to think of numbers as properties of sets than as properties of predicates, he enters into a careful examination of the relationship between sets and (co-extensive) predicates, and between sets of sets and predicates of predicates, showing that the basic concepts of set theory can be translated into the language of the extended function calculus.

Of course, this unrestrained approach is not without its risks, and the rest of the lectures are a dialectical working out of the technical consequences of the paradoxes. Hilbert first gives a careful exposition of three paradoxes (Russell's, the Liar, and Richard's), showing how each can be derived in his formal system. He traces the error to a logical vicious circle; the circle, he argues, is to be avoided by adopting (in effect) ramified type theory with Russell's axiom of reducibility, which can then be used to establish the beginnings of Dedekind's theory of real numbers (and, in particular, the least upper bound principle).

The notes end with the remark (p. 246):

> Thus it is clear that the introduction of the axiom of reducibility is the appropriate means to turn the calculus of levels into a system out of which the foundations for higher mathematics can be developed.

2

Hilbert thus already in 1917-18 was firmly in possession of the basic apparatus of modern mathematical logic: subject to a few refinements of the sort found in Bernays's *Habilitation*, he had isolated and formalized the system of first-order logic, distinguished it from higher-order systems, and posed the central metalogical questions. Nevertheless, it would be quite mistaken to attribute to Hilbert a fully modern understanding of first-order logic; and here we must be careful not to import into his lecture notes a conception that to be sure grew out of his logical investigations, but that only emerged some two decades later, and that depended on further technical advances at the hands of other researchers.

Indeed, it is important to observe that Hilbert was neither the only nor even the first logician to note the existence of what we today call 'first-order logic.' Such a discovery was made at least four (or possibly five) times, and each of these discoveries seems to have been independent of the others: certainly each was put to a different use.

The first clear distinction between first-order and higher-order quantification antedates Hilbert's lectures by more than 30 years. In the remarkable paper in which C. S. Peirce introduced his existential and universal quantifiers [9] he also distinguishes extremely clearly both between sentential and quantificational logic, and (following medieval precedent) between quantification over objects and quantification over predicates. (The latter he calls 'second-intentional logic.') He uses his second-intentional logic to define in the usual way the relation of identity: one object is identical to another just in case exactly the same predicates apply to each. Peirce's distinction was the clearest and most explicit drawing of the boundaries until the time of Hilbert: in this regard, far clearer than Frege or Russell. However, Peirce did not develop anything resembling an axiomatized formal system, still less subject it to metamathematical study; nor did he explore the relationship of higher-order logic to the paradoxes, or the suitability of its use in axiomatic set theory. The technical tools for these investigations were not yet to hand; and Peirce, for all his insight, was in no position to put his prescient distinction to its modern use.

Likewise, Hermann Weyl, as early as his 1909 *Habilitation* thesis (published as [14]) can be seen, at least in retrospect, as having isolated the basic idea of first-order logic. One of his aims was to scrutinize the notion of definability in set theory in the light of Richard's Paradox, and in particular to make precise Zermelo's notion of 'definite statement'; his proposal was in effect that 'definite statements' were those that could be generated out of a stock of basic properties by first-order principles. Weyl expanded these ideas in the opening pages of *Das Kontinuum (1918)*, arguing for a predicative grounding of analysis. In contrast both to Peirce and to Hilbert we find in Weyl an explicit drawing of connections between higher-order logic, the paradoxes, and axiomatic set theory, and also a sceptical stance towards higher-order systems. – Although Weyl had been a doctoral student of Hilbert's, it seems unlikely that Weyl's work on logic and predicativity exerted any direct influence on Hilbert's 1917-18 lectures. Weyl's name is not mentioned, and in any case *Das Kontinuum* was not published until 1918; as for the 1909 *Habilitation*, its treatment of first-order logic is sketchy and hard to follow. In neither text does Weyl possess a perspicuous quantificational notation, or draw a clear-cut distinction between sentential, first-order, and higher-order systems; nor does he present formal calculi or pose metamathematical questions about them. Although there are undoubtedly similarities in the conception of first-order quantification, the technical projects are quite different; perhaps more importantly, the entire thrust of Weyl's 1918 approach – the banishment of certain forms of mathematical inference because of worries about paradox – finds no echo in Hilbert's lectures. On the contrary, Hilbert's goal is to present in axiomatic form full ramified type theory. For him as for Russell this was the core system out of which mathematics was to be developed. If one is searching for intellectual influences, Russell (who figures explicitly in the lecture notes) is the more obvious candidate. It is true that Russell did not isolate a first-order subsystem of the logic of *Principia Mathematica*; but Russell (building on Frege) of course possessed the idea of the universe striated into types or orders, with a range of objects forming the base type. Hilbert's lectures show him systematically building up to type theory, one careful step at a time, beginning with the simplest calculi, and gradually adding more power. He was well aware that the paradoxes and the complexities of the axiom of reducibility arise only when one admits quantification of higher type; it was a natural step for him, as he built up his logical calculi, to pause to study first-order quantificational logic, and only later to proceed to the more elaborate details of the full system. In addition, as we saw, he explicitly noted that the first-order system was adequate to the purposes of formalizing mathematical inference: so he had ample reason to linger. Where Weyl is interested in distinguishing legitimate from illegitimate forms of mathematical reasoning, Hilbert is concerned with building formal axiomatic systems and subjecting them to metamathematical scrutiny; where Weyl is sceptical towards higher-order and impredicative systems, Hilbert still in [5, p. 113] says that 'the introduction of the axiom of reducibility is the appropriate means to shape the calculus of types [Stufen] into a system with whose help the inferences of higher mathematics can be won. A complete construction of the foundations of mathematics with the help of the calculus of types has been given by Whitehead and Russell.'

These points can be put another way. For Hilbert, both in 1917 and 1928, logic in its fullest sense encompassed ramified type theory; he spent a good deal of effort exploring the relationships between set theory and the theory of (higher-order) predication, and seems to have regarded both as falling properly within the ambit of logic. First-order logic he regarded as an important *subsystem* – as it were, a way-station on the path to his ultimate goal. (Recall that, in the terminology of *Hilbert & Ackermann*, first-order logic is called the 'restricted' function calculus.) This seems to have been the prevailing mainstream view throughout the 1920s. To be sure, there were some questioning voices, notably those of Weyl and of Skolem, who in [12] urged a reliance on first-order logic, at least for the purposes of formalizing axiomatic set theory. (Skolem's use of first-order logic to make precise Zermelo's axiom of separation is reminiscent of [14], but appears to have been independently arrived at.) But neither went so far as to reject higher-order logic altogether. The situation only began to change in the 1930s, as a result of two technical advances. To sketch them briefly: first, the Gödel completeness and incompleteness theorems made it clear that, although first-order logic can be formalized completely, type theory cannot: in other words, if one was interested in formalizing mathematical theories (as Hilbert and his followers certainly were) first-order logic offered distinct proof-theoretical advantages. Secondly, first-order formulations of set theory (at the hands of Bernays, Gödel, Quine, and others) gained widespread acceptance and popularity. These technical advances gave new impetus to the philosophically-based scepticism about higher-order systems, and by the end of the 1930s the attitude had begun to emerge that logic, *stricto sensu*, was to be identified, for all practical purposes, with first-order logic; higher-order logic, in contrast, was set theory in disguise, and as such belonged to the realm of mathematics.

This train of thought about first-order logic is familiar; my present point is that it stands in a complicated relationship to Hilbert. Indeed, this little episode shows the perils of attempting to assign a specific date to the birth of such a complex organism as first-order logic. At a minimum, one must be careful not to confuse several distinct developments: (1) the drawing of a clear distinction between first-order and higher-order logic; (2) the elaboration of formal calculi for various systems of logic; (3) the subjection of these calculi to metamathematical study; (4) the observation that first-order logic suffices for the formalization of ordinary mathematical inference; (5) the observation that higher-order logic, because of its relationship to the paradoxes, is *in some sense* less trustworthy than first-order logic; (6) the technical discovery that higher-order logic is metamathematically less well-behaved than first-order; and, (7) the philosophical argument that logic is to be identified with first-order logic, and that set theory in contrast is to be considered a branch of mathematics. Speaking very roughly, we can say that Peirce, in 1885, was the first to accomplish (1). Hilbert can be credited with (1), (2), (3) and (4). Weyl and Skolem are principally occupied with point (5), but also had independently accomplished (1). Point (7) only became an attractive option after the work of Gödel (6), which in turn grew out of (3), but which then gave renewed energy to point (5). Within this entire development, which spans fully half a century, Hilbert's lectures play a pivotal role, both introducing first-order logic in its modern guise, and proposing a research program that was in time to lead to a conception of first-order logic that displaced his own; but even the Hilbert lectures are far from the whole of the story, and in that sense one should perhaps resist the temptation to pin a specific date to the emergence of first-order logic.

3 Bibliography

[1] Paul Bernays. Axiomatische Untersuchung des Aussagen-Kalküls der 'Principia Mathematica'. *Mathematische Zeitschrift*, 25:305–320, 1926.

[2] William Ewald, editor. *From Kant to Hilbert; A Source Book in the Foundations of Mathematics*, volume I,II. Clarendon Press, Oxford, 1996.

[3] David Hilbert. Über die Grundlagen der Logik und der Arithmetik. In *Verhandlungen des Dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, Leipzig, 1905. Teubner. (English translation in [13], pp 129–138).

[4] David Hilbert. Axiomatisches Denken. *Mathematische Annalen*, 78:405–415, 1917. (English translation in [2], pp. 1105–1115).

[5] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik*. Springer, Berlin, 1928.

[6] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*, volume 1. Springer, Berlin, 1934.

[7] David Hilbert and Paul Bernays. *Grundlagen der Mathematik*, volume 2. Springer, Berlin, 1939.

[8] Paolo Mancosu. Between Russell and Hilbert: Behmann on the foundations of mathematics. *The Bulletin of Symbolic Logic*, 5:303–330, 1999.

[9] Charles S. Peirce. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics*, 7:180–202, 1885. (Reprinted in [2], pp. 608–632).

[10] Henri Poincaré. Les mathématiques et la logique. *Revue de métaphysique et de morale*, 13,14:815–835, 17–34, 294–317, 1905-06. (English translation in [2], pp. 1021–1071).

[11] Wilfried Sieg. Hilbert's programs, 1917-1922. *Bulletin of Symbolic Logic*, 5:1–44, 1999.

[12] Thoralf Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Matematikerkongressen i Helsingfors den 4-7 Juli 1922, Den femte skandinaviska matematikerkongressen, Redogörelse*, pp. 217–232, Helsinki, 1922. Akademiska Bokhandeln. (English translation in [13], pp 290–301).

[13] Jean van Heijenoort, editor. *From Frege to Gödel: A Source Book in Mathematical Logic*. Harvard, Cambridge, 1967.

[14] Hermann Weyl. Über die Definitionen der mathematischen Grundbegriffe. *Mathematische-naturwissenschaftliche Blätter*, 7:93–95, 109–113, 1910. (Reprinted in Weyl, Gesammelte Abhandlungen, vol. 1, Springer, Berlin, 1968).

[15] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Veit, Leipzig, 1918.

Fuzzy logic and arithmetical hierarchy IV

Petr Hájek
hajek@cs.cas.cz
Institute of Computer Science, Academy of Sciences, Prague

1 Introduction

Publication of the book "Grundzüge der theoretischen Logik" by Hilbert and Ackermann in 1928 [7] is properly considered the birthday of first-order predicate logic, as celebrated on the conference "FOL 75". Reading [7] is rather pleasing: one has *Aussagen* (propositions) that may be *richtig* or *falsch* (right of false); as an example of a false proposition one has "*Schnee is schwarz*" (snow is black). "*... ist schön*" (... is beautiful) is an example of a predicate. But we may ask: is the proposition "the snow is black" always absolutely false? (Come to Prague center in winter!) And isn't being beautiful a matter of degree, so that it is more true about one thing (person, ...) than about another one? We meet *vagueness*; and vague propositions need a *comparative* notion of truth. We are lead to *fuzzy logic* as a particular kind of many-valued logic; more precisely to fuzzy logic in the *narrow sense*, i.e. a formal logical system in distinction to what is called fuzzy logic *in the broad sense,* which is just almost everything concerning fuzziness. (This distinction was made by founder of the notion of a fuzzy set – Lotfi Zadeh.)

In [1] I formulated a *basic fuzzy propositional logic BL* and its *first order version BL*$\forall$, with the real unit interval $[0, 1]$ as its standard set of truth degrees, any continuous t-norm $*$ as its truth function of conjunction and its

Partial support of the grant No. A103004/00 of the Grant Agency of the Academy of Sciences of the Czech Republic is recognized.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 107–115

residuum $\to_*$ as its truth function of implication. Fixing a particular continuous t-norm $*$ one gets the corresponding propositional logic $PC(*)$ and predicate logic $PC\forall(*)$. There are three most important continuous t-norms: Łukasiewicz $*_Ł$, Gödel $*_G$, product $*_\Pi$. (If there is no danger of misunderstanding, we write just Ł for both $*_Ł$ and $PC(*_Ł)$, similarly for G, Π; we write Ł$\forall$ for $PC\forall(*_Ł)$ etc.)

The reader may find details in [1] or, as a survey, in [6] or [5]. In particular, a *standard interpretation* of a predicate language is a structure

$$\mathbf{M} = (M, (r_P)_{P\,predicate}(m_C)_{C\,constant})$$

where $M \neq \emptyset$ is a crisp set and for each n-ary predicate $P, r_P : M^n \to [0,1]$ is a fuzzy relation; $m_C \in M$. One defines, for each formula φ of predicate logic, each continuous t-norm $*$ and each evaluation v of object variables by elements of M, the truth value $\|\varphi\|^*_{M,v}$ in a natural Tarskian style. For φ closed, the v is superfluous; we write just $\|\varphi\|^*_M$. We restrict ourselves to closed formulas. For each $*$ we have four important sets of formulas: 1-tautologies, 1-satisfiable formulas, positive tautologies and positively satisfiable formulas.

$$1TAUT(*) = \{\varphi |\, \|\varphi\|^*_{\mathbf{M}} = 1 \text{ for all } \mathbf{M}\}$$

$$1SAT(*) = \{\varphi |\, \|\varphi\|^*_{\mathbf{M}} = 1 \text{ for some } \mathbf{M}\}$$

$$posTAUT(*) = \{\varphi |\, \|\varphi\|^*_{\mathbf{M}} > 0 \text{ for all } \mathbf{M}\}$$

$$posSAT(*) = \{\varphi |\, \|\varphi\|^*_{\mathbf{M}} > 0 \text{ for some } \mathbf{M}\}.$$

If $\mathcal{K}$ is a non-empty class of continuous t-norms, then 1-$TAUT(\mathcal{K})$ is the intersection of all 1-$TAUT(*)$ for all $*$ in $\mathcal{K}$, 1-$SAT(\mathcal{K})$ is the union of all 1-$SAT(*)$, similarly for $posTAUT(\mathcal{K})$, $posSAT(\mathcal{K})$. In particular, a formula is a (standard) 1-tautology of BL if $\varphi \in$ 1-$TAUT(\mathcal{K})$, $\mathcal{K}$ being the set of *all* continuous t-norms, etc.

Note in passing that BL ($BL\forall$) has also a general semantics with so-called BL-algebras as algebras of truth functions on a general domain; each continuous t-norm determines a standard BL-algebra on $[0,1]$. One has a natural notion of 1-$TAUT(\mathbf{L})$, $\mathbf{L}$ being an arbitrary BL-algebra and the axiomatic system of $BL\forall$ is complete with respect to 1-tautologies over arbitrary linearly ordered BL-algebras (general BL-tautologies). Similarly, one has general semantics for Ł∀, $G\forall$ and $\Pi\forall$. This semantic is not considered in the present paper; we restrict ourselves to standard semantics.

The complexity of sets of (standard) tautologies and (standard) satisfiable formulas of our fuzzy logic was studied in my previous papers with the title "fuzzy logic and arithmetical hierarchy" (see e.g. [9] for arithmetical hierarchy) very important results were obtained by Montagna. (See below.)

Main results of this paper concern arithmetical complexity of sets of formulas related to continuous t-norms different from the three famous ones. To be more precise let us briefly recall a basic fact on those t-norms (Mostert-Shields representation see e.g. [1].) Let $*$ be an arbitrary fixed continuous t-norm.

The set E of its idempotents is closed; its complement is the union of a countable system of non-overlapping open intervals. On the closure of such an interval, * is isomorphic either to Łukasiewicz t-norm or to product t-norm.

Define: $*$ *begins with* Ł if there is an idempotent $0 < e < 1$ such that $*$ on $[0, e]$ is isomorphic with Ł on $[0, 1]$. Similarly for beginning with Π.

Fact: The following are equivalent for each continuous t-norm $*$:

(1) $*$ is not Ł and $*$ does not begin with Ł;

(2) its negation is Gödel negation $(-)0 = 1, (-)x = 0$ for $x > 0$.

Define: $*$ begins with G if there is an idempotent $0 < e < 1$ such that all elements of $[0, e]$ are idempotents and there is a non-idempotent element above e.

Caution: There are continuous t-norms having no beginning of this kind, e.g. if positive idempotents are numbers $1/n, n$ positive natural.

2 Main results

The results contained in [1, 3, 8] are as follows:

Theorem 1. *The set of 1-tautologies is* Σ_1*-complete for* $G\forall, \Pi_2$*-complete for Ł∀, non-arithmetical for* $\Pi\forall$ *and* $BL\forall$*. The set of 1-satisfiable formulas is* Π_1*-complete for* $G\forall$ *and Ł∀, non-arithmetical for* $\Pi\forall$ *and* $BL\forall$*. The set of positive tautologies of Ł∀ is* Σ_1*-complete and the set of positively satisfiable formulas of Ł∀ is* Σ_1*-complete.*

Montagna generalized my construction showing that the set of 1-satisfiable formulas of $\Pi\forall$ is not arithmetical and constructed a sentence Ψ of $BL\forall$

and a translation associating with each sentence Φ of arithmetic a sentence Φ^0 of $BL\forall$ such that the following lemma.

Lemma 1. *[8] Let $\mathcal{K}$ be a set of continuous t-norms containing the product t-norm. Then, for each arithmetical Φ,*

(i) if Φ is true in the standard model $\mathbf{N}$ then $\Psi \to \Phi^0$ is a 1-tautology of $\mathcal{K}$ and $\Psi \& \Phi^0$ is a 1-satisfiable formula of $\mathcal{K}$.

(ii) if Φ is false in $\mathbf{N}$ then $(\Psi \to \Phi^0)$ is not a positive tautology of $\mathcal{K}$ and $\Psi \& \Phi^0$ is not positively satisfiable.

An easy checking shows that the assumption that $\mathcal{K}$ contains the product t-norm can be replaced by assuming that $\mathcal{K}$ contains a t-norm whose first component is product. This gives us the following

Theorem 2. *If $\mathcal{K}$ is a class of continuous t-norms containing a t-norm whose first component is product then the sets 1-$TAUT(\mathcal{K})$, 1-$SAT(\mathcal{K})$, pos-$TAUT(\mathcal{K})$, pos-$SAT(\mathcal{K})$ are all non-arithmetical.*

In particular $\mathcal{K}$ may contain all continuous t-norms ($BL\forall$), or a single element which is product t-norm or a t-norm whose first component is product.

Now let us study continuous t-norm $*$ beginning with $\mathcal{C}$, $\mathcal{C}$ being Ł,Π or G; for simplicity assume that the first positive idempotent is $\frac{1}{2}$. This can always be always be achieved up to an isomorphism. Furthermore, we may assume without loss of generality that the isomorphism of the restriction of $*$ to $[0, \frac{1}{2}]$ and $\mathcal{C}$ on $[0, 1]$ is just the mapping sending x to $2x$. Let us say that $*$ begins well with $\mathcal{C}$.

Definition 1. *Let h be the following mapping of $[0,1]$ onto itself: $h(x) = 2x$ for $x \leq \frac{1}{2}$, $h(x) = 1$ for $x \in [\frac{1}{2}, 1]$. Let $\mathbf{M} = (M, (r_P)_{P\,predicate}, (m_c)_{c\,constant})$ be a fuzzy structure of the language in question. Then $h(\mathbf{M})$ is the structure $(M, (r'_P)_{P\,pred.}, (m_c)_{c\,const.})$ where for each P (n-ary) and each tuple $a_1, \ldots a_n \in M$, $r'_P(a_1, \ldots, a_n) = h(r_P(a_1, \ldots, a_n))$. Further, let $\mathbf{M}/2$ be the structure $(M, (r_P/2)_{P\,pred.}, (m_c)_{c\,const.})$ where $(r_P/2)(a_1, \ldots, a_n) = r_P(a_1, \ldots, a_n)/2$.*

Lemma 2. *Let $*$ begin well with $\mathcal{C}$. Then h is a homomorphism of the structure $([0,1], *, \to_*, 0, 1)$ onto $([0,1], \mathcal{C}, \to_{\mathcal{C}}, 0, 1)$ preserving infinite joins and meets. Consequently, for each sentence φ,* *(1)* $h(\|\varphi\|^*_{\mathbf{M}}) = \|\varphi\|^{\mathcal{C}}_{h(\mathbf{M})}$,

(2) $\|\varphi\|^{\mathcal{C}}_{\mathbf{M}} = h(\|\varphi\|^*_{\mathbf{M}/2})$.

Proof: Clearly h commutes with $*$ and with finite and infinite joins and meets. We check $\to$. If $x \leq y$ then $h(x) \leq h(y)$ and $h(x \to_* y) = h(x) \to_{\mathcal{C}} h(y) = 1$. Assume $\frac{1}{2} \geq x > y$ and recall that h is an isomorphism of $[0, \frac{1}{2}]$ with $*$ to $[0,1]$ with $\mathcal{C}$; thus $h(x \to_* y) = h(x) \to_{\mathcal{C}} h(y)$ by the definition of residuum. Next let $x > \frac{1}{2} \geq y$; then $x \to_* y = y$ and $h(x) \to_{\mathcal{C}} h(y) = 1 \to_{\mathcal{C}} h(y) = h(y)$, thus again $h(x \to_* y) = h(x) \to_{\mathcal{C}} h(y)$. Finally let $x > y \geq \frac{1}{2}$, then $x \to y_* \geq \frac{1}{2}, h(x \to_* y) = 1 = 1 \to_{\mathcal{C}} 1 = h(x) \to_{\mathcal{C}} h(y)$.

The rest follows by induction on complexity of formulas.

Theorem 3. *Let $*$ begin with $\mathcal{C}$, $\mathcal{C}$ being Ł, G or* Π. *Then φ is a positive tautology of $*$ iff it is a positive tautology of $\mathcal{C}$. Moreover, φ is positively $*$-satisfiable iff φ is positively $\mathcal{C}$-satisfiable. In symbols: $posTAUT(*) = posTAUT(\mathcal{C})$, $posSAT(*) = posSAT(\mathcal{C})$.*

Proof: This is immediate from the preceding lemma which shows that there is an $\mathbf{M}$ with $\|\varphi\|^*_{\mathbf{M}} = 0$ iff there is an $\mathbf{M}$ with $\|\varphi\|^{\mathcal{C}}_{\mathbf{M}} = 0$, and the same for $\neq 0$.

Lemma 3. *Let $*$ begin with $\mathcal{C}$, $\mathcal{C}$ being Ł or G (nothing is claimed about* Π*!) For each (standard) structure* $\mathbf{M}$ *there is a structure* $\mathbf{M}'$ *such that for each* φ,

$$\|\varphi\|^{\mathcal{C}}_{\mathbf{M}} = 1 \textit{ iff } \|\varphi\|^*_{\mathbf{M}'} = 1.$$

(You may say that the $\mathcal{C}$-structure $\mathbf{M}$ *is elementarily 1-equivalent to the $*$-structure* $\mathbf{M}'$.)

Proof: Recall that if T is $\mathcal{C}$-consistent and $T \vdash (\exists x)\varphi(x)$ then $T + \varphi(c)$ is $\mathcal{C}$-consistent (c a new witnessing constant, see [1] 5.4.17 observing that the proof works also for $BL\forall$ and other logics). Let $T = Th(\mathbf{M}) = \{\varphi | \|\varphi\|^{\mathcal{C}}_{\mathbf{M}} = 1\}$; φ is any closed formula possibly containing names of elements of M. Let $\hat{T} \supseteq T$ be a consistent extension of T witnessing all existential closed formulas.

Recall that for $\mathcal{C}$ being Ł or G we know that each consistent theory (over $\mathcal{C}\forall$) has a standard $\mathcal{C}$-model (for G see the proof of [1] 5.3.3; for Ł see [1] 5.4.24).

Let $\mathbf{M}_1$ be a standard $\mathcal{C}$-model of $\hat{T}$. Let $f(x) = \frac{x}{2}$ for $x < 1$, $f(1) = 1$. Make $\mathbf{M}_1$ to a $*$-structure $\mathbf{M}'$ (with the $\mathcal{C}$-component on $[0, \frac{1}{2}]$) and with the same domain as $\mathbf{M}_1$ by defining $r_P^{M'}(a_1, \dots) = f(r_P^{M1}(a_1, \dots))$ for all P and $a_1, \dots$. Show by induction on the complexity of closed $\bar{T}$-formulas φ,

$$\|\varphi\|^*_{\mathbf{M}'} = f(\|\varphi\|^{\mathcal{C}}_{\mathbf{M}_1}).$$

This is evident for atoms and connectives (since $[0, \frac{1}{2}) \cup \{1\}$ is a $\mathcal{C}$-subalgebra of $[0,1]_*$) and for $\forall$ (since f preserves infinite meets); similarly for $\|(\exists x)\psi\|^{\mathcal{C}}_{\mathbf{M}_1} < 1$. For $\|(\exists x)\psi\|^{\mathcal{C}}_{\mathbf{M}_1} = 1$ use witnessing: there is a c such that $\|\psi(c)\|^{\mathcal{C}}_{\mathbf{M}_1} = 1$. In particular, if $\|\varphi\|^{\mathcal{C}}_{\mathbf{M}} = 1$ then $\|\varphi\|^{\mathcal{C}}_{\mathbf{M}_1} = 1$ and $\|\varphi\|^*_{\mathbf{M}'} = 1$.

Theorem 4. *(1) $posSAT(G) = 1SAT(G)$.*
(2) For $*$ beginning by G, $1SAT(*) = 1SAT(G)$.
(3) Similarly for $*$ beginning by Ł, $1SAT(*) = 1SAT(Ł)$.

Proof: (1) Clearly $1SAT(G) \subseteq posSAT(G)$. Conversely if $0 < r = \|\varphi\|^G_{\mathbf{M}} < 1$ for some $\mathbf{M}$ then taking a one-one increasing mapping of $[0,1]$ onto itself produce an isomorphic copy $\mathbf{M}'$ of $\mathbf{M}$ such that $\|\varphi\|^G_{\mathbf{M}'} = \frac{1}{2}$. Then apply the homomorphism h from Definition 1 and observe that it is a homomorphism of the G-structure $\mathbf{M}'$ to the G-structure $h(\mathbf{M}')$ sending $\frac{1}{2}$ to 1. Thus $\|\varphi\|^G_{h(\mathbf{M}')} = 1$.

(2) $1SAT(*) \subseteq 1SAT(G)$ by Lemma 2; $1SAT(G) \subseteq 1SAT(*)$ by Lemma 3. Similarly for (3).

Corollary 1. *If $*$ begins with G then*

$$posSAT(*) = posSAT(G) = 1SAT(G) = 1SAT(*)$$

Lemma 4. *If $*$ has Gödel negation (i.e. $*$ does not begin with Łukasiewicz) then for each φ, φ is $*$-positively satisfiable iff $\neg\neg\varphi$ is $*$-1-satisfiable; and φ is a $*$-positive tautology iff $\neg\neg\varphi$ is a $*$-1-tautology. Let $\varphi^{\neg\neg}$ result from φ by replacing each atom by its double negation. φ is a Boolean tautology iff $\varphi^{\neg\neg}$ is a 1-tautology of $*$. φ is not a Boolean tautology iff $\varphi^{\neg\neg}$ is not a positive tautology of $*$. (Evident.)*

Lemma 5. *If $*$ begins with Ł then for each φ, φ is a 1-tautology of Ł iff $\neg\neg\varphi$ is a 1-tautology of $*$; similarly for 1-satisfiability.*

Proof: By our Lemma 2, φ is a 1-tautology of Ł iff φ is a $[\frac{1}{2}, 1]$-tautology of $*$ (for each $\mathbf{M}$, $\|\varphi\|^*_{\mathbf{M}} \in [\frac{1}{2}, 1]$) iff $\neg\neg\varphi$ is a 1-tautology of $*$. Similarly for satisfiability.

The following theorem collects results of arithmetical complexity not stated till now

Theorem 5. *(1) $posTAUT(G)$ is Σ_1-complete, $posSAT(G)$ is Π_1-complete.*

(2) If $*$ begins with G then $1SAT(*) = posSAT(*)$ is Π_1-complete, $posTAUT(*)$ is Σ_1-complete and $1TAUT(*)$ is Σ_1-hard.

(3) If $*$ begins with Ł then $posTAUT(*)$ is Σ_1-complete, $posSAT$ is Σ_2-complete, $1SAT$ is Σ_1-complete and $1TAUT$ is Π_2-hard.

Proof: (1) $posTAUT$ is in Σ_1 by Lemma 4 and is Σ_1-hard by the same Lemma. $posSAT = 1SAT$ by Theorem 4 (1) and hence is Π_1-complete.

(2) By Corollary 1, $1SAT(*) = posSAT(*) = 1SAT(G)$, hence is Π_1-complete; $posTAUT(*) = posTAUT(G)$ by Theorem 3, hence Π_1-complete, and $1TAUT(*)$ is Σ_1-hard by Lemma 4.

(3) $posTAUT$ is Σ_1-complete and $posSAT$ is Σ_2-complete by Theorem 3. 1-SAT is Π_1-complete by Theorem 4; $1TAUT$ is Π_2-hard by Lemma 5.

The results are summarized in Table 1, where Ł$\oplus$ means any t-norm beginning by Łukasiewicz, similarly for $G\oplus, \Pi\oplus$. NA stands for "not arithmetical".

	1-$TAUT$	1-SAT	$posTAUT$	$posSAT$
BL	NA	NA	NA	NA
Ł	Π_2-comp.	Π_1-comp.	Σ_1-comp.	Σ_2-comp.
G	Σ_1-comp.	Π_1-comp.	Σ_1-comp.	Π_1-comp.
Π	NA	NA	NA	NA
Ł$\oplus$	Π_2-hard	Π_1-comp.	Σ_1-comp.	Σ_2-comp.
G$\oplus$	Σ_1-hard	Π_1-comp.	Σ_1-comp.	Π_1-comp.
$\Pi\oplus$	NA	NA	NA	NA

Tab. 1

3 Calculi with the Δ-projection

Finally, let us discuss our logics extended by the unary connective Δ (Baaz's delta.) Recall that the truth function of the Δ-connection (denoted also by Δ) satisfies $\Delta 1 = 1$, $\Delta x = 0$ for $x < 1$ ($\Delta\varphi$ says "φ is absolutely true").

First we modify the construction showing that tautologicity/satisfiability of the product logic is not arithmetical to work in Ł$\forall_\Delta$. Let Ψ be the finite conjunction of axioms saying that zero, successor, addition and multiplication are crisp and satisfy a sufficiently rich finite fragment of arithmetic and let U be a unary predicate satisfying (use $\emptyset$ for zero in arithmetic)

$$U(\emptyset) \equiv \neg U(\emptyset), (\forall x)(U(x) \equiv \neg\neg U(x)),$$

$$(\forall x)(U(x) \equiv (U(x+1) \& U(x+1)),$$

$$(\forall x, y)(x \leq y \rightarrow (U(x) \rightarrow U(y))),$$

$$(\forall x)\neg \Delta U(x).$$

Lemma 6. *Let $*$ be Łukasiewicz t-norm or a continuous t-norm beginning with Łukasiewicz. Then for each arithmetical Φ,*

(i) If Φ is true in the standard model $\mathbf{N}$ then $\Delta\Psi \rightarrow \Phi$ is a 1-tautology of $PC\forall_\Delta(*)$ (predicate calculus over $*$ with Baaz's Δ) and $\Delta\Psi \& \Phi$ is 1-satisfiable in $PC\forall_\Delta(*)$;

(ii) If Φ is false in the standard model then $\Delta\Psi \rightarrow \Phi$ is not a positive tautology and $\Delta\Psi \& \Phi$ is not positively satisfiable (w.r.t. $PC\forall_\Delta(*)$).

Theorem 6. *Let $*$ be as above (beginning by Łukasiewicz or just Łukasiewicz). Then the sets $1TAUT$, $1SAT$, $posTAUT$, $posSAT$ of $PC\forall_\Delta(*)$ are not arithmetical.*

Possibly the construction can be modified to show the non-arithmeticity of the four sets in question w.r.t. $PC\forall_\Delta(*)$ for each continuous t-norm different from Gödel t-norm. For Gödel t-norm the sets appear to be of the same complexity as they are Gödel logic without Δ; but this has to be checked.

4 Bibliography

[1] Hájek P.: Metamathematics of fuzzy logic. Kluwer, 1998.

[2] Hájek P.: Fuzzy logic and arithmetical hierarchy. Fuzzy Sets and Systems, 73 (1995), pp. 359-363.

[3] Hájek P.: Fuzzy logic and arithmetical hierarchy II. Studia Logica, 58 (1997), pp. 129-141.

[4] Hájek P.: Fuzzy logic and arithmetical hierarchy III. Studia Logica, 68 (2001), pp. 129-142.

[5] Hájek P.: Fuzzy logic. In: Standard Encyclopedia of Philosophy, http://plato.stanford.edu (see logic).

[6] Hájek P.: Why fuzzy logic. In: (D. Jacquette, ed.) A compagnon to philosophical logic, Blackwell, 2002, pp. 595-606.

[7] Hilbert D., Ackermann W.: Grundzüge der theoretischen Logik. Julius Springer, Berlin, 1928.

[8] Montagna F.: Three complexity problems in quantified fuzzy logic. Studia Logica 68 (2001), pp. 143-152.

[9] Rogers H. jr.: Theory of recursive functions and effective computability. McGraw-Hill, 1967.

What is the True Algebra of First-Order Logic?

Jaakko Hintikka
hintikka@bu.edu
Boston University

Things are no longer as they used to be in the good (or bad) old times. We used to teach our students that the rock bottom of all logic is the received first-order logic, alias quantification theory or lower predicate calculus. This logic is a part of the logics of Frege and of the *Principia Mathematica* and was extracted from them and made the basic layer of logic by Hilbert and Ackermann [6]. Even after the deaths of Quine and Dreben, there undoubtedly still are people who believe that the received first-order logic is *the* basic logic. By now it has nevertheless turned out that the usual first-order logic that goes back to Frege and Russell (as a part of their more comprehensive logic of quantification independent of the type level of quantifiers) does not fulfill its job description. (See here [8].) An important part of the semantics of quantifiers is their use in representing the relations of actual dependence and independence between variables by means of the relations of formal dependence and independence between the quantifiers to which the variables are bound. Now in the received notation for quantifiers such dependencies among quantifiers are represented by the nesting of their syntactical scopes. But such nesting cannot represent all possible patterns of dependence and independence. It is among other things transitive and antisymmetric, and hence cannot capture many interesting dependence patterns, for instance mutual dependence.

This defect is corrected by adding to the notation of first-order logic a slash / which exempts a quantifier, say (Q_2y), from being dependent on

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 117–128

another quantifier, say (Q_1x), in whose syntactic scope it occurs. This declaration of independence is expressed by writing (Q_2y) as (Q_2y/Q_1x). (This notation may be complemented by other conventions.) This amended notation implements what is called independence-friendly (IF) logic. It is the engine of the change I referred to above. Its semantics can be obtained from the well-known game-theoretical semantics for the ordinary first-order logic by allowing the move mandated by (Q_2y/Q_1) to be informationally independent of (Q_1x) in the usual game-theoretical sense of informational independence.

This notion of independence can – and should – also be applied to game moves associated with propositional connectives. It is understood that this is done in a full-fledged IF first-order logic. One might therefore look away from quantifiers and ask what kind of prepositional logic might be obtained by allowing independencies between truth functional correctives. There might be more than one way of doing so, but what is obtained can most naturally be considered as a part of the study of partiality logics. (See [16].) They are not without interest, but they do not seem to yield very much new that would be of general theoretical interest, especially if we concentrate on what can be said about when propositions are true rather than not false. This is perhaps to be expected. For, while the dependencies between quantifiers express important real relations between their variables, the dependencies between the moves associated with propositional connectives do not in general have an equally interesting interpretation.

More interesting structures are obtained by first noting the behavior of negation in IF logic. This behavior is in fact surprising. Perfectly classical game rules for negation turn out to yield a concept of negation which does not obey the law of excluded middle. Does this mean that IF logic should be called "nonclassical"? No, it should be taken to mean instead that *tertium non datur* is not an integral part of the classical conception of logic. From this point of view it is merely an accident that the law of excluded middle holds in the fragment of IF logic that the received first-order logic is. From a purely abstract logical point of view it is but another accident that the primary notion of negation that we employ in ordinary language is the contradictory negation that satisfies the principle of excluded middle.

This does not mean that we cannot incorporate the contradictory negation that obeys the law of excluded middle in our logical notation and study it. Such an extension of IF logic is necessary for its applicability, e.g. the negation employed in ordinary language is in most cases obviously intended to be the contradictory one. In order to understand our actual *Sprachlogik*, we therefore have to study logics which have two different negations, both the one $\sim$ that results from the classical rules for negation in game-theoretical semantics and the contradictory negation $\neg$. (See here [9, 7].) Since there cannot be any game rules that are more classical than the classical ones, contradictory negation cannot be introduced by any rules for semantical games. Without radically changing or extending IF logic, contradictory negation can only be introduced by a stipulation as to what it means for an entire closed sentence to be negated. If no further notions are introduced, $\neg$ can therefore occur only sentence-initially.

The logic that will be studied here is therefore what is sometimes called extended IF first-order logic. It is simply IF first-order logic with a sentence-initial contradictory negation $\neg$ added to is vocabulary. Its semantics is the same as for IF first-order logic, except for a metalogical rule that says that $\neg S$ is true if and only if S is *not* true, where the italicized metalogical negation has to be taken to be the contradictory one.

Since in propositional logic the components of sentences are themselves sentences (closed formulas without free variables), the contradictory negation can there be used without restrictions.

The distinction between the two negations means that we have corresponding difference between two conditionals in our hands. The conditional $(A \supset B)$ that goes together with $\sim$ is equivalent with

$$(\sim A \vee B) \tag{1}$$

while the one $(A \to B)$ that goes together with $\neg$ equals

$$(\neg A \vee B). \tag{2}$$

Furthermore, we have two equivalences amounting respectively to

$$(A \& B) \vee (\sim A \& \sim B) \tag{3}$$

and

$$(A \& B) \vee (\neg A \& \neg B) \tag{4}$$

By resorting to what might be called substitutional interpretation (see [7, 10, 9]), the extended IF first-order logic can be extended further so as to enable $\neg$ to occur within the scope of quantifiers. For instance

$$(\forall x)\neg F[x] \tag{5}$$

is interpreted by saying that it is true if and only if all sentences of the form $\neg F[b]$ are true, where b is a name of a member of the domain of individuals. This presupposes of course that there is a name available in the language in question for each member of the domain. The availability of names is not the essential nonelementary assumption here, however. What is nonelementary is the need of considering the (possibly infinite) domain as closed completed totality in our truth-conditions. No such appeals to closed infinite totalities of individuals is needed in the game-theoretical semantics for IF logic.

By using a mixture of game-theoretical and substitutional interpretation, one can obviously interpret sentences where $\neg$ occurs without any unusual restrictions. However, in this paper such substitutional or mixed interpretations will not be discussed, unless indicated in so many words. The need of considering closed infinite totalities makes an IF logic supplemented by the unlimited use of contradictory negation surprisingly strong. A measure of this surprise is that such a logic is as strong as the entire second-order logic. (See [7].)

In this paper, the propositional part of extended IF logic will receive special attention. What kind of structure is formed by the propositional part of extended IF first-order logic? For the purpose, let us first note that all the usual laws of propositional logic connecting $\sim$, & and $\vee$ hold, except for those depending on the necessary truth of sentences of the form $(S \vee \sim S)$. Moreover, the following laws hold:

$$\neg\sim(A \vee B) \leftrightarrow (\neg\sim A \vee \neg\sim B) \tag{6}$$

$$\neg\sim(A \& B) \leftrightarrow (\neg\sim A \& \neg\sim B) \tag{7}$$

$$\neg\sim\neg\sim A \leftrightarrow \neg\sim A \tag{8}$$

What kind of logic do we obtain in this way? Let us first consider this logic purely algebraically. What kind of algebraic structure do we have? Now $\neg$, & and $\vee$ form a Boolean algebra. But $\sim$ introduces an additional element. Since $\sim A$ equals $\neg(\neg\sim A)$, we can think of this additional structure as being introduced by the operator $\neg\sim$. Now (6), (7) show that this is in fact an operator in Tarski's sense. Hence all of Tarski's and Jónsson's results apply. (See [11, 12].) In particular, it follows that any extended IF first-order logic has a representation theorem. In Tarski's and Jónsson's words, "roughly speaking every such algebra is isomorphic to an algebra formed by a field of sets with the usual set-theoretical operations, and with [the additional] operators defined as images under certain relations between elements of the universal set (the largest set) of this field of sets, the notion of image under a relation being a generalization of the notion of the image under a function." [11, p. 372]

Perhaps even more roughly, it follows that every extended IF logic admits of an interpretation in set-theoretical terms where $\neg$, & and $\vee$ correspond to the usual Boolean operations or sets.

Tarski and Jónsson prove their representation theorem by first proving an extension theorem which also applies to extended IF first-order logics. It says that every Boolean algebra with operators can be imbedded in a complete atomistic Boolean algebra with an additional relational structure.

Tarski's and Jónsson's results enable us to treat extended IF logic as if it were an algebra of sets. For instance, we can consider measures defined on such an algebra. We will not avail ourselves of this possibility in this paper, however. Meanwhile (8) shows (in conjunction with (6)–(7)) that even more can be said here. Not only is the extended IF logic a Boolean algebra with an operator. It is a closure algebra. In other words, the structure of an extended IF first-order logic is a topology. We will call it IF topology. Hence all the results by Tarski, McKinsey and Jónsson on closure algebras apply to it. Moreover, many of the questions that can be asked about topologies in general can be asked about it. (See here [13, 14, 15].)

This provides an answer to the title question of this paper. The true algebra of logic is not a Boolean algebra but a closure algebra. Hence general topology is an eminently suitable framework for metalogic. Conversely, general topology thus turns out to be a useful tool for foundational studies in logic and mathematics in general.

Here only some of the most fundamental features of the IF topology will be listed.

i) The sentences of the received first-order logic correspond to open sets.

ii) The same sentences also correspond to closed sets, and are the only closed ones.

iii) The IF topology is compact.

iv) The IF topology has a countable base.

v) If A and B are disjoint sets in an IF topological space, then there are open sets U and V such that $A \subset U$, $B \subset V$.

The other properties of the IF topology remain to be investigated.

The compactness of IF topology (in the sense of general topology) has to be distinguished from the compactness of the extended IF logic in the logical sense. In the latter sense, even though IF first-order logic is compact, the extended IF logic is not. An example showing this noncompactness is obtained by noting that in IF logic you can formulate a sentence S that says that the domain is infinite. The following formula will do the trick:

$$(\forall x)(\forall y)(\exists z/\forall y)(\exists u/\forall x)((z \neq x)\&(u \neq y)\&((x = y) \leftrightarrow (z = u)) \qquad (9)$$

assuming that the domain is not empty. Hence $\neg S$ says that the domain is finite. Hence the following set of formulas is not satisfiable even though all its finite subsets are:

$$\begin{aligned}&\{\neg S, x_1 \neq x_2, x_1 \neq x_3, x_1 \neq x_4, \ldots\\&\ldots x_2 \neq x_3, x_2 \neq x_4, \ldots, x_3 \neq x_4, x_3 \neq x_5, \ldots\}\end{aligned} \qquad (10)$$

This apparently minor observation might turn out to be consequential. For, as von Neumann [18, pp. 233–234] emphasizes, in a noncompact space "a number of properties which are equivalent in compact spaces are not equivalent any longer". For instance, the continuity of a function does not imply the continuity of its inverse. Furthermore, noncompactness causes difficulties in the theory of operators on the spaces in question. What we have found is that IF *topological space* is still compact even though extended IF *logic* is not.

There is nevertheless more structure to extended IF logic than the usual topological one (or perhaps structure of a different kind). In order to see what this extra structure is, one might use the possibilities opened by Tarski's representation theorem. This theorem enables us to discuss an extended IF logic in analogy with set-theoretical structures. In this analogy, disjunction, conjunction and contradictory negation have their obvious geometrical (set-theoretical) counterparts in set union, set intersection and complementation. But what is the geometrical counterpart to the strong (dual) negation $\sim$? In order to answer this question, we may recall that Jaakko Hintikka [10] has shown that in the special case of Hilbert spaces the strong negation corresponds to orthocomplementation. Now orthocomplementation does not have a priori any counterpart in most logical spaces. However, we can turn the tables here and to propose to consider strong negation in general as implementing a generalization of the notion of orthogonality. On this suggestion, which we will adopt as our working definition, the sets corresponding to A and B are orthogonal if and only if it is the case that $(A \supset {\sim}B)$ or in other words if and only if $({\sim}A \vee {\sim}B)$.

This might at first sight look like a purely nominal extension of the notion of orthogonality. Its usefulness depends on to what extent this notion captures the characteristic behavior of the received notion of orthogonality. In order to see how it does so, let us examine some of the properties of the newly defined notion. Some of these properties are brought out by further defined notions. If we have a notion of orthogonality, we can define the dimensionality of a logical space as the largest number of pairwise orthogonal elements in it. Thus we can define and say that our logical space had d dimensions if and only if there are $B_1, B_2, \ldots, B_d$ such that

(i) $\neg{\sim}B_i$ $(i = 1, 2, \ldots, d)$

(ii) ${\sim}B_i \vee {\sim}B_j$ $(i \neq j,\ i, j = 1, 2, \ldots, d)$

(iii) for any sentence A, $\neg(({\sim}A \vee {\sim}B_1)\&({\sim}A \vee {\sim}B_2)\&\ldots\&({\sim}A \vee B_d))$

Here (i) says that the sentences ("vectors") spanning the space are not "null vectors", (ii) says that they are orthogonal to each other, and (iii) says that no element A is orthogonal to all of them. The condition (iii) can also be written

$$(\neg{\sim}A\&\neg{\sim}B_1) \vee (\neg{\sim}A\&\neg{\sim}B_2) \vee \ldots \vee (\neg{\sim}A\&\neg{\sim}B_d) \tag{11}$$

If we substitute here for A the sentence $(A \vee \neg A)$, we obtain from (9)

$$\neg{\sim}B_1 \vee \neg{\sim}B_2 \vee \ldots \vee \neg{\sim}B_d \tag{12}$$

But if (12) is true, then any A is equivalent with

$$(A\&\neg{\sim}B_1) \vee (A\&\neg{\sim}B_2) \vee \ldots \vee (A\&\neg{\sim}B_d) \tag{13}$$

This can be considered as a coordinate representation of A with the conjunctions $(A\&\neg{\sim}B_i)$ as the different components of A in the coordinate system $\langle B_1, B_2, \ldots, B_d\rangle$. The components of A, viz. $(A\&\neg{\sim}B_i)$, are mutually orthogonal if (ii) is assumed. However, it can be seen that the "coordinate representation" (13) is independent of (ii). Hence the "coordinate system" $B_1, B_2, \ldots, B_d$ need not be orthogonal.

One obvious but nevertheless noteworthy observation here is that disjunction or, set-theoretically speaking, set union behaves like vector addition. For the coordinate representation of $(D_1 \vee D_2)$ is

$$\bigcup_i ((D_1 \vee D_2)\&\neg{\sim}B_i) \tag{14}$$

which equals

$$\bigcup_i ((D_1\&\neg{\sim}B_i) \vee (D_2\&\neg{\sim}B_i)) \tag{15}$$

Hence the coordinates of the union of two propositions are obtained by adding to each other their coordinates one by one.

Since Tarski's representation theorem establishes complete additivity, the notion of dimension can also be extended to infinite-dimensional spaces.

This suffices to show that the extension of the notion of orthogonality by means of the notion of strong negation captures some of the characteristic properties of orthogonality.

These observations lead to further problems. As is well known, Brouwer has characterized the dimension of a class recursively in set-theoretical terms. (See here e.g. [1].) According to his definition, a topological space has d dimensions if and only if any two disjoint closed sets A, B in it can be separated by a set of dimension at most $d-1$, in the sense that there is a closed subset C whose complement is the sum of two disjoint open sets C_1, C_2 one of which contains A and the other B. The problem now is whether the notion of dimension can be related to Brouwer's. For this purpose, our definition of dimension must first be extended from coordinate systems to arbitrary sets.

It needs to be added that the algebraic structure formed by extended IF formulas with one free variable has to be studied separately. The reason is pointed out above. When contradictory negation is prefixed to an open formula, it does no admit of a game-theoretical interpretation, but instead has to be independent substitutionally.

These observations have significant repercussions for the philosophy of logic and mathematics. For one thing, it shows that there is no clear boundary between logical and geometrical concepts. More generally, it calls into question whether any meaningful distinction can be made between logical and nonlogical concepts.

Further perspectives are opened by a comparison between the algebraic structure we have reached and other related structures. Boolean algebras with operators include suitable modal logics, algebraically viewed. In particular, closure algebras correspond to the modal logic known as S4. This modal logic has a number of interesting interpretations. It is arguably the propositional part of epistemic logic (logic of knowledge). This is philosophically interesting in that game-theoretical semantics involves an epistemic element in the form of players' information sets, which codify certain aspects of their knowledge. However, it might appear prima facie that we have not taken this epistemic element into account. The relationship between extended IF logic and S4 nevertheless suggests that this impression is not correct and that we have after all managed tacitly to take into account the epistemic (informational) ideas needed here.

Even more interestingly, propositional S4 is known to be tantamount to intuitionistic propositional logic in Heyting's formulation, in the sense that each of them is interpretable in the other. (This was first pointed out in [3, (1933)].) Hence extended prepositional IF logic is in the same sense tantamount to intuitionistic propositional logic. It has been asked in the literature, "What kind of logic is IF logic?" (See [2].) If we want to use only logics familiar from earlier discussions, the shortest (but admittedly somewhat over-simplified) answer therefore is: intuitionistic logic. This may not be an entirely surprising result in view of the failure of *tertium non datur* in IF logic, for this failure is often considered the most important distinguishing mark of intuitionistic logic. The possibility of considering IF first-order logic as a kind of intuitionistic logic also helps to illustrate the sense in which IF logic is elementary in contradistinction to logics using unrestricted contradictory negation which violate the law of excluded middle. The logics relying on the substitutional interpretation of quantifiers which were mentioned above are cases in point.

An explanation is perhaps needed here. It is usually thought and said that before the expressive resources of the ordinary first-order logic are extended, it is equivalent with the slash-free part of IF logic. This is nevertheless true only on the assumption that atomic sentences obey the law of excluded middle. If they do not, unextended IF first-order logic is weaker than the received one. It is in this sense that it is nearly equivalent with intuitionistic logic.

This relationship between IF logic and intuitionistic logic can of course be proved directly. The translation rules between extended IF logic and intuitionistic logic are complicated by the fact that in the former there are two negations but in the latter only one. However, even this discrepancy becomes natural when it is recalled that contradictory negation can only occur sentence-initially. Hence in a natural correspondence an intuitionistic negation prefixed to a negation-free formula can be taken to be paired with $\sim$, whereas what looks like a double negation will be paired with $\neg\sim$.

I will leave the details of a proof of the IF–intuitionism connection for the reader to work out as an exercise.

This connection cannot be extended to the entire first-order logic in any simple manner. However, the reason is not any inadequacy of IF logic but Heyting's treatment of quantifiers which does not do justice to their interplay with tacit or explicit epistemic operators.

The insight into a connection between extended IF logic and intuitionistic logic opens an interesting further perspective. According to the original intuitionists, the main logical principle that creates the unacceptable strength of classical mathematics is the unlimited use of *tertium non datur.* This is analogous with what was found earlier in this paper, viz. that by which we can create an extremely strong logic on the first-order level simply by starting from extended IF logic and then additionally allowing arbitrary occurrences of contradictory negation. Indeed, the result is a first-order logic that is as strong as the conventional second-order logic. This goes a long way toward vindicating Brouwer's emphasis on *tertium non datur* as the source of the excessive strength of classical mathematics.

The connection between intuitionistic logic and extended IF logic poses intriguing further questions. Intuitionistic logic is generally thought of as being weaker than the received first-order logic. This is in an agreement with the fact that fewer formulas are logically true according to intuitionistic logicians than the logical truths of the received first-order logic. But at the same time some of the most crucial mathematical concepts, concepts that could not be captured by means of the received first-order logic, can be defined with the help of IF logic. Among these concepts there are the notions of equicardinality and topological continuity. What is then to be said of such concepts from and intuitionistic point of view?

This question is but a special case of a larger complex of problems which is at the same time a complex of opportunities. Using IF logic as our Archimedean point, we can investigate by means of explicit logic what can and what cannot be done in intuitionistic logic and intuitionistic mathematics.

Thus after 75 years of the Hilbert-Ackermann first-order logic and after 73 years of its antithesis Heyting's intuitionistic logic, the two are ripe to be synthesized into IF first-order logic.

1 Bibliography

[1] R. Engelkind. *Dimension Theory.* North-Holland, Amsterdam, 1978.

[2] Solomon Feferman. What kind of logic is IF logic? In L. Hahn and R. Auxier, editors, *The Philosophy of Jaakko Hintikka*, Library of Living Philosophers. 2004.

[3] Kurt Gödel. An interpretation of the intuitionistic propositional calculus. In *Collected Works*, volume 1, pages 300–303. Oxford U.P., New York, 1986. Original 1933.

[4] Arend Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsbericht der preussischen Akademie der Wissenschaften, Berlin*, pages 42–56, 57–71, 158–169, 1930.

[5] Arend Heyting. *Intuitionism: An Introduction.* North-Holland, Amsterdam, 1956.

[6] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik.* Springer, Berlin, 1928.

[7] Jaakko Hintikka. Truth, negation and other basic notions of logic. forthcoming.

[8] Jaakko Hintikka. Hyperclassical logic (aka indepence-friendly logic) and its general significance. *Bulletin of Symbolic Logic*, 8:404–423, 2002.

[9] Jaakko Hintikka. Negation in logic and in natural language. *Philosophy and Linguistics*, 25:585–600, 2002.

[10] Jaakko Hintikka. Quantum logic as a fragment of independence-friendly logic. *Journal of Philosophical Logic*, 31:197–209, 2002.

[11] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators I. *American Journal of Mathematics*, 73:891–939, 1951. Also in [17, vol. 3, pp. 369-419].

[12] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators II. *American Journal of Mathematics*, 74:127–162, 1952. Also in [17, vol. 3, pp. 421-458].

[13] John L. Kelly. *General Topology.* Van Nostrand, Princeton and New York, 1955.

[14] J.C.C. McKinsey and Alfred Tarski. The algebra of topology. *Annals of Mathematics*, 45:141–191, 1944. Also in [17, vol. 2, pp. 607-659].

[15] J.C.C. McKinsey and Alfred Tarski. On closed elements in closure algebras. *Annals of Mathematics*, 47:122–162, 1946. Also in [17, vol. 3, pp. 19-61].

[16] Ahti-Veikko Pietarinen and Gabriel Sandu. Partiality and games: propositional logic. *Logic Journal of the IGPL*, 9:107–127, 2001.

[17] Alfred Tarski. *Collected Papers*. Birkhäuser, Basel, 1986. Ed. by Stevent R. Givant and Ralph N. McKenzie.

[18] John von Neumann. Unsolved problems in mathematics. In Miklos Redei and Michael Stöltzner, editors, *John von Neumann and the Foundations of Quantum Physics*, pages 231–246. Kluwer Academic, Dordrecht, 2001. Originally presented 1954.

The importance and neglect of conceptual analysis: Hilbert-Ackermann iii.3

Wilfrid Hodges
w.hodges@qmul.ac.uk
School of Mathematical Sciences,
Queen Mary, University of London,
Mile End Road,
London E1 4NS

The famous text of Hilbert and Ackermann, *Grundzüge der Theoretischen Logik* [13], appeared first in 1928, but it was closely based on Hilbert's lectures in Göttingen in the period 1917–1922. In section iii.3 ('Preliminary Orientation on the Use of the Predicate Calculus') we find the following passage:

> We begin with the sentence ... "If there is a son, then there is a father." The symbolic rendering of this assertion in the predicate calculus is (1)
>
> $$(Ex)S(x) \to (Ex)F(x).$$
>
> ... A proof of this statement is possible only if we analyze conceptually (begrifflich zerlegen) the meanings of the two predicates which occur.

This passage is interesting because it tells us something about what Hilbert thought logic is for.

One of the chief tasks of the history of ideas is to recreate the agendas of leading researchers from past generations. With mathematics there is a particular problem: A mathematical fact, once discovered, has its own

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 129–153

identity and takes on its own life. Later mathematicians construct their own proofs of it and use it in contexts that its first discoverer never imagined. The fact itself carries no traces of how people first came to it. For this reason the body of theorems proved by the logicians of the first half of the twentieth century can be a misleading guide to the contexts in which they proved them.

Hilbert's example above is not a deep contribution to mathematical logic, but it does raise questions about his agenda. For example in his independence proofs in geometry [9] in 1899 he had shown that certain axioms A are not logically derivable from certain other axioms $B_1, \ldots, B_n$. But he had never stopped to ask whether this is just a fact about the surface forms of the axioms, so that A might become derivable from the other axioms if we first made some conceptual analyses. Patricia Blanchette has claimed that this was Frege's chief difficulty with Hilbert's independence proofs. Thus [1] p. 336:

> [For Frege] The fact that knowledge of the members of [a set of thoughts] Σ suffices to ground knowledge of [a thought] α is entirely compatible with the *in*dependence, in Hilbert's sense, of the sentence expressing α from those expressing the members of Σ. This will be the case whenever ... the epistemological dependence turns on analysis of the constituent concepts of the thoughts in question. (2)

(Blanchette is discussing the correspondence [10], [4] between Hilbert and Frege in 1900.) On this account, Hilbert's agenda in 1899 missed a crucial point that Frege noticed. By 1928 Hilbert had noticed it too, whether or not he learned it from Frege.

I think this account is wrong. There clearly are differences between Frege and Hilbert-1899, and between Hilbert-1899 and Hilbert-1928. But they have nothing to do with the role of conceptual analysis; on this there is no reason to think that Hilbert's views changed. There is plenty of evidence that in 1900 he and Frege had no disagreement about the relationship between conceptual analysis and logical inference in geometry.

I will discuss first the 1928 text, and then Frege's views in his debate with Hilbert (including his 1906 paper [7] responding to Korselt). Finally I will sketch a certain paradigm for mathematical work. Hilbert, Frege and Tarski (before the Second World War) all shared this paradigm, and many of their opinions fall neatly into place within it. Along the way I discuss a recent paper of Jamie Tappenden [17] on Frege's notion of logical independence; I think Tappenden has misreported what Frege says, but it gives us an excuse to put flesh on some of Frege's abstract remarks.

The word 'system' will appear in many of the quotations below. During this period it had two main meanings: (1) a system of axioms, i.e. a theory, (2) a system of things, i.e. a structure. It would be wise not to confuse these.

My warm thanks and appreciation to the organisers and my fellow participants in 'First Order Logic 75' at the Humboldt University, Berlin in September 2003. I thank Patricia Blanchette and Jamie Tappenden for corrections and comments.

1 What did Hilbert and Ackermann have in mind?

The passage quoted from Hilbert and Ackermann above refers back to a remark in an earlier section ([13] iii.1):

> "If there is a son, then there is a father," is certainly a logically self-evident [logisch selbstverständliche] assertion, and we may demand of any satisfactory logical calculus that it make obvious [in Evidenz setzt] this self-evidence, in the sense that the asserted connection will be seen, by means of the symbolic representation, to be a consequence of simple logical principles. (3)

(I thank William Ewald for confirming that this and the next quotation from the 1928 edition of [13] both correspond almost verbatim to material in Hilbert's Göttingen lectures.) There are some problems with this passage.

First, the example is not entirely convincing. Are Hilbert and Ackermann claiming that it is self-evident that every son has a father? Apparently so, from the analysis that they give for $S(x)$ ("x is a son") later in iii.3:

> In the concept "son" is contained the property "male," on the one hand, and, on the other, the relation of child to parents; in the concept "father," the relation to wife and child. Accordingly, if we introduce for "x is male" the symbol $M(x)$ and render the predicate "x and y are the parents of z" (or more exactly, "x and y as husband and wife have z as their child") by the symbol $C(x, y, z)$, then we define $S(x)$ by (4)
>
> $$M(x) \ \& \ (Eu)(Ev)C(u, v, x).$$

This is surely a mistake. There is nothing conceptually wrong in talking about a mother having a biological son without the intervention of a father. I don't know if it has ever happened, but in the present state of genetic engineering it probably will have done by the time you read this. If the example is about legal fathers and sons, then it is even less convincing.

That shows only that Hilbert and Ackermann chose a bad example. Curiously the example is very old; it was a stock-in-trade of the medieval western logicians, who took it from Boethius [2] Book 3, 1198A4. The usual medieval version was the other way round: If there is a father then there is a son. (We should congratulate the twelfth century *Introductiones Montane Minores* [16] p. 68 for noticing that this requires that sons include daughters—most of the medievals missed this.) The best logicians analysed the example in pretty much the same way as Hilbert and Ackermann did, and added that the resulting proof is not an aristotelian syllogism but a 'topic'.

The second problem is a little subtler. According to Hilbert and Ackermann, if S and C have the meanings above, then

$$\forall x(S(x) \leftrightarrow (Eu)(Ev)C(u, v, x)) \tag{5}$$

is a conceptual truth. But what is the domain that the quantifiers (Eu) and (Ev) range over? A son can outlive his parents, so it should include past as well as present humans.

We find an answer to this question in the second edition of [13]. Discussing the application of first-order logic to sentences that express facts about a particular subject matter, Hilbert and Ackermann comment (iii.11):

> When interpreting the formulas as regards content, we must bear in mind that the individual variables no longer refer, in general, to a domain of individuals which is left indeterminate; rather, the latter is usually more or less definitely determined by the nature of the premises, so that the individual members of the domain may perhaps be the integers, the real numbers, the points in a plane, or any other things whatsoever. (6)

Nothing corresponds to this text in the first edition. But the treatment of natural numbers in iii.3 of the first edition is entirely in line with (6), so that it seems the second edition makes explicit an idea that was implicit in the first. (Hilbert was hardly active in 1938; he died in 1943. I think we can assume that the new material in the 1938 edition bears the voice of Paul

Bernays, who was a very reliable expositor of Hilbert's ideas and sometimes had a clearer vision than Hilbert himself. But below I will keep to the fiction that Hilbert made these changes.)

The quotation (6) is also close to what Tarski says in [18], published in Polish two years before the second edition of [13]; see for example his treatment of the Universe of Discourse in chapter 4. Presumably when the subject is human relationships, one should take the domain to consist of all human beings at any time.

The connection between (6) and logical inference runs as follows. If a first-order sentence is deducible by a formal proof, then it is true when interpreted in any (nonempty) domain. Section iii.5 of the second edition of [13] explicitly sets up the logical calculus so as to ensure this, though no soundness theorem is stated. Again one guesses that the same idea is below the surface in the first edition; it peeps up briefly in the comments on the 'Individuenbereich' in iii.9. In any case, if a sentence is true when interpreted in any domain, then in particular it is true when interpreted in the domain that is 'more or less definitely determined by the nature of the premises'. Hence the sentence

$$(Ex)[M(x)\&(Eu)(Ev)C(u,v,x)] \rightarrow (Ex)(Ey)(Ez)C(x,y,z), \tag{7}$$

which has a formal proof in first-order logic, is true when M and C are interpreted as previously in the domain of human beings.

Recall that Hilbert and Ackermann's problem was to prove 'If there is a son, then there is a father'. This sentence is symbolised as

$$(Ex)S(x) \rightarrow (Ex)F(x). \tag{8}$$

But instead of proving this sentence, Hilbert and Ackermann prove (7). How does this solve the original problem?

To plug the gap, they first observe that

$$S(x) \text{ has the same meaning as } M(x)\&(Eu)(Ev)C(u,v,x)) \tag{9}$$

Here they use a symbol for 'has the same meaning as' which they introduced in i.2, explaining there that it is not one of the logical symbols. From (9) they deduce

$$(Ex)S(x) \rightarrow (Ex)(Eu)(Ev)C(u,v,x) \tag{10}$$

and the rest is plain sailing.

By modern practice (or indeed Frege's) this is rather sloppy. The status of (9) needs to be made clear. Today we would make it a formal definition, together with a similar definition for $F(x)$:

$$\begin{aligned} &\forall x(S(x) \leftrightarrow M(x) \wedge \exists u \exists v C(u,v,x)), \\ &\forall x(F(x) \leftrightarrow \exists y \exists z C(x,y,z)). \end{aligned} \tag{11}$$

One can deduce (8) logically from these two definitions (11) together with (7).

We are not quite finished with the Hilbert-Ackermann example. In (3) they claimed that (8) is logically self-evident; so why do they want a proof of it? Why not adopt it as an axiom about family relationships? As Tarski says ([18], p. 93f of original, §39 of English translations):

> [When] we decide to select a certain system of primitive terms and axioms ... it may be desirable to get along with as few of them as possible ... (12)

Since Hilbert and Ackermann have nothing to say about this, I leave it till the final section below.

2 Axiomatic theories and models

Hilbert and Ackermann might seem to contradict themselves when they say first that the domain of individuals 'is usually more or less definitely determined by the nature of the premises', and then add that the individual members of the domain 'may perhaps be ... any other thing'. But this is an ungenerous reading of their text. They must mean that while a given set of premises determines a particular domain of individuals, different sets of premises could include anything whatever in the domains that they determine.

But then how can Hilbert and Ackermann claim that the domain of a model of a set of first-order sentences is 'more or less definitely determined by' the first-order sentences? They can't. Hilbert was already clear about this in his correspondence with Frege ([10] 29 December 1899):

> In other words, each and every theory can always be applied to infinitely many systems of basic elements. For one merely has to apply a univocal and reversible one-to-one transformation and stipulate that the axioms for the transformed things be correspondingly similar. ... The totality of assertions of a theory of electricity does of course hold of every other system of things substituted in place of the concepts magnetism, electricity, ... [sic] just as long as the required axioms are fulfilled. (13)

Hilbert is certainly right about this. So doesn't it explicitly refute his later claim that the premises determine the domain?

A closer reading shows that in both 1899 and 1938 Hilbert has in mind a set of formal axioms written down *as a set of premises for reasoning about a particular subject-matter*. To this extent his formal sentences are not in an uninterpreted formal language; they are meaningful statements about a determinate topic. It's this subject matter, not the syntactic form of the sentences, that determines the domain of individuals. A model theorist today would explain that Hilbert's axioms have an *intended model*.

Hilbert seems not to have had the modern notion of a model of a set of axioms. Two other notions take its place in different contexts. The first he explains as follows. Here he is talking about a system of axioms that presumably already has an intended subject matter ([10] 29 December 1899).

> If I think of my points as some system or other of things, e.g. the system of love, of law, or of chimney sweeps ... [sic] and then conceive of all my axioms as relations between these things, then my theorems, e.g. the Pythagorean one, will hold of these things as well. (14)

Hilbert says not that the axioms can be true of structures made up of any individuals; he says that *if we think of the axioms as statements about* these structures, then (under appropriate conditions) they will be true. The interpretation is a matter of our intentions, our thoughts.

The second notion appears when Hilbert (Bernays?) discusses consistency and validity of a set of axioms, for example in iii.5 of the second edition of [13]. Here instead of interpreting the primitive symbols, he *replaces* them by names or predicates, in general from another language. In model theory we call the result not a model of the axioms but an *interpretation* of them.

Hilbert's language in both 1899 and 1938 is again close to Tarski's in 1936 [18]. Tarski requires that a set of axioms for a deductive theory should be meaningful and evidently true (see §32 of 1936, §36 of the English). Tarski realises (as Hilbert seems not to) that this account is nonsensical when applied to axiomatic theories like group theory, where we have no particular group in mind. He admits that his treatment is inadequate here, but offers no repairs ([18] §33 of original, §38 of English translation).

> Admittedly, sometimes we develop a deductive theory without ascribing a definite meaning to its primitive terms, thus dealing with the latter as with variables; under such circumstances we say that we treat the theory as a FORMAL SYSTEM. But this kind of situation (which was not taken into account in our general characterization of deductive theories ...) arises only if several interpretations for the axiom system of this theory are available to us, that is, if we are concerned with several ways of ascribing concrete meaning to the terms occurring in the theory, but we do not desire to give preference in advance to any one of these ways. (15)

By the mid 1950s model theory was well equipped to describe how axioms for groups work. In [14] I trace the steps which led to the breakdown, during the 1940s, of the standard approach to axiom systems before that date. Basically, the model theorists (Mal'tsev, Tarski, Abraham Robinson) had started to prove results that it was difficult to state in the old framework without tiresome ad hoc adjustments. In reading what Hilbert and Tarski said before the mid century we should beware of the anachronism of supposing that they were talking model theory.

In later editions of the English translation of [18] the word 'interpretations' in (15) was expanded to 'models or interpretations'. This is correct in the modern use of the word 'model'. But it is wrong as Tarski defined 'model' in [18]; in fact that is part of the point Tarski is making here.

3 Frege on conceptual analysis

As Blanchette rightly mentions, for Frege an inference is primarily between thoughts and only derivatively between sentences expressing the thoughts.

She also says ([1] p. 325):

> Because each thought is generally expressible by a number of different sentences, there is much more to the relation of *provability* than is evidenced by the relation of *derivability*. Where p is a thought and s a sentence expressing it, Π a set of thoughts and Σ a set of sentences expressing Π: The derivability of s from Σ guarantees that p is a consequence of Π, but the fact that s is not derivable from Σ is no guarantee that p is not a consequence of Π. For s's nonderivability from Σ is entirely compatible with the existence of some s' and Σ' expressing p and Π, respectively, such that s' is derivable from Σ'. (16)

Frege doesn't say this in the passages she cites. Did he believe it?

On page 210 of *Logic in Mathematics* [8], Frege addresses this issue in the case where s' and Σ' differ from s and Σ by using a definition of some concept occurring in the thoughts p, Π. He distinguishes two cases.

> (1) We construct a sense out of its constituents and introduce an entirely new sign to express this sense. (17)

For example this case would arise if Hilbert and Ackermann had first proved (7) with M and C having the meanings assigned to them, and then *introduced* $S(x)$ *as an abbreviation of* $M(x)\&(Eu)(Ev)C(u, v, x)$. In this case (8) would express the same thought as (7) by stipulation. But then we should be allowed to use the definition of S in deductions, otherwise what would be the point of the definition? So presumably (8) is no less provable than (7).

The second case is where

> (2) We have a simple sign with a long established use. We believe that we can give a logical analysis of its sense. ... what we should here like to call a definition is really to be regarded as an axiom. (18)

This is the case discussed by Hilbert and Ackermann. We analyse 'son' and 'father', and we see ('by an immediate insight', as Frege puts it) that the definitions of these concepts in terms of M and C are true. In this case it is not at all clear to me why Frege should regard (8) and (7) as expressing the same thoughts, since the two definitions (11), which are needed to get from one of (7) and (8) to the other, express significant thoughts and not just a notational convenience. But the key point in this case is that Frege regards the conceptual analysis not as providing new sentences s' and Σ' to express the same thoughts as before, but as supplying a new premise in the form an axiom expressing a relation between concepts.

On the previous page of [8], Frege suggests that a conceptual analysis can 'reduce the number of axioms', and waits a page before admitting that it will normally add a new axiom as well. (Again compare Tarski's (12).)

Blanchette's claim (2) is about the conceptual analysis of concepts that we already have, so (18) applies rather than (17). The only remark of Frege about conceptual analysis that she cites ([1] p. 324) from the geometric work of 1900–1906 is the third sentence of the following passage from Frege's answer to Korselt ([7] p. 303):

> The real importance of a definition lies in its logical construction out of primitive elements. And for that reason we should not do without it, not even in a case like this. The insight it permits into the logical structure is not only valuable in itself, but also is a condition for insight into the logical linkage of truths. A definition is a constituent of the system of a science. As soon as the stipulation it makes is accepted, the explained sign becomes known and the proposition explaining it becomes an assertion. The self-evident truth it contains will now appear in the system as a premise of inferences. (19)

Here he says 'stipulation', and the preceding half-page makes it clear that he is talking about new concepts being introduced as abbreviations of longer phrases. So here (17) applies—there is no reference to analysis of existing concepts. The 'insight into the logical linkage of truths' is presumably the insight given by a well-chosen notation—something that writers on mathematics have often commented on.

Another passage from the reply to Korselt is relevant ([7] p. 423):

> We must first ask what is here to be understood by 'indepen- (20) dence' ... If we take the words "point" and "straight line" in Hilbert's so-called Axiom II.1 in the proper Euclidean sense, and similarly the words "lie" and "between," then we obtain a proposition that has a sense, and we can acknowledge the thought expressed therein as a real axiom. Let us designate it by "[II.1]". Let [II.2] emerge in a similar way from Hilbert's II.2. Now if one has acknowledged [II.1] as true, one has grasped the sense of the words "point," "straight line," "lie," "between"; and from this the truth of [II.2] immediately follows, so that one will be unable to avoid acknowledging the latter as well. Thus one could call [II.2] dependent upon [II.1]. Of course, we do not have an inference here; and it seems inexpedient to use the word "dependent" in this way, even though linguistically it might be possible.

Frege seems to be saying here that one could speak of an epistemological 'dependence' between two axioms, but he finds this an inexpedient use of the word. Presumably he would allow us to call [II.2] dependent on [II.1] if the conceptual relations between 'point' etc. that are used in reaching [II.2] were stated explicitly as axioms and [II.2] was deduced from them together with [II.1]. But it might be that [II.2] is already a statement of these relations, so [II.2] makes a satisfactory axiom itself.

From Frege's writings of the period 1899-1906, (19) is the only passage that Blanchette cites as evidence of Frege's interest in conceptual analysis. In context it is part of an attack on Hilbert's notion of definitions, and not a request for conceptual analyses of 'point' etc. The fact that in the whole correspondence with Hilbert and the sixty-odd pages of [6] and [7] Frege never once says that he finds Hilbert's independence proofs unacceptable because they ignore the possibilities of conceptual analysis is one strong argument against Blanchette's thesis. Another is Frege's own account of how he would like to see independence proofs carried out. We turn to this.

4 Frege's independence proofs

Frege finishes his response to Korselt [7] by sketching how he thinks a proof of the independence of a geometric axiom from other geometric axioms should go. First he defines what he means by 'dependent' (p. 423f).

> Let Ω be a group of true thoughts. Let a thought G follow from one or several of the thoughts of this group by means of a logical inference such that apart from the laws of logic, no proposition not belonging to Ω is used. Let us now form a new group of thoughts by adding the thought G to the group Ω. Call what we have just performed a logical step. Now if through a sequence of such steps, where every step takes the result of the preceding one as its basis, we can reach a group of thoughts that contains the thought A, then we call A dependent upon group Ω. If this is not possible, then we call A independent of Ω. The latter will always occur when A is false. (21)

In particular dependence can never rely on any conceptual analysis that is not already included as a thought in Ω. As we saw, conceptual analysis gives us new consequences only by adding a new thought to the group Ω, and here Frege forbids us to do that.

Frege remarks that in order to show that A is not in this sense a consequence of Ω, we need some facts about consequences. One fact that we can call on is (p. 426)

> If the thought G follows from the thoughts A, B, C by a logical inference, then G is true. (22)

(Recall that for Frege a logical inference must have true premises.) Facts of this type won't allow us to prove the independence of a true conclusion.

So Frege proposes a further law. Suppose we choose a logically perfect language $\mathcal{L}$ in which Ω and A are expressible, say by sentences Σ and s (to follow Blanchette's notation). Suppose that in the language $\mathcal{L}$ we can choose, for each 'word' w occurring in Σ or s, a corresponding 'word' $f(w)$ of the same grammatical form, in such a way that f is one-to-one. Suppose also that $f(w) = w$ whenever w is a word 'whose reference belongs to logic' (deren Bedeutungen der Logik angehören). Let Σ' and s' be the result of replacing each occurrence of a word w in Σ and s by an occurrence of the word $f(w)$. Since $\mathcal{L}$ is logically perfect, Σ' and s' will express a set of thoughts and a thought. Suppose we can do all this in such a way that the thoughts expressed by Σ' are true and the thought expressed by s' is false. Then as noted in (22), there is no inference from Σ' to s'. It follows that there is no inference from Σ to s either, since otherwise the steps in the proof would translate into the steps of a proof of s' from Σ'.

Several things in this account are unclear. Frege himself notes that in order to use this method one should first define 'logical inference' and what concepts 'belong to logic', and then one should tidy up the account of the further law above.

Let me add another thing to tidy up. Frege seems to assume that one could never get a false independence proof by choosing the wrong language $\mathcal{L}$. This is perplexing. Suppose for example that Ω consists of three thoughts, which we shall express by sentence letters p, q, r, and A is a thought which we express by the sentence letter s. Then in $\mathcal{L}$ the inference from Ω to A becomes

$$p, q, r \ \vdash \ s \tag{23}$$

and we easily show by Frege's method that this inference is invalid, regardless of what Ω and A were.

We are in the dark about how Frege would block this. Since Frege says nothing about it, I suspect he has in mind a solution that doesn't use any radically new ideas. One possibility is that Frege believes something along the following lines:

> If sentences S and T in logically perfect languages express the same thought, then one can set up a bijection β between some components of S and some components of T, in such a way that (1) if two components are related by β, then either they are synonymous words, or one of them is a compound expression and the other is a word that by stipulation means the same as the compound expression, and (2) both S and T are built up in the same way from the components related by β. (24)

If he believes this, then all that's needed to repair the 'further law' above is a requirement that all stipulative abbreviations have to be expanded before the map f is found. But for myself I am out of sympathy with Frege's talk of 'thoughts', so I leave it there.

5 Did Frege's views on independence change?

According to Tappenden [17] p. 283:

> Prior to 1906, Frege takes "independence" to involve assuming an axiom to be false. Frege himself defines "independence" this way in the correspondence of 1899–1900, and this definition also informs the discussion of independence in the *Grundlagen* ... sixteen years earlier. When Frege turns, in 1906, to sketch how he has come to think "independence" should *properly* be defined to carry out independence arguments among thoughts, he defines it differently: A thought T is independent of others Ω if there is no sequence of logical steps leading from Ω to T. (25)

So according to Tappenden, Frege has two different 'notions of independence' (Tappenden's phrase); we can call them the pre-1906 notion and the 1906 notion.

It seems to me that Tappenden's statement about the correspondence of 1899-1900 is false; Frege makes no reference to assumptions anywhere in this correspondence. On page 273 Tappenden [17] attributes to Frege (writing to Liebmann in 1900) a definition of independence, and it does refer to what "can be assumed". But Tappenden introduced these three words himself; Frege's original says nothing about assumptions.

True, there is a difference between the definition Frege sent to Liebmann and the one he gave in 1906. To Liebmann he said ([5] 29 July 1900)

> The independence of an axiom A from other axioms is the freedom from contradiction of the negation of A together with the other axioms. (26)

In other words, if the other axioms are $B_1, \ldots, B_n$, then to say that A is independent of $B_1, \ldots, B_n$ is to say that there is no logical proof of

$$((\neg A) \land B_1 \land \ldots \land B_n) \to (\chi \land \neg\chi) \tag{27}$$

for any statement χ. The definition in 1906 says that A is independent of $B_1, \ldots, B_n$ if there is no logical proof of

$$(B_1 \land \ldots \land B_n) \to A. \tag{28}$$

These two definitions are equivalent in the sense that it takes only a small amount of propositional logic to deduce either definition from the other.

Any argument that established independence in one sense would count as establishing independence in the other sense too.

So Tappenden has given no evidence that Frege is using different notions of independence in 1900 and in 1906. But Tappenden also refers to a passage of the *Grundlagen* of 1884. I quote Tappenden again ([17] p. 273):

> Whatever "conceptual thought" is, *independence* as Frege understands it [in the *Grundlagen*] involves showing that a set consisting of several axioms plus the negation of another can be used in reasoning without turning up contradictions. Also ... he says *directly* that the fact that "it is possible to deny any of the axioms" of Euclidean geometry "shows (*zeigt*) that the axioms of geometry are independent of one another and of the primitive laws of logic ...". (29)

Here is what Frege says in the place Tappenden is referring to ([3] §14):

> Empirical propositions hold good of what is physically or psychologically actual, the truths of geometry govern all that is spatially intuitable, whether actual or product of our fancy. ... Conceptual thought alone can after a fashion shake off [the yoke of intuition], when it assumes [annimt], say, a space of four dimensions or positive curvature. To study such conceptions is not useless by any means; but it is to leave the ground of intuition entirely behind. ... For purposes of conceptual thought we can always assume [annehmen] the contrary of some one or other [diesem oder jenem] of the geometrical axioms, without entangling ourselves [mit sich selbst verwickelt] in any self-contradictions when we proceed to our deductions, despite the conflict between our assumptions and our intuition. The fact that this is possible shows that the axioms of geometry are independent of one another and of the primitive laws of logic, and consequently are synthetic. (30)

Frege admittedly needs some interpretation here. But straight away we note two mathematical points that Tappenden's statement (29) misconstrues.

First, Frege doesn't say that the axioms of geometry are independent of one another. How could he? It depends on what set of axioms one chooses. The phrase 'the axioms of geometry' in his last sentence must refer back to the 'some one or other of the geometrical axioms' in the preceding sentence, which Tappenden has mistranslated as 'any of the axioms'. Frege is only claiming that this method for proving independence works in the cases where it works.

Second, the fact that it is possible to deny an axiom A can't conceivably establish anything about the independence of A. Words are cheap; anybody can deny anything. It's equally hopeless to try to establish the independence of A by denying it 'without turning up contradictions'. So Frege can't have meant what (29) ascribes to him. What Frege must have meant—from the logical facts of the case—is that if we adopt the negation of A, together with the other axioms, we *can't* deduce a contradiction from them. What he actually says is more ambiguous and metaphorical than this, to the effect that adopting the negation of A together with the other axioms won't 'entangle us' with contradictions.

This second point is important. If we want to *prove* that A is independent of $B_1, \ldots, B_n$ along the lines Frege is discussing, we shall need to *prove* that the assumption of $\neg A$ together with $B_1, \ldots, B_n$ never leads to a contradiction. It is quite obvious that we could never prove such a thing just by assuming these propositions and failing to deduce a contradiction. Something else is needed.

And before we examine Frege's text to see what that something else is, we should note that the question whether $\neg A$ together with $B_1, \ldots, B_n$ leads to a contradiction has nothing to do with whether any human being happens to assume any of them. We can't make a contradictory set consistent, or vice versa, just by not thinking about it. So Frege's brief reference to 'assuming' must be a turn of phrase and nothing essential to his argument.

It's time to look at the earlier parts of Frege's text quoted above. Frege notes that the truths of geometry are true propositions about the space of our geometrical intuitions, let me call it *intuitive space*. We can safely assume that the truths that Frege has in mind either include or imply the statement

$$\text{The number of dimensions is three.} \tag{31}$$

Then he speaks of 'assuming a space of four dimensions'. If a mathematical colleague asked me to 'assume a space of four dimensions', my first guess would be that he wanted me to think of an example of such a space. Like most mathematicians I would think first of euclidean 4-space $\mathbb{R}^4$; any serious European geometer in the 1880s would have been familiar with this space, though not in this notation. (I think in this context 'takes' is a more idiomatic rendering of 'annimt' than Austin's 'assumes'; but 'assumes' will do.)

Frege is right that we can't work out the properties of $\mathbb{R}^4$ by any kind of mental inspection or intuition. We have to reason algebraically, abstractly, or as Frege puts it, by using 'conceptual thought'. Conceptual thought shows for example that the standard axioms of vector spaces (which again were familiar at the time, though not under that name) hold in $\mathbb{R}^4$. Presumably they also hold in intuitive space, though Frege thinks we establish this by intuition.

Since $\mathbb{R}^4$ has four dimensions, the following statement about it is also true:

$$\neg(\text{The number of dimensions is three}). \tag{32}$$

Here (32) both is and isn't the negation of (31). Syntactically it is the negation; but it doesn't express the negation of what (31) expresses, because (31) expressed a truth about intuitive space and (32) expresses a truth about $\mathbb{R}^4$. Because (32) and the vector space axioms are all of them true statements about $\mathbb{R}^4$, they form a syntactically consistent set. This establishes that (31) is independent of the vector space axioms. Frege's passage (30) is pretty clearly about examples like this.

Is this a different notion of independence from that in 1906? Plainly not; it establishes independence in the 1906 sense. Is it a different *method* for proving independence? Again I think not. Let Ω be the vector space axioms as true propositions about intuitive space, and let A be (31) as a true proposition about intuitive space. Now take the language used to express Ω and A, and add to it a corresponding set of concepts and terms that are about $\mathbb{R}^4$ instead of intuitive space. Replace each geometric expression in the statements of Ω and A by their counterparts; this turns Ω into the true statement that $\mathbb{R}^4$ obeys the vector space axioms, and A into the false statement that $\mathbb{R}^4$ is three-dimensional. Ergo.

For emphasis let me repeat that 'assuming a four-dimensional space' doesn't involve assuming a false proposition. It does involve accepting a *true* proposition that is syntactically identical to the negation of a true proposition about something else. As Frege says in his letter of 6 January 1900 to Hilbert [4],

> Only the wording is the same; the thought-content is different in each different geometry. (33)

Hence Tappenden's concerns in [17] about arguments from false assumptions are beside the point. They have nothing to do with Frege's views on independence, at any date.

We should mention Frege's doubt, expressed to Hilbert [4] and Liebmann [5] in 1900, about whether it would be possible to prove the mutual independence of the axioms of euclidean geometry by this method. He is reserved about the reasons for his doubt, but his statement to Liebmann does give some clues. His worry is that it might be impossible to find a structure in which (a statement syntactically the same as) A is false while (statements syntactically the same as) $B_1, \ldots, B_n$ are true, even if there is no logical proof of A from $B_1, \ldots, B_n$, because the form of $B_1, \ldots, B_n$ might impose a 'Beschränkung' that rules out the required countermodels. In modern terms, he might have had in mind a second-order completeness axiom that rules out models with nonstandard elements. It would have been very interesting to know more, because Frege could have been feeling his way towards the question whether semantic entailment in the Tarski sense necessarily implies syntactic entailment. (Today we know that it can't in second-order logic, because the set of valid sentences of second-order logic is not computably enumerable. But in 1900 Gödel was still thirty-one years in the future.)

6 The broader picture

Frege, Hilbert and Tarski were all deeply interested in the question how one should formulate mathematical theorems, with a view both to rigour and to fruitful research. All three adopted essentially the same paradigm, which Tarski ([18] §36 of the English version) traced back to Pascal [15]; though Tarski himself became reticent about this paradigm after the Second World War. Today the paradigm is largely forgotten, and so we easily misconstrue remarks that refer to it. Hilbert spoke of the paradigm in terms of 'theories' and 'axiomatic thought' [10], while Tarski wrote of 'deductive theories' and 'deductive method' [18]. Frege sometimes used the expression 'system of a science'.

(a) The primitives. We assume we know what subject matter we intend to discuss. The first step is to decide what objects we want to talk about (this will determine the domain of individuals, cf. Hilbert and Ackermann (6) above), and what concepts we need in order to talk about them. We adopt expressions for these concepts; Pascal and Tarski call these expressions *prim-*

itive. The meanings of the primitive expressions should be self-explanatory (Tarski: They should 'seem to us to be immediately understandable'). Frege ([4] 6 January 1900) invites Hilbert to agree with him that it is unacceptable to leave the primitive concepts only partly specified (nicht fertig) in hopes of clarifying them as the work proceeds. (Frege adds that he had thought he was alone in this opinion.)

(b) The axioms. The second step is to write down self-evident truths that relate these concepts; all three logicians describe the sentences expressing these truths as *axioms*. (Hilbert: These concepts and the stated relations between them form the 'framework of concepts' for our theory.)

In [10] 29 December 1899, Hilbert says 'As soon as I posit an axiom, it is available and "true" '. This is absurd; no wonder Frege objected. Hilbert was more circumspect in later writings. But the position is complicated by the fact that Hilbert wanted to include physical theories under the head of axiomatic theories. 'Self-evidence' is not a criterion commonly used for choosing basic principles in physics.

(c) Internal work. All three logicians, and Pascal too, draw a distinction between work within the theory and work that changes the theory. Thus for example Pascal says he is going to discuss how we prove truths, but not how we discover them. Introducing my own terminology here, let me distinguish between internal work within the theory and external work on the theory.

Work within the theory is severely constrained. We are allowed to do just two things. As Pascal puts it, we are allowed first to use terms whose meaning has been clearly stated, and second to assert propositions that we deduce from truths already known. That's all. The processes are cumulative; a definition can use terms introduced by earlier definitions and be justified by propositions already proved, and a proof can use terms already introduced and be proved from propositions already proved. Pascal also requires that each proof should be 'conforme aux règles qu'on connaît'.

Frege and Tarski bring this up to date: We are allowed only to introduce new terms or concepts by definition, and to introduce new propositions by sequences of steps in a formal deductive calculus; the deductive calculus would be what today we would count as a logic of finite order, not necessarily first order. Hilbert is less explicit, at least in [11]. His programme for metamathematics requires that it should in principle be possible to formalise any mathematical argument in a suitable calculus (‘Everything that previously made up mathematics is to be rigorously formalized’, [12]). But he clearly feels that the threats to rigour come more from choice of inappropriate axioms than from unacceptable inference steps. Hence his concern with the consistency of the axioms of arithmetic and set theory. Both Tarski and Hilbert experiment with rules of logic that have infinitely many premises.

Frege’s (17) above is about definitions internal to a theory.

(d) Deepening the theory. So we settle down to prove theorems. Inevitably we find that the concepts we first wrote down are too crude; to carry the proof through we need to analyse one of them. Or maybe we have a proof, but we see a way of doing it better, using different concepts and axioms. Either way, alteration of the primitives or the axioms is not an allowed operation within the theory; so at this point we step outside the theory and revise it. Hilbert [11] describes this kind of reconstruction of the theory as ‘deepening of the foundations’.

A common feature of deepenings is that we introduce new concepts and define one of the primitives of our theory in terms of them. This is the situation that Frege describes at (18), and it is also the situation of Hilbert and Ackermann (1). It leaves us with three choices. The first is to stick with our original primitives and abandon the deepening; the second is to add the new concepts as primitives and include the definition of the old primitive as a new axiom; the third is to add the new concepts as primitives, strike out the old primitive and then introduce the old primitive by a definition as described under (c) above. Frege [8] takes a strong stand against the third option: It implies that we could have added the original primitive by stipulation, contradicting the fact that the primitive term already had a sense attached to it.

Whether we prefer the second or the third option, there remains the first. This brings us back to the question at the end of section 1 above: What do Hilbert and Ackermann gain by adding primitives M and C and defining S and F in terms of them, rather than simply adding (8) as a new axiom? One can show that the two definitions (11) have no more consequences than (8) in the original primitives S and F, so there is no real gain in theorems. In such questions I suppose one has to follow one’s nose. But I think we can say a bit more. Suppose that our theory is going to be developed with further axioms. The definitions (11) will allow us to replace S and F by other primitives in any sentence, precisely because they are definitions. So they allow us to derive new statements using S and F from axioms using other primitives. The statement (8) is not a definition, so it allows us to get new information in terms of S and F only when these terms occur in suitable contexts. Hence, broadly speaking, an added definition is a better investment than a new axiom of another form.

This is the wrong place to pursue such thoughts further. But it's the right place to point out a likely misunderstanding on Hilbert's part. Suppose we follow the second option described above, and add (11) as new axioms. Then the definitions have the function of stating a relationship between the primitives. Hilbert seems to have thought it was safe to generalise this idea, and consider any axiom as a definition; after all, any axiom states a relationship between the primitives. Ideas of this kind seem to be one component of his confused notion of definitions in 1899/1900; Frege was merciless.

(e) Dependence and independence. Already Pascal mentions the question of determining what is and what is not derivable from the axioms of a theory. This is an external question, but Pascal notes that at least when something is derivable, a proof within the theory will demonstrate this. Pascal has no suggestions on how to demonstrate underivability.

We have seen that this question mattered to Frege and Hilbert, and anybody familiar with Tarski's work on logical consequence [19] knows that it mattered to him too. Hilbert's most distinctive contribution was to show that one can sometimes prove underivability by a purely syntactic argument. This was a central theme of his metamathematics, but it came some years later than his debate with Frege. In [4] 6 January 1900, Frege was still able to say

> What means do we have for proving that certain properties or requirements ... do not contradict one another? The only way I know of is to present an object that has all of these properties, to exhibit a case where all these requirements are fulfilled. Surely it is impossible to prove consistency in any other way. (34)

Apart from Hilbert's syntactic breakthrough, all three logicians had just the one method for showing that a meaningful sentence A is not logically derivable from meaningful sentences $B_1, \ldots, B_n$. The method was to find a set of meaningful sentences $A', B'_1, \ldots, B'_n$ that are syntactically isomorphic to $A, B_1, \ldots, B_n$ and equal in their logical expressions, such that A' is false and $B'_1, \ldots, B'_n$ are true. Today we would say that this establishes that A is not a semantic consequence of $B_1, \ldots, B_n$, and hence it is not a syntactic consequence in any sound proof calculus.

From this point onwards there are some significant differences. Hilbert in 1899/1900 is happy if $A', B'_1, \ldots, B'_n$ are syntactically the same sentences as $A, B_1, \ldots, B_n$ but reinterpreted. We saw that Frege toyed with the same possibility in [3] §14. But Frege had a very strong antipathy to using the same symbols in different meanings on different occasions, and by the time he came to write [7] in 1906 he made it explicit that A' etc. should be distinct from A etc. (To be precise, he required the symbols in A' etc. to be in the same language as those in A etc. but with different senses.)

Frege's later insistence on using different symbols might have been provoked in part by Hilbert's idea that one could switch the meaning of a symbol just by having different thoughts in connection with it (see (14) above), and Korselt's unclarity about how 'interpretations' are made. If we reckon that the non-validity of $B_1 \wedge \ldots \wedge B_n \to A$ establishes the underivability of A from $B_1, \ldots, B_n$, then Hilbert (Bernays?) in 1938 ([13] iii.5 of the second edition) goes with Frege and makes substitutions in $A, B_1, \ldots, B_n$ to find suitable $A', B'_1, \ldots, B'_n$, though without Frege's unnecessary requirement that the substitutions are one-to-one.

Tarski's account of semantic consequence in [18] and [19] also agrees with Frege in that Tarski substitutes new expressions for the non-logical expressions. But here he diverges onto a new track: The substituted expressions are variables, and the replacement concepts need not be in a language at all. Rather they are objects, and Tarski's requirement is that some assignment of these objects to the new variables satisfies the relevant formulas. Also Tarski drops Frege's requirement that the substitutions should be one-to-one.

Why did Frege require the substitution to be one-to-one? He was thinking syntactically: If the sentences on both sides are syntactically isomorphic, then any valid proof using the first group sentences will translate into a valid proof using the second. If the substitution wasn't one-to-one, this might no longer be true, because some rules of proof require that certain symbols are distinct.

Finally, Tarski would have to disagree with Hilbert and Ackermann about (8). For him, (8) is not a logical truth at all, because we can find interpretations of S and F that make it false. The fact that S and F are definitional abbreviations as in (11) is neither here nor there. So he could hardly have described (8) as 'logically self-evident', even though it is only a notational variant of 'If there is a son then there is a father' (3). As far as I can see, this is a purely terminological difference between Tarski and Hilbert. Tarski wishes to restrict 'logical consequence' to consequence internally within a deductive theory, whereas Hilbert reads it more broadly. The term 'logical consequence' doesn't have a long enough pedigree for it to make sense to ask which of them was right. In fact one is hard pressed to find any occurrences of this precise phrase before Tarski's 1936 paper.

As we saw in our discussion of (23), Frege's position on the issue is unclear. Possibly he would have relaxed Tarski's position by requiring us to expand all stipulative definitions before making the substitutions; this would increase the number of logical consequences by including (8), for example.

7 Bibliography

[1] Patricia A. Blanchette, 'Frege and Hilbert on Consistency', *Journal of Philosophy* 93 (1996) 317–336.

[2] Manlius Boethius, *De Topicis Differentiis*, ed. Eleonore Stump, Cornell University Press, Ithaca 1978.

[3] Gottlob Frege, *Die Grundlagen der Arithmetik*, Koehner, Breslau 1884; translated by J. L. Austin as *The Foundations of Arithmetic*, Blackwell, Oxford 1959.

[4] Gottlob Frege, Frege an Hilbert 27.12.1899, 6.1.1900, in *Gottlob Freges Briefwechsel*, ed. G. Gabriel, F. Kambartel and C. Thiel, Felix Meiner, Hamburg 1980, pp. 6–20.

[5] Gottlob Frege, Frege an Liebmann 29.7.1900, 25.8.1900, in *Gottlob Freges Briefwechsel*, ed. G. Gabriel, F. Kambartel and C. Thiel, Felix Meiner, Hamburg 1980, pp. 25–29.

[6] Gottlob Frege, 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematiker-Vereinigung* 12 (1903) 319–324, 368–375; translated as 'On the Foundations of Geometry' in *Gottlob Frege: On the Foundations of Geometry and Formal Theories of Arithmetic*, ed. Eike-Henner W. Kluge, Yale University Press, New Haven 1971, pp. 22–37.

[7] Gottlob Frege, 'Über die Grundlagen der Geometrie', *Jahresbericht der deutschen Mathematiker-Vereinigung* 15 (1906) 293–309, 377–403, 423–430; translated as 'On the Foundations of Geometry' in *Gottlob Frege: On the Foundations of Geometry and Formal Theories of Arithmetic*, ed. Eike-Henner W. Kluge, Yale University Press, New Haven 1971, pp. 49–112.

[8] Gottlob Frege, 'Logic in Mathematics' (unpublished 1914), translated in *Gottlob Frege, Posthumous Writings*, ed. Hans Hermes, Friedrich Kambartel and Friedrich Kaulbach, Blackwell, Oxford 1979, pp. 203–250.

[9] David Hilbert, *Grundlagen der Geometrie*, Teubner, Leipzig 1899.

[10] David Hilbert, Hilbert an Frege 29.12.1899, 15.1.1900, 22.9.1900, in *Gottlob Freges Briefwechsel*, ed. G. Gabriel, F. Kambartel and C. Thiel, Felix Meiner, Hamburg 1980, pp. 11–23.

[11] David Hilbert, 'Axiomatisches Denken', *Mathematische Annalen* 78 (1918) 405–415; translated as 'Axiomatic Thought' in *From Kant to Hilbert* Vol. II, ed. William Ewald, Clarendon Press, Oxford 1996, pp. 1107–1115.

[12] David Hilbert, 'Die logische Grundlagen der Mathematik', *Mathematische Annalen* 88 (1923) 151–165; translated as 'Foundations of Mathematics' in *From Kant to Hilbert* Vol. II, ed. William Ewald, Clarendon Press, Oxford 1996, pp. 1136–1148.

[13] David Hilbert and Wilhelm Ackermann, *Grundzüge der Theoretischen Logik*, Springer, Berlin 1928; translated from the 2nd edition 1938 as *Principles of Mathematical Logic*, Chelsea, New York 1950.

[14] Wilfrid Hodges, 'What languages have Tarski truth definitions?', *Annals of Pure and Applied Logic* 126 (2004) 93–113.

[15] Blaise Pascal, *De l'esprit géométrique*, in *Oeuvres Complètes*, Éditions du Seuil, Paris 1963, pp. 348–355.

[16] L. M. de Rijk, *Logica Modernorum* ii 2, Van Gorcum, Assen 1967.

[17] Jamie Tappenden, 'Frege on axioms, indirect proof, and independence arguments in geometry: Did Frege reject independence arguments?', *Notre Dame Journal of Formal Logic* 41 (2000) 271–315.

[18] Alfred Tarski, *O logice matematycznej i metodzie dedukcyjnej*, Biblioteczka Mat. 3–5, Książnica-Atlas, Lwów and Warsaw 1936; English revised translation 'Introduction to Logic and to the Methodology of the Deductive Sciences', Oxford University Press, New York 1941.

[19] Alfred Tarski, 'O pojęciu wynikania logicznego', *Przegląd Filozoficzny* 39 (1936) 58–68; reprinted pp. 186–202 of Alfred Tarski, *Pisma logiczno-filozoficzne, Tom 1 'Prawda'*, ed. Jan Zygmunt, Wydawnictwo Naukowe PWN, Warsaw 1995; English translation 'On the concept of logical consequence', pp. 409–420 of Alfred Tarski, *Logic, Semantics, Metamathematics: papers from 1923 to 1938*, ed. John Corcoran, Hackett Publishing Company, Indianapolis, Indiana 1983.

Notes on Substitution in First–Order Logic

Marcus Kracht
kracht@humnet.ucla.edu
Department of Linguistics, UCLA, 3125 Campbell Hall, PO Box 951543, Los Angeles, California, CA 90095-1543, USA

1 Introduction

Axiomatisations of first–order logic (henceforth FOL) are given by means of a finite list of axioms or axiom schemes, each of which represents an infinite set of actual formulae. This latter set is either derived by means of a rule of substitution into an actual axiom or by properly instantiating the axiom scheme, putting actual formulae in place of the metavariables. Either way, the presentation rests on an understanding of how one properly replaces formulae (or terms) by other formulae (terms). Though any of these operations can be rigorously defined, we shall ask whether the grammar of FOL actually suggests a notion of replacement to begin with. The ideas behind this derive from [5], where we scrutinize the notion of substitution in linguistic theory and the role it plays in structural linguistics as well as modern logic. One of the underpinnings of this research was the assumption that there is no independent notion of substitution — substitution is canonically defined on the basis of the grammar and the analyses it provides for linguistic expressions. The outcome for FOL is that substitution is simply string replacement of one constituent by another. This substitution is also simpler than the standard one. This result raises the question why substitution (and not even general substitution, which allows for a change of bound variables, as in λ–terms) is chosen for FOL. For with some extra effort it can be seen that this makes

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 155–172

no real difference for the axiomatization. While changing to string substitution seems like an unnecessary complication, the present result vindicates to a certain extent the instinctive mistrust of a novice in FOL against using a variable in the same formula both free and bound. And it may help in understanding why these difficulties exist.

2 Motivation

Consider the following sentence

`Achilles is faster than the tortoise.`

Suppose we replace `fast` by `wise`, `good` or `bad`. Then we expect to find the following.

```
Achilles is wiser than the tortoise.
Achilles is better than the tortoise.
Achilles is worse than the tortoise.
```

Thus, we do *not* get *`wiseer`, *`gooder`, nor *`bader`. So, in none of these cases the substitution is mere string substitution. As for the first, the fact that we have one `e` rather than two is due to a general writing convention. In the case of `good` the stem changes its form in the comparative. In the case of `bad`, it is both the stem and the comparative morpheme that change form. So, we find that sometimes the thing that we put into the hole ('filler') changes form, that sometimes only the container changes form, and sometimes both. The explanation is as follows. We analyze the comparative form as a combination of stem plus comparative suffix. We can represent this as $\sigma \bullet \gamma$, where $\bullet$ is a binary symbol of combination. Here σ and γ are not necessarily strings but they can be more complex. In this case, however, they are strings. Moreover, $\bullet$ can be something else than mere string concatenation. In the present case, the actual form of $\sigma \bullet \gamma$ sometimes is the concatenation of σ and γ (`fast`$^\frown$`er`), sometimes the concatenation with one `e` removed (`wise`$^\frown$`r`). Or σ is changed before concatenation, or it is γ that changes, or even both.

Each string has a term associated with it, a so–called **analysis**, of which the string is a **representative**. The term defines the string uniquely, but the relation between terms and strings may be many–to–one. If it is one–to–one, the language is said to be **uniquely readable**. This is a property that logical languages are required to have. While both stem and suffix are generally considered to be strings, this is in actual fact only a simplification (see [5]). On a more abstract level of analysis, however, the situation is simply as follows. The 'container' has an analysis $\mathfrak{s}(x)$ in which there is a hole x (which may contain several occurrences of x). The filler has the analysis $\mathfrak{t}$ which we put in place of x and obtain $\mathfrak{s}(\mathfrak{t})$. The term $\mathfrak{s}(\mathfrak{t})$ both determines the form of the expression (the string) and the meaning. Substitution is the replacement of $\mathfrak{t}$ by a different term $\mathfrak{u}$, giving $\mathfrak{s}(\mathfrak{u})$ in place of $\mathfrak{s}(\mathfrak{t})$.

Logical languages are man made. Therefore we do not expect the pathological examples of natural languages to exist. In particular, we expect that the objects that we manipulate are simply strings, and that the operation that forms constituents is simply concatenation. This is the case in propositional logic. For example, look at

`((p0∧(¬p01))→p0)`

We can think of this string as being obtained from

`((p0∧(¬____))→p0)`

by inserting the string `p01`. If we substitute something else for `p01`, we just replace the string occurrence of `p01` by whatever substitutes it. Similarly, replacing one or two occurrences of `p0` by another variable is simply string replacement.

We notice here right away that in addition to the official definition there exist 'dialects' of logical languages obtained by changing more or less drastically the strings that represent a formula (not to speak of the famous Begriffsschrift of [2]). For example, brackets are often dropped, variables are denoted by metavariables, and more function symbols are added. While logicians abstract from these changes and deal with the strings as imperfect representatives (in place of the correct ones), we shall treat these strings as objects in their own right. The approach is thus not normative, it is descriptive. It is not our aim to prescribe which strings constitute the language, we are interested in what happens if they are one way or another.

3 Sign Systems

Definition 1. *A **sign** is a triple $\sigma = \langle e, c, m\rangle$, where e is the **exponent** of σ, c its **category**, and m its **meaning**. A **language** is a set of signs.*

A **signature** is a function $\Omega\colon F \to \mathbb{N}$ for an arbitrary set F of function symbols. A (partial) Ω–algebra is a pair $\mathfrak{A} = \langle A, \mathcal{I}\rangle$, where A is a set and for every $f \in F$, $\mathcal{I}(f)$ is a partial $\Omega(f)$–ary function on A.

Definition 2. *A **grammar** consists of*

1. *a finite set F of **modes**,*

2. *a signature $\Omega\colon F \to \mathbb{N}$,*

3. *for every $f \in F$, partial $\Omega(f)$–ary functions f^{ε} on E, f^{γ} on C, f^{μ} on M*

*A grammar G is a **grammar for** $\Delta \subseteq E \times C \times M$ if the functions $\langle f^{\varepsilon}, f^{\gamma}, f^{\mu}\rangle$, $f \in F$, generate Δ from the empty set.*

[5] introduces a few more conditions on grammars. One is that functions are not allowed to destroy material of the exponents, the second that all functions are computable. A third condition is that the categories are distribution classes. These requirements need comment. First, the requirement of computability derives from the notion of compositionality; if we understand the meaning of a complex expression by applying a certain function to the meaning of its parts then we cannot strictly speaking understand the meaning of a complex expression if it is derived using a noncomputable function. This is awkward and we cannot go into the ramifications of this, but it has certainly been a concern in the foundation of mathematics (for example in intuitionism and constructivism). We shall not insist on computability. However, the other constraints shall be met.

To give an example, let Δ be the set of pairs $\langle \vec{x}, T, n\rangle$, where $\vec{x} \in \{\mathtt{0}, \mathtt{1}\}^*$ is a binary sequence and n is the number that $\vec{x}$ represents in binary. This set can be generated from the zeroary functions $f_0 := \langle \mathtt{0}, T, 0\rangle$ and $f_1 := \langle \mathtt{1}, T, 1\rangle$ and two unary functions, f_2 and f_3.

$$\begin{aligned} f_2(\langle \vec{x}, T, n\rangle) &:= \langle \vec{x}^\frown\mathtt{0}, T, 2n\rangle \\ f_3(\langle \vec{x}, T, n\rangle) &:= \langle \vec{x}^\frown\mathtt{1}, T, 2n+1\rangle \end{aligned}$$

So, put $F := \{f_0, f_1, f_2, f_3\}$, $\Omega : f_0, f_1 \mapsto 0; f_2, f_3 \mapsto 1$.

4 A Grammar for Generating the Exponents

We first deal with the **morphology** of predicate logic. The exponents of signs are also called (**well–formed**) **expressions** (**wfes**). Predicate logic is given by its well–formed expressions and their meanings. We assume that the expressions are strings and that the only operation to form complex expressions is concatenation (and no blank is inserted). It turns out that there is a context free grammar generating the wfes.

For concreteness' sake, there will be a unary predicate symbol `p`, a binary function symbol `+` and a binary relation symbol `=`. The exponents must be strings over a *finite* alphabet A. Thus, it is not possible to assume an infinite supply of primitive symbols. So we put

$$A := \{\neg, \to, +, \forall, (,), 0, 1, \texttt{p}, \texttt{x}, =\}$$

A **variable** is a sequence $\texttt{x}\vec{\alpha}$, where $\vec{\alpha}$ is a binary sequence (a decimal representation would be just as fine, but changes nothing in principle).

Definition 3. *A* ***context free grammar*** *is a quadruple* $\langle S, N, A, R\rangle$*, where* S *is the* ***start symbol****,* N *the set of* ***nonterminals****,* A *the* ***alphabet*** *and* R *the set of* ***rules****.*

A is as above, $N = \{$<variable>, <term>, <formula>$\}$, $S =$ <formula>. Here is now the set of rules.

<variable> ::= `x` | <variable>⌢`0` | <variable>⌢`1`
<term> ::= <variable> | `(`⌢<term>⌢`+`⌢<term>⌢`)`
<formula> ::= `p`⌢`(`⌢<term>⌢`)` | `(`⌢<term>⌢`=`⌢<term>⌢`)`
| `(`⌢`¬`⌢<formula>⌢`)`
| `(`⌢ <formula>⌢`→`⌢<formula>⌢`)`
| `(∀`⌢<variable>⌢`)`⌢<formula>

A **context** is a pair $C = \langle\vec{u}_1, \vec{u}_2\rangle$ of strings. If $\vec{y}$ is a string, then $C(\vec{y}) := \vec{u}_1\vec{y}\vec{u}_2$. Let $\vec{x} = \vec{u}_1\vec{y}\vec{u}_2$ be a string. We say that $\vec{y}$ is a **substring** of $\vec{x}$, and say that $C = \langle\vec{u}_1, \vec{u}_2\rangle$ is an **occurrence** of $\vec{y}$ in $\vec{x}$. Given a CFG $G = \langle S, N, A, R\rangle$, if there is a derivation $Y \Rightarrow_G^* \vec{u}_1 X \vec{u}_2$ and $\vec{y}$ is an X–string, we say that the occurrence $\langle\vec{u}_1, \vec{u}_2\rangle$ of $\vec{y}$ in $\vec{x}$ is a **constituent occurrence** (**of category** X). The **distribution classes** are the following sets (where S is the start symbol of G).

$$\mathrm{Dist}_G(\vec{y}) := \{\langle\vec{u}_1, \vec{u}_2\rangle : (\exists X \in N)(S \Rightarrow_G^* \vec{u}_1 X \vec{u}_2 \text{ and } X \Rightarrow_G^* \vec{y})\}$$

If $\vec{x}$ and $\vec{y}$ are X–strings, then they define the same distribution class. The converse need not hold. If the converse holds, the grammar is called **balanced**.

The grammar above is balanced. This grammar defines in total 3 different distribution classes: that of variables, of terms and of formulae. And they correspond exactly to the nonterminals used above. This contrasts with predicate logic as defined in textbooks, where only terms and formulae are recognized. The category of a variable is motivated on purely distributional grounds: only a variable can occur right after $\forall$. Indeed, it is generally assumed that a variable has no occurrence after a quantifier, so that substitution will not touch it. The grammar speaks a slightly different language: the variable occurs as a *variable*, but not as a *term*. And term substitution targets exclusively the term occurrences. (Often, variable substitutions are also considered, known as 'replacement of bound variables' or 'α–conversion'.)

5 A Grammar for FOL

Now we shall propose a grammar for FOL in the sense of our earlier definition. Put $E := A^*$ and $C := \{\nu, \tau, \varphi\}$. A **structure** is defined as usual, except that we fix the underlying domain to be a cardinal number. This makes the class of models of bounded cardinality a set. Clearly, every ordinary structure has an isomorphic structure of that kind so that we do not loose anything. Let $\mathbb{N}$ be the set of natural numbers. A **valuation** in a structure $\mathfrak{M}$ is a function from $\mathbb{N}$ to the carrier set of $\mathfrak{M}$. A **model** is a pair $\langle \mathfrak{M}, \beta \rangle$, where $\mathfrak{M}$ is a structure and β a valuation. An **index** is a function from models to truth values. (The set of truth values is as usual $\{0, 1\}$.) A **point** is a function from models to elements of the carrier set. $\mathbb{H}$ is the set of indices over countable models, $\mathbb{P}$ the set of points over countable models. Then

$$M := \mathbb{N} \cup \mathbb{P} \cup \mathbb{H}$$

(These three sets correspond to the three categories ν, τ and φ.) For a number m, let $\pi(m)$ be the point such that $\pi(m)(\langle \mathfrak{M}, \beta \rangle) = \beta(m)$. If p and q are points, $p + q$ is defined on $\mathfrak{M} = \langle M, \mathcal{I} \rangle$ by

$$(p + q)(\langle \mathfrak{M}, \beta \rangle) := \mathcal{I}(\mathtt{+})(p(\langle \mathfrak{M}, \beta \rangle), q(\langle \mathfrak{M}, \beta \rangle))$$

Further, $(=(p,q))(\langle\mathfrak{M},\beta\rangle) := 1$ iff $p(\langle\mathfrak{M},\beta\rangle) = q(\langle\mathfrak{M},\beta\rangle)$. We assume the following functions on truth values:

	$-$
0	1
1	0

$\to$	0	1
0	1	1
1	0	1

Then we put

$$\begin{array}{ll} (-i)(\langle\mathfrak{M},\beta\rangle) & := -(i(\langle\mathfrak{M},\beta\rangle)) \\ (i \to j)(\langle\mathfrak{M},\beta\rangle) & := i(\langle\mathfrak{M},\beta\rangle) \to j(\langle\mathfrak{M},\beta\rangle) \end{array}$$

Finally,

$$A(n,i)(\langle\mathfrak{M},\beta\rangle) := 1 \Leftrightarrow \text{ for all } \beta' \sim_n \beta : i(\langle\mathfrak{M},\beta\rangle) = 1$$

We propose a zeroary mode Z, unary modes N, E, T, Q and binary modes X, P, G, C.

$$\begin{array}{ll} \mathsf{Z} & := \langle \mathtt{x}, \nu, 1\rangle \\ \mathsf{N}(\langle\vec{x},\nu,m\rangle) & := \langle\vec{x}^\frown\mathtt{0}, \nu, 2m\rangle \\ \mathsf{E}(\langle\vec{x},\nu,m\rangle) & := \langle\vec{x}^\frown\mathtt{1}, \nu, 2m+1\rangle \\ \mathsf{T}(\langle\vec{x},\rho,m\rangle) & := \langle\vec{x}, \tau, \pi(m)\rangle \\ \mathsf{P}(\langle\vec{x},\tau,p\rangle, \langle\vec{y},\tau,q\rangle) & := \langle \mathtt{(}^\frown\vec{x}^\frown\mathtt{+}^\frown\vec{y}^\frown\mathtt{)}, \tau, p+q\rangle \\ \mathsf{G}(\langle\vec{x},\tau,p\rangle, \langle\vec{y},\tau,q\rangle) & := \langle \mathtt{(}^\frown\vec{x}^\frown\mathtt{=}^\frown\vec{y}^\frown\mathtt{)}, \varphi, =(p,q)\rangle \\ \mathsf{Q}(\langle\vec{x},\varphi,i\rangle) & := \langle \mathtt{(}^\frown\neg^\frown\vec{x}^\frown\mathtt{)}, \varphi, -i\rangle \\ \mathsf{C}(\langle\vec{x},\varphi,i\rangle, \langle\vec{y},\varphi,j\rangle) & := \langle \mathtt{(}^\frown\vec{x}^\frown\to^\frown\vec{y}^\frown\mathtt{)}, \varphi, i \to j\rangle \\ \mathsf{X}(\langle\vec{x},\nu,n\rangle, \langle\vec{y},\varphi,i\rangle) & := \langle \mathtt{(}\forall^\frown\vec{x}^\frown\mathtt{)}^\frown\vec{y}, \varphi, A(n,i)\rangle \end{array}$$

The grammar just defined is called PRED. Notice that the functions so defined are not computable. However, this is a deficit of predicate logic in general.

6 Substitution

We shall consider three types of substitution. String substitution is standard in propositional logic. We decompose the string $\vec{x} = C(\vec{y})$, then replacing the occurrence C of $\vec{y}$ by $\vec{z}$ is $C(\vec{z})$. This substitution however is a substitution that replaces proper occurrences of a subexpression only. It is not indiscriminate string substitution. We denote it by $\ulcorner t/x \urcorner$ (for the case of replacing a variable by a term, but the notation is analogously used in other cases).

The second type we find for example in [8]. We call it **standard substitution**, since it is the most widely used.

$$
\begin{array}{lll}
[t/x]y & := & \begin{cases} t & \text{if } y = x, \\ y & \text{else.} \end{cases} \\
[t/x]f(\vec{s}) & := & f([t/x]s_0, \ldots, [t/x]s_{n-1}) \\
[t/x]r(\vec{s}) & := & r([t/x]s_0, \ldots, [t/x]s_{n-1}) \\
[t/x](\neg\varphi) & := & (\neg[t/x]\varphi) \\
[t/x](\chi\to\chi') & := & ([t/x]\chi\to[t/x]\chi') \\
[t/x](\forall y)\chi & := & \begin{cases} (\forall y)\chi & \text{if } y = x, \\ (\forall y)[t/x]\chi & \text{if } y \text{ not in } t \text{ or } x \text{ not free in } \chi, \\ (\forall y)\chi & \text{else.} \end{cases}
\end{array}
$$

This is basically a substitution that replaces free occurrences of x by t. The third type, called **general substitution** in [7], executes a renaming of bound variables in the last clause of the definition above whenever y is free in t and t actually occurs free in χ. We denote by $\{t/x\}\varphi$ the result of applying the generalized substitution to φ. (To make this into a function, y has to be chosen according to a fixed procedure, for example, choosing the smallest binary sequence possible.) Notice right away that these operations are only substitutions of variables by terms. Occasionally, however, authors do look at substitutions of predicates by predicates (see for example [4], 155–162).

Given a grammar, substitution is defined as follows. Call a **structure term** a well–formed term over the signature. Given a structure term, we can compute the sign that it denotes — if it denotes a sign at all. For, as some operations may be partial, some structure terms fail to denote signs; when they do denote a sign, however, it is unique. Structure terms can contain variables. (If they don't they are called **constant**.) Structure terms shall be written in Polish Notation. (This is an arbitrary choice of no significance.) For example, PTNZTEZ is a structure term, which defines the sign $\langle \texttt{x0+x1}, \tau, \pi(2) + \pi(3)\rangle$. Not all structure terms denote a sign. If they do, they are called **definite**. PTNZEZ is a structure term but not definite.

Definition 4. *Let $\mathfrak{s}$, $\mathfrak{t}$, $\mathfrak{t}'$ be structure terms, and let $\mathfrak{s}$ contain a single occurrence of x. Denote by $[\mathfrak{t}/x]\mathfrak{s}$ the substitution of $\mathfrak{t}$ for the occurrence of x in $\mathfrak{s}$. $[\mathfrak{t}'/x]\mathfrak{s}$ is said to be the result of substituting the occurrence of $\mathfrak{t}$ named by x by $\mathfrak{t}'$ in $\mathfrak{s}$.*

Definition 5. *Let σ and τ be signs. We say that τ is a **part of** σ **under the analysis** $\mathfrak{s}$ if there is a constant term $\mathfrak{t}$ and a term $\mathfrak{u}$ with a single* *occurrence of a free variable, x, such that $[\mathfrak{t}/x]\mathfrak{u} = \mathfrak{s}$, and $\mathfrak{s}$ unfolds to σ and $\mathfrak{t}$ unfolds to τ.*

In predicate logic as defined above, the notion of part is straightforward. A part always is a certain subexpression. Substitution is the replacement of such subexpressions by others. Moreover, it is substitution of variables by variables, of terms by terms, and of formulae by formulae.

Theorem 1. *The substitution defined by the grammar* PRED *is string substitution of a variable by a variable, of a term by a term, and of a formula by a formula.*

The operation might perhaps better be called *replacement* to avoid collision with the ordinary substitution of predicate logic, but we wish to contend that what we call here substitution is *the* linguistically appropriate one. To wit, the substitution we obtain on the level of exponents goes as follows. (We omit some obvious clauses.)

$$\begin{array}{lll} \ulcorner t/x \urcorner \vec{u} & := [t/x]\vec{u}, & \text{if } \vec{u} \text{ is a term} \\ \ulcorner t/x \urcorner (\mathtt{(}^\frown\neg^\frown\vec{y}^\frown\mathtt{)}) & := \mathtt{(}^\frown\neg^\frown\ulcorner t/x \urcorner\vec{y}^\frown\mathtt{)} & \\ \ulcorner t/x \urcorner (\mathtt{(}^\frown\forall^\frown\vec{z}^\frown\mathtt{)}^\frown\vec{y}) & := \mathtt{(}^\frown\forall^\frown\vec{z}^\frown\mathtt{)}^\frown\ulcorner t/x \urcorner\vec{y} & \end{array}$$

(The first line means that if $\vec{u}$ is a string denoting a term, then the substitution is simply ordinary substitution on terms. We shall sometimes use $\vec{u}$ to remind ourselves that the object in question is a string. t and x are also strings but we did not write $\vec{t}$ or $\vec{x}$, for example.)

In particular, the operation $\ulcorner t/x \urcorner$ does not care about the distinction between free and bound variables. There is a similar substitution of formulae by formulae, which once again is simple string replacement, and an operation of variable replacement. The latter changes all occurrences of a variable (including the ones in the quantifier prefixes), again disregarding free and bound occurrences.

Substitution also figures in what is known as **Leibniz' Principle**. Let L be a logic, here identified with its set of tautologies.

Definition 6. *A logic L is* ***Leibnizian*** *with respect to a sign grammar iff for any two structure terms $\mathfrak{t}$ and $\mathfrak{u}$, $\mathfrak{t}^\mu = \mathfrak{u}^\mu$ iff for all $\mathfrak{s}$ containing at most x free:*

1. *$[\mathfrak{t}/x]\mathfrak{s}$ is a sign iff $[\mathfrak{u}/x]\mathfrak{s}$ is.*

2. If $[\mathfrak{t}/x]\mathfrak{s}$ *is a formula,* $[\mathfrak{t}/x]\mathfrak{s}^\varepsilon \in L$ *iff* $[\mathfrak{u}/x]\mathfrak{s}^\varepsilon \in L$.

This is a formalization of the informal statement saying that two things have equal meaning iff they can be substituted for each other in all contexts. Notice that this definition is relative to the grammar and not just the language. This is so since it relies on the notion of substitution, which is not available in a language, only in a grammar.

Proposition 1. *Every predicate logic with identity is Leibnizian with respect to* PRED.

Proof. Suppose that φ and ψ are formulae with different meaning. Then without loss of generality there is a model in which φ is true and ψ is false. Therefore $(\varphi\to\psi) \notin L$, but $(\varphi\to\varphi) \in L$. Consider the terms $\mathfrak{t}$, $\mathfrak{u}$ such that $\mathfrak{t}^\varepsilon = \varphi$ and $\mathfrak{u}^\varepsilon = \psi$. Then put $\mathfrak{s} := \mathsf{C}\mathfrak{t}x$. We have

$$\begin{aligned}[\mathfrak{t}/x]\mathfrak{s} &= (\varphi\to\varphi) \\ [\mathfrak{u}/x]\mathfrak{s} &= (\varphi\to\psi)\end{aligned}$$

This shows the claim for formulae. Let terms t and u be given. If their meaning is different, then $(t=u) \notin L$ but $(t=t) \in L$. Let $\mathfrak{t}$ and $\mathfrak{u}$ be structure terms with exponents t and u. Then put $\mathfrak{s} := \mathsf{G}\mathfrak{t}x$. Similarly with variables.

The above proof works in the presence of a predicate. If the signature is empty, the theorem holds since there are no formulae anyway. Consider the following strengthening of Leibniz' Principle: *For all structure terms* $\mathfrak{t}$ *and* $\mathfrak{u}$, *and for all* $\mathfrak{s}$ *such that* $[\mathfrak{t}/x]\mathfrak{s}$ *and* $[\mathfrak{u}/x]\mathfrak{s}$ *are formulae:*

$$\mathsf{CGtuC}[\mathfrak{t}/x]\mathfrak{s}[\mathfrak{u}/x]\mathfrak{s}^\varepsilon \in L$$

This means that for terms s, t (where $\varphi(s)$ results from replacing some appropriate variable z in φ by s — whatever substitution is actually used) the following holds.

$$((s=t)\to(\varphi(s)\to\varphi(t))) \in L$$

However, string substitution does not satisfy this property. Neither does standard substitution. Put $\varphi(z) :=$ `(∀x0)(¬(`z`=x0))`. The following is not valid in predicate logic.

```
(∀x)(∀x0)((x=x0)→((¬(∀x0)(x=x0))→(¬(∀x0)(x0=x0)))
```

Notice that generalized substitution satisfies the stronger version. However, Leibniz' Principle as defined above is about synonymy. It is clear that if $\mathfrak{t}$ and

$\mathfrak{u}$ are synonymous, then so are $[\mathfrak{t}/x]\mathfrak{s}$ and $[\mathfrak{u}/x]\mathfrak{s}$. It follows from Proposition 1 that the rules

$$\frac{(s{=}t)}{(\varphi(s/x){\to}\varphi(t/x))} \qquad \frac{(\chi{\to}\chi')\,,\,(\chi'{\to}\chi)}{(\varphi(\chi/p){\to}\varphi(\chi'/p))}$$

are admissible, where $\varphi(s/x)$ and $\varphi(\chi/p)$ are short for substitution of s for every occurrence of x and substitution of χ for the proposition (meta)variable p. For all three substitutions, predicate logics admit the above rules.

7 Uniqueness of the Analysis

We have presented a grammar which makes any predicate logic Leibnizian. The substitution it defines is string substitution. The question is whether there are grammars for which the corresponding substitution on the level of terms is either standard or generalized substitution. First, notice that the meaning of a formula depends on the meaning of those well–formed formulae which have been used to build it. It follows that if χ is a subformula on whose meaning the meaning of φ depends, then χ or an equivalent formula must actually be part of φ. Unfortunately, the converse need not hold. φ may be built using formulae that do not appear in it, and these might be formulae on whose meaning its meaning does not even depend. This is not an absurd statement. [1] makes that point clearly: the meaning of the formula $(\forall x)\,\varphi(x)$, x free, depends on more than just its subformulae. Its truth cannot be assessed by looking at the truth of the subformulae alone as in classical logic (though its meaning can be so found). Instead the entire model must be inspected. We may see this as a reflex of the fact that this formula actually contains all substitution instances as subformulae to begin with. Here is how. Let us add a ternary mode Y. For terms $\vec{u}$, variables $\vec{x}$, and formulae $\vec{z}$ let

$$\mathsf{Y}^{\varepsilon}(\vec{u},\vec{x},\vec{z}) := \vec{z}\,[\vec{u}/\vec{x}]$$

where $\vec{z}\,[\vec{y}/\vec{x}]$ denotes the result of replacing all free occurrences of $\vec{x}$ by $\vec{y}$. The interpretation of that function is also straightforward. Let m be a number, i an index and p a point.

$$\mathsf{Y}^{\mu}(p,m,i)(\langle\mathfrak{M},\beta\rangle) := i(\langle\mathfrak{M},\beta[p(\langle\mathfrak{M},\beta\rangle)/m]\rangle)$$

where $\beta[p(\langle\mathfrak{M},\beta\rangle)/m]$ is different from β only in assigning the value of p on $\langle\mathfrak{M},\beta\rangle$ to the number m. (If p is the meaning of the term t, $p(\langle\mathfrak{M},\beta\rangle)$ is nothing but the denotation of t in the model.) For completeness' sake we remark that

$$\mathsf{Y}^{\gamma}(a, b, c) = \begin{cases} \varphi & \text{if } a = \tau, b = \nu, c = \varphi \\ \text{undefined} & \text{else.} \end{cases}$$

This fully defines the operation of the mode Y on signs. Notice that $\mathsf{Y}xx\mathfrak{u}$ denotes the same sign as $\mathfrak{u}$. So, structure terms are no longer unique for formulae. Now, let $\mathfrak{x}$ be the structure term for x, $\mathfrak{t}$ the structure term for t and $\mathfrak{u}$ some (!) structure term for φ. Then

$$\mathsf{Y}\mathfrak{t}\mathfrak{x}\mathfrak{u}^{\varepsilon} = [t/x]\varphi, \qquad \mathsf{Y}\mathfrak{x}\mathfrak{x}\mathfrak{u}^{\varepsilon} = \varphi$$

Thus, with $\mathfrak{s} := \mathsf{Y}x\mathfrak{x}\mathfrak{u}$ we have a structure term such that standard substitution is nothing but substitution in the grammar sense. Yet, the proposed solution has a defect: it makes formulae into subformulae that have no string occurrence in it. For example, $\varphi(y)$ is a subformula of $\varphi(x)$ and vice versa. This is unacceptable.

It is clear that there are many grammars that generate predicate logic. Even if we exclude pathological cases and require, say, that the grammar is context free, there are infinitely many solutions. Let us therefore try to elaborate which further assumptions on the grammar make it unique. The sign based grammar PRED is unique on the following assumptions. The datum is pairs of well–formed expressions combined with their meanings.

[a] We assume that the exponents are strings over the alphabet, and that the modes can only concatenate them, possibly adding syncategorematic symbols.

[b] We assume that the system is **monotectonic** (or **unambigous**): every well–formed expression has a unique structure term.

[c] We assume that for every substring that is a well–formed expression of some type, that occurrence is the exponent of some subterm of the structure term. This is called **transparency**.

[d] Any two well–formed expressions have the same category iff they have the same distribution.

By [a] and [c] we get that every well–formed expression is composed from its immediate subexpressions. For example, `(∀x0)(x=x0)` has as its immediate subexpressions `x0` and `(x=x0)`. Hence, the additional brackets as well as `∀` are syncategorematic. Moreover, it makes `x` a subexpression of `x0`, which is a subexpression of `x01`, and so on. Notice that by transparency substitution can be formulated as mere string replacement, since 'accidental' occurrences of subterms cannot be mistaken for proper ones (as would be `p0→p1` in `p∧p0→p1`, when brackets are dropped). [b] is put in to make sure that there are no two modes that operate in the same way on signs. [d] ensures that we do not create more categories than absolutely necessary. By distributional analysis we get in fact three categories: variables (only they appear right after `∀`), terms and formulae.

It is interesting to note that different results obtain if the morphology is different. Suppose we drop brackets in conjunctive expressions as we do with additive terms, writing $\varphi\wedge\chi\wedge\psi$ in place of $(\varphi\wedge(\chi\wedge\psi))$. Then the resulting language is no longer transparent. The string `(x=x0)∧(x0=x1)∧(x1=10)` has two different overlapping well–formed substrings, `(x=x0)∧(x0=x1)` and `(x0=x1)∧(x1=x10)`. But not both of them can be part of one and the same analysis, because of [a]. Of course, the ambiguity is spurious: we may choose either analysis. On either analysis we get the same result. (As a note of clarification: here we do *not* view the omission of brackets as an abbreviatory convention, but we take strings without brackets as genuine, well–formed expressions in their own right. We are analyzing, so to speak, the actual usage of predicate logic rather than the official norm.)

Now, there are other conventions as well. The brackets around an equation are always omitted. One also assumes that `¬` is the strongest symbol, and brackets are dropped from $(\neg\varphi)$. Outer brackets are always dropped. However, notice that while `x=x0∧x0=x1` is unambigous, its negation is to be written `¬(x=x0∧x0=x1)` rather than `¬x=x0∧x0=x1`. All this only concerns the manipulation of the exponents of the grammar that generates this language. Substitution, however, remains the same operation on the level of structure terms, only that it projects differently onto the exponents (= wfes).

8 Axiomatization

The relevance of substitution is seen when we turn to axiomatization. There are two kinds of axiomatizations. The first uses an inbuilt rule of substitution, the second specifies axiom schemes, using metavariables for formulae. In propositional logics both approaches are possible, while FOL can use internal substitution only with respect to terms, since there are no variables for formulae. However, we should be aware of the fact that the employment of axiom schemes relies on a correct understanding of what is substituted and how. This includes a proper understanding of the morphology of the actual language. Let us give an example. In propositional logics substitution is taken to be replacement of occurrences of a (meta)variable by a string. Yet, this works only in the transparent case. If we do not write brackets matters are different. For example, replacing `p0` in `p0∧p1` by the disjunction `p01∨p11` will make it necessary to insert brackets, so that we get `(p0∨p11)∧p1` and not `p01∨p11∧p1`. On the other hand, if we have to substitute `p01∧p11` for `p0`, no insertion of brackets is needed. This shows clearly that substitution may require syntactic analysis. Instead, one should think of the strings as representatives of structure terms, and metavariables as proxy for structure term variables. For example, $\texttt{(¬((∀}x\texttt{)}\varphi\texttt{→(¬}\varphi\texttt{)))}$ is proxy for $\mathsf{QCX}xz\mathsf{Q}z$, where x is a variable for a structure term for variables, z a variable for a structure term for formulae. Given the grammar there is no question of what function substitution actually is, since it is universally and unequivocally determined at the level of structure terms. If, say, we instantiate x to ENZ and z to GENZEZ then the structure term becomes

$$\mathsf{QCXENZGENZEZQGENZEZ}$$

Its exponent is `(¬((∀x01)(x01=x1)→(¬(x01=x1)))).`

The first task we set ourselves is to axiomatize predicate logic using string substitution, denoted here by $\ulcorner t/x \urcorner$. Recall that predicate logic has three kinds of axioms. The first set is the propositional axioms, for example

$$\texttt{((}\varphi\texttt{→(}\psi\texttt{→}\chi\texttt{))→((}\varphi\texttt{→}\psi\texttt{)→(}\varphi\texttt{→}\chi\texttt{)))}$$

These rules are unproblematic. The given axioms can be instantiated by substituting any given structure term for formulae for the variables for terms x, y and z in

$$\mathsf{CC}x\mathsf{C}yz\mathsf{CC}xy\mathsf{C}xz$$

So these axioms are completely schematic. Likewise the rule

$$\text{(mp)} \quad \frac{\texttt{(}\varphi\texttt{→}\chi\texttt{)} \quad \varphi}{\chi}$$

The next set are the axioms for equality: reflexivity, transitivity and symmetry, and the rule of replacement of equals:

$$\texttt{(∀x)(∀x0)(x=x0→(p(x)→p(x0)))}$$

These axioms are the exponents of concrete structure terms. (This is enough given [2] and (gen) below.)

Finally, the following need to be added.

[1] $((\forall x)(\varphi \to \psi) \to ((\forall x)\varphi \to (\forall x)\psi))$

[2] $((\forall x)\varphi \to [t/x]\varphi)$

[3] $(\varphi \to (\forall x)\varphi)$ (x not free in φ)

and the rule

$$\text{(gen)} \quad \frac{\varphi}{(\forall x)\varphi}$$

Notice that [1] and the rule (gen) present no problem. [2] and [3] present two problems: they have side side conditions and they involve explicit substitutions.

Suppose that we change to the substitution $\ulcorner t/x \urcorner$.

Definition 7. *Call a formula* ***regimented*** *if there is no subformula which contains a variable both free and bound.*

Lemma 1. *Suppose that* $(\forall x)\varphi$ *is regimented. Then* $\ulcorner t/x \urcorner\varphi = [t/x]\varphi$.

The following is relatively straightforward to prove.

Proposition 2. *Let* φ *be regimented. If* φ *is provable in* FOL*, it has a proof using only regimented formulae.*

We replace the set of axioms by the following:

[i] $((\forall x)(\varphi \to \psi) \to ((\forall x)\varphi \to (\forall x)\psi))$

[ii] $((\forall x)\varphi \to \ulcorner t/x \urcorner\varphi)$ ($(\forall x)\varphi$ regimented)

[iii] $(\varphi \to (\forall x)\varphi)$ (x not free in φ)

[iv] $((\forall x)\varphi \to (\forall y)\ulcorner y/x \urcorner\varphi)$ (y not free in φ, x not bound in φ)

Thus, we have effectively restricted only the second axiom. We are guaranteed only to derive the regimented formulae. That is why we have added the formula [iv]. The axiomatization is complete. For suppose that φ contains a subformula containing a variable bound which occurs free outside of it. Then we can do a suitable replacement of that bound variable and get a formula φ'. The two are equivalent in predicate logic. φ' is regimented, and φ can be derived from it using [iv]. It is certainly valid, since under the given conditions it is identical to

$$((\forall x)\varphi\to(\forall y)[y/x]\varphi)$$

Notice that the axiomatization is not entirely algebraic. There are side conditions on the formulae and we have made use of explicit substitutions. The question arises whether using a different notion of substitution can make a difference here. This would mean to present a schematic axiomatization in which we use metavariables for variables, for terms and for formulae. Clearly, this is feasible, by simply listing them. This is the way in which the axiomatization of first–order logic is standardly taken. The set is decidable but infinite. The question is whether a finite subset is enough. Suppose for simplicity that we have no function symbols. Then we only need to worry about variables and formulae. It is known that substitution can be defined by quantification. In view of a result of [6] it is impossible to reduce the list to a finite one. Interestingly, [9] show that adding explicit substitution functions (of variables by variables) does not improve the situation.

We remark here only that if we use general substitution, the axiom [2] becomes

$$((\forall x)\varphi\to\{t/x\}\varphi)$$

without side conditions. Yet, the axiom [3] still needs the side condition x is not free in φ. Thus, none of the substitutions actually substantially simplifies the task of axiomatizing FOL.

9 Conclusion

The point of this paper was to argue that the grammar of FOL virtually forces us to assume a particular kind of substitution. The general question concerning this is: why is this a concern? And why should the logician care? To answer the second question first: we have shown that string substitution is actually no more and no less suited for the purpose than is standard substitution, but it is easier to use. The real complications do not arise from substitution, they arise from FOL itself. To answer the first question: in an artifical language we can make arbitrary decisions, but in a natural language we cannot. Still, in reasoning within natural language we wish to do the very same as we did for predicate logic: formulate the rules of reasoning within language. The Stoics used expressions like *the first* or *the second* as variables for sentences, something which does not require great care in formulation. But when it comes to syllogisms, matters are less straightforward. Consider *modus barbara* (see [3]):

$$\frac{\begin{array}{l}\texttt{Every man is mortal.}\\ \texttt{Every priest is a man.}\end{array}}{\texttt{Every priest is mortal.}}$$

It arises in the same way from insertion into a schematic expression:

$$\frac{\begin{array}{l}\texttt{Every } P \texttt{ is } Q\texttt{.}\\ \texttt{Every } R \texttt{ is } P\texttt{.}\end{array}}{\texttt{Every } R \texttt{ is } Q\texttt{.}}$$

Yet, when we turn to other languages, the substitution will necessitate changes. In Latin and French, the predicative adjective agrees in gender with the subject. Although we wish to consider the agreement patterns to be inessential, still it is important to set up the system in such a way that they really are taken care of. In other words, we wish to set up a grammar for French and Latin in such a way that it provides a schematic expression of the kind that does the agreement automatically. Otherwise the logical schemata need to make up for that (for example by devising different schemata for different genders).

10 Bibliography

[1] José Ferreirós. The road to modern logic — an interpretation. *The Bulletin of Symbolic Logic*, 7:441 – 484, 2001.

[2] Gottlob Frege. *Begriffsschrift.* Nebert, Halle, 1879.

[3] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik.* Springer, Berlin, Heidelberg, 5 edition, 1967.

[4] Stephen C. Kleene. *Introduction to Metamathematics.* North–Holland, Amsterdam, 1964.

[5] Marcus Kracht. *The Mathematics of Language.* Mouton de Gruyter, Berlin, 2003.

[6] Donald J. Monk. Nonfinitizability of classes of representable cylindric algebras. *Journal of Symbolic Logic*, 34:331 – 343, 1969.

[7] Arnold Oberschelp. *Elementare Logik und Mengenlehre*. Bibliographisches Institut, Mannheim, 1974.

[8] Wolfgang Rautenberg. *Einführung in die mathematische Logik*. Vieweg, Wiesbaden, 1996.

[9] Ildikó Sain and Richard S. Thompson. Finite Schema Axiomatization of Quasi–Polyadic Algebras. In Hajnal Andréka, Donald Monk, and István Németi, editors, *Algebraic Logic*, number 54 in Colloquia Mathematica Societatis János Bolyai, pages 539 – 571. János Bolyai Matematikai Társulat and North–Holland, Budapest/Amsterdam, 1991.

Logical Inquiries into a New Formal System with Plural Reference

Ran Lanzet
lanzetr@netvision.net.il
Philosophy Department, Tel Aviv University, Israel
Hanoch Ben-Yami
benyami@post.tau.ac.il
Philosophy Department, Tel Aviv University, Israel

1 Introduction: The Semantics of Natural Language

In this paper we develop a new formal system of logic, which consists of syntactic rules, derivation rules and a model-theoretic semantics. We then make some meta-logical inquiries into the nature of this system, comparing it with the first-order predicate calculus or logic (FOL).

Our formal system is based on an analysis of the semantics of natural language sentences, an analysis which departs in several basic respects from the semantic analyses one finds in the literature. All these use some version of FOL to analyze the semantic structure of natural language sentences; the semantic structure of these sentences, so it is assumed, can be transparently represented by their translation into some version of FOL. By contrast, we think that some semantic features of some natural language sentences *cannot* be captured by means of FOL, and that one distorts the semantic structure of these sentences if one tries to represent it by such translations.

This alternative analysis of natural language semantics, together with a criticism of the analysis suggested by FOL, are developed in detail in Ben-Yami's *Logic & Natural Language* (Ashgate, 2004). We shall mention some

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 173–223

of this book's central claims below, but – given the purpose and space of this paper – we shall not attempt to justify them here. We have to refer the reader to this book for the full justification and development of these claims.

This semantic analysis, which is the basis of our formal system, should also serve to clarify what we tried to achieve by this system's development, and what we *did not*. Usually, when one develops a new formal system of logic, one does that in order to capture some inferences that hitherto one could not capture, and frequently could not even express, in existing formal systems. *This was not our purpose.* Rather, we tried to show that the alternative analysis of the semantics of natural language can serve as the basis for a formal system which is as powerful as some version of FOL, in a sense to be made precise below. We wanted to show that one need not abandon the semantic structure of natural language if one wants to apply a deductive system of FOL's power.

With this in mind we can proceed to a concise presentation of some of the semantic claims made in *Logic & Natural Language*.

FOL distinguishes two kinds of expression which are not variables or logical constants: predicates on the one hand and individual constants (and possibly other closed terms as well) on the other. Individual constants translate the proper-names and other *singular* referring phrases or expressions of natural language, and can thus be said to refer to or designate particulars. Accordingly, FOL can be said to recognize only one kind of referring expressions: *singular* ones (but see the discussion of many-sorted logic below).

Natural language, by contrast, contains *plural* referring expressions as well. These include plural pronouns (in English, 'we', 'you', 'they' and their declined forms), plural demonstratives ('these', 'those'), plural definite descriptions (e.g., 'my children', 'the students'), some phrases that resemble both definite descriptions and proper-names ('the Knights of the Round Table', 'the Simpsons'), and conjunctions and disjunctions of singular and plural referring expressions (e.g., 'Peter and/or Jane', 'Mary and the children'). Such expressions may have other, non-referential uses as well; but they can all be used to refer to several particulars.

The italicized words and phrases below are examples of the referential use of expressions of these kinds:

> *We* saw *the Simpsons* in the supermarket. *These* are my books.
> *My children* are asleep. *Peter and Jane* should soon be here.

What is involved in plural reference, vis-à-vis singular reference, is straightforward. Whatever is achieved in referring to a single person or thing can be achieved with respect to several persons or things, and we then have plural reference.

When we talk about plural reference we mean referring to more than a single person or thing. We *do not* mean referring to a set with many members, to a complex individual, or to any other variation on these ideas. We mean achieving with relation to more than a single thing what is achieved by reference to a single one.

The great majority of existing attempts to translate sentences that contain plural referring expressions into FOL are reductive, in the sense of trying to analyze such expressions either as singular referring ones, or as involving an implicit structure that contains referring expressions of the singular kind only. But these analyses can be shown to be either mistaken or at least implausible. Moreover, they are not motivated by any linguistic phenomenon, but by the unjustified conviction that FOL must be capable of translating the relevant sentences. Yet FOL cannot adequately represent the semantics of natural language sentences containing plural referring expressions precisely because it lacks such expressions. (Again, for the full development and justification of these claims, and of some of the following, see *Logic & Natural Language*.)

Now, the careful analysis of the functioning of common nouns in natural language shows, that in many cases, *common nouns in quantified noun phrases are plural referring expressions.* For instance, in 'Some children are asleep', 'children' is used to refer to children. Similarly, in 'John met several members of my college', 'members of my college' is used to refer to persons, several of which John met. (*N.B.*: It refers not to those met by John, but to *all* members of my college.)

This is in marked contrast with the way FOL translates these expressions. Common nouns are taken to be predicative not only when they function as grammatical predicates, but when they appear in the grammatical subject position as well. Already Frege, and as early as in his *Begriffsschrift* (§12), has translated the subjects in the four Aristotelian quantified sentences by predicates, and several times in his later writings he argued for this analysis.

Let us demonstrate the difference between the two approaches by one standard example. The sentence

(1) All philosophers are wise

is translated into FOL by the sentence

$$\forall x(\text{Philosopher } (x) \longrightarrow \text{Wise } (x))$$

That is, the expression ‘philosophers’ is seen as contributing to the meaning of the natural language sentence in the same way that ‘wise’ does: they are both *predicative*; they are both used to say something about particulars referred to in some other way. By contrast, on our analysis, ‘philosophers’ in (1) is not predicative but *referential*; it is used to specify which are the particulars about which something is said (in this case, the philosophers). The same applies to the use of ‘philosophers’ in the sentences ‘Some/Seven/Many/Most philosophers are wise’.

The fact that his calculus did not contain plural referring expressions forced Frege to introduce quantification into it in a way that is far different from the way it functions in natural language. For Frege, and in FOL generally, quantifiers are operators that operate on sentential functions; they are second-order concepts. This is not the way quantifiers function in natural language, as we shall now explain.

When we quantify, we refer to a plurality of particulars, and say that specific quantities of them are such-and-such; quantification involves reference to a plurality. Natural language accomplishes this kind of reference by means of plural referring expressions, which designate the plurality, or pluralities, about which something is being said. And by using different expressions, natural language can refer to different pluralities. By contrast, since FOL uses concepts only as predicates, it has no plural referring expressions. The plurality about which something is said by its sentences has to be presupposed, and different sentences cannot specify different pluralities (but see again the notes on many-sorted logic below). In natural language, pluralities are introduced and specified by means of plural referring expressions; in FOL, a plurality, which is unspecified by the sentence, is introduced by presupposing a domain of discourse.

In order to speak of pluralities natural language sentences presuppose no domain of discourse, in the technical sense in which this concept is used in predicate logic semantics. A domain of discourse is a necessary component of the semantics of FOL, which has no parallel in the semantics of natural language. The idea of a domain of discourse may have important applications for formal systems, and we shall use it ourselves in that context below. But one distorts the semantics of natural language if one insists on finding a domain there.

This semantic difference results in a syntactic one as well. If the plurality is referred to by some plural referring expression, the quantifier has to be related in some syntactic way to the plural referring expression that indicates the plurality of which a quantified claim is made. Consequently, in natural language the quantifier is attached to a noun that is used to refer to a plurality, and together they form a noun phrase. However, if no expression is used to refer to a plurality, but the plurality is presupposed by the quantified construction, then the quantifier does not have to be attached to any specific component of the quantified sentence. Consequently, in FOL the quantifier operates on a sentential function.

This alternative semantic analysis of natural language can explain many features of language that create difficulties for attempts to analyze it by means of some version of FOL (including versions that use generalized quantifiers). Among other things, it explains away several alleged ambiguities of the copula; it explains some semantic features of natural kind terms and of empty concepts; it yields a natural classification of quantifiers (classifying 'many' and 'most', but not 'more', as belonging to the same family as 'every' and 'some'); it explains the semantic need for some linguistic devices like an affirmative and negative copulas, active versus passive voice, etc.; and more (cf. [1]).

Although in *Logic & Natural Language* a consistent deductive system for natural language sentences was developed on the basis of this semantic analysis, no attempt was there made to develop a rich artificial language, with rigorous rules for wffs, derivation rules and a model-theoretic semantics. This, as was said above, is our main purpose in this paper, to which we shall now proceed. In doing this we shall also demonstrate that the new analysis can be used as a basis for a formal system which resembles FOL in its power.

A note is in order here on the use of universal and existential sentences below. When we use such sentences in our proofs, we adopt the conventions customary in mathematics. In particular, we use 'Every A is B' as short for 'If anything is A, then it is B'. This is meant to enable a more fluent reading. Since we use these conventions consistently, the differences between this way of using sentences and the way they are commonly used in natural language should not bother us.

2 The Definitions of Our System

This part includes our definitions of a formal language and of a formula. It also includes our definitions of truth in a model and our deductive system.

2.1 Some basic definitions

Definition 1 (Formal Language). *A* formal language L *is a disjoint union of nine sets:* $\mathcal{P}$ *– a set of* one-place predicates*, one of which is the predicate Thing;* $\mathcal{R}$ *– a set of* relation-signs *or* many-place predicates *(to every one of which we assign a natural number* $n > 1$*, called its number of places);* $\mathcal{S}$ *– a denumerable set, whose members are called* singular referring expressions *(or: SREs);* $\mathcal{A} = \{a, a_1, a_2, \ldots\}$ *– the set of* anaphors*;* $\{1, 2, 3, \ldots\}$ *– the set of* indices*;* $\{\wedge, \vee, \neg, \longrightarrow\}$ *– the set of* sentence-connectives*;* $\{$*every, some*$\}$ *– the set of* quantifiers*;* $\{$*is, isn't*$\}$ *– affirmative and negative* copulas*;* $\{), (, \rangle, \langle, , \}$ *– parenthesis and comma. The members of* L *are its* signs*.*

Note: In order to fully determine a language L, it is enough to determine $\mathcal{P}$, $\mathcal{R}$ and $\mathcal{S}$; the rest of the constituents are the same for all languages.

As we shall explain below, one-place predicates function in our system also as plural referring expressions, as common nouns do in natural language. One might claim that the name 'predicates' is not appropriate for such expressions; 'concept-letters' might have been more suitable. However, since these expressions function also as predicates, and since the term 'n-place predicate' will be convenient to use as a collective name for both one-place and many-place predicates, we shall continue using this terminology in what follows.

We shall also see that the extension of *Thing* in every model will be the whole universe. We have added such a predicate to our system in order to obtain formulas that refer to the whole domain. As we shall see, this will help us translate formulas from FOL to our system. It should be noted, however, that there is no internal need for such a predicate in our formal system, and that the system can be developed without it, as indeed is the case with the related system developed in [1].

Definition 2 (Quantified Noun-Phrase, Noun-Phrase). *If* P *is a one-place predicate, then every* P *and some* P *are* quantified noun-phrases *(QNP). If* α *is a QNP or an SRE, then* α *is a* noun-phrase *(NP).*

In natural language there are quantified noun-phrases that contain a defining clause of some sort; for instance, 'every man who owns a Jaguar', as used in 'Every man who owns a Jaguar is rich'. Quantified noun-phrases composed in this manner are not dealt with in the present paper, and are not represented in the formal system developed below. We limit the system developed here to QNPs in which the referring expression is a simple (non-composed) one-place predicate.

The use of anaphors in our system resembles their use in natural language. As we shall see below, anaphors in our system will always relate to (an occurrence of) a noun-phrase, and their meaning will be determined with relation to that noun-phrase. The relation 'being anaphoric on' is syntactically defined as follows:

Definition 3 (Anaphors of a Noun-Phrase). *Let φ be a string of signs. An occurrence α of an anaphor in φ is* anaphoric on *an occurrence t of an NP δ in φ if the following conditions hold: t is to the left of α; the same index k appears in parenthesis both immediately to the left of t and immediately to the left of α; the string (k) does not occur immediately to the left of any sign that is not an anaphor between t and α. In this case, we may also say that α is* an anaphor of t*, and that t is the* source *of α.*

Example: In the string $(((1)s_1, (2)\ every\ P)\ is\ R) \longrightarrow (((2)a, (1)a)\ is\ L)$, the first (i.e. the leftmost) occurrence of a is anaphoric on the occurrence of *every P* ; the second – on the occurrence of s_1.

In natural language, a given relation can be represented in various forms: the sentences 'John kissed Mary' and 'Mary was kissed by John', for instance, represent the same relation, as do the sentences 'John gave this book to Mary', 'This book was given by John to Mary', 'To Mary was this book given by John', etc. We call such variations *transpositions*. To represent these in our system, we use the following definition:

Definition 4 (Transpositions). *Let R be an* n*-place predicate, $n > 1$, and let τ be a non-trivial permutation (i.e., not the identity permutation) of $\{1, \ldots, n\}$. Then the string $R\langle\tau(1), \ldots, \tau(n)\rangle$ is a* transposition *of R. (The symbol τ here does not belong to our formal language; it belongs to the meta-language.) Thus, if R is a 3-place predicate, its transpositions are $R\langle 1, 3, 2\rangle$, $R\langle 2, 1, 3\rangle$, etc.*

Note: For the sake of convenience, we shall sometimes refer to R as $R\langle\tau(1),\ldots,\tau(n)\rangle$, where τ is the identity permutation.

Note: If t is an occurrence of a certain sign, or string of signs, in a string φ, and α is a sign, or a string of signs, then we write $\varphi[t/\alpha]$ to denote the string that is the product of replacing t with α in φ. In case several occurrences $t_1,\ldots,t_n$ are replaced by $\alpha_1,\ldots,\alpha_n$, we write: $\varphi[t_1/\alpha_1,\ldots,t_n/\alpha_n]$. Sometimes we would like to replace *all* occurrences of a certain sign (or string of signs) α in a string φ by another sign or string β. To refer to the product of such a replacement we write: $\varphi[\alpha/\beta]$.

Note: If α is a sign, or a string of signs, that occurs in a string φ, then, in order to emphasize the fact that φ contains α, we shall sometimes refer to φ as $\varphi(\alpha)$.

2.2 Formulas

Our formation rules are somewhat more complex than those of FOL. We shall first give a brief sketch of these rules, and several examples of formulas together with the English sentences they translate. Only then shall we proceed to give the exact definition of a formula.

Our atomic formulas include strings of the form: $(s_1,\ldots,s_n)$ *is* R, which are meant to express a relation between n individuals, and: $(s_1,\ldots,s_n)$ *isn't* R, which are meant to deny such a relation. We allow the forming of new formulas from given ones by means of sentence connective in the usual manner.

Another thing we allow is the replacement of some occurrences of an SRE by anaphors of another occurrence of the same SRE. Thus, for instance, since (s,s) *is* L is a formula, $((1)s,(1)a)$ *is* L is also a formula. The first of these two can translate 'John loves John'; the second – 'John loves himself'. Anaphors are written with indices to their left, to indicate their being anaphoric on a certain occurrence of an NP.

Under certain conditions, we also allow the replacement of an SRE by a QNP. We thus have formulas such as (*every* M,s) *is* L (which can translate 'Every man loves John'), ((1) *every* $M,(1)a$) *is* L ('Every man loves himself') and also (*every* M, *some* W) *is* L ('Every man loves some women').

Let us now turn to the exact definitions. We shall start with a definition of atomic formula, and proceed by induction.

Definition 5 (Atomic Formula). *A string of signs φ in a language L is an* atomic formula *if it is of one of the following forms: s_1 is s_2; s_1 isn't s_2; $(s_1, \ldots, s_n)$ is R; $(s_1, \ldots, s_n)$ isn't R, where $s_1, \ldots, s_n$ are SREs and R is an* n*-place predicate ($n \geq 1$) or a transposition of such a predicate.*

Definition 6 ($\#\varphi$). *Let φ be a (finite) sequence. $\#\varphi$ is the length of φ. If φ is a string of signs in a language L, then $\#\varphi$ is the number of sign-occurrences in φ.*

Definition 7 (Formula, Sub-Formula, Main QNP). *Let φ be a string of signs in a language L.*

1. *If $\#\varphi \leq 3$, then φ is a formula iff it is an atomic formula.*

2. *Assume that $\#\varphi = n$, and that for any string ψ for which $\#\psi < n$, it is determined whether ψ is a formula.*

   Define: Let δ and ψ be strings of signs in L such that $\#\delta, \#\psi < n$. Then:

   (i) δ is a sub-formula *of ψ if the following conditions hold: ψ is a formula; $\#\delta < \#\psi$; δ is contained in ψ as a string; δ itself is a formula, or the product of one or more of the following operations on a formula: substitution of anaphors (with indices to their left) for SREs, addition of indices in parenthesis to the left of some NP occurrences, substitution of NPs for other NP occurrences.*

   (ii) An NP occurrence t in ψ is distributed *in ψ if there is no sub-formula of ψ that contains both t and all its anaphors.*

   Now, φ is a formula iff one of the following conditions holds:

   (a) φ is an atomic formula.

   (b) There are formulas α and β such that: $\#\alpha, \#\beta < n$; α, β do not contain anaphors of SRE occurrences; $\varphi \in \{\neg(\alpha), (\alpha) \vee (\beta), (\alpha) \wedge (\beta), (\alpha) \longrightarrow (\beta)\}$.

   (c) There is a formula ψ and an index k such that: $\#\psi < n$; $c_1, \ldots, c_n$ are occurrences of an SRE s in ψ, ordered from left to right; none of $c_1, \ldots, c_n$ has an index in parenthesis to the left of it; the string (k) does not occur between c_1 and c_n; if (k) occurs to the right of

c_n and is immediately followed by an anaphor, then this anaphor has a source that lies to the right of c_n ; $a_2, \ldots, a_n$ are anaphors; φ is $\psi[c_1/(k)s, c_2/(k)a_2, \ldots, c_n/(k)a_n]$.

(d) *There is a formula ψ, an SRE s and a QNP qP such that: $\#\psi < n$; c is a distributed occurrence of s in ψ; ψ does not contain distributed occurrences of QNPs to the left of c; other than c, no SRE occurrence in ψ has anaphors; φ is $\psi[c/qP]$. In this case, the occurrence of qP that replaced c is called* the main QNP *in φ.*

Note: We shall sometimes omit parenthesis, where this is unlikely to cause confusion. For instance, we shall refer to $((\alpha) \wedge (\beta)) \wedge (\gamma)$ as $\alpha \wedge \beta \wedge \gamma$; to (s) *is* P as s *is* P; and to (*every* Q) *is* P as *every* Q *is* P.

Theorem 1 (Induction on Formulas). *Let A be a set of formulas in a language L and assume that A satisfies the following conditions:*

1. *All the atomic formulas of L are members of A.*
2. *If $\alpha, \beta \in A$ do not contain anaphors of SRE occurrences, then $\neg(\alpha)$, $(\alpha) \wedge (\beta), (\alpha) \vee (\beta), (\alpha) \longrightarrow (\beta) \in A$.*
3. *If $\psi \in A$ and φ is the product of substituting anaphors for SRE occurrences in ψ as described in section 2c of the formula definition, then $\varphi \in A$.*
4. *If $\varphi(qP)$ is a formula in which an occurrence t of qP is the main QNP, then: if A contains every formula of the form $\varphi[t/s]$, where s is an SRE, then $\varphi \in A$.*

Then, A contains all the formulas in L.

To prove this theorem, one can prove, by induction on $\#\varphi$, that for any finite string φ, if φ is a formula, then $\varphi \in A$. We shall not give such a proof here.

2.3 Models, truth in a model

As we have already mentioned, our system is based upon the analysis of common nouns, in some of their uses, as referring expressions. The noun ‘whale’, for instance, is used referentially in sentences like ‘Every whale is a mammal'. 'Whale' refers here to whales; it does not refer to a *set* of whales, but to the whales themselves.

Now, for a referring expression to fulfill its task, there have to be some thing or things to which it can refer. For instance, if there are no whales, then 'whales' in the previous example cannot fulfill its semantic task. To exclude such failures of reference, we require that the extension of any one-place predicate be non-empty. This requirement guarantees that every component of our system fulfills its semantic task. It can be compared with the requirement in FOL, that any referring expression (i.e. any closed term; e.g., individual constants) be interpreted as designating some individual. This last demand, like the one stated above, excludes failures of reference. And while it yields the result that $\exists x(x = s)$ is true in every model in FOL, our demand concerning the extensions of predicates gives the same status to formulas of the form *some* P *is* P.

Definition 8 (Model). *A* model *for a language* L *is an ordered pair* $m = \langle M, \sigma\rangle$ *such that:*

1. M*, the* universe *of* m*, is a non-empty set.*

2. σ*, the* interpretation function*, is a function such that:*

   (a) *The domain of* σ *is the set of all SREs, predicates and predicate-transpositions of* L.

   (b) *If* s *is a singular referring expression, then* $\sigma(s) \in M$.

   (c) $\sigma(\mathit{Thing}) = M$.

   (d) *If* P *is a one-place predicate, then* $\sigma(P)$ *is a non-empty subset of* M.

   (e) *If* R *is an* n*-place predicate,* $n > 1$*, then* $\sigma(R) \subseteq M^n$.

   (f) *If* R *is an* n*-place predicate,* $n > 1$*, and* τ *is a non trivial permutation of* $\{1, \ldots, n\}$*, then*
   $\sigma(R\langle\tau(1), \ldots, \tau(n)\rangle) = \{\langle x_{\tau(1)}, \ldots, x_{\tau(n)}\rangle | \langle x_1, \ldots, x_n\rangle \in \sigma(R)\}$.

Note: in order to fully determine a model, it is enough to determine M and $\sigma(\alpha)$ for all SREs and predicates α.

It may be claimed that our requirement concerning the extensions of one-place predicates is more than is really needed: the extension of a predicate should be non-empty only if this predicate is used referentially, but in '*Every* S *is* P', for instance, P is not used in this way.

Our system would have been closer to natural language had we taken the following road: instead of excluding models that assign the empty set to some one-place predicates, we could have allowed them, and say that a formula containing a QNP of the form qP expresses a (true or false) proposition only in models in which $\sigma(P)$ is non-empty. This alternative approach may also be necessary if we would like to deal with quantified noun-phrases containing a defining clause, such as 'every man who owns a Jaguar'. It seems that the extension of 'man who owns a Jaguar' should be the set of all things that are both men and own a Jaguar. And requiring every two extensions to have a member in common seems to seriously limit our notion of model. The alternative approach would, however, result in a much more complicated system, and since we do not treat composed quantified noun-phrases to begin with, we shall stick to our original requirement: the extension of one-place predicates should never be empty.

Definition 9 (The Characteristic SRE). *For every Language L, let c_L be a new sign, not in L. L^* is defined as the language $L \cup c_L$, in which c_L is an SRE. c_L is* the characteristic SRE *of L.*

The above notion will be used in the definition of truth in a model. The idea is the following. Given a model m for L and a predicate P, we shall look at all the enrichments m' of m to the language L^* that interpret c_L as a member of the extension of P. These enrichments, which we shall call $\sigma(P)$-*enrichments*, will enable us to define the truth-conditions of quantified formulas: a formula $\varphi(qP)$, in which an occurrence t of qP is the main QNP, will be true in m iff $\varphi[t/c_L]$ is true in q of the $\sigma(P)$-enrichments of m. Let us now give the exact definitions.

Definition 10 (Enrichment, Restriction). *Let L_1, L_2 be formal languages, and assume $L_1 \subseteq L_2$.*[1] *Let $m_1 = \langle M_1, \sigma_1 \rangle$ and $m_2 = \langle M_2, \sigma_2 \rangle$ be models for L_1, L_2 respectively. m_2 is an* enrichment *of m_1 to L_2 if the following conditions hold: $M_1 = M_2$; $\sigma_1 \subseteq \sigma_2$ (i.e., for every predicate or SRE α in L_1, $\sigma_2(\alpha) = \sigma_1(\alpha)$). m_1, in this case, is a* restriction *of m_2 to L_1.*

Definition 11 (A-enrichment). *Let $m = \langle M, \sigma \rangle$ be a model for a language L, and let $A \subseteq M$. An enrichment $m' = \langle M', \sigma' \rangle$ of m to L^* is an A-*enrichment *of m if $\sigma'(c_L) \in A$.*

[1] We assume here that for every n, each n-place predicate of L_1 is an n-place predicate of L_2, each SRE of L_1 is an SRE of L_2, etc.

Note: In order to determine a specific A-enrichment m' of m, it is enough to choose a member $\alpha \in A$ and define $\sigma'(c_L) = \alpha$.

Definition 12 (Truth-conditions of Atomic Formulas). *Let φ be an atomic formula in a language L, and let $m = \langle M, \sigma \rangle$ be a model for L. The relation $m \models \varphi$ (φ is true in m) is defined as follows:*

1. *If s_1, s_2 are SREs, then: $m \models [s_1 \textit{ is } s_2]$ iff $\sigma(s_1) = \sigma(s_2)$; $m \models [s_1 \textit{ isn't } s_2]$ iff $\sigma(s_1) \neq \sigma(s_2)$.*

2. *If R is an n-place predicate ($n \geq 1$) or a transposition of such a predicate, and $s_1, \ldots, s_n$ are SREs, then:*
   $m \models [(s_1, \ldots, s_n) \textit{ is } R]$ iff $\langle \sigma(s_1), \ldots, \sigma(s_n) \rangle \in \sigma(R)$;
   $m \models [(s_1, \ldots, s_n) \textit{ isn't } R]$ iff $\langle \sigma(s_1), \ldots, \sigma(s_n) \rangle \notin \sigma(R)$.

Definition 13 (Truth-conditions of Formulas). *Let φ be a finite sequence, Let L be any language in which φ is a formula, and let $m = \langle M, \sigma \rangle$ be a model for L. The relation $m \models \varphi$ is defined by induction on $\#\varphi$:*

1. *If $\#\varphi \leq 3$, then φ is an atomic formula in L, and its truth-conditions in m are defined as in definition 12.*

2. *Let $n = \#\varphi$, and assume that for any $k < n$, if ψ is a string of length k, L' is a language in which ψ is a formula, and m' is a model for L', then it is already determined whether $m' \models \psi$. Let L be a language in which φ is a formula.*

   (a) *If φ is an atomic formula, then its truth-conditions in any model for L are as in definition 12.*

   (b) *If α and β are formulas in L that do not contain anaphors of SRE occurrences, then: $m \models \neg(\alpha)$ iff $m \not\models \alpha$; $m \models [(\alpha) \wedge (\beta)]$ iff $m \models \alpha$ and $m \models \beta$; $m \models [(\alpha) \vee (\beta)]$ iff $m \models \alpha$ or $m \models \beta$; $m \models [(\alpha) \longrightarrow (\beta)]$ iff it is not the case that $m \models \alpha$ and $m \not\models \beta$.*

   (c) *If φ is the product of substituting anaphors for SRE occurrences in a formula ψ as in section 2c of the formula definition, then $m \models \varphi$ iff $m \models \psi$.*

   (d) *If $\varphi(qP)$ is a formula that contains no anaphors of SRE occurrences, and in which an occurrence t of qP is the main QNP, then:*

i. If q is every, then: $m \models \varphi(every\ P)$ iff:
for every $\sigma(P)$-enrichment m' of m, $m' \models \varphi[t/c_L]$.

ii. If q is some, then: $m \models \varphi(some\ P)$ iff:
for some $\sigma(P)$-enrichment m' of m, $m' \models \varphi[t/c_L]$.

Note: If $m \models \varphi$, we also say that m *satisfies* φ, and that φ *holds in* m.

The only two quantifiers treated in our system are 'every' and 'some'. It should be noted, however, that our definition of the truth-conditions of quantified formulas can easily be extended to treat other quantifiers as well. Our basic idea was, that $\varphi(qP)$ is true in m iff $\varphi[t/c_L]$ is true in q $\sigma(P)$-enrichments of m. And this remains true for quantifiers such as 'seven', 'at least three' and 'most'. Our analysis of quantification gives a uniform account of all these quantifiers, as can be expected in view of the syntactic similarities between them in natural language. Such a uniform analysis is not available if we use standard versions of FOL as a tool for the analysis of natural language. As is well known, these versions cannot incorporate quantifiers such as 'most', which require restricted or binary quantification (cf. [1, section 6.4]; [4]).

Definition 14 (Theory). *A* theory *T in a language L is a set of formulas in L.*

Definition 15 (Model of a Theory). *Let T be a theory in a language L. m is a* model of *T if it is a model for L and $m \models \varphi$ for all $\varphi \in T$. In that case, we may also say that m* satisfies *T, etc.*

Definition 16 (Entailment). *A theory T entails a formula φ if φ is true in every model of T. In this case, we write: $T \models \varphi$.*

2.4 Deduction

We shall use a natural deduction system. Our way of writing proofs resembles the one found in Lemmon [5] and in Newton-Smith [6].

Definition 17 (Proof). *Let L be a formal language. A* proof *in L is a finite sequence of 4-tuples of the form $\langle \alpha, (k), \varphi, J \rangle$, called the* lines *of the proof, where:*

(a) *α is a finite (possibly empty) set of natural numbers, all of which are smaller than or equal to k. Lines $\langle \alpha', (k'), \varphi', J' \rangle$ in the proof for which $k' \in \alpha$ will be called* the lines on which the k-th line relies*. The formulas φ' in such lines will be called* the formulas on which the k-th line relies.

(b) *k,* the line's number*, is a natural number. The first line in a proof has $k = 1$, the second – $k = 2$, etc.*

(c) *φ is a formula in L.*

(d) *J,* the justification of the k-th line*, is written in accordance with one of the following rules.*

The following derivation rules allow the beginning of a proof and the addition of lines to a given proof. In fact, these rules complete definition 17 to a precise definition of proof, by induction on the number of lines.

For the sake of convenience, we occasionally drop the parenthesis '$\langle , \rangle$' or commas when referring to lines in a proof. Also, instead of writing the set of lines on which a certain line relies, we sometimes write the members of this set. In case this set is empty, we may not write anything. Another convenient convention is the following: a proof containing a single line is identified with that line.

17.1 (Premise). *If φ is a formula in L, then $\langle 1(1)\varphi$* Premise$\rangle$ *is a proof. Also, if D is a proof of length $k - 1$ (i.e., it has exactly $k - 1$ lines), then we may add to D the line: $\langle k(k)\varphi$* Premise$\rangle$ *(that is: the addition of such a line to D gives a proof).*

17.2 (*Thing* Introduction). *If s is an SRE, then $\langle (1)s$ is Thing Th I$\rangle$ is a proof. Also, if D is a proof of length $k - 1$, then we may add to D the line: $\langle (k)s$ is Thing Th I$\rangle$.*

17.3 (Identity Introduction). *If s is an SRE, then $\langle (1)s$ is s Id I$\rangle$ is a proof. Also, if D is a proof of length $k - 1$, then we may add to D the line $\langle (k)s$ is s Id I$\rangle$.*

17.4 (Identity Elimination). *Let D be a proof of length $k - 1$. Assume that s and s' are SREs, and that D includes the line: $\langle \alpha(i)s$ is $s' J_i \rangle$. Assume also that D includes a line of the form $\langle \beta(j)\varphi J_j \rangle$, where φ contains the occurrences $c_1, \ldots, c_n$ of s (φ may contain other occurrences*

of s as well).
Then we may add to D the line $\langle \alpha \cup \beta(k)\varphi[c_1/s', \ldots, c_n/s']$ *Id* $E, i, j\rangle$.[2]

17.5 (Propositional Calculus Rules). *We allow the usual propositional calculus derivation rules for formulas that do not contain anaphors of SRE occurrences. We shall give only two examples here:*

$\longrightarrow$ ***Introduction.*** *Let D be a proof of length $k-1$. If D contains the lines* $\langle i(i)\varphi$ Premise $\rangle$; $\langle\beta(j)\psi J\rangle$, *where φ and ψ do not contain anaphors of SRE occurrences, then we may add to D the line* $\langle\beta\backslash\{i\}\ (k)\varphi \longrightarrow \psi\ \longrightarrow\ I, i, j\rangle$.

$\vee$ ***Elimination.*** *Let D be a proof of length $k-1$. Assume that D contains the lines* $\langle\alpha(i)\varphi\ \vee\ \psi J_i\rangle$; $\langle j(j)\varphi$ Premise $\rangle$; $\langle\beta(l)\delta J_l\rangle$; $\langle m(m)\psi$ *Premise* $\rangle$; $\langle\gamma(n)\delta J_n\rangle$, *where φ and ψ do not contain anaphors of SRE occurrences, $j \notin \gamma$ and $m \notin \beta$. Then we may add to D the line* $\langle((\beta \cup \gamma)\backslash\{j, m\}) \cup \alpha(k)\delta\ \vee\ E, i, j, l, m, n\rangle$.

17.6 (Transposition). *Let D be a proof of length $k-1$, and let τ and ξ be any permutations of $\{1, \ldots, n\}$. Assume that D contains the line* $\langle\alpha(i)(s_{\tau(1)}, \ldots, s_{\tau(n)})$ *is* $R\langle\tau(1), \ldots, \tau(n)\rangle J\rangle$, *where $s_1, \ldots, s_n$ are SREs. Then we may add to D the line*
$\langle\alpha(k)(s_{\xi(1)}, \ldots, s_{\xi(n)})$ *is* $R\langle\xi(1), \ldots, \xi(n)\rangle$ *Tr* $, i\rangle$.

17.7 (Negative-Copula Introduction). *Let D be a proof of length $k-1$. Let R be an* n*-place predicate ($n \geq 1$) or a transposition of such a predicate, and let $s_1, \ldots, s_n$ be SREs. If D contains the line* $\langle\alpha(i)\neg((s_1, \ldots, s_n)$ *is* $R)J\rangle$, *then we may add to D the line*
$\langle\alpha(k)(s_1, \ldots, s_n)$ *isn't* R *NC* $I, i\rangle$.

17.8 (Negative-Copula Elimination). *Let D be a proof of length $k-1$. Let R be an* n*-place predicate ($n \geq 1$) or a transposition of such a predicate, and let $s_1, \ldots, s_n$ be SREs. If D contains the line* $\langle\alpha(i)(s_1, \ldots, s_n)$ *isn't* $R\ J\rangle$, *then we may add to D the line*
$\langle\alpha(k)\neg((s_1, \ldots, s_n)$ *is* $R)NC\ E, i\rangle$.

17.9 (Anaphors Introduction). *Let D be a proof of length $k-1$. Assume that D contains the line* $\langle\alpha(i)\psi J\rangle$. *If φ is the product of substituting*

[2] It is not hard to show that substituting SRE occurrences for SREs in a formula gives a formula. Therefore, $\varphi[c_1/s', \ldots, c_n/s']$ is a formula.

anaphors for SRE occurrences in ψ as in section 2c of the formula definition, then we may add to D the line $\langle \alpha(k)\varphi A\ I, i\rangle$.

17.10 (Anaphors Elimination). *Let D be a proof of length* $k-1$. *Assume that D contains the line* $\langle \alpha(i)\varphi J\rangle$. *If φ is the product of substituting anaphors for SRE occurrences in a formula ψ as in section 2c of the formula definition, then we may add to D the line* $\langle \alpha(k)\psi A\ E, i\rangle$.

17.11 (*every* Introduction). *Let $\varphi(every\ P)$ be a formula in which an occurrence t of every P is the main QNP, and assume that φ does not contain s. Let D be a proof of length $k-1$, and assume that D includes the lines* $\langle i(i)s\ is\ P\ \text{Premise}\rangle$; $\langle \beta(j)\varphi[t/s]J\rangle$. *Also assume that β does not contain any number different than i of a line in which s occurs. Then, we may add to D the line* $\langle \beta\backslash\{i\}(k)\varphi(every\ P)\ every\ I, i, j\rangle$.

17.12 (*every* Elimination). *Let $\varphi(every\ P)$ be a formula in which an occurrence t of every P is the main QNP, and let s be any SRE. Let D be a proof of length $k-1$, and assume that D includes the lines* $\langle \alpha(i)\varphi(every\ P)J_i\rangle$; $\langle \beta(j)s\ is\ PJ_j\rangle$. *Then, we may add to D the line* $\langle \alpha\cup\beta(k)\varphi[t/s]\ every\ E, i, j\rangle$.

17.13 (*some* Introduction). *Let $\varphi(some\ P)$ be a formula in which an occurrence t of some P is the main QNP. Let D be a proof of length $k-1$, and assume that D includes the lines* $\langle \alpha(i)\varphi[t/s]J_i\rangle$; $\langle \beta(j)s\ is\ PJ_j,\rangle$, *where s is an SRE. Then we may add to D the line*
$\langle \alpha\cup\beta(k)\varphi(some\ P)\ some\ I, i, j\rangle$.

17.14 (*some* Elimination). *Let $\varphi(some\ P)$ be a formula in which an occurrence t of some P is the main QNP. Assume that φ does not contain the SRE s, and that ψ is a formula that does not contain s. Let D be a proof of length $k-1$, and assume that D includes the lines* $\langle \alpha(i)\varphi(some\ P)J_i\rangle$; $\langle j(j)s\ is\ P\ \text{Premise}\rangle$; $\langle k(k)\varphi[t/s]\ \text{Premise}\rangle$; $\langle \beta(l)\psi J_l\rangle$. *Also assume that $j, k \notin \alpha$, and that β does not contain any number, other than j and k, of a line in which s occurs. Then we may add to D the line* $\langle (\alpha\cup\beta)\backslash\{j,k\}(m)\psi\ some\ E, i, j, k, l\rangle$.

17.15 (Referential Import). *Let $\varphi(every\ P)$ be a formula in which an occurrence t of every P is the main QNP. Let D be a proof of length $k-1$, and assume that D includes the line* $\langle \alpha(i)\varphi(every\ P)J\rangle$. *Then we may add to D the line* $\langle \alpha(k)\varphi[t/some\ P]\ RI, i\rangle$.

Referential Import captures the referential use of one-place predicates in our system. With it the list of derivation rules of our system was concluded.

Definition 18 ($T_D(\alpha)$). *If α is a set of numbers of lines in a proof D, then $T_D(\alpha)$ is the set of all formulas that appear in D in lines whose numbers belong to α.*

Definition 19 (Provability of Formulas). *Let T be a theory in a language L, and let φ be a formula in L. φ is* provable *from T in L, if there is a proof D in L such that:*

1. *The last line in D is of the form: $\langle \alpha, (k), \varphi, J \rangle$.*

2. *$T_D(\alpha) \subseteq T$ (in other words: the last line of D relies only on members of T).*

D, in this case, is called a proof of φ from T.

Definition 20 (Provability of Theories). *Let T_1, T_2 be theories in a language L. T_2 is provable from T_1 if $T_1 \vdash \varphi$ for any $\varphi \in T_2$. In this case we write: $T_1 \vdash T_2$.*

3 Some Examples of Formalization and Proofs

We shall now give a few examples of proofs in our formal system, so that the reader gets a feel of it. These examples will also supply us with an opportunity to comment on some of the characteristics of our system, mainly in relation to other formal systems.

Consider first the following inference (Contrariety):

> Every philosopher is wise; hence, it's not the case that every philosopher isn't wise.

These sentences translate into our system as:

> *every S is P*; $\neg$(*every S isn't P*)

And the inference can be proved as follows:

1	(1)	*every S isn't P*	*Premise*
2	(2)	*every S is P*	*Premise*
2	(3)	*some S is P*	$RI, 2$
4	(4)	*s is S*	*Premise*
5	(5)	*s is P*	*Premise*
1, 4	(6)	*s isn't P*	*every* $E, 1, 4$
1, 4	(7)	$\neg$(*s is P*)	$NC\ E, 6$
1, 4, 5	(8)	(*s is P*) $\wedge \neg$(*s is P*)	$\wedge\ I, 5, 7$
4, 5	(9)	$\neg$(*every S isn't P*)	$\neg\ I, 1, 8$
2	(10)	$\neg$(*every S isn't P*)	*some* $E, 3, 4, 5, 9$

Secondly, let us translate and prove the following inference (Darii):

> Some philosophers are Athenians; every Athenian is Greek; hence, some philosophers are Greek.

Its translation:

> *some S is M*; *every M is P*; *some S is P*

And its proof:

1	(1)	*some S is M*	*Premise*
2	(2)	*every M is P*	*Premise*
3	(3)	*s is S*	*Premise*
4	(4)	*s is M*	*Premise*
2, 4	(5)	*s is P*	*every* $E, 2, 4$
2, 3, 4	(6)	*some S is P*	*some* $I, 5, 3$
1, 2	(7)	*some S is P*	*some* $E, 1, 3, 4, 6$

As the reader would have noticed, these two inferences are part of the valid inferences of Aristotelian logic: the first belongs to the Square of Opposition, the second to the Syllogisms. All the other valid inferences of Aristotelian logic can also be proved in our system (cf. [1, chap. 10]). Our system thus contains Aristotelian logic. By contrast, on any acceptable translation of the four Aristotelian sentences (*every*/*some S is*/*isn't P*) into FOL, some of these inferences turn out invalid (unless some *ad hoc* axioms are added to the calculus; see below). We think this demonstrates the fact that the analysis of the semantics of natural language on which our system is constructed is more adequate than what a similar analysis, using only the semantic categories of FOL, can supply. A formal system whose verdict on natural language inferences coincides with what logicians considered valid for more than two millennia obviously has a desirable feature.

On the other hand, unlike Aristotelian logic, our system can also prove inferences that involve multiply quantified sentences. For instance:

> Some women are loved by every man; hence, every man loves some women.

Notice that these sentences use both the passive and the active form of the same verb. This is translated into our system as a relation-sign and its transposition. If we translate 'a loves b' as (a, b) *is* L, then 'b is loved by a' should be translated as (b, a) *is* $L\langle 2, 1\rangle$. The former sentences are thus translated as:

$$(some\ W,\ every\ M)\ is\ L\langle 2,1\rangle;\ (every\ M,\ some\ W)\ is\ L.$$

Let us now prove this inference:

1	(1)	$(some\ W, every\ M)\ is\ L\langle 2,1\rangle$	*Premise*
2	(2)	$s_1\ is\ W$	*Premise*
3	(3)	$(s_1,\ every\ M)\ is\ L\langle 2,1\rangle$	*Premise*
4	(4)	$s_2\ is\ M$	*Premise*
3, 4	(5)	$(s_1, s_2)\ is\ L\langle 2,1\rangle$	$every\ E, 3, 4$
3, 4	(6)	$(s_2, s_1)\ is\ L$	$Tr, 5$
2, 3, 4	(7)	$(s_2, some\ W)\ is\ L$	$some\ I, 6, 2$
2, 3	(8)	$(every\ M, some\ W)\ is\ L$	$every\ I, 7, 4$
1	(9)	$(every\ M, some\ W)\ is\ L$	$some\ E, 1, 2, 3, 8$

Moreover, we can prove in our system inferences that involve sentences with anaphors of quantified noun phrases, a capacity which greatly increases our system's power. We shall give one simple example:

> Every man loves every man; hence, every man loves himself.

Its translation:

$(every\ M, every\ M)\ is\ L$; $((1)\ every\ M, (1)a)\ is\ L$.

And its proof:

1	(1)	$(every\ M,\ every\ M)\ is\ L$	*Premise*
2	(2)	$s\ is\ M$	*Premise*
1, 2	(3)	$(s,\ every\ M)\ is\ L$	$every\ E, 1, 2$
1, 2	(4)	$(s, s)\ is\ L$	$every\ E, 3, 2$
1, 2	(5)	$((1)s, (1)a)\ is\ L$	$A\ I, 4$
1	(6)	$((1)\ every\ M, (1)a)\ is\ L$	$every\ I, 5, 2$

These examples demonstrate the nature and power of our system.

4 Many-sorted Logic

Our formal system resembles in some ways many-sorted logic. We shall therefore pause to discuss the relation between the two.

In many-sorted logic, different sorts of variables are used, each sort having its own domain. For instance, one occasionally uses x_1, x_2, x_3 etc. to range over particulars; e_1, e_2, e_3 etc. to range over events; t_1, t_2, t_3 etc. to range over times; and so on. Since in this case the variables determine the sort of pluralities about which something is said, and since different sorts of variables may determine different sorts of pluralities, it seems many-sorted logic can be considered a kind of logic with plural referring expressions, namely its variables.

Yet a significant logical distinction between many-sorted logic and our formal system (and natural language) still remains. In our system, one-place predicate letters can be used both as referring expressions and as predicates (correspondingly, in natural language some concept-words are used both as referring expressions and as one-place predicates). Consider, for instance, the two sentences:

Some Athenians are philosophers.
Every philosopher is wise.

'Philosophers' is used as a predicate in the first sentence and referentially in the second. These sentences translate into our system as, respectively:

some A is P
every P is W

Consequently, one can derive in our system the formula *some A is W* from these two formulas by means of syntactic derivation rules.

By contrast, many-sorted logic would either use, like ordinary one-sorted logic, the same variable when translating both sentences – in which case it would not mirror our use of different plural referring expressions in the two sentence; or it may use different variables in the two translations, e.g.:

$\exists x_1 P x_1$
$\forall y_1 W y_1.$

Here x_1 and y_1 are variables of different sorts. But in this latter case, as can be seen, the syntactic relation between the predicate P in the first formula, and the variable y_1 in the second, is lost. Consequently, one cannot derive in this case by means of syntactic derivation rules the translation of natural language's 'Some Athenians are wise' ($\exists x_1 W x_1$) from these two formulas. This is obviously an undesirable result.

To avoid this result, one may use, for instance, the same letters both as variables and as one-place predicate letters. Each one-place predicate will then be interpreted as designating its own domain of discourse. Appropriate syntactic derivation rules could then be introduced (which, although probably more complex than the usual ones, may be rather similar to those of our system).

But additional modifications of many-sorted logic should also be introduced. For instance, predicate letters in formulas of many-sorted logic usually combine with variables of specific sorts in order to form well formed formulas (see [2, pp. 295ff]). However, in order to translate both 'Every philosopher is wise' and 'Some Athenians are wise', the predicate W should combine both with the variable P and the variable A. Similarly, individual constants, which correspond in ordinary many-sorted logic to sorts, should not be classified into sorts in the modified version, in order to translate sentences like 'Socrates is an Athenian' and 'Socrates is a philosopher'. Moreover, in order to distinguish between sentences like, say, 'Every philosopher loves some philosopher' and 'Every philosopher is loved by some philosopher', several variable letters should be assigned, as usual, to each sort or predicate letter. If we use upper case letters for predicates, indexed lower case letters for variables, these sentences would be translated as, respectively:

$\forall p_1 \exists p_2$ Loves (p_1, p_2)
$\forall p_1 \exists p_2$ Loves (p_2, p_1).

This contrasts with our system, which uses no variables (or anaphors) for such sentences:

> $(every\ P,\ some\ P)\ is\ L$
> $(every\ P,\ some\ P)\ is\ L\langle 2,1\rangle$

As can also be seen, natural language's need for a distinction between active and passive voice, or some such linguistic device, which is preserved in our formal system, is lost in many-sorted logic, in its modified form as well. The same applies to the need for a distinction between an affirmative and a negative copula.

So many-sorted logic should be significantly modified to resemble our formal system, and even then some important distinctions would still remain.

Indeed, many-sorted logic, even in its usual form, can be shown to parallel our system in its deductive power, in the sense that this will be shown below for FOL. In fact, this follows immediately from the proofs that shall be given below, together with the fact that many-sorted logic can be reduced to one-sorted logic (see [2, pp. 296ff]). But remember that our purpose in the development of a new formal system was not to capture some new forms of inference, but to show that an alternative analysis of the semantics of natural language can serve as the basis for a formal system similar in its power to FOL. Consequently, we do not consider the fact that our system is similar in its power to FOL, on any of its versions – e.g., many-sorted logic – as a drawback, but rather as an advantage.

5 Properties of Our Formal System

In this part we shall prove some of the properties of our system. Our main goal will be to show that the system's deductive power is comparable to that of FOL. To be more precise, we shall prove that our system is equivalent to FOL, supplemented by all axioms of the form $\exists xPx$, which we shall call *axioms of existential import*. The set of all these axioms will be called *EI*. We shall correlate models in our system with models of *EI* in FOL, and define a translation of formulas in FOL into our system. We will prove this translation to be one-to-one, and to cover all the formulas in our system in the following sense: each of these formulas is both deductively and semantically equivalent to a translation of some formula of FOL. We shall also prove that the translation preserves truth in a model, entailment and provability. The existence of such a translation, together with the completeness of FOL will entail the completeness of our system.

Let us now turn to the exact definitions and proofs.

5.1 Some basic results

This section contains some elementary lemmas and theorems that will be used in the proofs below. Some of these results will be stated without proof.

Given a formula φ and a model m, we can change some of the SREs in φ. If the new SREs are interpreted by m in exactly the same way as the old ones, the replacement should not affect the truth of φ in m. It should not matter even if the new SREs belong to a language richer than the one we started with, as long as we enrich m accordingly. This is, more or less, the content of the following lemma. Its somewhat complex formulation is meant to enable an easier proof by induction.

Lemma 1. *Let φ be a formula in a language L, and assume that φ contains the occurrences $c_1, \ldots, c_n$ of SREs $s_1, \ldots, s_n$ respectively ($s_1, \ldots, s_n$ not necessarily different). Let $s'_1, \ldots, s'_n$ and $s''_1, \ldots, s''_n$ be SREs, some (or all) of which may not belong to L ($s'_1, \ldots, s'_n, s''_1, \ldots, s''_n$ not necessarily different).*
Let $m' = \langle M', \sigma' \rangle$ be a model for a language L' that contains $L \cup \{s'_1, \ldots, s'_n\}$, and let $m'' = \langle M'', \sigma'' \rangle$ be a model for a language L'' that contains $L \cup \{s''_1, \ldots, s''_n\}$. Assume that m' and m'' coincide with m in L (that is: $M' = M'' = M$, and $\sigma'(\alpha) = \sigma''(\alpha) = \sigma(\alpha)$ for any $\alpha \in \mathsf{dom}\ (\sigma)$).
If m' interprets $s'_1, \ldots, s'_n$ as m interprets $s_1, \ldots, s_n$ (i.e. $\sigma'(s'_i) = \sigma(s_i)$ for every i), and m'' interprets $s''_1, \ldots, s''_n$ as m interprets $s_1, \ldots, s_n$, then: $m' \models \varphi[c_1/s'_1, \ldots, c_n/s'_n] \iff m'' \models \varphi[c_1/s''_1, \ldots, c_n/s''_n]$.

It is obvious that lemma 1 is true: from the definition of truth in a model it can be seen that SREs contribute to the truth of φ in a given model only through the way in which they are interpreted in that model; and if two SREs are assigned the same object, then they ought to have the same contribution to the truth of φ.

Theorem 2 (Agreement of Models and their Enrichments). *Let m_1 be a model for L_1, and let m_2 be an enrichment of m_1 to L_2. If φ is a formula in L_1 (and therefore, also in L_2), then: $m_1 \models \varphi \iff m_2 \models \varphi$.*

The following theorem shows that the reliance on c_L in the definition of truth in a model is not necessary; in order to define the truth-conditions of a quantified formula φ, we could use any SRE of L that does not occur in φ. The notion of $\sigma(P)$-enrichment can be replaced by that of $\sigma(P)$-change:

Definition 21 (A-s-Change). *Let $m = \langle M, \sigma \rangle$ be a model for a language L, let s be an SRE in L, and let $A \subseteq M$. A model $m' = \langle M', \sigma' \rangle$ for L is an A-s-*change *of m if the following conditions hold:*

1. *$M' = M$.*
2. *$\sigma'(s) \in A$.*
3. *For any $\alpha \in$ dom $(\sigma) \backslash \{s\}$, $\sigma'(\alpha) = \sigma(\alpha)$.*

Note: In order to fully determine an A-s-change m' of m, it is sufficient to choose any $\alpha \in A$, and define $\sigma'(s)$ to be α .

Theorem 3 (Truth-conditions of Quantified Formulas). *Let $\varphi(qP)$ be a formula in a language L, and assume that an occurrence t of qP is the main QNP in φ. Let $m = \langle M, \sigma \rangle$ be a model for L, and let s be an SRE not occurring in φ.*

1. *If q is every, then: $m \models \varphi(every\ P)$ iff: for every $\sigma(P)$-s-change m' of m, $m' \models \varphi[t/s]$.*
2. *If q is some, then: $m \models \varphi(some\ P)$ iff: for some $\sigma(P)$-s-change m' of m, $m' \models \varphi[t/s]$.*

The following two theorems follow immediately from lemma 1 and theorem 2, respectively.

Theorem 4 (Interchangeability of Identicals). *Let φ be a formula in a language L, and let $m = \langle M, \sigma \rangle$ be a model for L. Assume that φ contains the occurrences $c_1, \ldots, c_n$ of an SRE s (and maybe some other occurrences of that SRE as well). If $m \models \varphi$ and $m \models [s \text{ is } s']$, then $m \models \varphi[c_1/s', \ldots, c_n/s']$.*

Theorem 5 (Agreement of Models and Their A-s-changes). *Let m be a model for a language L, let $A \subseteq M$, and let m' be an A-s-change of m. If φ is a formula (in L) that does not contain s, then: $m' \models \varphi \Longleftrightarrow m \models \varphi$.*

The following lemma says, roughly, that any proof can be replaced by a proof in which no premise appears twice.

Lemma 2. *Let D be a proof in our system, and assume that the last line in D is $\langle \alpha(n)\varphi J\rangle$. Then there exists a proof D' of φ from $T_D(\alpha)$ in which no premise appears more than once (i.e. if $\langle i(i)\psi$ Premise$\rangle$ and $\langle j(j)\psi$ Premise$\rangle$ are both lines of D' , then $i = j$). Moreover, we can assume that for any line $\langle \beta(i)\psi J_i\rangle$ in D, there exists a line $\langle \beta'(i')\psi J'_i\rangle$ in D' such that $T_D(\beta) = T'_D(\beta')$.*

This lemma follows from the following fact: whenever one of our inference rules allows us to rely on a given premise, it places no restriction on the place (the line-number) of that premise in the proof. Therefore, the need to write a premise that already appears in the proof never arises.

Lemma 3 (Concatenation of Proofs). *Let D_1, D_2 be proofs in a language L, and assume that D_2 does not contain the same premise twice. Assume that the last line of D_1 is $\langle \beta(k)\psi J\rangle$. Also assume that the last line of D_2 is $\langle \alpha(n)\varphi J'\rangle$, where $T_{D_2}(\alpha) \subseteq \{\psi\}$ (that is: D_2 is a proof of φ from ψ, which is the formula in the last line of D_1). Then, there exists a proof D in L such that:*

1. *The first $\#D_1$ lines of D are exactly those of D_1, in the same order.*
2. *The last line of D is $\langle \beta'(m)\varphi J''\rangle$, where $\beta' = \beta$ if $T_{D_2}(\alpha) = \{\psi\}$ (i.e. if the last line of D_2 indeed relied on ψ) and $\beta' = \emptyset$ if $T_{D_2}(\alpha) = \emptyset$ (i.e. if D_2 is a proof of φ from $\emptyset$).*

D is called the concatenation *of D_1 and D_2, and we write: $D = (D_1, D_2)$.*

We shall not give a precise proof of this lemma. The idea is the following: we start with D_1, and apply to it the rules applied in D_2, one by one. If, in doing that, if we have to rely on ψ, we rely on the last line of D_2.

Lemma 4. *Let ψ, ψ' be formulas in a language L. If $\psi' \vdash \psi$, then there exists a proof D that has a last line of the form $\langle \alpha(n)\psi J\rangle$, where $T_D(\alpha) = \psi'$.*

The following lemma asserts that we can replace a premise in a proof with a stronger premise, without significantly changing the rest of the proof.

Lemma 5. *Let D be a proof in our system, and assume that no premise appears in D twice. Also assume that D includes the line: $\langle i(i)\psi$ Premise $\rangle$. Let ψ' be a formula such that $\psi' \vdash \psi$. Then there exists a proof D' such that:*

1. *For any $j < i$, the j-th lines of D and D' are identical.*

2. *The i-th line of D' is $\langle i(i)\psi'$ Premise $\rangle$.*

3. *D' includes a line of the form $\langle i(k)\psi J_k\rangle$.*

4. *For any $j > i$, if D includes the line $\langle\alpha(j)\varphi J\rangle$, then D' includes a line $\langle\alpha'(j')\varphi J'\rangle$, where $T_{D'}(\alpha') = T_D(\alpha)$ up to the replacement of ψ with ψ' (that is: in case $\psi \in T_D(\alpha)$, we do not have the above equality, but instead: $T_{D'}(\alpha') = (T_D(\alpha)\backslash\{\psi\}) \cup \{\psi'\}$. Otherwise – we have equality).*

   A line that stands in the above relation to $\langle\alpha(j)\varphi J\rangle$ will be called a twin *of $\langle\alpha(j)\varphi J\rangle$.*

This lemma can be proved by induction on $\#D$. We shall not give such a proof here.

Theorem 6 (Provability of Theories is Transitive). *Let T_1, T_2, T_3 be theories in a language L. If $T_1 \vdash T_2$ and $T_2 \vdash T_3$, then $T_1 \vdash T_3$.*

This theorem follows from the fact that any proof uses only a finite number of premises, and from the following three lemmas, that hold for any φ_i, ψ_i and φ, and can be proved using our previous results:

1. If $\varphi_1 \vdash \varphi_2$ and $\varphi_2 \vdash \varphi_3$, then $\varphi_1 \vdash \varphi_3$.

2. $\varphi_1, \ldots, \varphi_n \vdash \varphi \Longleftrightarrow \varphi_1 \wedge \ldots \wedge \varphi_n \vdash \varphi$.

3. If $\varphi_1 \vdash \psi_1, \varphi_2 \vdash \psi_2, \ldots, \varphi_n \vdash \psi_n$, then: $\varphi_1 \wedge \ldots \wedge \varphi_n \vdash \psi_1 \wedge \ldots \wedge \psi_n$.

5.2 Soundness

Theorem 7 (Soundness). *Let T be a Theory in a language L, and let φ be a formula in L. If $T \vdash \varphi$, then $T \models \varphi$.*

Proof: To prove the theorem, it is convenient to prove the following proposition by induction on n: let D be a proof of length n. If the last line in D is $\langle\alpha(n)\varphi J\rangle$, then $T_D(\alpha) \models \varphi$.

The induction base is trivial. Assume now that the proposition holds for any $k < n$. To complete the proof, we need to check each of the possibilities for the justification of the n-th line in D , and prove that $T_D(\alpha) \models \varphi$ in each of them. We give the case of *some E* as an example.

If the last line is justified by *some E*, then D includes lines of the forms $\langle\alpha(i)\psi(\textit{some } P)J\rangle$; $\langle j(j)s \textit{ is } P \textit{ Premise }\rangle$; $\langle k(k)\psi[t/s] \textit{ Premise }\rangle$; $\langle\beta(l)\delta J_l\rangle$, where: $\psi(some\ P)$ is a formula in which an occurrence t of *some* P is the main QNP; neither $\psi(some\ P)$ nor δ contains s; $j, k \notin \alpha$; β does not contain any number, other than j and k, of a line in which s occurs. Also, the last line in D is $\langle(\alpha \cup \beta)\backslash\{j,k\}(n)\delta\ some\ E, i, j, k, l\rangle$. Let m be a model of $T_D((\alpha \cup \beta)\backslash\{j,k\})$. We shall prove that $m \models \delta$. Since $j, k \notin \alpha$, we have $T_D(\alpha) \subseteq T_D((\alpha \cup \beta)\backslash\{j,k\})$. Therefore: $m \models T_D(\alpha)$, and by the induction hypothesis: $m \models \psi(some\ P)$. Since ψ does not contain s, it follows, by theorem 3, that there exists a $\sigma(P)$-s-change m' of m such that $m' \models \psi[t/s]$. m', as a $\sigma(P)$-s-change, also satisfies s *is* P. That is: m' satisfies the formulas in lines j and k, or, in other words: $m' \models T_D(\{i,j\})$. Now, since $T_D(\beta\backslash\{j,k\}) \subseteq T_D((\alpha \cup \beta)\backslash\{j,k\})$, we have: $m \models T_D(\beta\backslash\{i,j\})$. And since none of the formulas in $T_D(\beta\backslash\{i,j\})$ contains s (β does not contain any number, other than j and k, of a line in which s occurs), we have: $m' \models T_D(\beta\backslash\{i,j\})$ (this follows from $m \models T_D(\beta\backslash\{i,j\})$, by theorem 5). Therefore: $m' \models T_D(\beta)$. From the induction hypothesis it now follows that $m' \models \delta$. And since δ does not contain s, it follows (by theorem 5) that $m \models \delta$.

■

5.3 The version of FOL that will be used below

We shall use a version of FOL with identity, without function signs, and without open formulas. The version of FOL to be defined and used below can be proved to be equivalent to standard versions found in the literature. This section contains the definitions of the relevant terms. Most of these terms were also used in defining our formal system. When we use a term below, we shall take care to specify, in cases where confusion might arise, whether it refers to FOL or to our system.

Definition 22 (Formal Language). *A* formal language L *is a disjoint union of eight sets:* $\mathcal{P}$ *– a set of* one-place predicates*;* $\mathcal{R}$ *– a set of* relation-signs *or* many-place predicates *(to every one of which we assign a natural number* $n > 1$*, called its number of places);* $\{=\}$ *– the* identity sign*;* $\mathcal{S}$ *– a denumerable set of* individual constants*;* $\{x_1, x_2, \ldots\}$ *– a denumerable set of* variables*;* $\{\neg, \wedge, \vee, \longrightarrow\}$ *–* connectives*;* $\{\forall, \exists\}$ *–* quantifiers*;* $\{), (\}$ *–* parenthesis.

Note: We use the notation $\varphi[\alpha/\beta]$ for FOL in the same way we use it for our system.

Definition 23 (Formula).

1. *Any string of the form:* $Rs_1 \ldots s_n$ *or:* $s_1 = s_2$*, where* $s_1, \ldots, s_n$ *are individual constants, and* R *– an* n*-place predicate, is a formula. Strings of these forms are also called* atomic formulas*.*

2. *If* α, β *are formulas, then so are* $\neg(\alpha)$*,* $(\alpha) \wedge (\beta)$*,* $(\alpha) \vee (\beta)$*,* $(\alpha) \longrightarrow (\beta)$*.*

3. *If* ψ *is a formula that contains an individual constant* s *and* x *is a variable that does not occur in* ψ*, then* $\forall x(\psi[s/x])$ *and* $\exists x(\psi[s/x])$ *are formulas.*

4. *Nothing else is a formula.*

Note: When referring to formulas, we shall sometimes omit parenthesis, for the sake of convenience.

The definition of truth in a model that will be used below is close to the one we use in our system. We begin by introducing the notion of characteristic constant:

Definition 24 (Characteristic Constant). *For every Language* L*, let* c_L *be a new sign, not in* L*.* L^* *is defined as the language* $L \cup \{c_L\}$*, in which* c_L *is an individual constant.* c_L *is* the characteristic constant *of* L*.*

Definition 25 (Model). *Let* L *be a formal language. A* model *for* L *is an ordered pair* $m = \langle M, \sigma \rangle$ *such that:* M*, the* universe *of* m*, is a non-empty set;* σ*, the* interpretation function*, is a function such that:*

1. *The domain of* σ *is the set of all constants and predicates of* L*.*

2. *If* s *is a constant, then* $\sigma(s) \in M$*.*

3. *If* R *is an* n*-place predicate, then* $\sigma(R) \subseteq M^n$*.*

Enrichment and restriction of a model are defined as in our system (definition 10 above).

Definition 26 (Truth in a Model). 1. *If* $s_1, \ldots, s_n$ *are individual constants and* R *an* n*-place predicate in a language* L*, and* $m = \langle M, \sigma \rangle$ *– a model for* L*, then:* $m \models [Rs_1 \ldots s_n]$ *iff* $\langle \sigma(s_1), \ldots, \sigma(s_n) \rangle \in \sigma(R)$*;* $m \models [s_1 = s_2]$ *iff* $\sigma(s_1) = \sigma(s_2)$*.*

2. *If α and β are formulas in a language L, and $m = \langle M, \sigma \rangle$ – a model for L, then:* $m \models \neg\alpha$ *iff* $m \not\models \alpha$; $m \models [\alpha \wedge \beta]$ *iff* $m \models \alpha$ *and* $m \models \beta$; $m \models [\alpha \vee \beta]$ *iff* $m \models \alpha$ *or* $m \models \beta$; $m \not\models [\alpha \longrightarrow \beta]$ *iff* $m \models \alpha$ *and* $m \not\models \beta$.

3. *Let ψ be a formula in a language L, and let $m = \langle M, \sigma \rangle$ be a model for L, assume that ψ contains the individual constant s and does not contain the variable x. Then:* $m \models \forall x(\psi[s/x])$ *iff* $m' \models \psi[s/c_L]$ *for every enrichment m' of m to L^*;* $m \models \exists x(\psi[s/x])$ *iff* $m' \models \psi[s/c_L]$ *for at least one enrichment m' of m to L^*.*

In section 5.1, we explained that the truth-conditions of quantified formulas in the system we defined can be determined without reference to c_L (by what we called *A*-*s*-changes). A similar theorem holds for FOL. First, we define:

Definition 27 (*s*-Change). *Let $m = \langle M, \sigma \rangle$ be a model for a language L in FOL, let s be an individual constant in L. A model $m' = \langle M', \sigma' \rangle$ for L is an* s-change *of m if the following conditions hold:*

1. $M' = M$.

2. *For any* $\alpha \in \textsf{dom}\,(\sigma)\backslash\{s\}$, $\sigma'(\alpha) = \sigma(\alpha)$.

Note: In order to fully determine an s-change m' of m, it is sufficient to choose any $\beta \in M$, and define $\sigma'(s)$ to be β.

Theorem 8 (Truth-conditions of Quantified Formulas). *Let $qx\varphi[s/x]$ be a formula in a language L in FOL. Let $m = \langle M, \sigma \rangle$ be a model for L, and let s be an SRE not occurring in φ.*

1. *If q is $\forall$, then:* $m \models \forall x\varphi(x)$ *iff:*
   for every s-change m' of m, $m' \models \varphi[x/s]$.

2. *If q is $\exists$, then:* $m \models \exists x\varphi(x)$ *iff:*
   for some s-change m' of m, $m' \models \varphi[x/s]$.

We shall not prove this theorem here.

Definition 28 (Theory). *A Theory in a language L in FOL is a set of formulas in L.*

Model of a theory and entailment are defined as in section 2.3 above (definitions 15, 16).

Definition 29 (Proof). *Let L be a formal language. A* proof *in L is a finite sequence of 4-tuples of the form $\langle \alpha, (k), \varphi, J \rangle$, called the* lines *of the proof, where:*

a. *α is a finite (possibly empty) set of natural numbers, all of which are smaller than or equal to k. Lines $\langle \alpha', (k'), \varphi', J' \rangle$ in the proof for which $k' \in \alpha$ will be called* the lines on which the k-th line relies. *The formulas φ' in such lines will be called* the formulas on which the k-th line relies.

b. *k,* the line's number, *is a natural number. The first line in a proof has $k = 1$, the second – $k = 2$, etc.*

c. *φ is a formula in L.*

d. *J,* the justification of the k-th line, *is written in accordance with one of the following rules.*

29.1 Premise. *If φ is a formula in L, then $\langle 1(1)\varphi$ Premise$\rangle$ is a proof. Also, if D is a proof of length $k-1$, then we may add to D the line: $\langle k(k)\varphi$ Premise$\rangle$.*

29.2 Identity Introduction. *If s is an individual constant, then $\langle (1)s = s\ Id\ I\rangle$ is a proof. Also, if D is a proof of length $k-1$, then we may add to it the line $\langle (k)s = s\ Id\ I\rangle$.*

29.3 Identity Elimination. *Let D be a proof of length $k-1$. Assume that s and s' are individual constants, and that D includes the line $\langle \alpha(i)s = s' J_i\rangle$. Assume also that D includes a line $\langle \beta(j)\varphi J_j\rangle$, where φ contains the occurrences $c_1, \ldots, c_n$ of s (and maybe other occurrences of s as well). Then, we may add to D the line $\langle \alpha \cup \beta(k)\varphi[c_1/s', \ldots, c_n/s']\ Id\ E, i, j\rangle$.*

29.4 Propositional Calculus Rules. *We allow the usual introduction and elimination rules for each connective. (The rules are similar to the ones we formulated for formulas with no SRE-anaphors in our system.)*

29.5 $\forall$ Introduction. *Let D be a proof of length $k-1$, and assume that D includes the line $\langle \alpha(i)\psi(s)J\rangle$, where s is an individual* *constant and ψ does not contain the variable x. Assume also that α does not contain any number of a line in which s occurs. Then, we may add to D the line $\langle \alpha(k) \forall x(\psi[s/x]) \forall\, I, i\rangle$.*

29.6 $\forall$ **Elimination.** *Let D be a proof of length $k-1$, and assume that D includes the line $\langle \alpha(i) \forall x(\psi[s'/x]) J\rangle$, where ψ is a formula containing an individual constant s'. If s is an individual constant, then we may add to D the line $\langle \alpha(k) \psi[s'/s] \forall\, E, i\rangle$.*

29.7 $\exists$ **Introduction.** *Let D be a proof of length $k-1$, and assume that D includes the line $\langle \alpha(i) \psi[s'/s] J\rangle$, where s and s' are individual constants, and ψ does not contain the variable x. Then, we may add to D the line $\langle \alpha(k) \exists x(\psi[s'/x]) \exists\, I, i\rangle$.*

29.8 $\exists$ **Elimination.** *Let D be a proof of length $k-1$, and assume that D includes the lines $\langle \alpha(i) \exists x(\psi[s/x]) J_i\rangle$; $\langle j(j) \psi(s)$ Premise$\rangle$; $\langle \beta(k) \delta J_k\rangle$. Assume also that $j \notin \alpha$, that δ does not contain s, and that β does not contain any number, other than j, of a line in which s occurs. Then we may add to D the line $\langle (\alpha \cup \beta) \backslash \{j\}(m) \delta\ \exists\, E, i, j, k\rangle$.*

Provability is defined as in our system (section 2.4 above); we also use the notation $T_D(\alpha)$ as defined there.

We state the following two theorems without proof:

Theorem 9 (Soundness and Completeness of FOL). *Let L be a formal language in FOL. Let T be a theory in L and φ a formula in L. Then: $T \models \varphi \Longleftrightarrow T \vdash \varphi$.*

5.4 Translation from FOL to our system

Definition 30 (Correlate of a Formal Language). *Let L be a formal language in our system. The* correlate *of L in FOL, L_π, is the formal language that satisfies the following conditions:*

1. *The individual constants of L_π are the SREs of L.*
2. (a) *The predicates of L_π are those of L, excluding Thing.*
   (b) *If a predicate is* n*-place in L, then it is* n*-place in L_π.*

Note: Given a language L, there is exactly one language L_π that satisfies the above conditions.

Definition 31 (Translation of Formulas). *Let L be a formal language in our system, and let φ be a formula in L_π (in FOL). The translation μ of formulas from L_π to L is defined by induction on formulas in L_π. First, we arrange the variables of L_π in an infinite list, in which each variable appears infinitely many times. (This list will enable us to make μ injective. It will be used in theorem 10 below.) Now:*

1. *The translation of atomic formulas is defined as follows:*
   (a) *If φ is $s_1 = s_2$, where s_1 and s_2 are individual constants, then $\mu(\varphi)$ is s_1 is s_2.*
   (b) *Let R be an* n-*place predicate ($n \geq 1$) in L_π. If φ is $Rs_1 \ldots s_n$, then $\mu(\varphi)$ is $(s_1, \ldots, s_n)$ is R.*
2. *If α, β are formulas in L_π, then: $\mu(\neg\alpha)$ is $\neg\mu(\alpha)$; $\mu(\alpha \wedge \beta)$ is $\mu(\alpha) \wedge \mu(\beta)$; $\mu(\alpha \vee \beta)$ is $\mu(\alpha) \vee \mu(\beta)$; $\mu(\alpha \longrightarrow \beta)$ is $\mu(\alpha) \longrightarrow \mu(\beta)$.*
3. *If $\varphi(s)$ is a formula that contains an individual constant s, and x a variable that does not occur in φ, then:*
   (a) *$\mu(\forall x\varphi[s/x])$ is $((l)$ every Thing is Thing$) \wedge (\mu(\varphi)[s/(l)a])$, where l is the least index not occurring in $\mu(\varphi)$ such that x is on the l-th place in the above mentioned list.*
   (b) *$\mu(\exists x\varphi[s/x])$ is $((l)$ some Thing is Thing$) \wedge (\mu(\varphi)[s/(l)a])$, where l is as above.*

Theorem 10 (The Translation μ is Injective). *Let L be a formal language in our system, and let φ be a formula in L_π. If ψ is a formula (in L_π) such that $\mu(\psi) = \mu(\varphi)$, then ψ is φ.*

To see that the theorem is true, we note that in each of the stages in the definition of μ , $\mu(\varphi)$ determines φ. A precise proof can be given by induction on $\#\varphi$.

Theorem 11 ($\mu(\varphi)$ Does Not Contain SRE-Anaphors). *If φ is a formula in L_π, then $\mu(\varphi)$ does not contain anaphors of SRE occurrences.*

The theorem is not hard to prove by induction on formulas in L_π.

As we already mentioned, universal quantification in FOL lacks existential import. Therefore, in order for the translations of formulas to be equivalent to the formulas they translate, we need to complement FOL with axioms of *existential import*. We define:

Definition 32 (EI). *Let L be a formal language in our system. Then* $\mathsf{EI}(L_\pi)$ *(in short:* EI*) is the set of all formulas in L_π that have the form: $\exists x P x$, where x is a variable and P is a one-place predicate.*

Theorem 12 (μ(EI) is Provable). *Let L be a formal language in our system. If T is a theory in L, then $T \vdash \mu(\mathsf{EI})$.*

Proof: It is sufficient to prove that $\emptyset \vdash \mu(\mathsf{EI})$. Let $\varphi \in \mathsf{EI}$. Then φ is of the form $\exists x P x$, and $\mu(\varphi)$ is: $((l)\ \textit{some Thing is Thing}\ ) \wedge ((l) a \textit{ is } P)$. That this formula is provable from $\emptyset$ follows from the existence of the following proof:

1	(1)	s *is* P	*Premise*
	(2)	*every* P *is* P	*every* $I, 1, 1$
	(3)	*some* P *is* P	$RI, 2$
	(4)	d *is Thing*	$Th\ I$
5	(5)	d *is* P	*Premise*
5	(6)	$(d \textit{ is Thing }) \wedge (d \textit{ is } P)$	$\wedge\ I, 4, 5$
5	(7)	$((l) d \textit{ is Thing }) \wedge ((l) a \textit{ is } P)$	$A\ I, 6$
5	(8)	$((l)\textit{some Thing is Thing }) \wedge ((l) a \textit{ is } P)$	*some* $I, 4, 7$
	(9)	$((l)\textit{some Thing is Thing }) \wedge ((l) a \textit{ is } P)$	*some* $E, 3, 5, 5, 8$

■

Definition 33 (Correlate of a Model). *Let L be a formal language in our system, and let $m = \langle M, \sigma \rangle$ be a model of* $\mathsf{EI}(L_\pi)$*. The* correlate *of m in our system, $\mu(m)$, is the model for L defined by: $\mu(m) = \langle M, \mu\sigma \rangle$, where:*

1. *For any individual constant s in L, $\mu\sigma(s) = \sigma(s)$.*
2. *$\mu\sigma(Thing) = M$.*
3. *For any other n-place predicate R in L ($n \geq 1$), $\mu\sigma(R) = \sigma(R)$.*

Note: The above definition indeed determines a model for L; the requirement that $\sigma(P) \neq \emptyset$ for all one-place predicates is fulfilled since m is a model of EI .

Theorem 13 (The Restriction of μ to Models is a Bijection). *Let L be a formal language in our system. Let A be the set of all models of* $\mathsf{EI}(L_\pi)$*, and let B be the set of all models for L. Then, the restriction of μ to A is a bijection from A to B.*

Proof: $\mu|_A$ is injective: if $m_1 = \langle M_1, \sigma_1\rangle$, $m_2 = \langle M_2, \sigma_2\rangle$ are models of EI such that $\mu(m_1) = \mu(m_2)$, then $\sigma_1(\mathit{Thing}) = M_1 = M_2 = \sigma_2(\mathit{Thing})$, and also: $\sigma_1(\alpha) = \mu\sigma_1(\alpha) = \mu\sigma_2(\alpha) = \sigma_2(\alpha)$ for any predicate or individual constant α in $L \backslash \{\mathit{Thing}\}$. Therefore: $m_1 = m_2$.

$\mu|_A$ is onto B: if $m = \langle M, \sigma\rangle$ is a model for L, then $\mu(m') = m$, where $m' = \langle M', \sigma'\rangle$ is the model for L_π determined by: $M' = M$; $\sigma'(\alpha) = \sigma(\alpha)$ for all predicates and individual constants α in L_π.

■

Theorem 14 (Truth under μ). *Let L be a formal language in our system, and let φ be a formula in L_π. If $m = \langle M, \sigma\rangle$ is a model of $\mathsf{EI}(L_\pi)$, then: $m \models \varphi \Longleftrightarrow \mu(m) \models \mu(\varphi)$.*

Proof: By induction on formulas in L_π.

1. Atomic formulas: $m \models [s_1 = s_2] \Longleftrightarrow \sigma(s_1) = \sigma(s_2) \Longleftrightarrow \mu\sigma(s_1) = \mu\sigma(s_2) \Longleftrightarrow \mu(m) \models [s_1 \textit{ is } s_2] \Longleftrightarrow \mu(m) \models \mu(s_1 = s_2)$.

   $m \models [Rs_1 \dots s_n] \Longleftrightarrow \langle\sigma(s_1), \dots, \sigma(s_n)\rangle \in \sigma(R) \Longleftrightarrow$
   $\langle\mu\sigma(s_1), \dots, \mu\sigma(s_n)\rangle \in \mu\sigma(R) \Longleftrightarrow \mu(m) \models [(s_1, \dots, s_n) \textit{ is } R] \Longleftrightarrow$
   $\mu(m) \models \mu(Rs_1 \dots s_n)$

2. If the theorem holds for α and β, then:
   $m \models [\alpha \wedge \beta] \Longleftrightarrow m \models \alpha$ and $m \models \beta \Longleftrightarrow \mu(m) \models \mu(\alpha)$ and $\mu(m) \models \mu(\beta) \Longleftrightarrow$ (since $\mu(\alpha)$, $\mu(\beta)$ do not contain anaphors of SRE occurrences) $\mu(m) \models [\mu(\alpha) \wedge \mu(\beta)] \Longleftrightarrow \mu(m) \models \mu(\alpha \wedge \beta)$.

   The proofs for $\neg\alpha$, $\alpha \vee \beta$ and $\alpha \longrightarrow \beta$ are similar.

3. If $\varphi(s)$ is a formula that contains an individual constant s, and x a variable that does not occur in φ, then:
   $m \models \forall x\varphi[s/x] \Longleftrightarrow$ every s-change m' of m satisfies $\varphi(s) \Longleftrightarrow$ (by the induction hypothesis) for every s-change m' of m, $\mu(m') \models \mu(\varphi(s))$ $\Longleftrightarrow$ (since the set of $\sigma(\mathit{Thing})$-s-changes of $\mu(m)$ is $\{\mu(m')\}|m'$ is an s-change of m) for every $\sigma(\mathit{Thing})$-s-change $(\mu(m))'$ of $\mu(m)$, $(\mu(m))' \models \mu(\varphi(s)) \Longleftrightarrow$ for every $\sigma(\mathit{Thing})$-s-change $(\mu(m))'$ of $\mu(m)$, $(\mu(m))' \models [(s \textit{ is Thing}) \wedge \mu(\varphi(s))] \Longleftrightarrow$ for every $\sigma(\mathit{Thing})$-s-change $(\mu(m))'$ of $\mu(m)$, $(\mu(m))' \models [((l)s \textit{ is Thing}) \wedge \mu(\varphi)[s/(l)a]] \Longleftrightarrow$ $\mu(m) \models [((l) \textit{ every Thing is Thing}) \wedge \mu(\varphi)[s/(l)a]] \Longleftrightarrow$
   $\mu(m) \models \mu(\forall x\varphi[s/x])$

4. The proof for $\exists x\varphi[s/x]$ is similar.

■

Theorem 15 (Entailment under μ). *Let L be a formal language (in our system), let T be a theory in L_π, and let φ be a formula in L_π. If $\mu(T) \models \mu(\varphi)$, then $T \cup \mathsf{EI} \models \varphi$.*

Proof: Assume that $\mu(T) \models \mu(\varphi)$. Let m be a model for $T \cup \mathsf{EI}$. m is a model of EI that satisfies T. Therefore, by theorem 14: $\mu(m) \models \mu(T)$, and it follows that $\mu(m) \models \mu(\varphi)$. Now, by theorem 14, it follows that $m \models \varphi$. Therefore: $T \cup \mathsf{EI} \models \varphi$.

■

Theorem 16 (Provability under μ). *Let φ be a formula in a language L_π, and let T be a theory in L_π. If $T \vdash \varphi$, then $\mu(T) \vdash \mu(\varphi)$.*

Proof: The idea behind the proof is the following: given a proof in FOL, our inference rules enable us to reconstruct it in our system.

In order to prove the theorem precisely, we prove the following proposition by induction on n: Let D be a proof of length n in L_π. If the last line in D is $\langle \alpha(n)\varphi J_n\rangle$, then $\mu(T_D(\alpha)) \vdash \mu(\varphi)$.

1. If $n = 1$, then α is either $\{n\}$ or $\emptyset$, and the justification J_n is either *Premise*, or *Id I* respectively. In either case, one application of *Premise* or *Id I* proves $\mu(\varphi)$ from $\mu(T_D(\alpha))$.

2. Assume that the above proposition is true for any $k < n$.

   (a) If the last line of D is justified by *Premise* or *Id I*, then the proof is as above.

   (b) If the last line of D is justified by $\longrightarrow$ *Introduction*, then D contains the lines $\langle i(i)\psi_1\ \textit{Premise}\rangle$; $\langle \beta(j)\psi_2 J_j\rangle$, and the last line in D is $\langle \beta\backslash\{i\}(n)\psi_1 \longrightarrow \psi_2 \quad \longrightarrow \quad I, i, j\rangle$. We shall show that $\mu(T_D(\beta\backslash\{i\})) \vdash \mu(\psi_1 \longrightarrow \psi_2)$. By the induction hypothesis, $\mu(T_D(\beta)) \vdash \mu(\psi_2)$. Therefore, there exists a proof D', in our system, the last line of which is $\langle \gamma(k)\mu(\psi_2)J\rangle$, where $T_{D'}(\gamma) \subseteq \mu(T_D(\beta))$.

Now, if D' includes a line of the form $\langle i(i)\mu(\psi_1)\ \textit{Premise}\rangle$, we can proceed as follows: we can assume that no formula appears in D' as a premise in two different lines (see lemma 2). We can now add to D' the line $\langle \gamma\backslash\{i\}(k+1)\mu(\psi_1) \longrightarrow \mu(\psi_2) \quad \longrightarrow I, i, k\rangle$.[3] Since $\mu(\psi_1)$ does not appear in D' as a primise in any line other than i, we have: $\mu(\psi_1) \notin \mu(T_{D'}(\gamma\backslash\{i\}))$. And since $T_{D'}(\gamma) \subseteq \mu(T_D(\beta))$ and line i of D is $\langle i(i)\psi_1\ \textit{Premise}\rangle$, it follows that $T_{D'}(\gamma\backslash\{i\}) \subseteq \mu(T_D(\beta\backslash\{i\}))$. Therefore: $\mu(T_D(\beta\backslash\{i\})) \vdash [\mu(\psi_1) \longrightarrow \mu(\psi_2)]$. In other words: $\mu(T_D(\beta\backslash\{i\})) \vdash \mu(\psi_1 \longrightarrow \psi_2)$, as we wanted to prove.

In case D' does not include a line of the form $\langle i(i)\mu(\psi_1)\ \textit{Premise}\rangle$, we have $\mu(\psi_1) \notin \mu(T_{D'}(\gamma))$, and therefore $T_{D'}(\gamma) \subseteq \mu(T_D(\beta\backslash\{i\}))$. It follows that: $\mu(T_D(\beta\backslash\{i\})) \vdash \mu(\psi_2)$, and hence: $\mu(T_D(\beta\backslash\{i\})) \vdash [\mu(\psi_1) \longrightarrow \mu(\psi_2)]$. That is: $\mu(T_D(\beta\backslash\{i\})) \vdash \mu(\psi_1 \longrightarrow \psi_2)$.

The proofs for the cases in which the last line of D is justified by some other propositional calculus derivation rule are similar to the one above, and will not be detailed here.

(c) If the last line of D is justified by $\forall$ *Introduction*, then D includes a line of the form $\langle \alpha(i)\psi(s)J_i\rangle$, where s is an individual constant, ψ does not contain the variable x, and α does not contain any number of a line in which s occurs. Also, the last line in D is $\langle \alpha(n)\forall x(\psi[s/x])\forall\ I, i\rangle$. We shall prove that $\mu(T_D(\alpha)) \vdash \mu(\forall x(\psi[s/x]))$ or, in other words, that:
$\mu(T_D(\alpha)) \vdash [((l)\ \textit{every}\ \ \textit{Thing is Thing}) \wedge \mu(\psi)[s/(l)a]]$.

We should note that since α does not contain numbers of lines in which s occurs, none of the formulas in $T_D(\alpha)$ contains s. And since for any formula δ, $\mu(\delta)$ contains exactly the same SREs/individual constants as δ does (this is easily proved by induction on formulas in L_π), none of the formulas in $\mu(T_D(\alpha))$ contains s.

By the induction hypothesis, $\mu(T_D(\alpha)) \vdash \mu(\psi(s))$. Therefore, there exists a proof D', in our system, the last line of which is $\langle \gamma(k)\mu(\psi)J\rangle$, where $T_{D'}(\gamma) \subseteq \mu(T_D(\alpha))$. It will be sufficient to prove that $T_{D'}(\gamma) \vdash [((l)\ \textit{every}\ \ \textit{Thing is Thing}) \wedge \mu(\psi)[s/(l)a]]$. To prove this, we add to D' the following lines:

[3]The $\longrightarrow$ Introduction rule requires that the formulas involved will not contain anaphors of SRE occurrences. This condition is fulfilled, since no $\mu(\psi)$ contains such anaphors (see theorem 10).

$k+1$	$(k+1)$	*s is Thing*	*Premise*
$\gamma \cup \{k+1\}$	$(k+2)$	$(s\ is\ Thing\ ) \wedge \mu(\psi)$	$\wedge\ I, k, k+1$
$\gamma \cup \{k+1\}$	$(k+3)$	$((l)s\ is\ Thing) \wedge \mu(\psi)[s/(l)a]$	$A\ I, k+2$
γ	$(k+4)$	$((l)every\ \ Thing\ is\ Thing) \wedge \mu(\psi)[s/(l)a]$	$every\ I, k+1, k+3$

(The *every Introduction rule* requires that $\gamma \cup \{k+1\}$ will not contain any number, other than $k+1$, of a line in which s occurs. This requirement is fulfilled here, since $T_{D'}(\gamma) \subseteq \mu(T_D(\alpha))$ and none of the formulas in $\mu(T_D(\alpha))$ contains s.)

The proof that results from the addition of the above lines to D' is a proof of $\mu(\forall x(\psi[s/x]))$ from $\mu(T_D(\alpha))$.

(d) If the last line of proof D is justified by $\forall$ *Elimination*, then D includes a line $\langle \alpha(i) \forall x(\psi[s'/x]) J_i \rangle$, and the last line in D is $\langle \alpha(n)\psi[s'/s] \forall\ E, i\rangle$. We shall prove that $\mu(T_D(\alpha)) \vdash \mu(\psi[s'/s])$.
By the induction hypothesis: $\mu(T_D(\alpha)) \vdash \mu(\forall x(\psi[s'/x]))$. That is: $\mu(T_D(\alpha)) \vdash [((l)\ every\ \ Thing\ is\ Thing\ ) \wedge \mu(\psi)[s'/(l)a]]$.
Therefore, there exists a proof D', whose last line is:
$\langle \gamma(k)((l)\ \ every\ \ Thing\ is\ Thing) \ \wedge\ \mu(\psi)[s'/(l)a]J\rangle$, such that $T_{D'}(\gamma) \subseteq \mu(T_D(\alpha))$.
We can add to D' the following lines:

	$(k+1)$	*s is Thing*	$Th\ I$
γ	$(k+2)$	$((l)s\ is\ Thing) \wedge \mu(\psi)[s'/(l)a]$	$every\ E, k, k+1$
γ	$(k+3)$	$(s\ is\ Thing) \wedge \mu(\psi)[s'/s]$	$A\ E, k+2$
γ	$(k+4)$	$\mu(\psi)[s'/s]$	$\wedge\ E, k+3$

The proof we got shows that $\mu(T_D(\alpha)) \vdash \mu(\psi)[s'/s]$. Now, since $\mu(\psi)[s'/s]$ is in fact $\mu(\psi[s'/s])$ (this can by proved by induction on formulas in L_π), we have: $\mu(T_D(\alpha)) \vdash \mu(\psi[s'/s])$, as required.

(e) If the last line of D is justified by $\exists$ *introduction*, then D includes a line: $\langle \alpha(i)\psi[s'/s]J_i\rangle$, and its last line is $\langle \alpha(n)\exists x(\psi[s'/x])\exists\ I, i\rangle$. We shall prove that $\mu(T_D(\alpha)) \vdash \mu(\exists x(\psi[s'/x]))$. By the induction hypothesis: $\mu(T_D(\alpha)) \vdash \mu(\psi[s'/s])$. Therefore, there exists a proof D', whose last line is: $\langle \gamma(k)\mu(\psi[s'/s])J\rangle$, such that $T_{D'}(\gamma) \subseteq \mu(T_D(\alpha))$.

We now add to D' the following lines:

	$(k+1)$	$s\ is\ Thing$	$Th\ I$
γ	$(k+2)$	$(s\ is\ Thing) \wedge \mu(\psi[s'/s])$	$\wedge\ I, k, k+1$

The formula in the last line above is, in fact: $(s\ is\ Thing) \wedge (\mu(\psi)[s'/s])$. We can therefore add the following lines:

γ	$(k+3)$	$((l)s\ is\ Thing) \wedge (\mu(\psi)[s'/(l)a])$	$A\ I, k+2$
γ	$(k+4)$	$((l)\ some\ \ Thing\ is\ Thing) \wedge (\mu(\psi)[s'/(l)a])$	$some\ I, k+1, k+3$

The proof we thus get is a proof of $\mu(\exists x(\psi[s'/x]))$ from $\mu(T_D(\alpha))$.

(f) If the last line of D is justified by $\exists$ *Elimination*, then D includes lines of the forms: $\langle \alpha(i)\exists x(\psi[s/x])J_i\rangle$; $\langle j(j)\psi(s)\ Premise\rangle$; $\langle \beta(k)\delta J_k\rangle$, where $j \notin \alpha$, δ does not contain s, and β does not contain any number, other than j, of a line in which s occurs. The last line in D is $\langle(\alpha \cup \beta)\backslash\{j\}(n)\delta\exists\ E, i, j, k\rangle$. We shall prove that $\mu(T_D((\alpha \cup \beta)\backslash\{j\})) \vdash \mu(\delta)$. By the induction hypothesis: $\mu(T_D(\alpha)) \vdash \mu(\exists x(\psi[s/x]))$, that is: $\mu(T_D(\alpha)) \vdash ((l)\ some\ Thing\ is\ Thing) \wedge (\mu(\psi)[s/(l)a])$, and also: $\mu(T_D(\beta)) \vdash \mu(\delta)$. If $T_D(\alpha)$ contains $\psi(s)$, then, since $j \notin \alpha$, α contains some other number j' of a line in D, in which $\psi(s)$ appears as a premise. Therefore, if we omit j from $\alpha \cup \beta$, $T_D(\alpha \cup \beta)$ will remain unchanged. That is: $T_D((\alpha \cup \beta)\backslash\{j\}) = T_D(\alpha \cup \beta)$. And since $\mu(\delta)$ is provable from $\mu(T_D(\beta))$, which is a subset of $T_D(\alpha \cup \beta)$, we have: $T_D((\alpha \cup \beta)\backslash\{j\}) = T_D(\alpha \cup \beta) \vdash \mu(\delta)$, as required.

Assume now that $\psi(s) \notin T_D(\alpha)$. Since: $\mu(T_D(\alpha)) \vdash [((l)\ some\ Thing\ is\ Thing) \wedge \mu(\psi)[s/(l)a]]$, and: $\mu(T_D(\beta)) \vdash \mu(\delta)$, there exists a proof D', whose last two lines are:

γ_1	(k)	$((l)\ some\ \ Thing\ is\ Thing) \wedge \mu(\psi)[s/(l)a]$	J_1
γ_2	$(k+1)$	$\mu(\delta)$	J_2

where $T_{D'}(\gamma_1) \subseteq \mu(T_D(\alpha))$ and $T_{D'}(\gamma_2) \subseteq \mu(T_D(\beta))$.[4]

[4]To construct such a proof, we can start with a proof of $\mu(\delta)$ from $\mu(T_D(\beta))$, and 'insert', so to speak, a proof of $((l)\ some\ \ Thing\ is\ Thing) \wedge (\mu(\psi)[s/(l)a])$ from $\mu(T_D(\alpha))$ between the last line of the proof we started with and the rest of that proof.

If $\mu(\psi) \notin T_{D'}(\gamma_2)$, then the above proof shows that $T_{D'}(\gamma_2)\backslash\{\mu(\psi)\} = T_{D'}(\gamma_2) \vdash \mu(\delta)$. We also have:

$$T_{D'}(\gamma_2)\backslash\{\mu(\psi)\} \subseteq \mu(T_D(\beta))\backslash\{\mu(\psi)\} \subseteq \mu(T_D(\beta\backslash\{j\})) \subseteq \mu(T_D((\alpha \cup \beta)\backslash\{j\})).$$

Therefore: $\mu(T_D((\alpha \cup \beta)\backslash\{j\})) \vdash \mu(\delta)$, as required.

Assume that $\mu(\psi) \in T_{D'}(\gamma_2)$. Then, D' contains a line of the form $\langle m(m)\mu(\psi(s))\ \textit{Premise}\rangle$ (it is not hard to show that every line in a proof, on which another line relies, is a premise). Since $T_{D'}(\gamma_2) \subseteq \mu(T_D(\beta))$, $T_{D'}(\gamma_2)$ contains no formula, other than $\mu(\psi)$, in which s occurs. Also, since $T_{D'}(\gamma_1) \subseteq \mu(T_D(\alpha))$, $\psi(s) \notin T_D(\alpha)$, and μ is injective (theorem 10) we have: $\mu(\psi(s)) \notin T_{D'}(\gamma_1)$. We proceed as follows: By lemma 2, we can assume that no premise appears in D' twice. By lemma 5, since $[((l)s\ \textit{is Thing}) \wedge \mu(\psi)[s/(l)a]] \vdash \mu(\psi(s))$,[5] there is a proof D'' that includes lines of the forms:

m	(m)	$((l)s\ \textit{is Thing}) \wedge \mu(\psi)[s/(l)a]$	*Premise*
γ_1'	(p)	$((l)\ \textit{some Thing is Thing}) \wedge (\mu(\psi)[s/(l)a])$	J_1'
γ_2'	(q)	$\mu(\delta)$	J_2'

where $T_{D''}(\gamma_1')$ and $T_{D''}(\gamma_2')$ are identical with $T_{D'}(\gamma_1)$ and $T_{D'}(\gamma_2)$ (respectively) up to the replacement of $\mu(\psi)$ by $((l)\ \textit{some Thing is Thing}) \wedge (\mu(\psi)[s/(l)a])$. By lemma 2, we can assume that D'' does not contain the same premise twice. γ_2' does not contain any number, other than m, of a line in which s occurs.[6] Also, γ_1' does not contain m.[7] We now add to D'' the following lines:

$q+1$	$(q+1)$	$s\ \textit{is Thing}$	*Premise*
$(\gamma_1' \cup \gamma_2')\backslash\{q+1, m\}$	$(q+2)$	$\mu(\delta)$	*some* $E, p, m, q+1, q$

Call the resulting proof D'''. We have $T_{D'''}((\gamma_1' \cup \gamma_2')\backslash\{q+1, m\}) \vdash \mu(\delta)$. It is now sufficient to show that $T_{D'''}((\gamma_1' \cup \gamma_2')\backslash\{q+1, m\}) \subseteq$

[5]The proof requires one application of AE, and one application of $\wedge\ E$.

[6]$T_{D'}(\gamma_2)$ contains no formula, other than $\mu(\psi)$, in which s occurs. And given the relation between $T_{D''}(\gamma_2')$ and $T_{D'}(\gamma_2)$, it follows that $T_{D''}(\gamma_2')$ does not contain any formula, other than $((l)s\ \textit{is Thing}) \wedge \mu(\psi)[s/(l)a]$, in which s occurs. And that formula appears as a premise in D'' only in line m.

[7]$T_{D''}(\gamma_1')$ is identical with $T_{D'}(\gamma_1)$ up to the replacement of $\mu(\psi)$ by $((l)\ \textit{some Thing is Thing}) \wedge (\mu(\psi)[s/(l)a])$. And $T_{D'}(\gamma_1)$ did not contain $\mu(\psi)$.

$\mu(T_D((\alpha \cup \beta)\backslash\{j\}))$. Assume that $\chi \in T_{D'''}((\gamma_1' \cup \gamma_2')\backslash\{q+1, m\})$. Then $\chi \in T_{D'''}(\gamma_1' \cup \gamma_2') = T_{D''}(\gamma_1' \cup \gamma_2') = T_{D''}(\gamma_1') \cup T_{D''}(\gamma_2')$. Also, since $\chi \notin T_{D'''}(\{q+1, m\})$, $\chi \neq ((l)s \textit{ is Thing}) \wedge \mu(\psi)[s/(l)a]$. Therefore:
$\chi \in (T_{D''}(\gamma_1') \cup T_{D''}(\gamma_2'))\backslash\{((l)s \textit{ is Thing }) \wedge \mu(\psi)[s/(l)a]\}$. And since $T_{D''}(\gamma_1')$ and $T_{D''}(\gamma_2')$ are identical to $T_{D'}(\gamma_1)$ and $T_{D'}(\gamma_2)$ up to the replacement of $\mu(\psi)$ by $((l) \textit{ some Thing is Thing}) \wedge (\mu(\psi)[s/(l)a])$, it follows that $\chi \in (T_{D'}(\gamma_1) \cup T_{D'}(\gamma_2))\backslash\{\mu(\psi)\} \subseteq [\mu(T_D(\alpha)) \cup \mu(T_D(\beta))]\backslash\{\mu(\psi)\} = \mu(T_D(\alpha) \cup T_D(\beta))\backslash\{\mu(\psi)\} = \mu(T_D(\alpha \cup \beta))\backslash\{\mu(\psi)\} \subseteq \mu(T_D((\alpha \cup \beta)\backslash\{j\}))$.
Therefore, we have
$T_{D'''}((\gamma_1' \cup \gamma_2')\backslash\{k+2, m\}) \subseteq \mu(T_D((\alpha \cup \beta)\backslash\{j\}))$, as required.

■

5.5 Paraphrases

In this section we show that every formula φ in our system is both semantically and deductively equivalent to a formula φ^*, which is the translation of some formula φ_π of FOL. We first correlate, with each formula, a set of paraphrases:

Definition 34 (Paraphrases). *Let L be a formal language in our system, and let φ be a formula in L. The* paraphrases *of φ are defined by induction on formulas:*

1. *Atomic formulas: if $s_1, \ldots, s_n$ are SREs and R is an n-place predicate ($n \geq 1$), then: the only paraphrase of s_1 is s_2 is itself; The only paraphrase of s_1 isn't s_2 is: $\neg(s_1$ is $s_2)$; the only paraphrase of $(s_1, \ldots, s_n)$ is R is itself; the only paraphrase of $(s_1, \ldots, s_n)$ isn't R is: $\neg((s_1, \ldots, s_n)$ is R).*

2. *If α and β are formulas that do not contain anaphors of SRE occurrences, then: the paraphrases of $\neg\alpha$ are all the formulas of the form $\neg(\alpha')$, where α' is a paraphrase of α. Similarly, the paraphrases of $\alpha \wedge \beta$ are the formulas of the form $\alpha' \wedge \beta'$ (where α' and β' are paraphrases of α and β respectively); those of $\alpha \vee \beta$ are the formulas of the form $\alpha' \vee \beta'$; and those of $\alpha \longrightarrow \beta$ are the formulas $\alpha' \longrightarrow \beta'$.*

3. If φ results from the substitution of anaphors for SRE occurrences in a formula ψ, as in section 2-c of the formula definition, then the paraphrases of φ are those of ψ.

4. Let $\varphi(qP)$ be a formula in which an occurrence t of the QNP qP is the main QNP. Let s be an SRE not occurring in φ, and let $\psi = \varphi[t/s]$.

(i) If q is every: the paraphrases of φ(every P) are all the formulas of the form: $((l)$ every Thing is Thing$)\wedge(((l)a$ is $P) \longrightarrow \psi'[s/(l)a])$, where ψ' is a paraphrase of ψ and l is an index that does not occur in ψ'.

(ii) If q is some: the paraphrases of φ(some P) are the formulas of the form: $((l)$ some Thing is Thing$) \wedge (((l)a$ is $P) \wedge \psi'[s/(l)a])$, where ψ' and l are as above.

Theorem 17 (Every Formula Has a Paraphrase that Translates a Formula of FOL). *Let L be a formal language in our system. If α is a formula in L, then there exist a paraphrase α' of α and formula α_π in L_π such that α' is $\mu(\alpha_\pi)$.*

Proof: by induction on formulas in L.

1. Atomic formulas: the (only) paraphrase of s_1 *is* s_2 is itself, and it is $\mu(s_1 = s_2)$; the paraphrase of s_1 *isn't* s_2 is $\neg(s_1$ *is* $s_2)$, and it is $\mu(\neg(s_1 = s_2))$; the paraphrase of $(s_1, \ldots, s_n)$ *is* R is itself, and it is $\mu(Rs_1 \ldots s_n)$; the paraphrase of $(s_1, \ldots, s_n)$ *isn't* R is $\neg((s_1, \ldots, s_n)$ *is* $R)$, and it is $\mu(\neg Rs_1 \ldots s_n)$.

2. If α, β do not contain anaphors of SRE occurrences, and $\alpha' = \mu(\alpha_\pi)$ and $\beta' = \mu(\beta_\pi)$ are paraphrases of α and β respectively, then: $\mu(\neg\alpha_\pi) = \neg\mu(\alpha_\pi) = \neg(\alpha')$, and this formula is a paraphrase of $\neg\alpha$. Similarly, $\mu(\alpha_\pi \wedge \beta_\pi) = \mu(\alpha_\pi) \wedge \mu(\beta_\pi) = (\alpha') \wedge (\ \beta')$, and this formula is a paraphrase of $\alpha \wedge \beta$. The proofs for $\alpha \vee \beta$ and $\alpha \longrightarrow \beta$ are similar.

3. If φ is the product of substituting anaphors for SRE occurrences in ψ, as in section 2c of the formula definition, and $\psi' = \mu(\psi_\pi)$ is a paraphrase of ψ, then it is also a paraphrase of φ.

4. Let $\varphi(qP)$ be a formula in which an occurrence t of qP is the main QNP, and assume that the theorem holds for any formula of the form

$\varphi[t/s]$, where s is an SRE. Let ψ be $\varphi[t/s]$, where s is an SRE that does not occur in φ (we thus have $\varphi(qP) = \psi[s/qP]$). Assume that $\psi' = \mu(\psi_\pi)$ is a paraphrase of ψ.

(i) If q is *every*: $\mu(\forall x((Ps \longrightarrow \psi_\pi)[s/x])) =$ $((l)\mathit{every}\ \ \mathit{Thing\ is\ Thing}) \wedge (\mu(Ps \longrightarrow \psi_\pi)[s/(l)a]) =$ $((l)\ \mathit{every}\ \ \mathit{Thing\ is\ Thing}) \wedge ((\mu(Ps) \longrightarrow \mu(\psi_\pi))[s/(l)a]) =$ $((l)\ \mathit{every}\ \ \mathit{Thing\ is\ Thing}) \wedge (((s\ \mathit{is}\ P) \longrightarrow \mu(\psi_\pi))[s/(l)a]) =$ $((l)\mathit{every}\ \ \mathit{Thing\ is\ Thing}) \wedge (((l)a\ \mathit{is}\ P) \longrightarrow \mu(\psi_\pi)[s/(l)a]) =$ $((l)\mathit{every}\ \ \mathit{Thing\ is\ Thing}) \wedge (((l)a\ \mathit{is}\ P) \longrightarrow \psi'[s/(l)a])$. And this formula is a paraphrase of $\varphi(\mathit{every}\ P)$.

(ii) The proof for the case in which q is *some* is similar.

■

Lemma 6 (Paraphrases Do Not Contain SRE-Anaphors). *Let φ be a formula in a language L. If φ' is a paraphrase of φ, then φ' does not contain any anaphors of SRE occurrences.*

This lemma can be easily proved by induction on formulas in L.

It will be convenient to correlate with each formula in our system a unique paraphrase φ' which translates some formula of FOL.

Definition 35 (φ^*). *Let L be a formal language in our system. With each formula φ in L, we correlate a paraphrase φ^* of φ, which is the translation of some formula in L_π.*[8]

Theorem 18 (Equivalence of φ and φ^*). *Let φ a formula in a language L in our system. Then:*

(1) φ and φ^* are semantically equivalent. That is: $m \models \varphi \Longleftrightarrow m \models \varphi^*$ for any model m for L.

(2) φ and φ^* are deductively equivalent. That is: $\varphi \vdash \varphi^*$ and $\varphi^* \vdash \varphi$.

---

[8]In case L is denumerable, we can arrange its formulas in lexicographic order, and define φ^* as the first paraphrase of φ that translates some formula in L_π. Otherwise, we can use the axiom of choice.

Proof: (1) follows from (2) and the soundness of our deductive system (theorem 7): If $\varphi \vdash \varphi^*$ and $\varphi^* \vdash \varphi$, then $\varphi \models \varphi^*$ and $\varphi^* \models \varphi$. That is: every model of $\{\varphi\}$ satisfies φ^*, and every model of $\{\varphi^*\}$ satisfies φ.

It remains to prove (2). We shall prove the following proposition by induction on $\#\varphi$:
(3) Let φ be a string. If φ is a formula in a language L (in our system), and φ' is any paraphrase of φ, then $\varphi \vdash \varphi'$ and $\varphi' \vdash \varphi$.

1. If $\#\varphi \leq 3$, then φ is an atomic formula in L. If φ is s_1 *is* s_2, then $\varphi' = \varphi$ and (3) holds trivially. If φ is s_1 *isn't* s_2, then α' is $\neg(s_1 \textit{ is } s_2)$. To see that (3) holds in this case, one needs only to apply the rules *NC E* and *NC I*. The proofs for $(s_1, \ldots, s_n)$ *is* R and $(s_1, \ldots, s_n)$ *isn't* R are similar.

2. Let $\#\varphi = n$, and assume that (3) holds for any string ψ for which $\#\psi < n$.

   (a) If φ is atomic, the proof is as above.

   (b) Assume that $\varphi \in \{\neg\alpha, \alpha \wedge \beta, \alpha \vee \beta, \alpha \longrightarrow \beta\}$, where α and β contain no anaphors of SRE occurrences. We shall prove (3) for $\alpha \vee \beta$ (The proofs for $\alpha \wedge \beta$, $\neg\alpha$ and $\alpha \longrightarrow \beta$ are similar). By definition 34, $(\alpha \vee \beta)'$ is $\alpha' \wedge \beta'$, where α' and β' are paraphrases of α and β respectively. By theorem 6, α' and β' do not contain anaphors of SRE occurrences. Since, by the induction hypothesis, α and β are deductively equivalent to α' and β' respectively, we have: $\alpha \vdash \alpha'$, $\beta \vdash \beta'$. It follows that $\alpha \vdash \alpha' \vee \beta'$ and $\beta \vdash \alpha' \vee \beta'$. Therefore, it is not hard to prove that $\alpha \vee \beta \vdash \alpha' \vee \beta'$. That is: $\alpha \vee \beta \vdash (\alpha \vee \beta)'$. The proof for $(\alpha \vee \beta)' \vdash \alpha \vee \beta$ is similar.

   (c) Let φ be the product of substituting anaphors for SRE occurrences in a formula ψ as in section 2c of the formula definition. If (3) holds for ψ, then it obviously holds for φ; for φ is deductively equivalent to ψ (to show this, one can apply the rules *A I* and *A E*), and they both have the same paraphrases.

   (d) Let $\varphi(qP)$ be a formula in which an occurrence t of qP is the main QNP. Let ψ be $\varphi[t/s]$, where s is an SRE that does not occur in φ. We thus have $\varphi(qP) = \psi[s/qP]$.

      i. If q is *every*, then φ' is of the form:
      $((l) \textit{ every Thing is Thing}) \wedge (((l)a \textit{ is } P) \longrightarrow \psi'[s/(l)a])$

where ψ' is a paraphrase of ψ. We first show that $(\varphi(every\ P))' \vdash \varphi(every\ P)$.
Since $(\varphi(every\ P))'$ contains no anaphors of SRE occurrences (lemma 6), we have he following proof:

1	(1)	$((l)\ every\ \ Thing\ is\ Thing)$ $\wedge(((l)a\ is\ P) \longrightarrow \psi'[s/(l)a])$	*Premise*
2	(2)	$s\ is\ P$	*Premise*
	(3)	$s\ is\ Thing$	$Th\ I$
1	(4)	$((l)s\ is\ Thing) \wedge (((l)a\ is\ P) \longrightarrow \psi'[s/(l)a])$	$every\ E, 1, 3$
1	(5)	$(s\ is\ Thing) \wedge ((s\ is\ P) \longrightarrow \psi')$	$A\ E, 4$
1	(6)	$(s\ is\ P) \longrightarrow \psi'$	$\wedge\ E, 5$
1, 2	(7)	ψ'	$\longrightarrow\ E, 2, 6$

Now from the induction hypothesis it follows that $\psi' \vdash \psi$. And according to lemmas 4, 2 and 3, we can add lines to the above proof until we get:

		...	
1, 2	(k)	ψ	J_k

And since ψ is $\varphi[t/s]$, we can add the line:

1	$(k+1)$	$\varphi(every\ P)$	$every\ I, 2, k$

It remains to prove that $\varphi(every\ P) \vdash (\varphi(every\ P))'$.
Since ψ is $\varphi[t/s]$, we have the following proof:

1	(1)	$\varphi(every\ P)$	*Premise*
2	(2)	$s\ is\ Thing$	*Premise*
3	(3)	$s\ is\ P$	*Premise*
1, 3	(4)	ψ	$every\ E, 1, 3$

Now, from the induction hypothesis it follows that $\psi \vdash \psi'$. And according to lemmas 4, 2 and 3, we can add lines to the above proof until we get:

		...	
1, 3	(k)	ψ'	J_k

We continue the proof:

1	$(k+1)$	$(s\ is\ P) \longrightarrow \psi'$	$\longrightarrow I, 3, k$
1, 2	$(k+2)$	$(s\ is\ Thing) \wedge ((s\ is\ P) \longrightarrow \psi')$	$\wedge\ I, 2, k+1$
1, 2	$(k+3)$	$((l)s\ is\ Thing) \wedge (((l)a\ is\ P) \longrightarrow \psi'[s/(l)a])$	$A\ I, k+2$
1	$(k+4)$	$((l)\ every\ \ Thing\ is\ Thing) \wedge$ $(((l)a\ is\ P) \longrightarrow \psi['s/(l)a])$	$every\ I, 2, k+3$

From the existence of the above proof it follows that $\varphi(every\ P) \vdash (\varphi(every\ P))'$.

ii. If q is *some*, then $(\varphi(some\ P))'$ is of the form: $((l)\ some\ \ Thing\ is\ Thing) \wedge (((l)a\ is\ P) \wedge \psi'[s/(l)a])$, where ψ' is a paraphrase of ψ. We first prove that $(\varphi(some\ P))' \vdash \varphi(some\ P)$. Since $(\varphi(some\ P))'$ contains no anaphors of SRE occurrences, we have the following proof:

1	(1)	$((l)\ some\ \ Thing\ is\ Thing)$ $\wedge(((l)a\ is\ P) \wedge \psi'[s/(l)a])$	*Premise*
2	(2)	$s\ is\ Thing$	*Premise*
3	(3)	$((l)s\ is\ Thing) \wedge (((l)a\ is\ P) \wedge \psi'[s/(l)a])$	*Premise*
3	(4)	$(s\ is\ Thing) \wedge ((s\ is\ P) \wedge \psi')$	$A\ E, 3$
3	(5)	$(s\ is\ P) \wedge \psi'$	$\wedge\ E, 4$
3	(6)	$s\ is\ P$	$\wedge\ E, 5$
3	(7)	ψ'	$\wedge\ E, 5$

From the induction hypothesis it follows, as before, that $\psi' \vdash \psi$. And according to lemmas 4, 2 and 3, we can add lines to the above proof until we get:

...

3 (k) ψ J_k

Since ψ is $\varphi[t/s]$, we can apply *some* I, and add the following lines to the proof:

3	$(k+1)$	$\varphi(some\ P)$	$some\ I, 6, k$
1	$(k+2)$	$\varphi(some\ P)$	$some\ E, 1, 2, 3, k+1$

It follows that $(\varphi(some\ P))' \vdash \varphi(some\ P)$.
It remains to prove that $\varphi(some\ P) \vdash (\varphi(some\ P))'$. We start with the following proof:

1	(1)	$\varphi(some\ P)$	*Premise*
2	(2)	$s\ is\ P$	*Premise*
3	(3)	ψ	*Premise*

From the induction hypothesis it follows that $\psi \vdash \psi'$, and we can add lines to the above proof until we get:

		...	
3	(k)	ψ'	J_k

We proceed:

2, 3	$(k+1)$	$(s\ is\ P) \wedge \psi'$	$\wedge\ I, 2, k$
	$(k+2)$	$s\ is\ Thing$	$Th\ I$
2, 3	$(k+3)$	$(s\ is\ Thing) \wedge ((s\ is\ P) \wedge \psi')$	$\wedge\ I, k+1, k+2$
2, 3	$(k+4)$	$((l)s\ is\ Thing) \wedge (((l)a\ is\ P) \wedge \psi'[s/(l)a])$	$A\ I, k+3$
2, 3	$(k+5)$	$((l)\ some\ \ Thing\ is\ Thing)$ $\wedge(((l)a\ is\ P) \wedge \psi'[s/(l)a])$	$some\ I, k+2, k+4$
1	$(k+6)$	$((l)\ some\ \ Thing\ is\ Thing)$ $\wedge(((l)a\ is\ P) \wedge \psi'[s/(l)a])$	$some\ E, 1, 2, 3, k+5$

It follows that $\varphi(some\ P) \vdash (\varphi(some\ P))'$, as we wanted to prove.

■

Definition 36 (T^*). *Let L be a formal language in our system. If T is a theory in L, then T^* is $\{\varphi^* | \varphi \in T\}$.*

Theorem 19 (Invariance of Entailment and Provability under $\alpha \longmapsto \alpha^*$). *Let T be a theory in a language L, and let φ be a formula in L. Then:*

1. $T \models \varphi \Longleftrightarrow T^* \models \varphi^*$

2. $T \vdash \varphi \Longleftrightarrow T^* \vdash \varphi^*$

Proof: 1. Assume that $T \models \varphi$. Let m be a model of T^*. If $\psi \in T$, then $m \models \psi^*$, and by theorem 18: $m \models \psi$. We therefore have: $m \models T$, and it follows that $m \models \varphi$. Therefore, by theorem 18: $m \models \varphi^*$.

The proof for $T^* \models \varphi^* \Longrightarrow T \models \varphi$ is similar.

2. Assume that $T \vdash \varphi$. Then there exists a proof D of φ from T. By lemma 2, we can assume that no premise appears in D twice. Let $\langle \alpha(n)\varphi J\rangle$ be the last line in D , and let $T_D(\alpha) = \{\varphi_1, \ldots, \varphi_n\} \subseteq T$.

Since (by theorem 18) $\varphi_1^* \vdash \varphi_1$, we have: $\{\varphi_1^*, \ldots, \varphi_n^*\} \vdash \{\varphi_1, \varphi_2^*, \ldots, \varphi_n^*\}$. And since $\varphi_2^* \vdash \varphi_2$, we have: $\{\varphi_1^*, \ldots, \varphi_n^*\} \vdash \{\varphi_1, \varphi_2, \varphi_3^*, \ldots, \varphi_n^*\}$. We continue by induction and get: $\{\varphi_1^*, \ldots, \varphi_n^*\} \vdash \{\varphi_1, \ldots, \varphi_n\}$. We also have: $\{\varphi_1, \ldots, \varphi_n\} \vdash \varphi$, and since provability for theories is transitive (theorem 6), we get: $\{\varphi_1^*, \ldots, \varphi_n^*\} \vdash \varphi$. Theorem 18 ensures that $\varphi \vdash \varphi^*$. And from the transitivity of $\vdash$ for theories: $\{\varphi_1^*, \ldots, \varphi_n^*\} \vdash \varphi^*$. Now, since $\varphi_1^*, \ldots, \varphi_n^* \in T^*$, it follows that $T^* \vdash \varphi^*$. The proof for $T^* \vdash \varphi^* \Longrightarrow T \vdash \varphi$ is similar.

■

5.6 The equivalence between our system and FOL+EI

Theorem 20 (Equivalence). *Let L be a formal language in our system, let F be the set of all formulas in L, and F_π the set of all formulas in L_π. μ, as a mapping from F_π to F, has the following properties:*

(1) *μ is injective.*

(2) *μ covers F in the following sense: for each formula $\varphi \in F$ there exists a formula $\varphi^* \in \mu(F_\pi)$ that is both semantically and deductively equivalent to φ (that is: φ and φ^* are true in exactly the same models, and they are provable from each other). Also, if $T \subseteq F$, $\psi \in F$, and $T^* = \{\alpha^* | \alpha \in T\}$, then:*

  (a) $T \models \psi \Longleftrightarrow T^* \models \psi^*$

  (b) $T \vdash \psi \Longleftrightarrow T^* \vdash \psi^*$.

(3) *μ preserves entailment and provability: for each theory $T_\pi \subseteq F_\pi$ and formula $\varphi_\pi \in F_\pi$:*

(a) $T_\pi \cup \mathsf{EI} \models \varphi_\pi \Longleftrightarrow \mu(T_\pi) \models \mu(\varphi_\pi)$

(b) $T_\pi \cup \mathsf{EI} \vdash \varphi_\pi \Longleftrightarrow \mu(T_\pi) \vdash \mu(\varphi_\pi)$.

Proof: (1) is theorem 10. (2) immediately follows from theorems 18 and 19. It remains to prove (3). Let $T_\pi \subseteq F_\pi$ and let $\varphi_\pi \in F_\pi$. We have:
$\mu(T_\pi) \models \mu(\varphi_\pi) \Longrightarrow$ (by theorem 15) $T_\pi \cup \mathsf{EI} \models \varphi_\pi$
$\Longrightarrow$ (by the completeness of the predicate calculus) $T_\pi \cup \mathsf{EI} \vdash \varphi_\pi$
$\Longrightarrow$ (by theorem 16) $\mu(T_\pi \cup \mathsf{EI}) \vdash \mu(\varphi_\pi) \Longrightarrow \mu(T_\pi) \cup \mu(\mathsf{EI}) \vdash \mu(\varphi_\pi)$
$\Longrightarrow$ (by theorem 12 and the transitivity of provability) $\mu(T_\pi) \vdash \mu(\varphi_\pi)$
$\Longrightarrow$ (since our system is sound) $\mu(T_\pi) \models \mu(\varphi_\pi)$. The required equivalences follow.

■

Theorem 21 (Completeness). *Let T be a theory in a language L, and let φ be a formula in L. If $T \models \varphi$, then $T \vdash \varphi$.*

Proof: Let F and F_π be as in theorem 20. We have: $T^* \subseteq \mu(F_\pi)$ and $\varphi^* \in \mu(F_\pi)$. Therefore, there exist $T_\pi \subseteq F_\pi$ and $\varphi_\pi \in F_\pi$ such that $\mu(T_\pi) = T^*$ and $\mu(\varphi_\pi) = \varphi^*$. Now, by theorem 20, and by the completeness of the predicate calculus, we get: $T \models \varphi \Longrightarrow T^* \models \varphi^* \Longrightarrow \mu(T_\pi) \models \mu(\varphi_\pi) \Longrightarrow T_\pi \cup \mathsf{EI} \models \varphi_\pi \Longrightarrow T_\pi \cup \mathsf{EI} \vdash \varphi_\pi \Longrightarrow \mu(T_\pi) \vdash \mu(\varphi_\pi) \Longrightarrow T^* \vdash \varphi^* \Longrightarrow T \vdash \varphi$.
■

Compactness is a consequence of completeness:

Theorem 22 (Compactness). *Let T be a theory in a language L. Then, T has a model iff every finite $T_1 \subseteq T$ has a model.*

Proof: If T has a model m, then m is a model of any finite subset of T. Conversely, assume that every finite subset T_1 of T has a model. If T does not have a model, then $T \models$ [*s isn't s*], and from the completeness of our system we get: $T \vdash$ [*s isn't s*]. Now, since any proof uses only a finite number of premises, there exists a finite $T_1 \subseteq T$ such that $T_1 \vdash$ [*s isn't s*]. Since our deductive system is sound, we have: $T_1 \models$ [*s isn't s*]. Therefore: T_1 is a finite subset of T that does not have a model. A contradiction.

■

6 Conclusion

An overview of the paper is due in this place. We started by describing, in brief outline, a new semantic analysis of natural language, according to which common nouns in noun phrases are often plural referring expressions, and not – *pace* Frege – logical predicates. We then described, again in outline, the implications of this analysis for the analysis of quantification in natural language.

This introductory discussion lead to the development of a new formal system, built on the basis of the mentioned semantic analysis of natural language. This system, unlike FOL but similarly to natural language, uses concept-letters both as plural referring expressions and as predicates; it combines quantifiers with concept-letters to form noun phrases, which occupy in sentences the same place as singular referring expressions do; the way anaphors are used in it is closer to the way anaphors are used in natural language than to that in which variables are used in FOL; and more. We have also compared and contrasted our system with many-sorted logic.

We defined formulas, derivation rules and models for our system, and proved it to be sound. We then turned to inquire its relation to FOL. For that purpose, we added to FOL a set of axioms, EI . On the other hand, while developing our system, we had introduced, having these future inquiries in mind, a special predicate to our system, *Thing*, to which any interpretation function assigns the whole domain. We correlated models in our system with models of EI in FOL. Relying on all this, we showed how to translate formulas of FOL into our system, and proved the translation to have the following properties: first, it is one-to-one. Secondly, it covers all the formulas in our system in the following sense: every formula is both semantically and deductively equivalent to a translation of some formula of FOL. Thirdly, the translation preserves truth in a model, entailment, and provability. The completeness and compactness of our system followed immediately.

Our system can be proved to be sound, complete and compact even without the predicate *Thing*; this, however, was not done here.

Accordingly, we have demonstrated that the new analysis of the semantics of natural language can be used as a basis for the construction of a powerful formal system, sound and complete, which parallels FOL, in the sense specified above. We have thus accomplished what we set out to do in this paper.[9]

7 Bibliography

[1] Hanoch Ben-Yami. *Logic & Natural Language: On Plural Reference and Its Semantic and Logical Significance*. Ashgate, Aldershot, 2004.

[2] B. Herbert Enderton. *A Mathematical Introduction to Logic*. Harcourt/Academic Press, second edition, 2001.

[3] Gottlob Frege. *Begriffsschrift: Eine der Arithmetischen nachgebildete Formelsprache des reinen Denkens*. Verlag von Louis Nebert, Halle a.S., 1879.

[4] Edward L. Keenan. Some properties of natural language quantifiers: generalized quantifier theory. *Linguistics and Philosophy*, 25:627–654, 2002.

[5] E. J. Lemmon. *Beginning Logic*. Nelson, London, 1965.

[6] W. H. Newton-Smith. *Logic: An Introductory Course*. Routledge, London, 1985.

[9] A part of the proof of theorem 16 (p. 208) was erroneously omitted. This part is the one dealing with the case in which the last line in the proof is justified by *Id E*.

On generalizing the logic-approach to space-time towards general relativity: first steps[1]

Judit X. Madarász **István Németi**
madarasz@renyi.hu nemeti@renyi.hu
Alfréd Rényi Institute of Mathematics, Budapest, H-1364 Pf. 127
Csaba Tőke
tcs@phys.psu.edu
Pennsylvania State University, 104 Davey Lab, State College, 16802 PA

1 Introduction

In this and related work, we study relativity theory, or theory of space-time, as a theory of first-order logic. It is important for our approach that we work in the framework of (mathematical) logic and within that in (many-sorted) first-order logic (FOL). The reasons for the latter choice can be found in, e.g., [2, Appendix], [3], [22], [23].[2] The aims of our project are summarized in the introduction of [2] available on the Internet (cf. also [3], [1]), here we briefly mention only aims (i) and (ii) below; (i) to do work on the logical foundation of space-time theories, and (ii) to elaborate a logic based conceptual analysis of relativity theories. For both of these goals, we want to start out with the so-called observational (in the sense of, e.g., Reichenbach [19]) or "bottom-up"

[1]Research supported by the Hungarian National Foundation for scientific research grants No's T30314, T43242, T35192, as well as by COST grant No. 274.

[2]This work is a self-contained part of a larger project, cf. e.g., [4], [2], [13], [3]. In the present introduction, "we" occasionally refers to the larger project and occasionally to the present paper only.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 225–268

versions of (kinematics of) relativity theories as opposed to the "monolithic", theoretically oriented "top-down" approaches. Of course, in due time we arrive at the theoretical versions, too, e.g., in [13], but by that time they will be well motivated, cf. e.g., [13], [14].

First we build up (the kinematics of) special relativity theory in FOL obtaining the finitely axiomatized FOL-theory Specrel. We put emphasis on making the axioms of Specrel streamlined, transparent, and intuitively convincing. Then we elaborate a conceptual analysis of special relativity, its variants, and its generalizations. This analysis is based on the FOL axiom system Specrel, on variants and fragments of Specrel and their generalizations. Among other things, we analyze Specrel both from the logical point of view (model theory, proof theory, "reverse mathematics" etc.) and from the physico-philosophical relativity theoretic point of view. Much of this is done in [2], [13], [1], [4]. As a natural continuation, we also experiment with generalizing Specrel in the direction of general relativity.

The first two steps in this generalization are (I) and (II) below. (I) We extend Specrel to accommodate accelerated observers, which, via Einstein's equivalence principle, enables us to discuss some features of gravity. E.g., the Twin Paradox and the Tower Paradox (gravity slows time down) become provable in the accelerated observers version Acc(Specrel) of Specrel, cf. e.g., [3]. (II) As a second step in this direction, we make Acc(Specrel) *local*, where local is understood in the sense of general relativity. We do this via the so-called method of localization which can be applied basically to any version of Specrel and Acc(Specrel). The localized versions of these theories are also built up in FOL (we make special efforts to ensure this) for methodological reasons mentioned earlier. Since localization turns out to be such a general procedure, we can denote the thus obtained theories as Loc(Specrel), Loc(Acc(Specrel)) etc.[3]

It is explained in the classic textbook [17, pp.163-5] on general relativity that by suitably combining accelerated observers and localization one can safely move towards general relativity by starting out from special relativity, cf. e.g., Box 6.1 on p.164 therein. This motivates our study of the FOL-theory Loc(Acc(Specrel)) and its variants. The investigation of Loc(Acc(Specrel)) is analogous with that of Specrel, i.e. after introducing the theory and proving theorems about its basic properties comes a fine-scale conceptual analysis

[3]So, Loc(-) can be regarded as some kind of a general "operator" applicable to theories which are variants of Specrel.

both from the logic point of view and from the relativity-theoretic point of view.

In the present paper we concentrate on illustrating the process of localization, i.e. step (II) towards general relativity mentioned above. To make the essential ideas stand out more, we concentrate on describing Loc(Specrel), since extending the localization procedure from Specrel to Acc(Specrel) goes the natural way. All the same, we would like to emphasize that the FOL-theory which brings us closer to having a FOL-based version of general relativity is Loc(Acc(Specrel)) and not Loc(Specrel) in itself. But if we keep this fact in mind, it is more useful to study the method and effects of localization first on the example of Loc(Specrel). For Acc(Specrel) and the definition of Loc(Acc(Specrel)) we refer to [3] available on the net. Besides the research school represented here, localization was used for moving towards general relativity in, e.g., Latzer [11], and Buseman [6].

Here, we introduce the theory Loc(Specrel) and prove some theorems about it. E.g., it turns out that already a small fragment of Loc(Specrel) proves distinguished predictions of relativity theory in the local setting. For lack of space the present paper gives only a small sample from the theory. More on Loc(Specrel) can be found in [14] where Loc(Specrel) is called partial domain relativity theory and is denoted as LocStd. Cf. also [3]. Here, a formal definition of Loc(Specrel) is given in §2 above Theorem 3, where Loc(Specrel) is denoted as **LocRel**. Its fragments and versions are denoted as $\mathbf{LocRel}^-$, $\mathbf{LocRel}_0^-$ etc.

In passing we note that the process of localization is related to that of relativization used in areas of logic related to algebraic logic, cf. e.g., [3], [5] or the volume [16], [18].

Intuitively, Loc(Specrel) is obtained from Specrel in two steps. These are: (A) We relax the condition in Specrel that all events "seen" by one observer are "seen" by the others. This is implemented by permitting observers not to put any event to points of their coordinate-systems too far from the origin (of the coordinate-system). I.e. we allow observers to use subsets of their coordinate-systems for coordinatizing events instead of using the whole coordinate-system. (B) Axioms of Specrel are rephrased in the local spirit (in the topological sense) which is something like the following: If Axi is an axiom of Specrel, then instead of Axi we say that for each point p of space-time (whatever this may mean) there is a small enough open neighborhood D of p such that Axi is true in D.

Latzer [11] pointed out a problem with (this kind of) localization of special relativity, as follows. In studying global theories like Specrel, one can rely on the so-called Alexandrov-Zeeman type theorems, e.g., in the style followed in the book of Goldblatt [9], or in [1], [4]. Because of their just mentioned usefulness, the Alexandrov-Zeeman theorems have been thoroughly generalized in various directions, cf. e.g., [3], [12], [10]. Latzer [11, p.237 lines 8-12, p.255 lines 5-8] writes that some kind of a generalization of the Alexandrov-Zeeman type theorems to the local approach to relativity would be needed for studying local versions of relativity, hence for moving towards general relativity. However, Lester [12, p.929] points out that the Alexandrov-Zeeman theorem does not generalize to the local setting. This fact slowed down progress with the logical analysis of local relativity theories. We address this problem in Theorems 1 and 2 way below. Namely, we prove two theorems in fragments of Loc(Specrel), which can be regarded as Alexandrov-Zeeman type results in the local setting. To illustrate their usefulness in analyzing local relativity, we state a theorem to the effect that a quite small fragment of Loc(Specrel) already proves the nonexistence of faster-than-light observers (NoFTL) in the local sense, cf. Theorems 3–5 in this connection.[4] This is proved via Theorems 1 and 2. In related work we also use our localized Alexandrov-Zeeman type results for establishing various distinguished predictions of local relativity, e.g., predicting the behavior of fast moving clocks, meter-rods, etc. We also indicate some of the global peculiarities of the localized theory, cf. [14].

2 The FOL-theory Loc(Specrel) of localized relativity

In this paper we will deal with kinematics of relativity, i.e. we will deal with motion of *bodies* (e.g., of *test-particles*). The motivation for our choice of vocabulary (for special relativity and its generalizations) is summarized as follows. We will represent motion as changing spatial location in time. To do so, we will have reference-frames for coordinatizing events and, for simplicity, we will associate reference-frames with special bodies which we will call *observers*. There will be another special kind of bodies which we will call *photons*. For coordinatizing events we will use an arbitrary *ordered field* in place of the field of the real numbers. Thus the elements of this field

[4]This can be interpreted as follows. In Theorems 1 and 2 we propose a kind of solution to the problem (mentioned, e.g., by Latzer) of extending the Alexandrov-Zeeman style approach to local versions of relativity theory. Then by Theorems 3–5 and ones in [14] we illustrate that this solution really works, in some sense.

will be the "*quantities*" which we will use for marking time and space.

Let us fix a natural number $n > 1$. n will be the number of space-time dimensions. In most works $n = 4$, i.e. one has 3 space dimensions and one time dimension.

Motivated by the above, our first-order language contains the following symbols:

- unary relation symbols $\mathsf{B}, \mathsf{Ob}, \mathsf{Ph}, \mathsf{F}$ (for bodies, observers, photons, and quantities, i.e. elements of the field, respectively),
- binary function symbols $+, \cdot$ and a binary relation symbol $<$ (for the field-operations and ordering on F),
- a $2+n$-ary relation symbol W (for coordinatizing events, i.e. for the world-view relation).

We will read "$\mathsf{B}(x), \mathsf{Ob}(x), \mathsf{Ph}(x), \mathsf{F}(x)$" as "$x$ is a body", "x is an observer", "x is a photon", "x is a field-element", and we will read "$\mathsf{W}(x, y, z_1, z_2, \ldots, z_n)$" as "observer x sees (or observes) the body y at time z_1 at spatial location $(z_2, \ldots, z_n)$". This "seeing" or "observing" has nothing to do with seeing via photons or observing via experiments, it simply means that, according to x's coordinate-system or reference frame, y is present at coordinates $(z_1, \ldots, z_n)$.

The following axiom will *always* be assumed and will be part of every axiom system we propose, without mentioning.

AxFrame $\mathsf{Ob} \cup \mathsf{Ph} \subseteq \mathsf{B}$, $+$ and $\cdot$ are binary operations on F, $<$ is a binary relation on F, and $\langle \mathsf{F}, +, \cdot, < \rangle$ is a Euclidean ordered field, i.e. an ordered field in which positive elements have square roots.

In pure first-order logic the above axiom would look like $(\mathsf{Ob}(x) \vee \mathsf{Ph}(x)) \rightarrow \mathsf{B}(x)$ etc. We do not write out the purely first-order logic translations of our axioms since they are straightforward to obtain.

Let $\mathbf{M}$ be a model of **AxFrame**. Let $\mathbf{F} = \langle \mathsf{F}, +, \cdot \rangle$ denote the field reduct of $\mathbf{M}$. We will use the following notation and terminology:

$-, /, 0, 1$ are the usual field operations. ${}^n\mathsf{F}$ denotes the set of all n-tuples of elements of F. If a is an n-tuple, then we will assume that $a = \langle a_1, \ldots, a_n \rangle$, i.e. a_i denotes the i-th member of the n-tuple a (for $0 < i \leq n$). We will use the vector-space structure of ${}^n\mathsf{F}$. I.e. if $p, q \in {}^n\mathsf{F}$ and $\lambda \in \mathsf{F}$, then $p+q, -p, \lambda p \in {}^n\mathsf{F}$, and $\bar{0} = \langle 0, \dots, 0\rangle$ is the *null vector.* Let $p \in {}^n\mathsf{F}$. Then $p_t := p_1$ denotes the time component of p and $p_s := \langle 0, p_2, \dots, p_n\rangle$ denotes the space component of p. $|p| := p_1^2 + \dots + p_n^2$ is the (square of the) length of p. The (square of the) *speed* of p is defined as $\mathsf{speed}(p) := |p_s|/p_t^2$ if $p_t \neq 0$, $\mathsf{speed}(p) := \infty$ otherwise. Here we require that $\infty \notin \mathsf{F}$. We extend the ordering $<$ on F to an ordering on $\mathsf{F} \cup \{\infty\}$ in the usual way, i.e. $(\forall x \in \mathsf{F})$ $x < \infty$.

Let $q, v \in {}^n\mathsf{F}, v \neq \bar{0}$. The (straight) *line* going through q and $q+v$ is $\{q + \lambda v : \lambda \in \mathsf{F}\}$. The set of lines is then

$$\mathsf{Lines} := \{\, \{\, q + \lambda v : \lambda \in \mathsf{F}\} : q, v \in {}^n\mathsf{F}, v \neq \bar{0}\,\}.$$

If ℓ is a subset of a line and has at least two elements, then

$$\mathsf{speed}(\ell) := \mathsf{speed}(p - q) \text{ for some (and then for all) } p, q \in \ell, p \neq q.$$

We say that a line ℓ is slower than $\lambda \in {}^+\mathsf{F}$ iff $\mathsf{speed}(\ell) < \lambda$.

$\|$ is the binary relation of *parallelism* on the set Lines, i.e.

$$\ell \parallel \ell' \qquad \Leftrightarrow \qquad (\exists p, q \in \ell)(\exists p', q' \in \ell')\ p - q = p' - q' \neq \bar{0}.$$

Coll is the ternary relation of *collinearity* on ${}^n\mathsf{F}$, i.e.

$$\mathsf{Coll}(p, q, r) \qquad \Leftrightarrow \qquad (\exists \ell \in \mathsf{Lines})\ \{p, q, r\} \subseteq \ell.$$

Let $q, u, v \in {}^n\mathsf{F}$, $\neg\mathsf{Coll}(q, q+u, q+v)$. The *plane* that contains $q, q+u, q+v$ is $\{\, q + \lambda u + \mu v \ :\ \lambda, \mu \in \mathsf{F}\,\}$. The set of planes is then

$$\mathsf{Planes} := \{\, \{\, q + \lambda u + \mu v \ :\ \lambda, \mu \in \mathsf{F}\,\} \ :\ q, u, v \in {}^n\mathsf{F}, \neg\mathsf{Coll}(q, q+u, q+v)\}.$$

${}^+\mathsf{F} := \{\lambda \in \mathsf{F} \ :\ \lambda > 0\}$ is the set of positive elements of F.

The (open) *ball* with center $p \in {}^n\mathsf{F}$ and radius $\varepsilon \in {}^+\mathsf{F}$ is

$$S(p, \varepsilon) := \{q \in {}^n\mathsf{F} \ :\ |p - q| < \varepsilon^2\}.$$

A set $N \subseteq {}^n\mathsf{F}$ is a *neighborhood* of $p \in {}^n\mathsf{F}$ iff $(\exists \varepsilon \in {}^+\mathsf{F})\ S(p, \varepsilon) \subseteq N$. A set $D \subseteq {}^n\mathsf{F}$ is *open* iff $(\forall p \in D)(\exists \varepsilon \in {}^+\mathsf{F})\ S(p, \varepsilon) \subseteq D$.

$\mathcal{A} := \langle {}^n\mathsf{F}, \mathsf{Coll}\rangle$ is the n-dimensional *affine structure* over the field $\mathbf{F}$. $\mathcal{A}$ can be extended to an n-dimensional *projective structure* $\mathcal{P} = \langle \mathsf{P}^n\mathsf{F}, \mathsf{PColl}\rangle$ over $\mathbf{F}$ in the usual way, i.e. as follows: The relation of parallelism $\|$ is an equivalence relation on the set Lines. For every $\ell \in \mathsf{Lines}$ let ℓ^∞ denote the equivalence class of ℓ under the relation $\|$. Intuitively ℓ^∞ is the point of line ℓ at infinity. For every $P \in \mathsf{Planes}$ let $P^\infty := \{\ell^\infty \ :\ \ell \in \mathsf{Lines},\ \ell \subseteq P\}$. Intuitively P^∞ is the line in plane P at infinity. The set of *points* of the projective structure is defined as

$$\mathsf{P}^n\mathsf{F} := {}^n\mathsf{F} \cup \{\ell^\infty \ :\ \ell \in \mathsf{Lines}\}.$$

The set of *(projective) lines* is defined as

$$\mathsf{PLines} := \{\, \ell \cup \{\ell^\infty\} \; : \; \ell \in \mathsf{Lines}\,\} \cup \{\, P^\infty \; : \; P \in \mathsf{Planes}\,\}.$$

Finally, the ternary relation PColl of *collinearity* on $\mathsf{P}^n\mathsf{F}$ is defined as

$$\mathsf{PColl}(a,b,c) \qquad \Leftrightarrow \qquad (\exists \ell \in \mathsf{PLines})\ \{a,b,c\} \subseteq \ell.$$

By the above, the *n-dimensional projective structure*

$$\mathcal{P} := \langle\, \mathsf{P}^n\mathsf{F},\ \mathsf{PColl}\,\rangle$$

over the field **F** has been defined. We note that the affine structure $\mathcal{A}$ is a strong sub-model of the projective structure $\mathcal{P}$.

An *$\mathcal{A}$-collineation* is an automorphism of the affine structure $\mathcal{A}$. In other words, an $\mathcal{A}$-collineation is a permutation of ${}^n\mathsf{F}$ that takes lines onto lines. A *$\mathcal{P}$-collineation* is an automorphism of the projective structure $\mathcal{P}$.

${}^n\mathsf{F}$ is the *coordinate-system* (of each observer) and we will refer to its elements as *coordinate-points*.

The *life-line*, or the *trace* of a body b in observer m's coordinate-system, or as seen by m, is the set of coordinate-points at which m sees b:

$$\mathsf{tr}_m(b) := \{\, p \in {}^n\mathsf{F} \; : \; \mathsf{W}(m,b,p)\,\}.$$

The set of bodies observer m sees at a given coordinate-point $p \in {}^n\mathsf{F}$ is the *event* happening for m at p:

$$\mathsf{ev}_m(p) := \{\, b \in \mathsf{B} \; : \; \mathsf{W}(m,b,p)\,\}.$$

The *coordinate-domain* of observer m is the set of the coordinate-points p where m sees non-empty events:

$$\mathsf{cd}(m) := \{\, p \in {}^n\mathsf{F} \; : \; \mathsf{ev}_m(p) \neq \emptyset\,\}.$$ [5]

The *world-view transformation* between the coordinate-systems of observers m and k is defined as:

$$\mathsf{f}_{mk} := \{\langle p,q\rangle \in {}^n\mathsf{F} \times {}^n\mathsf{F} \; : \; \mathsf{ev}_m(p) = \mathsf{ev}_k(q) \neq \emptyset\}.$$

Note that f_{mk} is a binary relation. f_{mk} will turn out to be an injective partial function assuming axiom **Ax∃Ob** below, cf. Proposition 1.

[5] $\mathsf{ev}_m(p) = \emptyset$ does not mean that space-time would be empty at point p (as seen by m). Instead, it means that observer m does not use point p for coordinatization. I.e. the world-view function $\mathsf{ev}_m : {}^n\mathsf{F} \longrightarrow \mathsf{Events}$ is partial. Part of the explanation of this meaning of $\mathsf{ev}_m(p) = \emptyset$ is that our bodies are only potential bodies. Hence $b \in \mathsf{ev}_m(p)$ means that, potentially, a body b could be present at point p for m. Hence $\mathsf{ev}_m(p) = \emptyset$ implies that nothing could be present, even in principle, at p for m.

If $R \subseteq A \times B$ is a binary relation, then $\mathsf{Dom}(R)$ and $\mathsf{Rng}(R)$ denote the *domain* and *range* of R, respectively, i.e. $\mathsf{Dom}(R) := \{a \in A \ : \ (\exists b \in B)\ \langle a, b\rangle \in R\}$ and $\mathsf{Rng}(R) := \{b \in B \ : \ (\exists a \in A)\ \langle a, b\rangle \in R\}$.

Now everything is ready to state further axioms.

AxLine The traces of observers and photons are subsets of lines, but they must be restrictions of lines to the coordinate-domain (or empty), i.e.

$$(\forall m \in \mathsf{Ob})(\forall k \in \mathsf{Ob} \cup \mathsf{Ph})(\exists \ell \in \mathsf{Lines})$$
$$(\, \mathsf{tr}_m(k) = \ell \cap \mathsf{cd}(m) \quad \text{or} \quad \mathsf{tr}_m(k) = \emptyset\,).$$

The above axiom motivates the definition: if $m \in \mathsf{Ob}$ and $\ell \in \mathsf{Lines}$, then ℓ is called an *observer line for* m iff $(\exists k \in \mathsf{Ob})\ \mathsf{tr}_m(k) = \ell \cap \mathsf{cd}(m)$; and ℓ is called a *photon line for* m iff $(\exists \mathsf{ph} \in \mathsf{Ph})\ \mathsf{tr}_m(\mathsf{ph}) = \ell \cap \mathsf{cd}(m)$.

Ax∃Ob Each point in the coordinate-domain has a neighborhood and a "speed threshold" λ such that each line slower than λ that intersects the neighborhood is an observer line, i.e.

$$(\forall m \in \mathsf{Ob})(\forall p \in \mathsf{cd}(m))(\exists \varepsilon, \lambda \in {}^{+}\mathsf{F})(\forall \ell \in \mathsf{Lines})$$
$$\Big((\mathsf{speed}(\ell) < \lambda \ \wedge \ \ell \cap S(p,\varepsilon) \neq \emptyset) \quad \rightarrow \quad \ell \text{ is an observer line for } m\Big).$$

AxOpen $(\forall m, k \in \mathsf{Ob})(\mathsf{Dom}(\mathsf{f}_{mk})$ is an *open* subset of ${}^{n}\mathsf{F})$.

AxOpen implies that $\mathsf{cd}(m)$ is an open set for every observer m since $\mathsf{cd}(m) = \mathsf{Dom}(\mathsf{f}_{mm})$.

The theorem below says that, locally, the world-view transformations are $\mathcal{P}$-collineations in models of **AxLine**, **Ax∃Ob**, **AxOpen**. This, we think, is a rather strong Alexandrov-Zeeman type theorem. For the Alexandrov-Zeeman theorem cf. Goldblatt [9] or Lester [12] or [4] in the present volume. It says that any bijection from ${}^{n}\mathsf{F}$ to ${}^{n}\mathsf{F}$ that takes lines of speed 1 onto lines of speed 1 is an $\mathcal{A}$-collineation if we assume that $\mathbf{F}$ is the field of reals and $n = 4$. Lester shows in [12, p.929] that this statement does not hold for partial injections in place of bijections.[6]

[6] The theorem on p.929 of Lester [12] and the discussion preceding it are strongly relevant to our Theorems 1 and 2 herein. The general relativistic generalization of Zeeman's theorem in Malament [15] is also relevant. Cf. also Guts [10, §26].

As usual, functions are binary relations. Thus $\mathcal{P}$-collineations are binary relations. We say that the binary relations R and R' *agree on a set* D iff $R \cap (D \times \mathsf{Rng}(R)) = R' \cap (D \times \mathsf{Rng}(R'))$.

Theorem 1. *Assume* **AxLine**, **Ax∃Ob**, **AxOpen**. *Then (i), (ii) below hold.*

(i) For every $m, k \in \mathsf{Ob}$ *and* $p \in \mathsf{Dom}(\mathsf{f}_{mk})$ *there is a unique* $\mathcal{P}$*-collineation, denoted by* C^p_{mk}*, that agrees with* f_{mk} *on some neighborhood of* p.

(ii) For every $m, k \in \mathsf{Ob}$ *and* $\langle p, q\rangle \in \mathsf{f}_{mk}$ *the* $\mathcal{P}$*-collineations* C^p_{mk} *and* C^q_{km} *are inverses of each other.*

The proof of Theorem 1 is in §3.

By item (i) of the above theorem, the world-view transformations preserve Coll and $\neg\mathsf{Coll}$ locally in models of **AxLine**, **Ax∃Ob**, **AxOpen**.

Conjecture 1. *Theorem 1 above remains true if we omit the assumption that our ordered field is Euclidean. Furthermore, the proof given in* §*3 works for the non-Euclidean case if we use cubes instead of balls in the proof.*

Question 1 *Does Theorem 1 above remain true if we replace the assumption* **Ax∃Ob** *by the weaker* **Ax∃Ob$^-$** *below?*

Ax∃Ob$^-$ For each point p in the coordinate-domain there is a "speed threshold" λ such that each line slower than λ that passes through point p is an observer line, i.e.

$$(\forall m \in \mathsf{Ob})(\forall p \in \mathsf{cd}(m))(\exists \lambda \in {}^{+}\mathsf{F})(\forall \ell \in \mathsf{Lines})$$
$$\Big((\mathsf{speed}(\ell) < \lambda \ \wedge\ p \in \ell) \quad\rightarrow\quad \ell \text{ is an observer line for } m\Big).$$

If we replace **AxOpen** by the much stronger **AxFull** below, then **Ax∃Ob** can be replaced by **Ax∃Ob$^-$** in Theorem 1. Beyond that, the world-view transformations will turn out to be $\mathcal{A}$-collineations in models of **AxLine**, **Ax∃Ob$^-$**, **AxFull**.

AxFull below is a typical example of a potential assumption which does *not* have the status of an axiom in the present work. It is a typical postulate which distinguishes special relativity from our more general theories studied herein and in [3]. We formulate **AxFull** to make it sure that we do not assume it in our generalized theories, not even by chance or even implicitly.

AxFull $(\forall m, k \in \mathsf{Ob})\ \mathsf{Dom}(\mathsf{f}_{mk}) = {}^n\mathsf{F}$.

Roughly speaking, **AxFull** says that every observer sees all the events and sees something everywhere in his coordinate-system. From the point of view of general relativity theory, **AxFull** is a too strong assumption, therefore we will not include **AxFull** in our localized relativity theories.

In Theorem 1 above, $\mathcal{P}$-collineation cannot be replaced by $\mathcal{A}$-collineation, see [14]. We get $\mathcal{A}$-collineation, however, if we add axioms about photons. Notice that so far, nothing has been used about photons. We will assume that the photon traces form an "upright" cone, called *light-cone*, at each point, however the angle (or "openness" or "width") of the light-cone may differ from point to point. We are going to formalize this, the result of which will be axiom **AxPh** below. We note that, assuming **AxLine**, **AxPh** is equivalent with $(\star)$ below.

$(\star)$ $(\forall m \in \mathsf{Ob})(\forall p \in \mathsf{cd}(m))(\exists \lambda \in {}^+\mathsf{F})$
$(\forall \ell \in \mathsf{Lines})\,[\, p \in \ell \quad \rightarrow \quad (\ell \text{ is a photon line for } m \ \leftrightarrow\ \mathsf{speed}(\ell) = \lambda)\,]$.

Therefore we could have stated Theorem 2 below here without introducing any further definitions and axioms.

The set of *(spatial) directions* dir is defined as

$$\mathsf{dir} := \{\, d \in {}^n\mathsf{F} \ :\ d_t = 0,\ |d| = 1 \,\}\,.$$

Assume $m \in \mathsf{Ob}$, $b \in \mathsf{B}$, $d \in \mathsf{dir}$. We say that b *moves in direction d as seen by m* iff $\quad (\forall p, q \in \mathsf{tr}_m(b))(\exists \lambda \in \mathsf{F})\ [p_s - q_s = \lambda d \wedge (p_t > q_t \ \rightarrow\ \lambda \geq 0)]$.

The *speed* of a body b as seen by an observer m is $\mathsf{speed}_m(b) := \mathsf{speed}(\mathsf{tr}_m(b))$ if $\mathsf{tr}_m(b)$ is a subset of a line and it has at least two elements, otherwise $\mathsf{speed}_m(b)$ is undefined. Note that $\mathsf{speed}_m(b) = \infty$ is possible. Furthermore, assuming **AxLine** and **AxOpen**, for any $h \in \mathsf{Ob} \cup \mathsf{Ph}$, $\mathsf{speed}_m(h)$ is defined or $\mathsf{tr}_m(h) = \emptyset$. Whenever we use $\mathsf{speed}_m(b)$ in an axiom, we will assume that the axiom states the existence of $\mathsf{speed}_m(b)$, too. Cf. e.g., **AxIstr** below.

Assume $m \in \mathsf{Ob}$, $b \in \mathsf{B}$ and $\mathsf{speed}_m(b)$ is defined. We note the following. If $\mathsf{speed}_m(b) \in {}^+\mathsf{F}$, then b moves in exactly one direction; if $\mathsf{speed}_m(b) = \infty$, then b moves in exactly two directions, i.e. b moves both in d and $-d$ for some direction d; and if $\mathsf{speed}_m(b) = 0$, then b moves in every direction as seen by m.

Ax∃Ph From any point $p \in \mathsf{cd}(m)$ in any direction there is a photon moving in that direction, i.e.

$$(\forall m \in \mathsf{Ob})(\forall p \in \mathsf{cd}(m))(\forall d \in \mathsf{dir})(\exists \mathsf{ph} \in \mathsf{Ph}) \Big(p \in \mathsf{tr}_m(\mathsf{ph}) \;\wedge\; (\mathsf{ph} \text{ moves in direction } d \text{ as seen by } m)\Big).$$

AxIstr below abbreviates Axiom of Isotropy.

AxIstr The speed of light is direction-independent at each point $p \in \mathsf{cd}(m)$, i.e.

$$(\forall m \in \mathsf{Ob})(\forall \mathsf{ph}, \mathsf{ph'} \in \mathsf{Ph}) (\mathsf{tr}_m(\mathsf{ph}) \cap \mathsf{tr}_m(\mathsf{ph'}) \neq \emptyset \quad \rightarrow \quad \mathsf{speed}_m(\mathsf{ph}) = \mathsf{speed}_m(\mathsf{ph'})).$$

AxFin The speed of each photon is nonzero and finite, i.e.

$$(\forall m \in \mathsf{Ob})(\forall \mathsf{ph} \in \mathsf{Ph})\,(0 < \mathsf{speed}_m(\mathsf{ph}) < \infty \;\text{ or }\; \mathsf{tr}_m(\mathsf{ph}) = \emptyset).$$

AxPh := Ax∃Ph ∧ AxIstr ∧ AxFin.

In effect, the photon traces that cross a given $p \in \mathsf{cd}(m)$ show an "upright" cone-like shape, called *light-cone*. Notice that the speed of light—the angle of the light-cone—may differ from point to point.

We note that assuming **AxLine**, **AxOpen**, **AxPh**, the speed of light is constant locally, i.e. $(\forall m \in \mathsf{Ob})(\forall p \in \mathsf{cd}(m))(\exists \varepsilon, \lambda \in {}^{+}\mathsf{F})(\forall \mathsf{ph} \in \mathsf{Ph})$ $(\mathsf{tr}_m(\mathsf{ph}) \cap S(p,\varepsilon) \neq \emptyset \quad \rightarrow \quad \mathsf{speed}_m(\mathsf{ph}) = \lambda)$.

The theorem below says that, locally, the world-view transformations are $\mathcal{A}$-collineations in models of **AxLine**, **Ax∃Ob**, **AxOpen**, **AxPh**.

Theorem 2. *Assume* **AxLine**, **Ax∃Ob**, **AxOpen**, **AxPh**. *Then for every* $m, k \in \mathsf{Ob}$ *and* $p \in \mathsf{Dom}(\mathsf{f}_{mk})$ *there is a unique* $\mathcal{A}$*-collineation that agrees with* f_{mk} *on some neighborhood of* p.

The proof of Theorem 2 is in §3.

By the above theorem, the f_{mk}'s preserve parallelism, Coll and $\neg\mathsf{Coll}$ locally under certain assumptions.

If, in Theorem 2, $n > 2$ and we replace the assumption **AxOpen** by the much stronger **AxFull**, the assumption **Ax∃Ob** becomes superfluous. Moreover, the world-view transformations are $\mathcal{A}$-collineations in models of

AxLine, **AxFull**, **AxPh** if $n > 2$ by the proof of the Alexandrov-Zeeman theorem. Despite of this fact, the assumption **Ax∃Ob** cannot be omitted from Theorem 2 even if $n > 2$ is assumed. This is so because the Alexandrov-Zeeman theorem does not generalize to the local approach pursued herein, as it is shown in Lester [12, p.929].

Question 2 *Does Theorem 2 above remain true if we replace the assumption* **Ax∃Ob** *by the much weaker* **Ax∃Ob^{--}** *below?*

Ax∃Ob^{--} Each line of speed 0 that intersects the coordinate-domain is an observer line, i.e.

$$(\forall m \in \mathsf{Ob})(\forall \ell \in \mathsf{Lines})$$
$$\Big((\mathsf{speed}(\ell) = 0 \ \wedge \ \ell \cap \mathsf{cd}(m) \neq \emptyset) \rightarrow \ \ell \text{ is an observer line for } m\Big).$$

We note that, in Theorem 1, **Ax∃Ob** cannot be replaced by **Ax∃Ob^{--}**. We conjecture that the assumption **Ax∃Ob** can be replaced by **Ax∃Ob^{-}** in Theorem 2.

Finally, we are going to state theorems concerning faster than light observers. To this end we introduce further axioms.

AxSelf Observers can see themselves only on the time-axis, i.e.

$$(\forall m \in \mathsf{Ob})\ \mathsf{tr}_m(m) \subseteq \{\langle t, 0, \ldots, 0\rangle \ : \ t \in \mathsf{F}\}.$$

There may be points on the time-axis where an observer can see nothing. Intuitively, such a point may be after the "Big Crunch"; or for an observer falling into a Schwarzschild black hole, it may be the point (measured by his own clock, i.e. his proper time) where his life-line intersects the singularity.

Assume $k, h \in \mathsf{Ob}$. Then we say that k is a *brother* of h iff $(\forall m \in \mathsf{Ob})\ \mathsf{tr}_m(k) = \mathsf{tr}_m(h)$.

AxEvent If m sees an event happening to k, some brother of k sees it, too,

$$(\forall m, k \in \mathsf{Ob})(\forall p \in \mathsf{tr}_m(k))(\exists h \in \mathsf{Ob})$$
$$[h \text{ is a brother of } k \text{ and } p \in \mathsf{Dom}(\mathsf{f}_{mh})].$$

Intuitive motivation for **AxEvent** above: Consider the life-line of Earth in general relativity. It is an infinitely long spiral. Therefore we cannot approximate the world-view of Earth by a single, long *inertial* frame (such does not exist).[7] On the other hand, we can hope for approximating the world-view of Earth by an infinite sequence of relatively small (hence also "short") inertial frames. Formally, this amounts to decomposing Earth to infinitely many observers whose body part is the same, namely Earth, but whose coordinate-domains correspond to different bounded pieces of Earth's history, so to speak. These versions "... Earth_{-2}, Earth_{-1}, Earth_0, Earth_1, Earth_2,..." of Earth will be brothers in our sense with different, small, coordinate-domains. The union of these domains covers the whole life-line of Earth. This is why in our **AxEvent** above we had to talk about some brother h of k instead of k itself.

Our axiom system **LocRel**, roughly speaking, consists of all axioms introduced so far, except **AxFull**. We note that **LocRel** is a concretely specified version of the axiom system Loc(Specrel) promised in the introduction.[8] **LocRel** excludes faster than light observers if $n > 2$ by Theorem 3 below.

$$\textbf{LocRel} := \{\textbf{AxLine}, \textbf{Ax}\exists\textbf{Ob}, \textbf{AxOpen}, \textbf{Ax}\exists\textbf{Ph}, \textbf{AxIstr}, \textbf{AxFin}, \textbf{AxSelf}, \textbf{AxEvent}\}.$$

FTL abbreviates "faster than light". Let $k, m \in \mathsf{Ob}$. We call k FTL w.r.t. m iff there is a $\mathsf{ph} \in \mathsf{Ph}$ such that k and ph move in the same direction as seen by m, they meet, i.e. $\mathsf{tr}_m(k) \cap \mathsf{tr}_m(\mathsf{ph}) \neq \emptyset$, and $\mathsf{speed}_m(k) > \mathsf{speed}_m(\mathsf{ph})$. **noFTL** abbreviates the formula saying that no observer k can move *faster than light* relative to any other observer, i.e. it abbreviates the formula $\neg(\exists m, k \in \mathsf{Ob})[k\ \mathsf{FTL}\ \text{w.r.t.}\ m]$.

Theorem 3. $\textbf{LocRel} \setminus \{\textbf{AxFin}\} \models \textbf{noFTL}$ *if* $n > 2$.

The proof of Theorem 3 is in §3.

[7] In general relativity, it is the so-called local inertial frames (LIF's) that correspond, roughly, to the world-views of inertial observers in our present local version of relativity, cf. Rindler [20] for LIF's and their such role.

[8] To be precise, **LocRel** is a streamlined, slightly generalized version of the result Loc(Specrel) of applying the localization procedure Loc(-), described in the introduction, to Specrel mechanically. As illustrated in [2], general procedures like Loc(-) are always meant to be applied in this way: we first apply the procedure Loc(-) "mechanically" and then streamline the thus obtained theory.

For $n = 2$, FTL observers do become possible even in models of axiom system $\mathsf{Specrel}$ mentioned in the introduction.

We are going to replace axiom **AxIstr** in **LocRel** by the much weaker **AxP1** below.

AxP1 The speed of light is unique and well-defined in each direction at each point $p \in \mathsf{cd}(m)$. In particular, it does not depend on the movement of the source: photons are unlike bullets. Basically, this is the first-order logic formalization of Friedman's principle (P1) in [8, p.159],

$$(\forall m \in \mathsf{Ob})(\forall \mathsf{ph}, \mathsf{ph'} \in \mathsf{Ph})$$
$$\Big(\mathsf{ph} \text{ and } \mathsf{ph'} \text{ move in the same direction as seen by } m \quad \rightarrow$$
$$(\mathsf{speed}_m(\mathsf{ph}) = \mathsf{speed}_m(\mathsf{ph'}) \text{ or } \mathsf{tr}_m(\mathsf{ph}) \cap \mathsf{tr}_m(\mathsf{ph'}) = \emptyset)\Big).$$

Let the axiom system $\mathbf{LocRel^-}$ be obtained from **LocRel** by replacing **AxIstr** by **AxP1**, i.e.

$$\mathbf{LocRel^- := \{AxLine, Ax\exists Ob, AxOpen,}$$
$$\mathbf{Ax\exists Ph, AxP1, AxFin, AxSelf, AxEvent\}.}$$

Question 3 *Assume* $n > 2$. *Does* $\mathbf{LocRel^-} \models \mathbf{noFTL}$ *hold?*

Let $\mathbf{LocRel^{--}}$ be obtained from $\mathbf{LocRel^-}$ by replacing $\mathbf{Ax\exists Ob}$ by $\mathbf{Ax\exists Ob^{--}}$. The following theorem is due to Gergely Székely [21].

Theorem 4. $\mathbf{LocRel^{--}} \cup \mathbf{\{AxFull\}} \not\models \mathbf{noFTL}$ *if* $n \in \{3, 4\}$.

We will weaken-and-strengthen $\mathbf{Ax\exists Ob}$ in $\mathbf{LocRel^-}$ to requiring that the observer-traces "fill" the light-cones. The thus obtained axiom system will exclude FTL observers if $n > 2$.

AxOb There are observers on lines which are slower than light, i.e.

$$(\forall m \in \mathsf{Ob})(\forall \mathsf{ph} \in \mathsf{Ph}, p \in \mathsf{tr}_m(\mathsf{ph}))(\forall 0 \le \lambda < \mathsf{speed}_m(\mathsf{ph}))(\exists k \in \mathsf{Ob})$$
$$[p \in \mathsf{tr}_m(k),\ \mathsf{speed}_m(k) = \lambda, \text{ and } \mathsf{ph}, k \text{ move in the same direction as seen by } m].$$

Let $\mathbf{LocRel_0^-}$ be obtained from $\mathbf{LocRel^-}$ by replacing $\mathbf{Ax\exists Ob}$ by **AxOb**, i.e.

$$\mathbf{LocRel_0^- := \{AxLine, AxOb, AxOpen,}$$
$$\mathbf{Ax\exists Ph, AxP1, AxFin, AxSelf, AxEvent\}.}$$

Theorem 5. $\mathbf{LocRel_0^-} \models \mathbf{noFTL}$ *if* $n > 2$.

The proof of Theorem 5 is in §3.

The assumption **AxFin** cannot be omitted from $\mathbf{LocRel_0^-}$ in the above theorem.

3 Proofs

The proof of Theorem 1 is based on Desargues' theorem and on Propositions 1, 2 below.

$f : A \longrightarrow B$ denotes that f is a function from A to B, i.e. $\mathsf{Dom}(f) = A$ and $\mathsf{Rng}(f) \subseteq B$.

$f : A \overset{\circ}{\longrightarrow} B$ denotes that f is a *partial function* from A to B; this means that f is a function $\mathsf{Dom}(f) \subseteq A$ and $\mathsf{Rng}(f) \subseteq B$.

Proposition 1. *Assume* $\mathbf{Ax\exists Ob^-}$.
Then for every $m, k \in \mathsf{Ob}$, $\mathsf{f}_{mk} : {}^n\mathsf{F} \overset{\circ}{\longrightarrow} {}^n\mathsf{F}$ *is an injective partial function.*

Proof: Assume the assumptions. Due to the definition of the world-view transformation, it is enough to prove that for every $m \in \mathsf{Ob}$ and distinct $p, q \in \mathsf{cd}(m)$, $\mathsf{ev}_m(p) \neq \mathsf{ev}_m(q)$. Let $m \in \mathsf{Ob}$ and $p, q \in \mathsf{cd}(m)$, $p \neq q$. Let $\ell \in \mathsf{Lines}$ and $k \in \mathsf{Ob}$ be such that $p \in \ell$, $q \notin \ell$, and $\mathsf{tr}_m(k) = \ell \cap \mathsf{cd}(m)$. They exist by $\mathbf{Ax\exists Ob^-}$. Now, $k \in \mathsf{ev}_m(p)$ but $k \notin \mathsf{ev}_m(q)$. Thus $\mathsf{ev}_m(p) \neq \mathsf{ev}_m(q)$.
QED (Prop.1)

In the remaining part of the present paper we use the following notation and definitions.

- If $a, b \in {}^n\mathsf{F}$ with $a \neq b$, then ab denotes the unique element of Lines that contains a and b.
- If $a, b \in \mathsf{P}^n\mathsf{F}$ with $a \neq b$, then $\mathsf{P}ab$ denotes the unique element of PLines that contains a and b.
- Bw is the ternary relation of *strict betweenness* on ${}^n\mathsf{F}$, i.e. $\mathsf{Bw}(p, q, r)$ iff p, q, r are distinct collinear points and q is between p and r. This can be formalized as
  $\mathsf{Bw}(p, q, r) \quad \Leftrightarrow \quad (\exists \lambda \in {}^+\mathsf{F})\,(q = p + \lambda(r - p) \ \wedge\ \lambda < 1).$

- If $a, b \in {}^n\mathsf{F}$, then $[a, b]$ denotes the closed segment determined by a and b, i.e. $[a, b] := \{\, c \in {}^n\mathsf{F} \;:\; \mathsf{Bw}(a, c, b) \;\vee\; c \in \{a, b\}\,\}$.

- Points $p, q, r, s \in {}^n\mathsf{F}$ are *coplanar* iff $(\exists P \in \mathsf{Planes})\, p, q, r, s \in P$.

- $P \in \mathsf{Planes}$ is a *vertical plane* iff $(\exists \ell \in \mathsf{Lines})(\ell \subseteq P \;\wedge\; \mathsf{speed}(\ell) = 0)$.

Next we recall Desargues' theorem from the literature, cf. e.g., Goldblatt [9]. To do so we need the following definitions:

Consider the projective structure $\mathcal{P} = \langle \mathsf{P}^n\mathsf{F}, \mathsf{PColl}\rangle$. A *triangle* is a triple of non-collinear points from $\mathsf{P}^n\mathsf{F}$. These points are the *vertices*, and the (projective) lines connecting two of the vertices are the *sides* of the triangle.

Triangles a', b', c' and a'', b'', c'' are *centrally perspective* iff there is $p \in \mathsf{P}^n\mathsf{F}$ such that $\mathsf{PColl}(p, a', a'')$, $\mathsf{PColl}(p, b', b'')$ and $\mathsf{PColl}(p, c', c'')$, see Fig.1. Triangles a', b', c' and a'', b'', c'' are *axially perspective* iff there are $a, b, c \in$

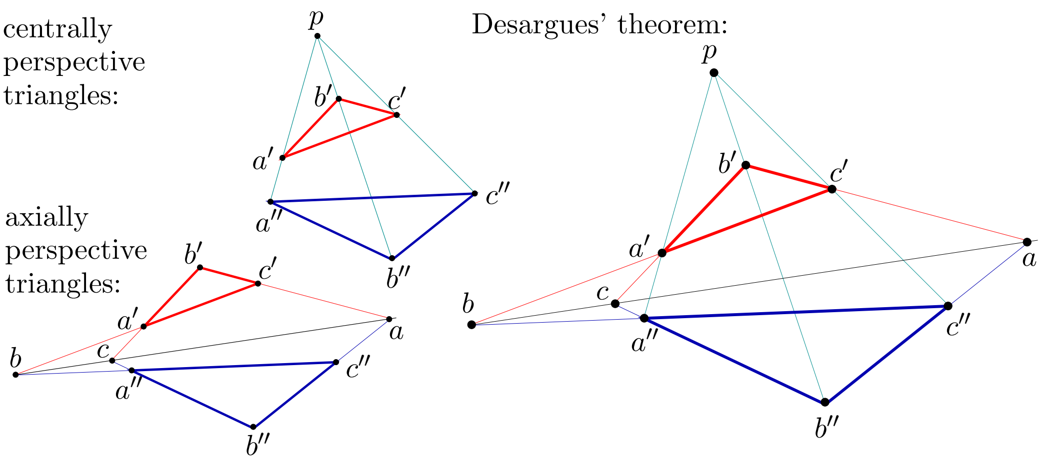


Figure 1: Desargues' theorem

$\mathsf{P}^n\mathsf{F}$ such that $\mathsf{PColl}(a, b, c)$, $\mathsf{PColl}(a, b', c')$, $\mathsf{PColl}(a, b'', c'')$, $\mathsf{PColl}(b, a', c')$, $\mathsf{PColl}(b, a'', c'')$, $\mathsf{PColl}(c, a', b')$, and $\mathsf{PColl}(c, a'', b'')$, see Fig.1.

Desargues' Theorem *Two triangles are centrally perspective if and only if they are axially perspective. Cf. Fig.1.*

<u>*Definition:*</u> Let $f : A \overset{\circ}{\longrightarrow} A$ be a partial function and let R be a ternary relation on A. We say that f *preserves* R *on a set* H iff $H \subseteq \mathsf{Dom}(f)$ and $(\forall x, y, z \in H)[R(x, y, z) \ \rightarrow \ R(f(x), f(y), f(z))]$. Furthermore, f *preserves* R (or f is *R-preserving*) iff f preserves R on $\mathsf{Dom}(f)$.

Proposition 2. *Let* $f : {}^n\mathsf{F} \overset{\circ}{\longrightarrow} {}^n\mathsf{F}$ *be a partial function. Assume* f *preserves* Coll *and* $\neg\mathsf{Coll}$*, and* $\mathsf{Dom}(f)$ *is a ball.*

Then there is a unique PColl*-preserving function* $g : \mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ *extending* f *(*$f \subseteq g$*).*[9] *Furthermore, this unique* g *is injective.*

Question Does g in Prop.2 above preserve $\neg\mathsf{PColl}$?

Proof of Prop.2: Assume $f : {}^n\mathsf{F} \overset{\circ}{\longrightarrow} {}^n\mathsf{F}$ satisfies the assumptions.

Let $\mathsf{L} := \{\, \ell \in \mathsf{PLines} \ : \ \ell \cap \mathsf{Dom}(f) \neq \emptyset \,\}$. For every $\ell \in \mathsf{L}$ there is a unique element of PLines that contains the f-image of ℓ. We will denote this unique element of PLines by $f(\ell)$.

<u>*Definition:*</u> Lines ℓ_1, ℓ_2, ℓ_3 are *concurrent* iff $\ell_1 \cap \ell_2 \cap \ell_3 \neq \emptyset$.

<u>*Claim:*</u> For any distinct and concurrent $\ell_1, \ell_2, \ell_3 \in \mathsf{L}$, the lines $f(\ell_1), f(\ell_2), f(\ell_3)$ are distinct and concurrent.

<u>*Proof:*</u> Assume $\ell_1, \ell_2, \ell_3 \in \mathsf{L}$ are distinct and concurrent. Since f preserves $\neg\mathsf{Coll}$, the lines $f(\ell_1), f(\ell_2), f(\ell_3)$ are distinct. It remains to prove that they are concurrent. We will prove this by Desargues' theorem.

Let $a', a'' \in \ell_1 \cap \mathsf{Dom}(f)$, $b', b'' \in \ell_2 \cap \mathsf{Dom}(f)$, $c', c'' \in \ell_3 \cap \mathsf{Dom}(f)$ be distinct points such that a', b', c' and a'', b'', c'' are triangles and the points of intersection of the corresponding sides of these triangles are in $\mathsf{Dom}(f)$, see Fig.2. It is explained in the caption of Fig.2 why such triangles exist. These triangles are centrally perspective. Thus, by Desargues' theorem, they are axially perspective, i.e. the points of intersection of the corresponding sides are collinear. Since f preserves Coll and $\neg\mathsf{Coll}$, $f(a'), f(b'), f(c')$ and $f(a''), f(b''), f(c'')$ are axially perspective triangles. But then, by Desargues' theorem, they are centrally perspective. Thus the lines $f(\ell_1), f(\ell_2)$ and $f(\ell_3)$ are concurrent. QED (Claim)

We are going to define a function $g : \mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$. Let $p \in \mathsf{P}^n\mathsf{F}$. By the above claim, there is a unique $p' \in \mathsf{P}^n\mathsf{F}$ such that $(\forall \ell \in \mathsf{L})\,(p \in \ell \ \rightarrow \ p' \in$

[9] In particular, $\mathsf{Dom}(g) = \mathsf{P}^n\mathsf{F}$.

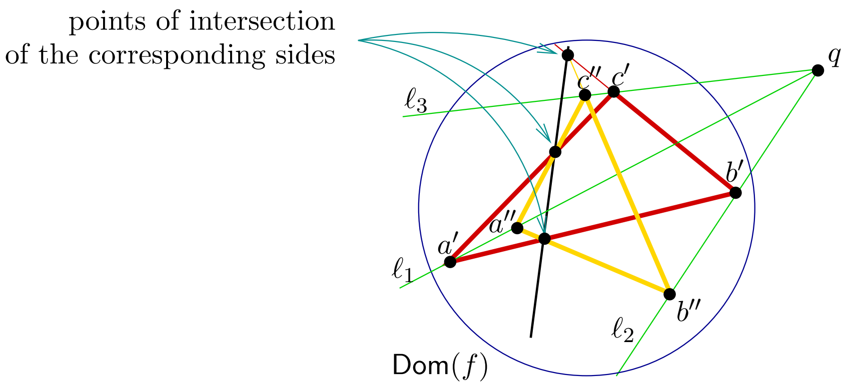


Figure 2: Let $q \in \ell_1 \cap \ell_2 \cap \ell_3$. Choose triangle a', b', c' arbitrarily. Choose a'', b'' such that $\mathsf{Bw}(a', a'', q)$ and $\mathsf{Bw}(b'', b', q)$. Then $a'b' \cap a''b''$ is a point in $\mathsf{Dom}(f)$. Choose c'' so "close" to c' that $a'c' \cap a''c''$ and $b'c' \cap b''c''$ are points in $\mathsf{Dom}(f)$.

$f(\ell))$. We define $g(p)$ to be this unique p'. Clearly, g extends f. Note that g is the unique $\mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ function with the property

$$(\forall \ell \in \mathsf{L})\,(p \in \ell \;\rightarrow\; g(p) \in f(\ell)). \tag{1}$$

If $g' : \mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ is a PColl-preserving function extending f, then g' satisfies (1) above, hence $g' = g$.

It remains to prove that g preserves PColl and that g is injective.

To prove that g is injective let $a, b \in \mathsf{P}^n\mathsf{F}$ be distinct points. Since g extends f and f preserves $\neg\mathsf{Coll}$, g is injective on $\mathsf{Dom}(f)$. Thus, there is $c \in \mathsf{Dom}(f)$ such that $g(a) \neq g(c) \neq g(b)$. Fix such a c. By (1), $g(a), g(c) \in f(\mathsf{P}ac)$ and $g(b), g(c) \in f(\mathsf{P}bc)$. But $f(\mathsf{P}ac), f(\mathsf{P}bc)$ are distinct because $\mathsf{P}ac, \mathsf{P}bc$ were such and f preserves $\neg\mathsf{Coll}$. Hence $g(a) \neq g(b)$.

We will use Desargues' theorem to prove that g preserves PColl. By (1), g preserves collinearity on elements of L, i.e. for any $\ell \in \mathsf{L}$ and $a, b, c \in \ell$, $\mathsf{PColl}(g(a), g(b), g(c))$.

To prove that g preserves PColl, let $a, b, c \in \mathsf{P}^n\mathsf{F}$ be such that $\mathsf{PColl}(a, b, c)$. We can assume that a, b, c are distinct. Let $a', b', c' \in \mathsf{Dom}(f)$ and $a'', b'', c'' \in \mathsf{Dom}(f)$ be triangles such that the corresponding sides meet in a, b, c, respectively, see Fig.3. It is explained in the caption of Fig.3 why such triangles exist. The two triangles are axially perspective. By Desargues' theorem, they are centrally perspective. Furthermore, the lines connecting the corresponding vertices are in L. Therefore, since g preserves $\neg\mathsf{Coll}$ on $\mathsf{Dom}(f)$ and

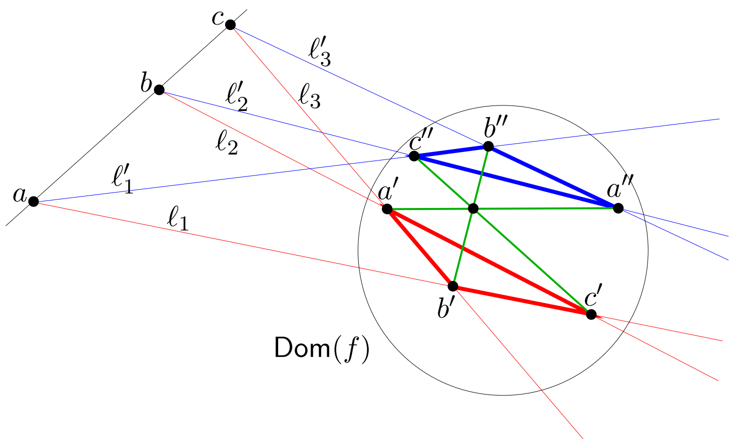


Figure 3: Let $\ell_1, \ell_2, \ell_3 \in \mathsf{L}$ be non-concurrent lines such that $a \in \ell_1,\ b \in \ell_2,\ c \in \ell_3$ and pairwise they meet in $\mathsf{Dom}(f)$. Let $c' \in \ell_1 \cap \ell_2$, $a' \in \ell_2 \cap \ell_3$, $b' \in \ell_1 \cap \ell_3$. Then a', b', c' is a triangle in $\mathsf{Dom}(f)$. The triangle a'', b'', c'' in $\mathsf{Dom}(f)$ is obtained analogously by using $\ell_1', \ell_2', \ell_3'$ distinct from ℓ_1, ℓ_2, ℓ_3.

g preserves PColl on elements of L, $g(a'), g(b'), g(c')$ and $g(a''), g(b''), g(c'')$ are centrally perspectivive triangles and the corresponding sides meet in $g(a), g(b), g(c)$, respectively. But then, by Desargues' theorem, the two triangles are axially perspective, which means that $\mathsf{PColl}(g(a), g(b), g(c))$.

QED (Prop.2)

Proof of Theorem 1:

Assume **AxLine**, **Ax∃Ob**, **AxOpen**. Recall that the world-view transformations (f_{mk}) are injective partial functions by Proposition 1. Furthermore, for every $m, k \in \mathsf{Ob}$, f_{mk} and f_{km} are inverses of each other. We will use these facts tacitly throughout the present proof.

Notation: Assume $m, k \in \mathsf{Ob}$. Then for every $a \in \mathsf{f}_{mk}$, a_m denotes the first component of a, while a_k denotes the second component of a, i.e. $a = \langle a_m, a_k \rangle$. Furthermore, if $a_m \in \mathsf{Dom}(\mathsf{f}_{mk})$, then a_k denotes $\mathsf{f}_{mk}(a_m)$ and if $a_k \in \mathsf{Rng}(\mathsf{f}_{mk}) = \mathsf{Dom}(\mathsf{f}_{km})$, then a_m denotes $\mathsf{f}_{km}(a_k)$.

Claim 1: Assume $m, k \in \mathsf{Ob}$ and $a, b \in \mathsf{f}_{mk}$, $a \neq b$. Then (i), (ii) below hold.

(i) ($a_m b_m$ is an observer line for m) $\Leftrightarrow$ ($a_k b_k$ is an observer line for k).

(ii) f_{mk} preserves Coll and $\neg\mathsf{Coll}$ between three points if the line determined by two of the points is an observer line. Formally: Assume $a_m b_m$ is an observer line for m. Then for every $c \in \mathsf{f}_{mk}$,
$\mathsf{Coll}(a_m, b_m, c_m) \Leftrightarrow \mathsf{Coll}(a_k, b_k, c_k)$, or equivalently $c_m \in a_m b_m \Leftrightarrow c_k \in a_k b_k$.

We omit the easy *proof.*

Claim 2: The world-view transformations preserve Coll locally, i.e. for every $m, k \in \mathsf{Ob}$ and $p \in \mathsf{Dom}(\mathsf{f}_{mk})$, there is a ball S with center p such that f_{mk} preserves Coll on S.

Proof: Let $m, k \in \mathsf{Ob}$. To prove that f_{mk} preserves Coll locally, let $p \in \mathsf{Dom}(\mathsf{f}_{mk})$. We need a ball with center p such that f_{mk} preserves Coll on that ball.

Let $\varepsilon, \lambda \in {}^{+}\mathsf{F}$ be such that $S := S(p, \varepsilon) \subseteq \mathsf{Dom}(\mathsf{f}_{mk})$ and any line slower than λ that intersects S is an observer line for m. Such ε, λ exist by **Ax∃Ob** and **AxOpen**.

Let S' be a ball with center p such that S' is a proper subset of S. For any $H \subseteq {}^{n}\mathsf{F}$, the "vertical cylinder" $c(H)$ of H is defined as
$c(H) := \{\, q \in {}^{n}\mathsf{F} \,:\, (\exists r \in H)\ q_s = r_s \,\}$.

Let $S'' \subseteq S'$ be a ball with center p such that S'' is small enough to satisfy $(*)$ below. See Figure 4.

$(*)$ Any line that intersects both S'' and $c(S'') \setminus S'$ is slower than λ.

Let $X := (S \setminus S') \cap c(S'')$, $X^{+} := \{\, q \in X \,:\, q_t > p_t \,\}$, and $X^{-} := \{\, q \in X \,:\, q_t < p_t \,\}$.

We will use Desargues' theorem and Claim 1 to prove that f_{mk} preserves Coll on S''. Let $a_m, b_m, c_m \in S''$ be such that $\mathsf{Coll}(a_m, b_m, c_m)$. We will prove that $\mathsf{Coll}(a_k, b_k, c_k)$. We can assume that a_m, b_m, c_m are distinct. Let $a'_m, b'_m, c'_m \in X^{+} \subseteq S$ and $a''_m, b''_m, c''_m \in X^{-} \subseteq S$ be triangles such that the corresponding sides meet in a_m, b_m, c_m, respectively, and $b'_m b''_m \cap c'_m c''_m$ is a point in S, see Fig.4. It is explained in the caption of Fig.4 why such triangles exist.

By $(*)$ above, all the sides and the lines connecting the corresponding vertices of these triangles are slower than λ. Thus all these lines are observer lines for m. Furthermore, these triangles are axially perspective. By Desargues' theorem, they are centrally perspective. Moreover, $a'_m a''_m \cap b'_m b''_m \cap c'_m c''_m$ is a point in S. Therefore, by Claim 1 (ii), a'_k, b'_k, c'_k and a''_k, b''_k, c''_k are centrally perspective triangles and the corresponding sides of these triangles meet in

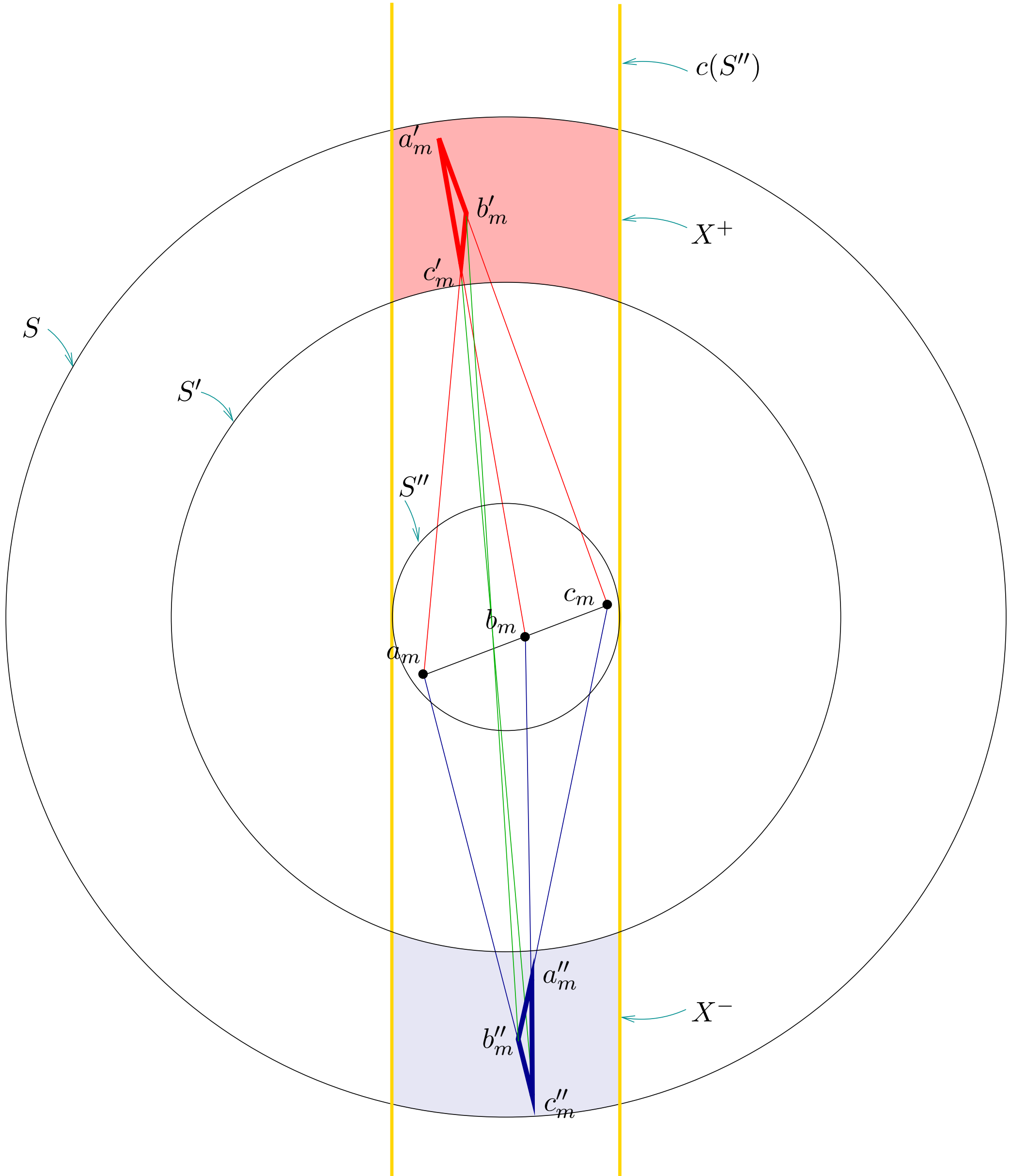


Figure 4: Choose $c'_m, b'_m \in X^+$ such that $\mathsf{Bw}(a_m, c'_m, b'_m)$ and $b_m c'_m \cap c_m b'_m$ is a point in X^+. The latter can be achieved by choosing c'_m and b'_m "close" to each other. Let a'_m be $b_m c'_m \cap c_m b'_m$. Choose $c''_m, b''_m \in X^-$ such that $\mathsf{Bw}(a_m, b''_m, c''_m)$, $b_m c''_m \cap c_m b''_m$ is a point in X^-, $a_m b'_m \neq a_m b''_m$, $b_m c'_m \neq b_m c''_m$ and $c_m b'_m \neq c_m b''_m$. Let a''_m be $b_m c''_m \cap c_m b''_m$. Then, by $\mathsf{Bw}(a_m, c'_m, b'_m)$ and $\mathsf{Bw}(a_m, b''_m, c''_m)$, $c'_m c''_m \cap b'_m b''_m$ is a point in S.

a_k, b_k, c_k, respectively. By Desargues' theorem, we conclude $\mathsf{Coll}(a_k, b_k, c_k)$.
QED (Claim 2)

<u>*Claim 3:*</u> Assume $m, k \in \mathsf{Ob}$, $a, b, c, d \in \mathsf{f}_{mk}$, $d \notin \{a, b, c\}$ and $a_m d_m$, $b_m d_m$, $c_m d_m$ are observer lines for m. Then

$$a_m, b_m, c_m, d_m \text{ are coplanar} \quad \Leftrightarrow \quad a_k, b_k, c_k, d_k \text{ are coplanar.}$$

<u>*Proof:*</u> Assume m, k, a, b, c, d satisfy the assumptions. By Claim 1 (i), it is enough to prove one direction of "$\Leftrightarrow$" in the present claim. We will prove, e.g., the "$\Leftarrow$" direction. Assume a_k, b_k, c_k, d_k are coplanar. Let $S \subseteq \mathsf{Dom}(\mathsf{f}_{km})$ be a ball with center d_k such that f_{km} preserves Coll on S. S exists by Claim 2. See the left hand side of Figure 5. Let $a'_k \in S \cap a_k d_k$, $b'_k \in S \cap b_k d_k$

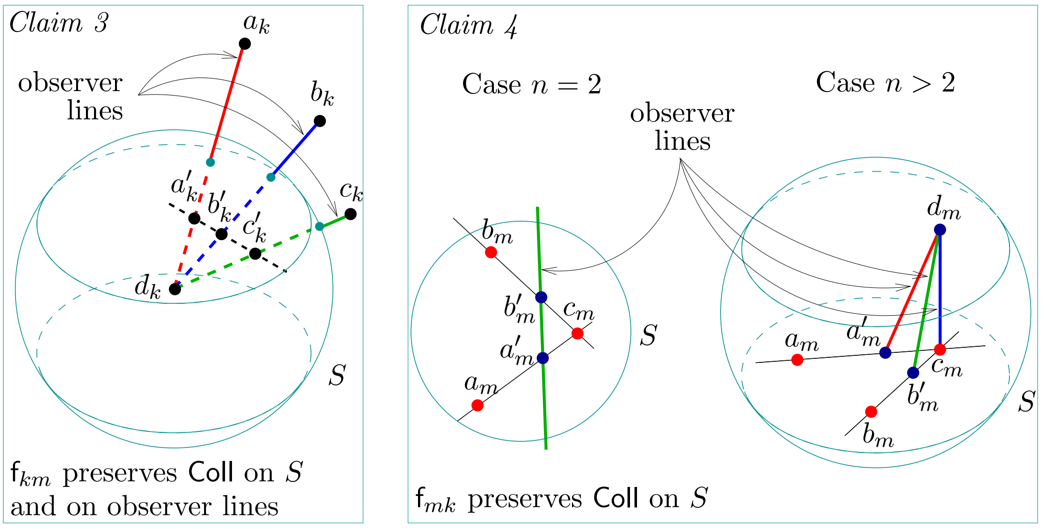


Figure 5: Illustrations for the proofs of Claims 3,4.

and $c'_k \in S \cap c_k d_k$ be such that $\mathsf{Coll}(a'_k, b'_k, c'_k)$ and $d_k \notin \{a'_k, b'_k, c'_k\}$, cf. Fig.5. Clearly, $d_m \notin \{a'_m, b'_m, c'_m\}$. Since f_{km} preserves Coll on S, we have $\mathsf{Coll}(a'_m, b'_m, c'_m)$. By Claim 1 (ii), $a'_m \in a_m d_m$, $b'_m \in b_m d_m$ and $c'_m \in c_m d_m$, since $a_m d_m$, $b_m d_m$, $c_m d_m$ are observer lines. Therefore a_m, b_m, c_m, d_m are coplanar.
QED (Claim 3)

<u>*Claim 4:*</u> Assume $m, k \in \mathsf{Ob}$ and S is a ball such that f_{mk} preserves Coll on S. Then f_{mk} preserves $\neg\mathsf{Coll}$ on S.

<u>*Proof:*</u> Assume m, k, S satisfy the assumptions. Let $a_m, b_m, c_m \in S$ be such that $\neg\mathsf{Coll}(a_m, b_m, c_m)$. We want to prove $\neg\mathsf{Coll}(a_k, b_k, c_k)$. We distinguish two cases, the case of $n = 2$ and the case of $n > 2$. See the right hand side of Fig.5.

Case of $n = 2$*:* Assume $n = 2$. Let $a'_m \in a_m c_m \cap S$ and $b'_m \in b_m c_m \cap S$ be such that $c_m \notin \{a'_m, b'_m\}$ and $a'_m b'_m$ is an observer line for m. a'_m, b'_m exist by **Ax∃Ob**. Clearly, $\neg\mathsf{Coll}(a'_m, b'_m, c_m)$. Then, by Claim 1 (ii), $\neg\mathsf{Coll}(a'_k, b'_k, c_k)$ since $a'_m b'_m$ is an observer line. By $\neg\mathsf{Coll}(a'_k, b'_k, c_k)$ and the assumption that f_{mk} preserves Coll on S, we have $\neg\mathsf{Coll}(a_k, b_k, c_k)$.

Case of $n > 2$*:* Assume $n > 2$. Let $a'_m \in a_m c_m \cap S$, $b'_m \in b_m c_m \cap S$ and $d_m \in S$ be such that a'_m, b'_m, c_m, d_m are not coplanar and $a'_m d_m$, $b'_m d_m$, $c_m d_m$ are observer lines for m, cf. Fig.5. a'_m, b'_m, d_m exist by **Ax∃Ob**. Then, by Claim 3, we have that a'_k, b'_k, c_k, d_k are not coplanar. But then $\neg\mathsf{Coll}(a'_k, b'_k, c_k)$. By this and by the assumption that f_{mk} preserves Coll on S, we have $\neg\mathsf{Coll}(a_k, b_k, c_k)$.

QED (Claim 4)

Let $m, k \in \mathsf{Ob}$, $p \in \mathsf{f}_{mk}$ be fixed until the proof is complete. Furthermore, let a ball S with center p_m be fixed such that f_{mk} preserves Coll on S. S exists by Claim 2. Then f_{mk} preserves $\neg\mathsf{Coll}$ on S by Claim 4.

Now, by Prop.2, there is a unique PColl-preserving $\mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ function that agrees with f_{mk} on S. Denote this function by g. g is injective by Prop.2. We will prove that g is a $\mathcal{P}$-collineation.

<u>*Claim 5:*</u> Assume $H \subseteq \mathsf{Dom}(\mathsf{f}_{mk})$ is an open set and g agrees with f_{mk} on H. Assume $e \in \mathsf{f}_{mk}$ and ℓ, ℓ' are observer lines for m such that $\ell \cap \ell' = \{e_m\}$ and $\ell \cap H \neq \emptyset \neq \ell' \cap H$. Then $e \in g$.

<u>*Proof:*</u> Note that for any line ℓ and open set H, we have $(\ell \cap H \neq \emptyset) \Rightarrow (\ell \cap H$ is an infinite set$)$. Now, assume H, e, ℓ, ℓ' satisfy the assumptions. Let $a_m, b_m \in H \cap \ell$ and $c_m, d_m \in H \cap \ell'$ be distinct points, cf. the left hand side of Figure 6. Then, by Claim 1 (ii), $a_k b_k \cap c_k d_k = \{e_k\}$. g takes a_m, b_m, c_m, d_m to a_k, b_k, c_k, d_k, respectively, by the assumption that g agrees with f_{mk} on H. Since g preserves PColl, it takes e_m to e_k. Thus $e \in g$.

QED (Claim 5)

Let $c(S) := \{\, q \in {}^n\mathsf{F} \,:\, (\exists r \in S)\ q_s = r_s \,\}$ be the "vertical cylinder" of S, cf. the right hand side of Fig.6.

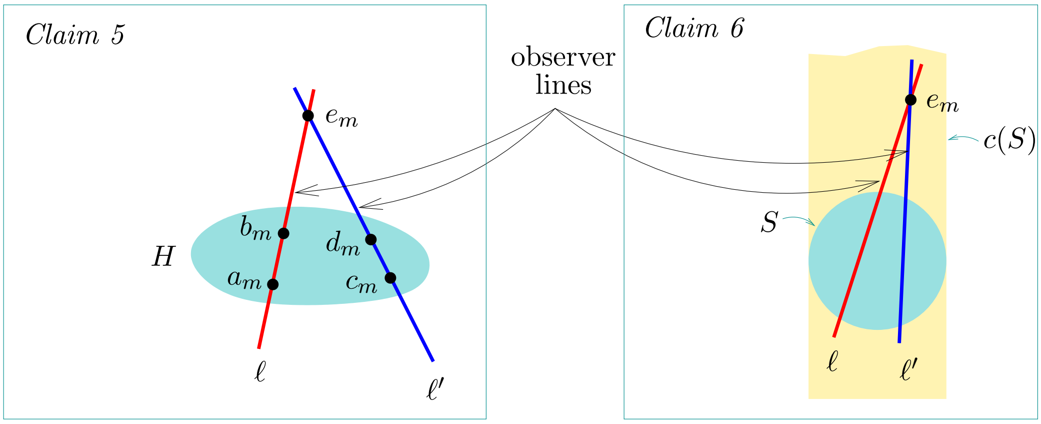


Figure 6: Illustrations for proofs of Claims 5,6.

<u>*Claim 6:*</u> g agrees with f_{mk} on $c(S) \cap \mathsf{Dom}(\mathsf{f}_{mk})$, i.e. for every $e \in \mathsf{f}_{mk}$, $e_m \in c(S) \;\Rightarrow\; e \in g$.

<u>*Proof:*</u> Let $e \in \mathsf{f}_{mk}$ be such that $e_m \in c(S)$. By **Ax∃Ob**, there are two observer lines ℓ, ℓ' for m such that they meet in e_m and both of them intersect S, cf. the right hand side of Fig.6. Then, by Claim 5 and by the fact that g agrees with f_{mk} on S, we have $e \in g$. QED (Claim 6)

Let a ball S_k with center p_k and $\lambda \in {}^+\mathsf{F}$ be fixed until the proof is complete such that

- f_{km} preserves Coll on S_k, and
- each line slower than λ that intersects S_k is an observer line for k.

Such S_k exists by **Ax∃Ob** and Claim 2.

<u>*Claim 7:*</u> Assume $e \in \mathsf{f}_{mk}$ is such that $e_k \neq p_k$ and $\mathsf{speed}(e_k p_k) < \lambda$. Then $e \in g$.

<u>*Proof:*</u> Assume e satisfies the assumptions. See Figure 7. Let $q_m \in S$ be such that $p_m \neq q_m$ and $\mathsf{speed}(p_m q_m) = 0$. Note that $p_m q_m \subseteq c(S)$ and, by **Ax∃Ob**, $p_m q_m$ is an observer line for m. Choose $a_k \in p_k q_k \cap S_k$ such that $a_k \neq p_k$ and $\mathsf{speed}(e_k a_k) < \lambda$. $\mathsf{speed}(e_k a_k) < \lambda$ can be achieved by choosing a_k "very close" to p_k.

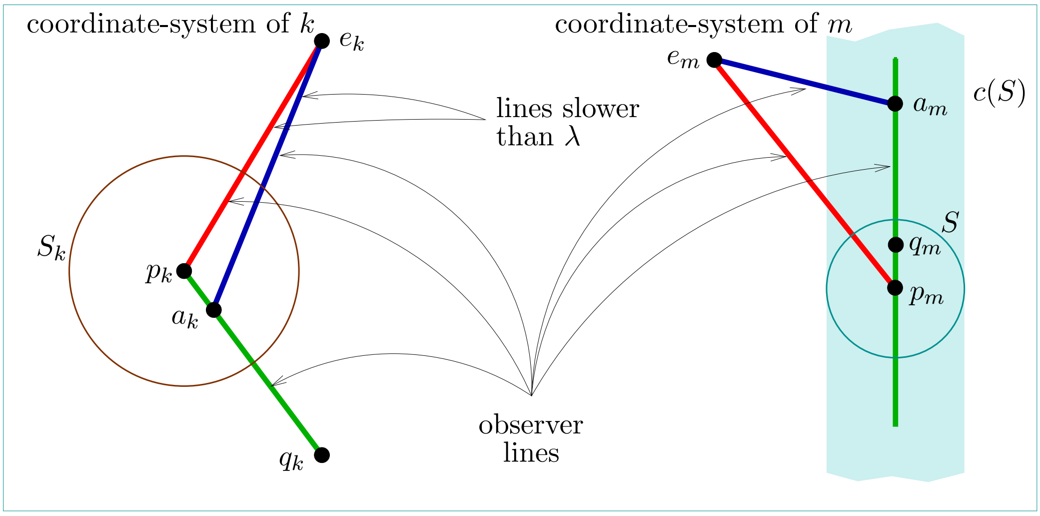


Figure 7: Illustration for the proof of Claim 7.

Now, by our choice of S_k and λ, both $e_k a_k$, $e_k p_k$ are observer lines for k. Thus, by Claim 1 (i), $e_m a_m$ and $e_m p_m$ are observer lines for m. Furthermore, by Claim 1 (ii), $a_m \in p_m q_m$.

Assume $e_m \in p_m q_m$. Then $e_m \in c(S)$. Thus, by Claim 6, $e \in g$.

Assume $e_m \notin p_m q_m$. Then the two observer lines $e_m a_m$, $e_m p_m$ meet in e_m and both of them intersect the open set $c(S) \cap \mathsf{Dom}(\mathsf{f}_{mk})$. But g agrees with f_{mk} on this open set by Claim 6. Therefore, by Claim 5, $e \in g$. QED (Claim 7)

g^{-1} denotes the inverse of g. We note that $g^{-1} : \mathsf{P}^n\mathsf{F} \xrightarrow{\circ} \mathsf{P}^n\mathsf{F}$ is a partial function.

Claim 8: $S_k \subseteq \mathsf{Rng}(g)$ and f_{km} and g^{-1} agree on S_k, i.e.

$$(e \in \mathsf{f}_{mk} \wedge e_k \in S_k) \ \Rightarrow \ e \in g.$$

Proof: Assume $e \in \mathsf{f}_{mk}$ and $e_k \in S_k$. Let lines ℓ, ℓ' be slower than λ such that $\ell \cap \ell' = \{p_k\}$ and e_k is in the plane determined by ℓ, ℓ'. See the left hand side of Figure 8.

If $e_k \in \ell \cup \ell'$, then $e \in g$ by Claim 7. So we can assume $e_k \notin \ell \cup \ell'$. Let a_k, b_k, c_k, d_k be distinct points such that $a_k, b_k \in \ell$, $c_k, d_k \in \ell'$ and

$$a_k c_k \cap b_k d_k = \{e_k\}.$$

Note that, by our choice of S_k and Claim 4, f_{km} preserves Coll and $\neg\mathsf{Coll}$ on

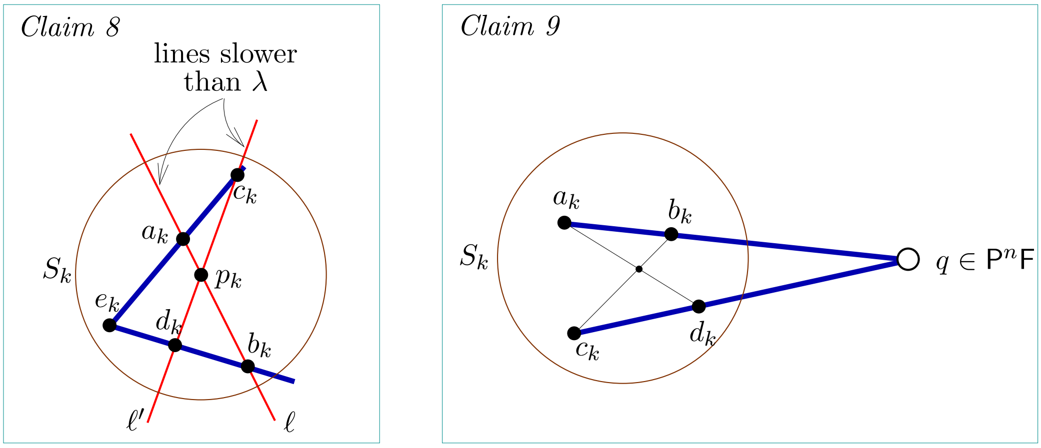


Figure 8: Illustrations for the proofs of Claims 8,9

S_k. Thus

$$a_m c_m \cap b_m d_m = \{e_m\}.$$

By Claim 7, g takes a_m, b_m, c_m, d_m to a_k, b_k, c_k, d_k, respectively. Since g preserves PColl, g takes e_m to e_k, i.e. $e \in g$. QED (Claim 8)

Claim 9: g is surjective, i.e. $\mathsf{Rng}(g) = \mathsf{P}^n\mathsf{F}$. Hence g is a bijection.

Proof: Let $q \in \mathsf{P}^n\mathsf{F}$. Let $a_k, b_k, c_k, d_k \in S_k$ be distinct points such that

$$\mathsf{P}a_k b_k \cap \mathsf{P}c_k d_k = \{q\},$$

cf. the right hand side of Figure 8. Note that the points a_k, b_k, c_k, d_k are coplanar and the lines $a_k b_k$, $c_k d_k$ are distinct. Since f_{km} preserves Coll and $\neg\mathsf{Coll}$ on S_k, a_m, b_m, c_m, d_m are coplanar and the lines $a_m b_m$ and $c_m d_m$ are distinct. But then there is $r \in \mathsf{P}^n\mathsf{F}$ such that

$$\mathsf{P}a_m b_m \cap \mathsf{P}c_m d_m = \{r\}.$$

By Claim 8, g takes a_m, b_m, c_m, d_m to a_k, b_k, c_k, d_k respectively. Since g preserves PColl it takes r to q. Thus $q \in \mathsf{Rng}(g)$. QED (Claim 9)

Claim 10: g is a $\mathcal{P}$-collineation.

Proof: By Claim 9, $g : \mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ is a PColl-preserving bijection. But any PColl-preserving bijection $f : \mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ is a $\mathcal{P}$-collineation.

QED (Claim 10)

<u>*Claim 11:*</u> Assume g' is a $\mathcal{P}$-collineation and S' is a neighborhood of p_m, and g' agrees with f_{mk} on S'. Then $g' = g$.

<u>*Proof:*</u> Assume g', S' satisfy the assumptions. Let S'' be a ball with center p_m such that $S'' \subseteq S \cap S'$. Then both g, g' agree with f_{mk} on S'', and f_{mk} preserves Coll and $\neg\mathsf{Coll}$ on S''. But then, by the "uniqueness" part of Prop.2, $g = g'$. QED (Claim 11)

At this point, item (i) of the theorem has been proven. Item (ii) of the theorem follows from Claims 8, 10 above and from item (i) of the theorem.
QED (Theorem 1)

<u>*Definition:*</u> Assume **Ax∃Ph**, **AxIstr**. Assume $m \in \mathsf{Ob}$ and $p \in \mathsf{cd}(m)$. Then there is a unique $\lambda \in {}^{+}\mathsf{F} \cup \{0, \infty\}$ such that

$$(\forall \mathsf{ph} \in \mathsf{Ph})(p \in \mathsf{tr}_m(\mathsf{ph}) \;\rightarrow\; \mathsf{speed}_m(\mathsf{ph}) = \lambda).$$

This unique λ is called the *speed of light at* p *for* m.

Proof of Theorem 2: Assume the assumptions. Recall that the f_{mk}'s are injective partial functions by Proposition 1.

Let $m, k \in \mathsf{Ob}$ and $p \in \mathsf{Dom}(\mathsf{f}_{mk})$ be fixed. Let g be a $\mathcal{P}$-collineation and let S be a ball with center p such that f_{mk} agrees with g on S. Such g and S exist by Theorem 1. Note that $S \subseteq \mathsf{Dom}(\mathsf{f}_{mk}) \subseteq \mathsf{cd}(m)$. We will prove that the restriction of g to ${}^{n}\mathsf{F}$ is an $\mathcal{A}$-collineation.

<u>*Claim 1:*</u> Assume $x, y \in \mathsf{Dom}(\mathsf{f}_{mk})$, $x \neq y$. Then
(xy is a photon line for m) $\Leftrightarrow$ ($\mathsf{f}_{mk}(x)\mathsf{f}_{mk}(y)$ is a photon line for k).

Here we omit the easy *proof*. If one wants to obtain a proof, one has to use **AxLine**.

<u>*Claim 2:*</u> Assume $h \in \mathsf{Ob}$, $\ell \in \mathsf{Lines}$ and $x \in \ell \cap \mathsf{cd}(h) \neq \emptyset$. Let η be the speed of light at x for h. Then

$$\mathsf{speed}(\ell) = \eta \quad \Leftrightarrow \quad (\ell \text{ is a photon line for } h).$$

Here we omit the easy *proof*. If one wants to obtain a proof one has to use **AxLine** and **Ax∃Ph**.

Let λ be the speed of light at p for m. By **AxFin**, $\lambda \in {}^{+}\mathsf{F}$.

Definition: We call the elements of $\mathsf{P}^n\mathsf{F} \setminus {}^n\mathsf{F}$ *infinite points* and the elements of ${}^n\mathsf{F}$ *finite points*.

Recall that if $\ell \in \mathsf{Lines}$, then $\ell^\infty \in \mathsf{P}^n\mathsf{F}$ is "the point of ℓ at infinity".

Claim 3: Assume $\ell \in \mathsf{Lines}$ is such that $p \in \ell$ and $\mathsf{speed}(\ell) = \lambda$. Then $g(\ell^\infty)$ is an infinite point.

Proof: Assume ℓ satisfies the assumptions. Let $a, b, c \in S$ be such that $a \notin \ell$,

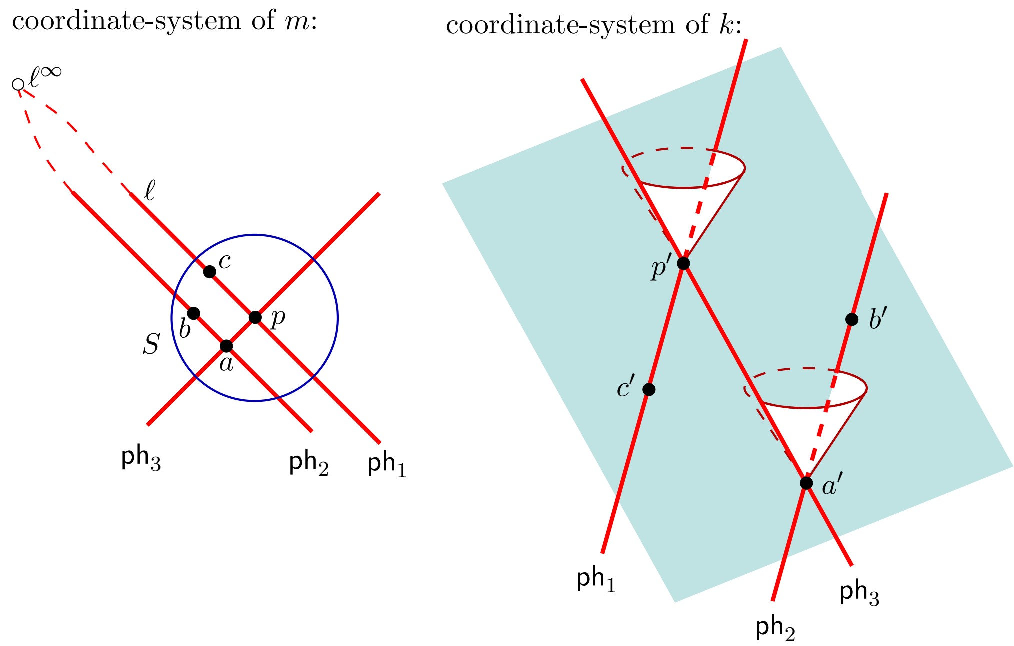


Figure 9: Illustration for the proof of Claim 3.

$\mathsf{speed}(pa) = \lambda$, $b \neq a$, $ab \parallel \ell$, $c \in \ell$ and $c \neq p$. See Figure 9. Then $\mathsf{speed}(ab) = \lambda$, too. Since the speed of light is λ at p for m, pc and pa are photon lines for m, by Claim 2. Since pa is a photon line, the speed of light at a is $\mathsf{speed}(pa) = \lambda$ for m. Thus, since $\mathsf{speed}(ab) = \lambda$, ab is a photon line for m by Claim 2. Note that a, b, c, p are coplanar and no three of them are collinear.

Let a', b', c', p' be the f_{mk} images of a, b, c, p, respectively. See Figure 9. Recall that f_{mk} agrees with a $\mathcal{P}$-collineation on S. Thus a', b', c', p' are coplanar and no three of them are collinear since a, b, c, p are such. By Claim

1, $p'c', p'a', a'b'$ are photon lines for k since pc, pa, ab are photon lines for m. Thus the speed of light is the same at p' and a' for k, which is $\mathsf{speed}(p'a')$; the speed of light at p' is $\mathsf{speed}(p'c')$; and the speed of light at a' is $\mathsf{speed}(a'b')$. Hence $\mathsf{speed}(p'c') = \mathsf{speed}(a'b')$. But then, $p'c' \parallel a'b'$. Now, since g preserves PColl and agrees with f_{mk} on S, it takes $\mathsf{P}pc \cap \mathsf{P}ab = \{\ell^\infty\}$ to $\mathsf{P}p'c' \cap \mathsf{P}a'b'$. Hence g takes ℓ^∞ to an infinite point. QED (Claim 2)

Claim 4: g takes infinite points to infinite points.

Proof: Let q be an infinite point. Then $q = \ell^\infty$ for some $\ell \in \mathsf{Lines}$ with $p \in \ell$. Let such an ℓ be fixed. Let P be a vertical plane that contains ℓ. See Figure 10. Let ℓ_1, ℓ_2 be lines of speed λ such that $\ell_1 \cap \ell_2 = \{p\}$ and

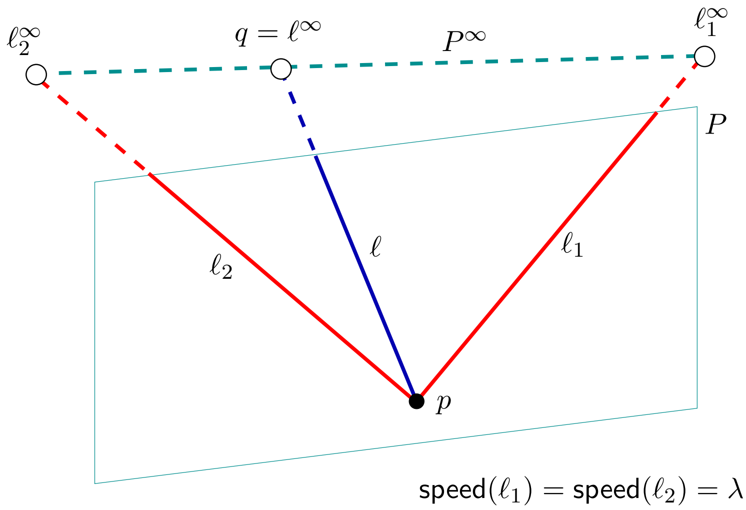


Figure 10: Illustration for the proof of Claim 4.

$\ell_1, \ell_2 \subseteq P$. Then, since $\ell^\infty, \ell_1^\infty, \ell_2^\infty \in P^\infty$, we have $\mathsf{PColl}(\ell^\infty, \ell_1^\infty, \ell_2^\infty)$. But then $\mathsf{PColl}(g(\ell^\infty), g(\ell_1^\infty), g(\ell_2^\infty))$. By Claim 3, $g(\ell_1^\infty)$ and $g(\ell_2^\infty)$ are infinite points. Furthermore, they are distinct. Since no projective line contains two infinite points and a finite point, we conclude that $g(\ell^\infty)$ is an infinite point.
QED (Claim 4)

Claim 5: $g \cap ({}^n\mathsf{F} \times {}^n\mathsf{F})$ is an $\mathcal{A}$-collineation.

Proof: Since g is a $\mathcal{P}$-collineation it is enough to prove that $g \cap ({}^n\mathsf{F} \times {}^n\mathsf{F})$ is a permutation. By Claim 4, it is enough to prove that g takes finite points to finite points since g is a permutation on $\mathsf{P}^n\mathsf{F}$. g takes p to a finite point, i.e. to $\mathsf{f}_{mk}(p)$. To prove that g takes finite points to finite points let q be a finite point, $p \neq q$. Since g preserves PColl, we have $\mathsf{PColl}(g(p), g(q), g(pq^\infty))$. Since $g(pq^\infty)$ is an infinite point and $g(p)$ is a finite point, we conclude that $g(q)$ is a finite point. QED (Claim 5)

By this the "existence" part of our theorem has been proven. The "uniqueness" part of the theorem follows from Theorem 1 and from the fact that any $\mathcal{A}$-collineation can be extended to a $\mathcal{P}$-collineation.

QED (Theorem 2)

We will use Lemma 1 in the proof of Theorem 3.

Lemma 1. *Assume* **LocRel $\setminus$ {AxIstr, AxFin}**. *If there is an observer trace in a plane passing through a point, then there is a photon trace in the plane passing through the point.*

Formally: Assume $m, k \in \mathsf{Ob}$ *and* $p \in \mathsf{tr}_m(k) \subseteq P \in \mathsf{Planes}$. *Then there is a* $\mathsf{ph} \in \mathsf{Ph}$ *such that* $p \in \mathsf{tr}_m(\mathsf{ph}) \subseteq P$.

Proof: Assume $m, k \in \mathsf{Ob}$ and $p \in \mathsf{tr}_m(k) \subseteq P \in \mathsf{Planes}$. f_{mk} is an injective partial function by Proposition 1. We can assume $p \in \mathsf{Dom}(\mathsf{f}_{mk})$ since, by **AxEvent**, k has a brother h such that $p \in \mathsf{Dom}(\mathsf{f}_{mh})$. Let $p' := \mathsf{f}_{mk}(p)$. By Theorem 1, there are a $\mathcal{P}$-collineation g and balls S, S' with centers p, p', respectively, such that f_{mk} agrees with g on S and f_{km} agrees with g^{-1} on S', where g^{-1} denotes the inverse of g. Let such g, S, S' be fixed. See Figure 11. By **AxLine** and $S \subseteq \mathsf{Dom}(\mathsf{f}_{mk}) \subseteq \mathsf{cd}(m)$, $\ell \cap S = \mathsf{tr}_m(k) \cap S$, for some $\ell \in \mathsf{Lines}$. Thus there is $q \in \mathsf{tr}_m(k) \cap S$ such that $p \neq q$. Let such a q be fixed and let $r \in S \cap P$ be such that p, q, r are non-collinear points. Let q' and r' be the f_{mk} images of q and r, respectively. The $\mathcal{P}$-collineation g takes p, q, r to p', q', r', respectively. Thus p', q', r' are non-collinear, too. Let P' be the plane that contains p', q', r'. We have $p', q' \in \mathsf{tr}_k(k)$ since $p, q \in \mathsf{tr}_m(k)$. By **AxSelf**, $\mathsf{speed}(p'q') = 0$. Thus P is a vertical plane. Hence, by **Ax∃Ph**, there is a photon ph such that $p' \in \mathsf{tr}_k(\mathsf{ph}) \subseteq P'$. Let such a ph be fixed. Let $a' \in \mathsf{tr}_k(\mathsf{ph}) \cap S'$ be such that $a' \neq p'$. Such an a' exists by **AxLine**. Let $a := \mathsf{f}_{km}(a')$. Since g^{-1} agrees with f_{km} on S' it takes a' to a. Since $\mathcal{P}$-collineation g^{-1} takes p', q', r', a' to p, q, r, a, respectively, we

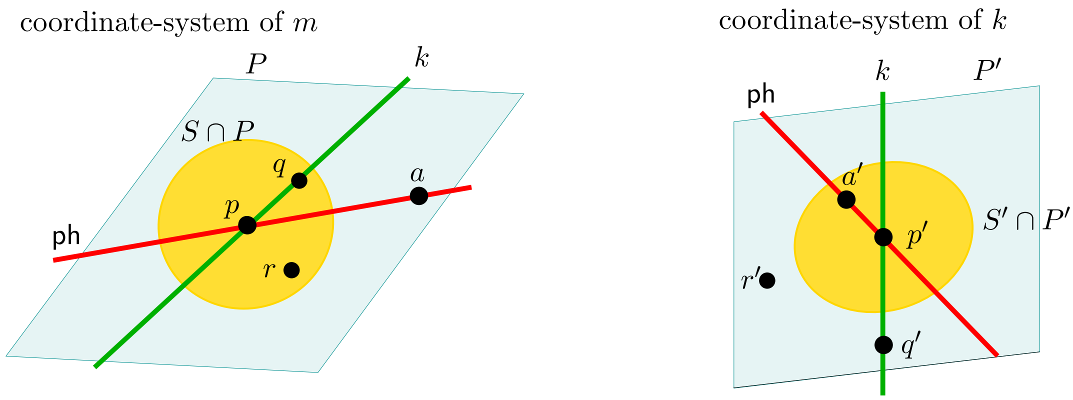


Figure 11: Illustration for the proof of Lemma 1.

conclude that p, q, r, a are coplanar, i.e. $a \in P$. Furthermore, $a, p \in \mathsf{tr}_m(\mathsf{ph})$ since $a', p' \in \mathsf{tr}_k(\mathsf{ph})$. But then, by **AxLine**, $\mathsf{tr}_m(\mathsf{ph}) \subseteq P$.

QED (Lemma 1)

Proof of Theorem 3: Assume $n > 2$ and **LocRel** $\setminus$ **{AxFin}**. Assume there is an FTL observer, i.e. there are $k, m \in \mathsf{Ob}$ such that k is FTL w.r.t. m. Let such m, k be fixed. Then there is $p \in \mathsf{tr}_m(k)$ such that $\mathsf{speed}_m(k) >$ (speed of light at p for m). Let such a p be fixed and let λ be the speed of light at p for m. Let P be a plane such that $\mathsf{tr}_m(k) \subseteq P$ and

$$(\forall \ell \in \mathsf{Lines})(\ell \subseteq P \ \rightarrow \ \mathsf{speed}(\ell) > \lambda). \tag{2}$$

See Figure 12. Such a plane exists since $\mathsf{speed}_m(k) > \lambda$ and $n > 2$. Now, by Lemma 1, there is a photon ph such that $p \in \mathsf{tr}_m(\mathsf{ph}) \subseteq P$. For this $\mathsf{ph} \in \mathsf{Ph}$, by (2), we have $\mathsf{speed}_m(\mathsf{ph}) > \lambda$. This contradicts the fact that the speed of light at p for m is λ.

QED (Theorem 3)

Now, we turn to the proof of our last "**noFTL**" theorem, Theorem 5. Propositions 3–6 and Lemmas 2–4 below are needed for the proof of this theorem.

Proposition 3. *Assume* **AxLine, Ax∃Ph, AxFin**.

Then for every $m, k \in \mathsf{Ob}$, $\mathsf{f}_{mk} : {}^n\mathsf{F} \xrightarrow{\circ} {}^n\mathsf{F}$ *is an injective partial function.*

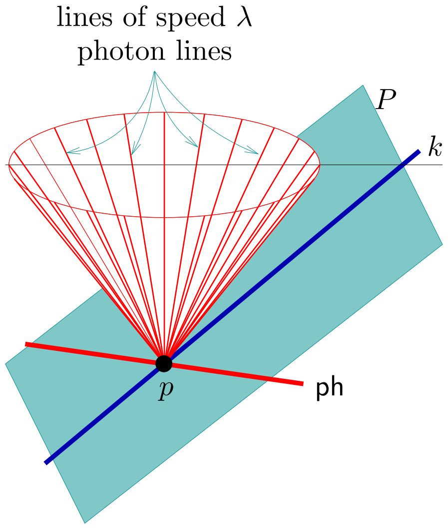


Figure 12: Illustration for the proof of Theorem 3.

Outline of proof: The proof of the proposition is similar to that of Proposition 1. Instead of observer m in that proof one has to use a photon to obtain a proof for the present proposition.

QED (Prop.3)

Ax∃Ob* below is a weaker version of **Ax∃Ob**.

Ax∃Ob* For every vertical plane P, each point in the coordinate-domain has a neighborhood and a "speed threshold" λ such that each line in plane P slower than λ and intersecting the neighborhood is an observer line, i.e.

$$(\forall m \in \mathsf{Ob})(\forall \text{ vertical plane } P)(\forall p \in \mathsf{cd}(m))(\exists \varepsilon, \lambda \in {}^{+}\mathsf{F})$$
$$(\forall \ell \in \mathsf{Lines})\Big((\ell \subseteq P \ \wedge\ \mathsf{speed}(\ell) < \lambda \ \wedge\ \ell \cap S(p,\varepsilon) \neq \emptyset) \quad \rightarrow \quad \ell \text{ is an observer line for } m\Big).$$

Proposition 4. $\mathbf{LocRel}_0^- \setminus \{\mathbf{AxSelf}, \mathbf{AxEvent}\} \models$ **Ax∃Ob***.

Proof:

<u>*Claim 1:*</u> Assume P is a vertical plane, $p \in P$, S' is a ball with center p, $\lambda \in {}^{+}\mathsf{F}$, and ℓ is a line such that $p \in \ell \subseteq P$ and $\lambda < \mathsf{speed}(\ell)$. Then there is a ball $S \subseteq S'$ with center p such that each line ℓ' in plane P slower than λ and intersecting S meets ℓ within S' (i.e. $\emptyset \neq \ell \cap \ell' \subseteq S'$).

<u>*Proof:*</u> Assume P, λ, p, S', ℓ satisfy the assumptions. Let $a, b \in \ell \cap S'$ be such that $\mathsf{Bw}(a, p, b)$. See the left hand side of Figure 13. Let $\ell_a, \ell'_a, \ell_b, \ell'_b$ be lines in P of speed λ such that $\{a\} = \ell_a \cap \ell'_a$ and $\{b\} = \ell_b \cap \ell'_b$. Let S be a ball with center p such that circle $S \cap P$ is inside the parallelogram determined by $\ell_a, \ell'_a, \ell_b, \ell'_b$, i.e. such that $S \cap (\ell_a \cup \ell'_a \cup \ell_b \cup \ell'_b) = \emptyset$. This S has the desired properties, cf. Figure 13. QED (Claim 1)

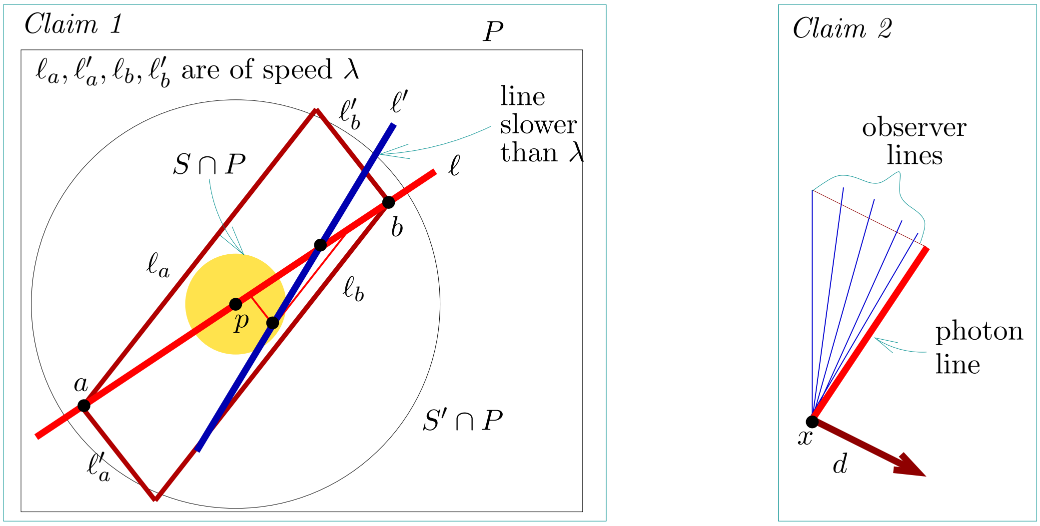


Figure 13: Illustrations for Claims 1 and 2.

<u>*Definition:*</u> Let $\ell \in \mathsf{Lines}$ and $d \in \mathsf{dir}$. We say that ℓ *is in direction* d iff $\mathsf{speed}(\ell) = 0$ or $(\exists p, q \in \ell)(p_t > q_t \ \wedge \ p_s - q_s = d)$.

Now, assume $\mathbf{LocRel}_0^- \setminus \{\mathbf{AxSelf, AxEvent}\}$.

<u>*Claim 2:*</u> Assume $m \in \mathsf{Ob}$, $x \in \mathsf{cd}(m)$ and $d \in \mathsf{dir}$. Then there is exactly one photon line for m in direction d passing through x. Let ℓ be this photon line. Then $0 < \mathsf{speed}(\ell) < \infty$ and any line slower than $\mathsf{speed}(\ell)$ in direction

d passing through x is an observer line for m. See the right hand side of Figure 13.

We omit the easy *proof.*

To prove that **Ax∃Ob*** holds, let $m \in \mathsf{Ob}$, let P be a vertical plane and $p \in \mathsf{cd}(m)$. We need a ball S with center p and a "speed threshold" λ such that each line in plane P slower than λ and intersecting S is an observer line for m. We can assume that $p \in P$. Let $d \in \mathsf{dir}$ be such that each line in P is in direction d or $-d$. Let ℓ_1, ℓ_2 be photon lines for m passing through p in directions $d, -d$, respectively. ℓ_1, ℓ_2 exist by Claim 2. $\mathsf{speed}(\ell_1), \mathsf{speed}(\ell_2) \in {}^+\mathsf{F}$ by Claim 2. Furthermore, $\ell_1, \ell_2 \subseteq P$. Let $\lambda \in {}^+\mathsf{F}$ be such that $\lambda < \mathsf{speed}(\ell_1), \mathsf{speed}(\ell_2)$ and let $S' \subseteq \mathsf{cd}(m)$ be a ball with center p. By Claim 1, there is a ball $S \subseteq S'$ with center p such that any line in P slower than λ and intersecting S meets both ℓ_1 and ℓ_2 within S'. Let such an S be fixed.

Claim 3: Any photon line ℓ for m in P intersecting S is of speed $\geq \lambda$.

Proof: Assume ℓ is a line in P slower than λ and intersecting S. Then ℓ is in direction d or $-d$. Assume, e.g., ℓ is in direction d. See Figure 14. By our choice of S, $\emptyset \neq \ell \cap \ell_1 \subseteq S' \subseteq \mathsf{cd}(m)$. Let $q \in \ell \cap \ell_1$. By Claim 2, there is exactly one photon line in direction d passing through $q \in \mathsf{cd}(m)$. This photon line is ℓ_1. Thus ℓ cannot be a photon line. QED (Claim 3)

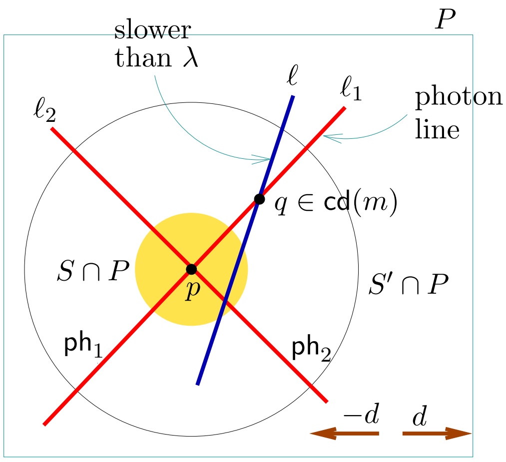


Figure 14: Illustration for the proof of Claim 3.

Claim 4: Any line in plane P slower than λ and intersecting S is an observer line for m.

Proof: The claim follows from Claims 2 and 3. For completeness we include a detailed proof. Let ℓ' be a line in plane P slower than λ and intersecting S. Let $q \in \ell' \cap S \subseteq \mathsf{cd}(m)$. ℓ' is in direction d or $-d$. Assume, e.g., ℓ' is in direction d. Let ℓ be the photon line passing through q in direction d. ℓ exists by Claim 2. Clearly $\ell \subseteq P$. By Claim 3, $\mathsf{speed}(\ell) \geq \lambda$. Thus, $\mathsf{speed}(\ell') < \mathsf{speed}(\ell)$. Hence, ℓ' is an observer line by Claim 2.

QED (Prop.4)

Proposition 5. *Assume* $\mathbf{LocRel_0^-} \setminus \{\mathbf{AxSelf, AxEvent}\}$ *or* $\{\mathbf{AxLine, Ax\exists Ob^*, AxOpen}\}$. *Then for every* $m, k \in \mathsf{Ob}$, *vertical plane* P *and* $p \in \mathsf{Dom}(\mathsf{f}_{mk}) \cap P$ *there is a neighborhood* N *of* p *and a* $\mathcal{P}$*-collineation that agrees with* f_{mk} *on* $N \cap P$.

Proof: Assume the assumptions. Then, by Proposition 4, $\mathbf{Ax\exists Ob^*}$ holds.

Notation: Assume $m, k \in \mathsf{Ob}$. Then for every $a \in \mathsf{f}_{mk}$, a_m denotes the first component of a, while a_k denotes the second component of a, i.e. $a = \langle a_m, a_k \rangle$. Furthermore, if $a_m \in \mathsf{Dom}(\mathsf{f}_{mk})$, then a_k denotes $\mathsf{f}_{mk}(a_m)$ and if $a_k \in \mathsf{Rng}(\mathsf{f}_{mk}) = \mathsf{Dom}(\mathsf{f}_{km})$, then a_m denotes $\mathsf{f}_{km}(a_k)$.

Let $m, k \in \mathsf{Ob}$, a vertical plane P and $p_m \in \mathsf{Dom}(\mathsf{f}_{mk}) \cap P$ be fixed until the proof is complete.

Claim 1: Assume $m, k \in \mathsf{Ob}$ and $a, b \in \mathsf{f}_{mk}$, $a \neq b$. Then (i), (ii) below hold.
(i) $(a_m b_m$ is an observer line for $m) \Leftrightarrow (a_k b_k$ is an observer line for $k)$.
(ii) f_{mk} preserves Coll and $\neg\mathsf{Coll}$ between three points if the line determined by two of the points is an observer line. Formally: Assume $a_m b_m$ is an observer line for m. Then for every $c \in \mathsf{f}_{mk}$,
$\mathsf{Coll}(a_m, b_m, c_m) \Leftrightarrow \mathsf{Coll}(a_k, b_k, c_k)$, or equivalently $c_m \in a_m b_m \Leftrightarrow c_k \in a_k b_k$.
We omit the easy *proof.*

Claim 2: There is a ball $S \subseteq \mathsf{Dom}(\mathsf{f}_{mk})$ with center p_m such that f_{mk} preserves Coll on $S \cap P$.

Proof: A proof can be obtained from the proof of Claim 2 for the proof of Theorem 1 (p.244), in the following way: One uses $\mathbf{Ax\exists Ob^*}$ in place of $\mathbf{Ax\exists Ob}$. QED (Claim 2)

Let $S \subseteq \mathsf{Dom}(\mathsf{f}_{mk})$ be a ball with center p_m such that f_{mk} preserves Coll on $S \cap P$.

Claim 3: f_{mk} preserves $\neg\mathsf{Coll}$ on $S \cap P$.

Proof: A proof can be obtained from the proof of Claim 4 for the case $n = 2$ in the proof of Theorem 1 (p.247), in the following way: One uses **Ax∃Ob*** in place of **Ax∃Ob**. QED (Claim 3)

Recall that for $B \in \mathsf{Planes}$, B^∞ denotes the "line of B at infinity".

f_{mk} preserves Coll and $\neg\mathsf{Coll}$ on $S \cap P$ by the choice of S and by Claim 3. Thus, by the proof of Proposition 2, there is a unique PColl-preserving function $g : (P \cup P^\infty) \longrightarrow \mathsf{P}^n\mathsf{F}$ that agrees with f_{mk} on $S \cap P$. Furthermore, this g is injective. Let g be fixed. By **Ax∃Ob***, there is a neighborhood of p_m and a "speed threshold" λ such that each line in P slower than λ and intersecting the neighborhood is an observer line for m. Let such a λ be fixed.

Claim 4: g agrees with f_{mk} on the set

$$\{ e_m \in \mathsf{Dom}(\mathsf{f}_{mk}) \cap P \ : \ \mathsf{speed}(p_m e_m) < \lambda \}.$$

Proof: Let $e_m \in \mathsf{Dom}(\mathsf{f}_{mk}) \cap P$ be such that $\mathsf{speed}(p_m e_m) < \lambda$. Then by the choice of λ, $p_m e_m$ is an observer line for m. Let $q_m \in S \cap P$ be such that $\neg\mathsf{Coll}(q_m, p_m, e_m)$ and $e_m q_m$ is an observer line for m, too. q_m exists by the choice of λ. See the left hand side of Figure 15. Let $a_m \in e_m p_m \cap S$ and $b_m \in q_m e_m \cap S$ be such that $b_m \neq q_m$ and $a_m \neq p_m$. Hence $p_m a_m \cap q_m b_m = \{e_m\}$. By Claim 1 (ii), $p_k a_k \cap q_k b_k = \{e_k\}$. Since g agrees with f_{mk} on $S \cap P$, it takes p_m, a_m, b_m, q_m to p_k, a_k, b_k, q_k, respectively. Since g preserves PColl, it takes e_m to e_k. QED (Claim 4)

Since $g : (P \cup P^\infty) \longrightarrow \mathsf{P}^n\mathsf{F}$ preserves PColl, there is $Q \in \mathsf{Planes}$ such that $\mathsf{Rng}(g) \subseteq Q \cup Q^\infty$. Let such a Q be fixed.

Claim 5: $\mathsf{Rng}(g) = Q \cup Q^\infty$. Thus g is a bijection between projective planes $P \cup P^\infty$ and $Q \cup Q^\infty$.

Proof: Let $a_m, b_m, c_m \in S \cap P$ be such that $p_m \notin \{a_m, b_m, c_m\}$ and $p_m a_m$, $p_m b_m$, $p_m c_m$ are distinct lines slower than λ. Then by our choice of λ, $p_m a_m$, $p_m b_m$, $p_m c_m$ are observer lines for m. See the right hand side of Figure 15. Since g agrees with f_{mk} on $S \cap P$ and $\mathsf{Rng}(g) \subseteq Q \cup Q^\infty$, $\mathsf{P}p_k a_k, \mathsf{P}p_k b_k, \mathsf{P}p_k c_k \subseteq Q \cup Q^\infty$.

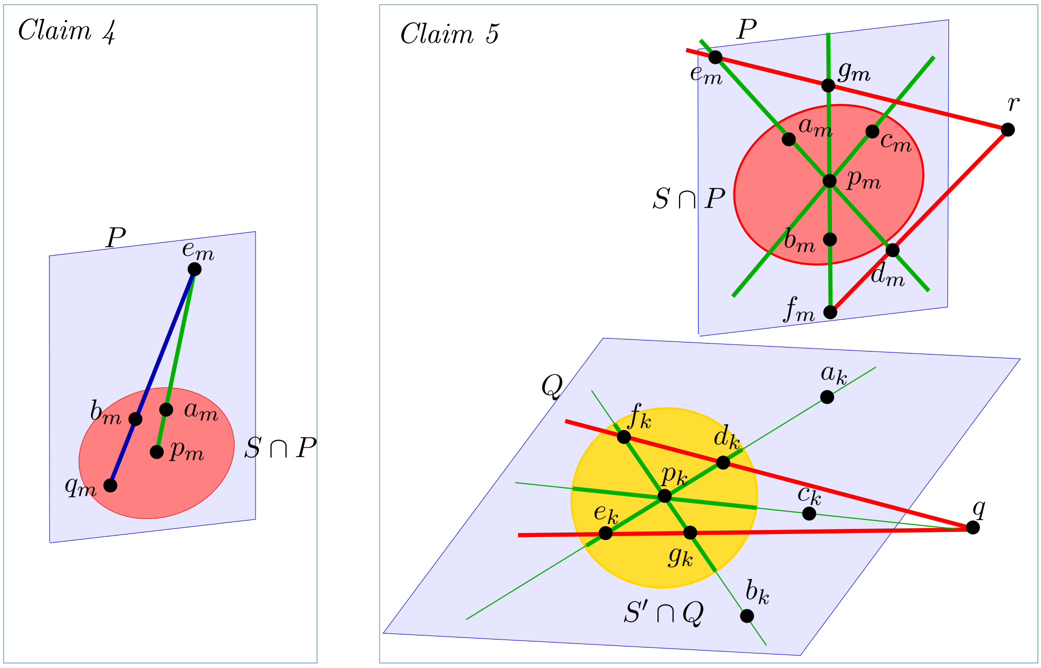


Figure 15: Illustration for the proofs of Claims 4 and 5

To prove $\mathsf{Rng}(g) = Q \cup Q^\infty$, let $q \in Q \cup Q^\infty$. By Claim 3, $\mathsf{P}p_k a_k$, $\mathsf{P}p_k b_k$, $\mathsf{P}p_k c_k$ are distinct. Thus two of these lines do not contain q. We can assume $q \notin \mathsf{P}p_k a_k \cup \mathsf{P}p_k b_k$. Let S' be a ball with center p_k such that $S' \subseteq \mathsf{Dom}(\mathsf{f}_{km}) = \mathsf{Rng}(\mathsf{f}_{mk})$. S' exists by **AxOpen**. Let $d_k, e_k \in p_k a_k \cap S'$ and $f_k, g_k \in p_k b_k \cap S'$ be such that $\mathsf{P}d_k f_k \cap \mathsf{P}e_k g_k = \{q\}$. See the right hand side of Figure 15. By Claim 1, $d_m, e_m \in p_m a_m$ and $f_m, g_m \in p_m b_m$ since $p_m a_m$, $p_m b_m$ are observer lines. d_m, e_m, f_m, g_m are distinct because d_k, e_k, f_k, g_k were such. Let $r \in \mathsf{P}^n\mathsf{F}$ be such that $\{r\} = \mathsf{P}d_m f_m \cap \mathsf{P}e_m g_m$. By Claim 4, g takes d_m, e_m, f_m, g_m to d_k, e_k, f_k, g_k, respectively. Since g preserve PColl, it takes r to q. Thus $q \in \mathsf{Rng}(g)$. QED (Claim 5)

Any PColl preserving bijection between two projective planes can be extended to a $\mathcal{P}$-collineation, cf. e.g. [7, 4.4.11, p.40]. Thus, there is a $\mathcal{P}$-collineation f such that $f \supseteq g$. Now, f_{mk} agrees with such a $\mathcal{P}$-collineation of $S \cap P$.

QED (Prop.5)

Proposition 6. *Assume f is a $\mathcal{P}$-collineation and $a, b \in {}^n\mathsf{F}$ are such that $(\forall p \in [a,b]) f(p) \in {}^n\mathsf{F}$. Then f takes $[a,b]$ onto $[f(a), f(b)]$.*

Outline of proof: Consider the $(n+1)$-dimensional vector space ${}^{n+1}\mathbf{F} := \langle {}^{n+1}\mathsf{F}, \bar{0}, +, \ldots \rangle$ over the field $\mathbf{F}$. Let us introduce an equivalence relation on the set ${}^{n+1}\mathsf{F} \setminus \{\bar{0}\}$ as follows. $u, v \in {}^{n+1}\mathsf{F} \setminus \{\bar{0}\}$ are equivalent iff there exists $\lambda \in \mathsf{F} \setminus \{0\}$ such that $u = \lambda v$. The set of equivalence classes is called the projective space associated with the vector space ${}^{n+1}\mathbf{F}$ and is denoted by $\mathbf{F}\mathsf{P}^n$ according to [7]. We will denote the equivalence class of a vector $\bar{0} \neq v \in {}^{n+1}\mathsf{F}$ by $[v]$. The collinearity relation Pcoll on $\mathbf{F}\mathsf{P}^n$ is defined as follows. $\mathsf{Pcoll}([u],[v],[z])$ iff $\bar{0}, u, v, z$ are coplanar. The following is known from projective geometry.

<u>*Fact 1:*</u> Structures $\mathcal{P} = \langle \mathsf{P}^n\mathsf{F}, \mathsf{PColl} \rangle$ and $\langle \mathbf{F}\mathsf{P}^n, \mathsf{Pcoll} \rangle$ are isomorphic. Moreover, there is a unique isomorphism i between the two structures such that
$(*)$ $\qquad (\forall p \in {}^n\mathsf{F})\; i(p) = [\langle 1, p_1, p_2, \ldots, p_n \rangle]$.

Let $f : \mathbf{F}\mathsf{P}^n \longrightarrow \mathbf{F}\mathsf{P}^n$ be a function. We say that f is induced by a bijective linear transformation iff there is a bijective linear transformation A of the vector space ${}^{n+1}\mathbf{F}$ such that $f([v]) = [Av]$ for all $v \in {}^{n+1}\mathsf{F} \setminus \{\bar{0}\}$. We say that f is induced by a field automorphism iff there is an automorphism ψ of the field $\mathbf{F}$ such that $f([v]) = [\langle \psi(v_1), \psi(v_2), \ldots, \psi(v_{n+1}) \rangle]$ for all $v \in {}^{n+1}\mathsf{F} \setminus \{\bar{0}\}$.

From now on we identify $\mathcal{P} = \langle \mathsf{P}^n\mathsf{F}, \mathsf{PColl} \rangle$ with $\langle \mathbf{F}\mathsf{P}^n, \mathsf{Pcoll} \rangle$ by the unique isomorphism i that satisfies $(*)$ in Fact 1 above, consequently, we treat $\mathsf{P}^n\mathsf{F}$ and $\mathbf{F}\mathsf{P}^n$ as they were identical. Then, $\mathsf{P}^n\mathsf{F} \longrightarrow \mathsf{P}^n\mathsf{F}$ functions induced by bijective linear transformations and the ones induced by automorphisms are $\mathcal{P}$-collineations. The following is known from projective geometry, cf. [7, 6.3, p.60].

<u>*Fact 2:*</u> Any $\mathcal{P}$-collineation is a composition of a $\mathcal{P}$-collineation induced by a bijective linear transformation and a $\mathcal{P}$-collineation induced by a field automorphism.

<u>*Claim 1:*</u> Assume f is a $\mathcal{P}$-collineation induced by a field automorphism. Then f takes ${}^n\mathsf{F}$ onto ${}^n\mathsf{F}$ and $(\forall a, b \in {}^n\mathsf{F})$ (f takes $[a,b]$ onto $[f(a), f(b)]$).

Proof: Since $\mathbf{F}$ is a Euclidean field, any automorphism ψ of $\mathbf{F}$ preserves $<$, i.e. $x < y \Rightarrow \psi(x) < \psi(y)$. By this fact, one can easily check that the statement holds. QED (Claim 1)

Claim 2: Assume f is a $\mathcal{P}$-collineation induced by a bijective linear transformation and $a, b \in {}^n\mathsf{F}$ are such that $(\forall p \in [a,b]) f(p) \in {}^n\mathsf{F}$. Then f takes $[a,b]$ onto $[f(a), f(b)]$.

We omit the easy *proof.*

Now, the proposition follows from Fact 2 and Claims 1 and 2.

QED (Prop.6)

Lemma 2. *Assume* $\mathbf{LocRel_0^-} \setminus \{\mathbf{AxSelf, AxEvent}\}$. *Assume* $m, k \in \mathsf{Ob}$. *Then for every* $\ell \in \mathsf{Lines}$ *and* $p \in \ell \cap \mathsf{Dom}(\mathsf{f}_{mk})$ *there is* $q \in \ell \cap \mathsf{Dom}(\mathsf{f}_{mk})$ *such that* $p \neq q$ *and* f_{mk} *takes* $[p,q]$ *onto* $[\mathsf{f}_{mk}(p), \mathsf{f}_{mk}(q)]$.

Proof: The lemma follows from Propositions 5 and 6. QED (Lemma 2)

Lemma 3. *Assume* $\mathbf{LocRel_0^-}$. *Then the conclusion of Lemma 1 holds, i.e. if there is an observer trace in a plane passing through a point, then there is a photon trace in the plane passing through the point.*

Formally: Assume $m, k \in \mathsf{Ob}$ *and* $p \in \mathsf{tr}_m(k) \subseteq P \in \mathsf{Planes}$. *Then there is a* $\mathsf{ph} \in \mathsf{Ph}$ *such that* $p \in \mathsf{tr}_m(\mathsf{ph}) \subseteq P$.

Proof: Assume $\mathbf{LocRel_0^-}$. Let $m, k \in \mathsf{Ob}$, $p \in \mathsf{tr}_m(k)$ and $P \in \mathsf{Planes}$ be such that $\mathsf{tr}_m(k) \subseteq P$. Then f_{mk} is an injective partial function by Proposition 3. We can assume $p \in \mathsf{Dom}(\mathsf{f}_{mk})$ since, by **AxEvent**, k has a brother h such that $p \in \mathsf{Dom}(\mathsf{f}_{mh})$.

Let $q \in P \cap \mathsf{Dom}(\mathsf{f}_{mk})$ be such that $q \notin \mathsf{tr}_m(k)$ and f_{mk} takes $[p,q]$ onto $[\mathsf{f}_{mk}(p), \mathsf{f}_{mk}(q)]$. Such a q exists by Lemma 2. See Figure 16. Let p' and q' be the f_{mk} images of p and q, respectively. Then $p' \in \mathsf{tr}_k(k)$ and $q' \notin \mathsf{tr}_k(k)$. Let P' be the plane determined by $\mathsf{tr}_k(k)$ and q'. P' is vertical by **AxSelf** (and **AxLine**, **AxOpen**).

Let S be a ball with center p' and let g be a $\mathcal{P}$-collineation such that g agrees with f_{km} on $S \cap P'$. They exist by Proposition 5. Let $\mathsf{ph} \in \mathsf{Ph}$ be such that $p' \in \mathsf{tr}_k(\mathsf{ph}) \subseteq P'$. Such a ph exists by $\mathbf{Ax\exists Ph}$. Let $a' \in S \cap \mathsf{tr}_k(\mathsf{ph})$, $b' \in S \cap [p', q']$ and $c' \in \mathsf{tr}_k(k) \cap S$ be such that $p' \notin \{a', b', c'\}$. Such a', c' exist by **AxLine** and by $S \cap P' \subseteq \mathsf{Dom}(\mathsf{f}_{mk}) \subseteq \mathsf{cd}(m)$. Let a, b, c be the

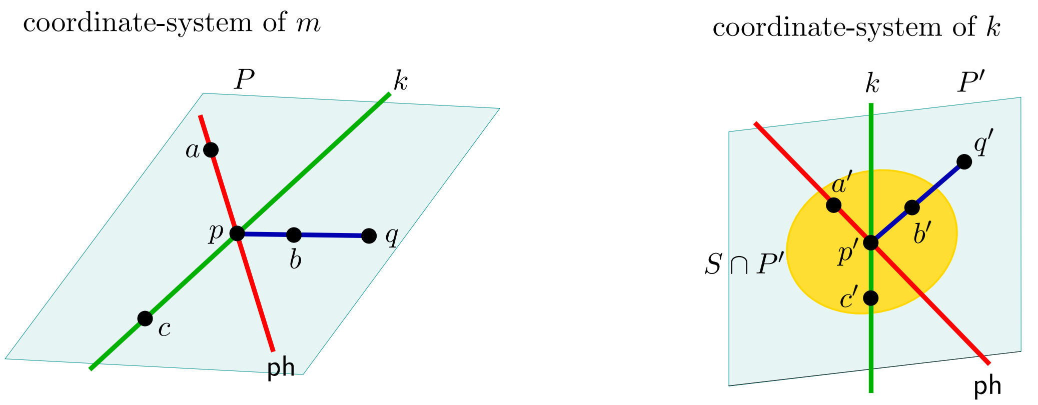


Figure 16: Illustration for the proof of Lemma 3.

f_{km} images of a', b', c', respectively. Then $p \notin \{a, b, c\}$, $c \in \mathsf{tr}_m(k)$, $b \in [p, q]$ and $a, p \in \mathsf{tr}_m(\mathsf{ph})$. Since $\mathcal{P}$-collineation g agrees with f_{km} on $S \cap P$, g takes p', a', b', c' to p, a, b, c, respectively. But then p, a, b, c are coplanar since p', a', b', c' are such. Thus $a \in P$. But then, by **AxLine** and $a, p \in \mathsf{tr}_m(\mathsf{ph})$, we have $\mathsf{tr}_m(\mathsf{ph}) \subseteq P$. QED (Lemma 3)

Lemma 4. *Assume* $\mathbf{LocRel}_0^-$. *Assume* $m \in \mathsf{Ob}$, $p, a, b \in {}^n\mathsf{F}$ *are non-collinear points,* $p \in \mathsf{cd}(m)$*, the plane that contains* p, a, b *is vertical,* pa *is an observer line for* m *and* $(\nexists q \in [a, b])(pq$ *is a photon line for* $m)$.

Then pb *is an observer line for* m.

Proof: Assume $\mathbf{LocRel}_0^-$. Assume m, p, a, b satisfy the assumptions. Then $pa \cap \mathsf{cd}(m) = \mathsf{tr}_m(k)$, for some $k \in \mathsf{Ob}$. Let such a k be fixed. We are in the coordinate-system of m. Let P be the vertical plane that contains p, a, b. Let S be a ball with center p such that f_{mk} agrees with a $\mathcal{P}$-collineation on $S \cap P$. Such an S exists by Proposition 5.

Let $c, d \in S \cap P$ be such that $\mathsf{Bw}(p, c, a)$ and $\mathsf{Bw}(p, d, b)$. See Figure 17. Then

$$(\nexists q \in [c, d])(pq \text{ is a photon line for } m). \tag{3}$$

By **AxLine** and $S \cap P \subseteq \mathsf{Dom}(\mathsf{f}_{mk}) \subseteq \mathsf{cd}(m)$, $p, c \in \mathsf{tr}_m(k)$.

Let us switch over to the coordinate-system of k. Let p', c', d' be, respectively, the f_{mk} images of p, c, d. Since f_{mk} agrees with a $\mathcal{P}$-collineation on

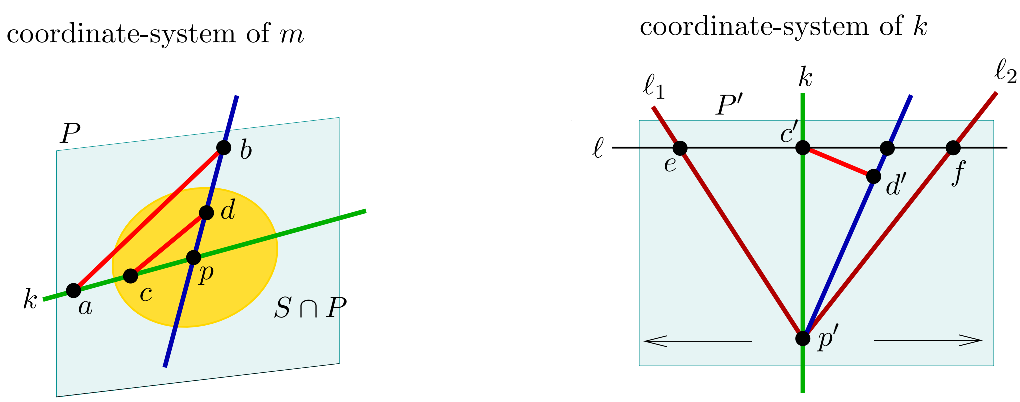


Figure 17: Illustration for the proof of Lemma 4.

$S \cap P$, p', c', d' are non collinear. Furthermore, f_{mk} takes $[c, d]$ onto $[c', d']$ by Proposition 6. Therefore, by (3),

$$(\nexists q \in [c', d'])(p'q \text{ is a photon line for } k). \tag{4}$$

Let P' be the plane that contains p', d', c'. $p', c' \in \mathsf{tr}_k(k)$ by $p, c \in \mathsf{tr}_m(k)$. By **AxSelf**, $\mathsf{speed}(p'c') = 0$. Thus P' is a vertical plane. Let $\ell \in \mathsf{Lines}$ be such that $\mathsf{speed}(\ell) = \infty$ and $c' \in \ell \subseteq P'$. By **Ax∃Ph**, **AxP1**, **AxFin**, there are exactly two photon-lines for k in plane P' passing through p, cf. Claim 2 on p.257. Let ℓ_1, ℓ_2 be these photon lines. $\mathsf{speed}(\ell_1), \mathsf{speed}(\ell_2) \in {}^+\mathsf{F}$ by **AxFin**. Let $e \in \ell \cap \ell_1$ and $f \in \ell \cap \ell_2$. Then $\mathsf{Bw}(e, c', f)$ by **AxP1**. By (4), neither ℓ_1 nor ℓ_2 intersects $[c', d']$. But then, by **AxOb**, $p'd'$ is an observer line for k, cf. Claim 2 on p.257. Thus, $pd = pb$ is an observer line for m, too.

QED (Lemma 4)

Proof of Theorem 5: Assume $n > 2$ and $\mathbf{LocRel}_0^-$. We will show that if there is an FTL observer, then there is a photon with infinite speed. This will contradict **AxFin**.

Assume there is an FTL observer, i.e. there are $m, k \in \mathsf{Ob}$, $\mathsf{ph} \in \mathsf{Ph}$, $d \in \mathsf{dir}$ and $p \in {}^n\mathsf{F}$ such that k and ph move in direction d as seen by m, $p \in \mathsf{tr}_m(k) \cap \mathsf{tr}_m(\mathsf{ph})$ and $\mathsf{speed}_m(k) > \mathsf{speed}_m(\mathsf{ph})$. Let such m, k, ph, d, p be fixed. See Figure 18.

By **AxLine** and **AxOpen**, there are unique $\ell_k, \ell_{\mathsf{ph}} \in \mathsf{Lines}$ such that $\mathsf{tr}_m(k) = \ell_k \cap \mathsf{cd}(m)$ and $\mathsf{tr}_m(\mathsf{ph}) = \ell_{\mathsf{ph}} \cap \mathsf{cd}(m)$. Let such ℓ_k, ℓ_{ph} be fixed. Let

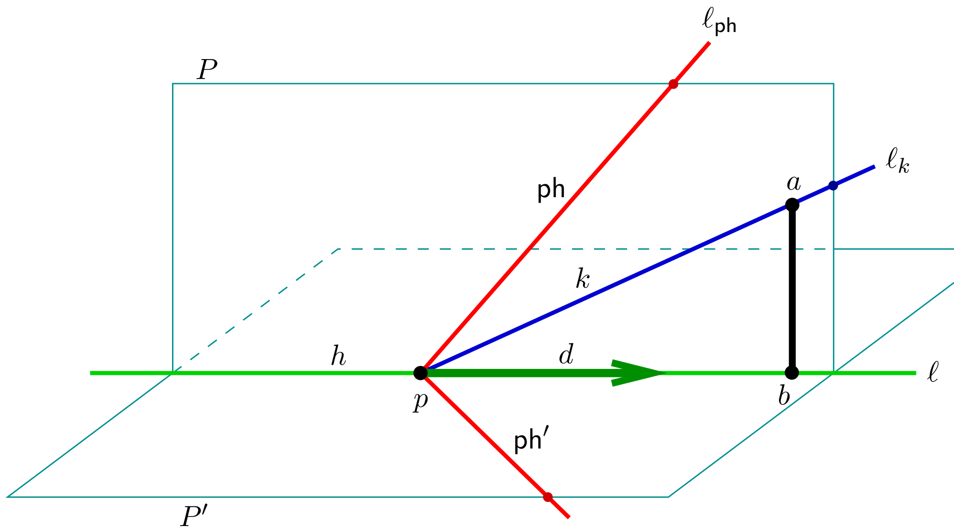


Figure 18: Illustration for the proof of Theorem 5.

P be the plane that contains ℓ_k and ℓ_{ph}. P is vertical since k and ph move in the same direction. Let $\ell \in \mathsf{Lines}$ be such that $p \in \ell \subseteq P$ and $\mathsf{speed}(\ell) = \infty$. We will show that ℓ is an observer line for m. Since ℓ_k is an observer line, we can assume that $\ell \neq \ell_k$. Since k is an FTL observer (w.r.t. m), $\mathsf{speed}(\ell_k) \neq 0$. Let $a \in \ell_k$ and $b \in \ell$ be such that $p \notin \{a, b\}$ and $\mathsf{speed}(ab) = 0$. Since k moves faster than ph (as seen by m), $\ell_{\mathsf{ph}} \cap [a, b] = \emptyset$. Hence, by **AxP1**, for any photon line ℓ' for m that passes through p, we have $\ell' \cap [a, b] = \emptyset$. Thus, by Lemma 4, $\ell = pb$ is an observer line for m.

Let $h \in \mathsf{Ob}$ be such that $\mathsf{tr}_m(h) = \ell \cap \mathsf{cd}(m)$. Let P' be a plane such that $\ell \subseteq P'$ and any line contained in P' has infinite speed. There is such a plane by $n > 2$. Clearly, $p \in \mathsf{tr}_k(h) \subseteq P'$. By Lemma 3, there is a photon ph' such that $p \in \mathsf{tr}_m(\mathsf{ph}') \subseteq P'$. Then, by **AxLine** and **AxOpen**, $\mathsf{speed}_m(\mathsf{ph}') = \infty$ for this ph'. But this contradicts **AxFin**.

QED (Theorem 5)

4 Bibliography

[1] H. Andréka, J. X. Madarász, and I. Németi. Logical analysis of special relativity. In J. Gerbrandy, M. Marx, M. de Rijke, and Y. Venema, editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*, http://www.wins.uva.nl/˜j50/cdrom, 1999. Vossiuspers, Amsterdam University Press. CD-ROM, ISBN: 90 5629 104 1, 23pp.

[2] H. Andréka, J. X. Madarász, and I. Németi. On the logical structure of relativity theories. Research report, Alfréd Rényi Institute of Mathematics, Hungar. Acad. Sci., Budapest, July 5, 2002. With contributions from A. Andai, G. Sági, I. Sain and Cs. Tőke. http://www.math-inst.hu/pub/algebraic-logic/Contents.html. 1312 pp.

[3] H. Andréka, J. X. Madarász, and I. Németi. Logical axiomatizations of space-time. Samples from the literature. In: *Non-Euclidean geometries (János Bolyai Memorial Volume).* Kluwer Scientific Publishers, to appear. http://www.math-inst.hu/pub/algebraic-logic/Contents.html

[4] H. Andréka, J. X. Madarász, and I. Németi. Logical analysis of relativity theories. In the present volume.

[5] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 65:217–274, 1998.

[6] H. Busemann. Time-like spaces. *Dissertationes Math. (Rozprawy Math.)*, 53, 1967.

[7] B. Csikós. Projective geometry. Lecture notes. http://www.cs.elte.hu/geometry/csikos/proj/proj.html.

[8] M. Friedman. *Foundations of Space-Time Theories. Relativistic Physics and Philosophy of Science.* Princeton University Press, 1983.

[9] R. Goldblatt. *Orthogonality and Spacetime Geometry.* Springer-Verlag, 1987.

[10] A. K. Guts. Axiomatic relativity theory. *Russian Math. Survey*, 37(2):41–89, 1982.

[11] R. W. Latzer. Nondirected light signals and the structure of time. *Synthese*, 24:236–280, 1972.

[12] J. A. Lester. Distance preserving transformations. In F. Buekenhout, editor, *Handbook of Incidence Geometry*, pages 921–944. Elseview Science B. V., 1995.

[13] J. X. Madarász. *Logic and Relativity (in the light of definability theory).* PhD thesis, ELTE, Budapest, http://www.math-inst.hu/pub/algebraic-logic/Contents.html., 2002.

[14] J. X. Madarász and Cs. Tőke. Partial relativity theories. Research report, Alfréd Rényi Institute of Mathematics, Hungar. Acad. Sci., Budapest, July, 2003.

[15] D. B. Malament. The class of continuous timelike curves determines the topology of space-time. *J. Mathematical Phys.* 18 (1977), 1399-1404.

[16] M. Marx, L. Pólos, and M. Masuch, editors. *Arrow logic and multi-modal logic.* CSLI Publications and FOLLI, Stanford, 1996.

[17] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation.* W. H. Freeman, San Francisco, 1973.

[18] I. Németi, A fine-structure analysis of first-order logic. In [16], pp.221-247.

[19] H. Reichenbach. *Axiomatization of the theory of relativity.* University of California Press, Berkeley, 1969. Translated by M. Reichenbach. Original German edition published in 1924.

[20] W. Rindler. *Essential relativity. Special, general and cosmological.* Springer-Verlag, Berlin, 1969, 1977.

[21] G. Székely. Personal communication. 2003 November.

[22] J. Väänänen. Second-order logic and foundations of mathematics. *The Bulletin of Symbolic Logic*, 7(4):504–520, 2001.

[23] J. Woleński. First-order logic: (philosophical) pro and contra. In the present volume.

Constructive Predicate Logic and Constructive Modal Logic. Formal Duality versus Semantical Duality

S.P. Odintsov
odintsov@math.nsc.ru
Institute of Mathematics, SD RAS, Novosibirsk
H. Wansing
Heinrich.Wansing@mailbox.tu-dresden.de
Institute of Philosophy, Dresden University of Technology

1 Introduction

The relationships between modal propositional logic and predicate logic are tight and diverse. It is well-known that modal propositional logic can be seen as a fragment of second-order predicate logic [14], and that classical first-order logic may be viewed as an extension of a 'minimal' predicate logic derived from the multi-modal version of the smallest normal modal logic K [2].

In this paper, we consider constructive modal and first-order logic with strong negation $\sim$. We say that a modal logic L satisfies *semantical duality* if the modal operators $\Box$ ("it is necessary that") and $\Diamond$ ("it is possible that") are interpreted with respect to the same accessibility relation R. The logic L satisfies *formal* (*syntactical*) *duality* if $\sim \Box A \leftrightarrow \Diamond \sim A$ and $\sim \Diamond A \leftrightarrow \Box \sim A$ are provable.

In the case of modal logics defined over *classical* logic, it is quite natural to assume that $\Box$ and $\Diamond$ are dependent both semantically and formally, that

The work of S.P. Odintsov was supported by Alexander von Humboldt-Stiftung.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 269–286

is, mutually definable using Boolean negation. This is not the case, if we consider *intuitionistic* modal logics. In intuitionistic predicate logic (QInt), the universal and the existential quantifier are not interdefinable by means of intuitionistic negation $\neg$. Since $\Box$ and $\Diamond$ are restricted universal, respectively existential quantifiers over states (possible words), a natural consideration is that $\Box$ and $\Diamond$ are independent in intuitionistic modal logic. M. Božič and K. Došen [3] tried to combine semantical and formal duality with intuitionistic propositional logic as the underlying non-modal base logic. However, the resulting system lacks the Disjunction Property (DP) and therefore can hardly be considered as an intuitionistic modal logic.

A.K. Simpson [12] supposes that a genuine intuitionistic propositional modal logic (IML) should satisfy the following properties:

1. IML is conservative over intuitionistic propositional logic (IPL);
2. IML contains all substitution instances of theorems of IPL and is closed under modus ponens;
3. $IML + A \vee \neg A$ yields a system of classical modal logic;
4. IML has the DP (if $\vdash A \vee B$, then $\vdash A$ or $\vdash B$);
5. $\Box$ and $\Diamond$ are independent in IML (as well as $\vee, \wedge, \rightarrow$, and $\neg$);
6. There exists an intuitionistically comprehensible explanation of the meaning of the modalities, relative to which IML is sound and complete.

Whereas Simpson rejects formal duality, he takes it that semantical duality should hold. But there are also authors who reject both forms of duality in IML. V. Sotirov [13] defined the system $\mathsf{IK}_0(\Box\Diamond)$ using three accessibility relations: one for interpreting intuitionistic implication, one for $\Box$, and another one for $\Diamond$. IML can be considered as a result of fibring classical modal logic and intuitionistic logic. In developing the idea of fibring, D. Gabbay [7] also prefers to consider $\Box$ and $\Diamond$ as semantically independent, and Sotirov's approach to IML is also taken as basic in [8]. But even in the absence of intuitionistic implication, in the context of an (absolutely) positive modal logic, it has been pointed out by M. Dunn [4] that it is "instructive to consider frames with two accessibility relations $R^{\Box}$ and $R^{\Diamond}$" so that the relation between the modal operators required to obtain semantical duality must be enforced and made explicit by suitable interaction rules.

Fuhrmann [6] defined modal relevance logics satisfying both formal and semantical duality, where negation is interpreted using the 'Routley star'. This negation is different from Nelson's strong negation.

The justification for Simpson's Condition 5 is not given, if we pass from intuitionistic negation to Nelson's negation. In Nelson's constructive logic with strong negation, the quantifiers are duals, and the De Morgan Laws hold:

$$\sim \forall x A(x) \leftrightarrow \exists x \sim A(x), \quad \sim \exists x A(x) \leftrightarrow \forall x \sim A(x),$$

$$\sim (A \vee B) \leftrightarrow (\sim A \wedge \sim B), \quad \sim (A \wedge B) \leftrightarrow (\sim A \vee \sim B).$$

In fact, formal duality between $\Box$ and $\Diamond$ was the main motivation for S. Akama [1] to introduce a modal extension of Nelson's three-valued system $\mathsf{N3}$. In [15] a modal epistemic extension of Nelson's four-valued logic $\mathsf{N4}$ is defined, in which formal duality holds but semantical duality fails. On the other hand, in [10], we have found a natural example of a modal extension of Nelson's paraconsistent logic $\mathsf{N4}$, where the modalities are semantically but not formally dual.

These systematical and historical observations motivate the desire to study the interrelations between modalities and strong negation. We shall emphasize that semantical and formal duality are independent of each other, by defining four axiomatic systems of constructive modal logic with strong negation. First of all, we shall consider positive constructive logic with and without semantical duality, systems PK and PK^d. Next, we shall add the strong negation axioms for the non-modal connectives to obtain the modal logics NK^- and NK^{d-}. To these systems we then add the duality axioms for $\Box$ and $\Diamond$, yielding the logics NK and NK^d. In each case we prove soundness and completeness. For the logics without semantical duality we use tri-relational models $\langle W, \leq, R_\Box, R_\Diamond, v^+, v^- \rangle$, for the remaining logics we use bi-relational models $\langle W, \leq, R, v^+, v^- \rangle$.

In a third step, we shall introduce certain extensions of NK^-, NK^{d-}, and NK^d. These extensions are motivated by considering four translations from the modal language into the language of $\mathsf{QN4}$, the first-order extension of Nelson's four-valued constructive logic with strong negation, see [10]. We shall observe that the translation for NK^d is a *faithful* embedding into $\mathsf{QN4}$, returning the system FSK^d and that in the remaining cases, the translations are embeddings. We do not introduce an extension of NK and conjecture that the corresponding translation faithfully embedds this logic into $\mathsf{QN4}$. In spite of their transparent axiomatizations, the possible worlds semantics for the logics NK^{-} and NK^{d-}, which lack formal duality, is unsatisfactory, because the falsification clauses for formulas $\Diamond A$ and $\Box A$ treat them as atoms. The translation-driven extensions FSK^{-} and FSK^{d-} yield perhaps more natural examples of constructive modal logics with strong negation that fail to satisfy formal duality. Now the falsification clauses for formulas $\Diamond A$ and $\Box A$ are recursive and the clause for $\Diamond A$ refers to an accessibility relation $R^{\sim}$, which means that *impossibility* is treated as an independent modal operator. In [10], the multi-modal version of FSK^{d} is referred to as $\mathcal{CALC}^{\mathsf{N4}d}$.

2 Positive constructive modal logic

Define the logic PK in the positive modal language $\mathcal{L}^{+} := \langle \vee, \wedge, \to, \Box, \Diamond \rangle$ with the set of propositional variables $Prop$ via the axioms:

1. Axioms of intuitionistic positive logic
2. $\Box A \wedge \Box B \to \Box(A \wedge B)$
3. $\Box(A \to A)$
4. $\Diamond(A \vee B) \to \Diamond A \vee \Diamond B$

and the rules

$$\text{MP.}\ \frac{A \quad A \to B}{B} \qquad \text{R}_{\Box}.\ \frac{A \to B}{\Box A \to \Box B} \qquad \text{R}_{\Diamond}.\ \frac{A \to B}{\Diamond A \to \Diamond B}$$

By an *extension of* PK we mean a set of formulas in the language $\mathcal{L}^{+}$ containing all PK-theorems and being closed under the three rules of PK. In this way, a logic is identified with the set of its theorems and $A \in \mathsf{PK}$ means that A is a PK-theorem. On the other hand, the relation $\Gamma \vdash_{\mathsf{PK}} A$, where Γ is a set of formulae and A a formula, holds iff A can be obtained from PK-theorems and elements of Γ with the help of *modus ponens* only. We put $Cl(\Gamma) := \{A \mid \Gamma \vdash_{\mathsf{PK}} A\}$. Corresponding definitions are assumed for all other logics considered below. It is clear that any PK-extension is closed under the replacement rule

$$\frac{A \leftrightarrow B}{C(A) \leftrightarrow C(B)},$$

i.e., provable equivalence has the congruence properties in all PK-extensions.

A PK-frame is a tuple $\mathcal{W} = \langle W, \leq, R_{\Box}, R_{\Diamond} \rangle$, where $\leq, R_{\Box}, R_{\Diamond}$ are binary relations on the set W such that:

1. $\leq$ is reflexive and transitive (i.e., is a quasi-order).
2. $\leq^{-1} \circ R_\Diamond \subseteq R_\Diamond \circ \leq^{-1}$.

For a quasi-order $\langle W, \leq \rangle$, denote by $\langle W, \leq \rangle^+$ the set of its cones (the set of all $X \subseteq W$ such that if $u \in X$ and $u \leq w$, then $w \in X$). A PK-model $\mathcal{M} = \langle W, \leq, R_\Box, R_\Diamond, v \rangle$ is a PK-frame augmented with a valuation function $v : Prop \to \langle W, \leq \rangle^+$.

The forcing relation $\mathcal{M}, t \models A$, where $t \in W$, is defined for propositional variables and positive connectives as in intuitionistic logic and for modalities as follows:

$$\mathcal{M}, t \models \Box A \quad \text{iff} \quad \forall u \geq t \forall v (u R_\Box v \text{ implies } \mathcal{M}, v \models A)$$
$$\mathcal{M}, t \models \Diamond A \quad \text{iff} \quad \exists u (t R_\Diamond u \text{ and } \mathcal{M}, u \models A)$$

A discussion of these forcing conditions can be found in [12]. The notions of validity on a model, $\mathcal{M} \models A$, on a frame, $\mathcal{W} \models A$, and of PK-validity, $\models_{\mathsf{PK}}$, are defined in the usual way.

Lemma 1. *(Persistence) For any* PK*-model $\mathcal{M} = \langle W, \leq, R_\Box, R_\Diamond, v \rangle$, $s, t \in W$, and formula A, if $s \leq t$, then $\mathcal{M}, s \models A$ implies $\mathcal{M}, t \models A$.*

Theorem 1. *For any A, $A \in$ PK iff $\models_{\mathsf{PK}} A$.*

The proof can be obtained as a combination of proofs of Theorems 1 and 4 from [3].

The logic PK^d is obtained from PK by adding the following axioms stating an interaction between $\Box$ and $\Diamond$:[1]

5. $\Diamond(A \to B) \to (\Box A \to \Diamond B)$
6. $(\Diamond(A \to A) \to \Box B) \to \Box B$

A PK^d-frame is a tuple $\mathcal{W} = \langle W, \leq, R \rangle$, where $\leq$ and R are binary relations on the set W such that:

[1] PK^d differs from Dunn's [4] positive modal logic $\mathbf{K}_+$. The intention of M. Dunn was to construct an absolutely positive modal logic; therefore he deleted from the language not only negation, but also implication, which can be considered as a negative operator wrt the first argument. The interaction between $\Box$ and $\Diamond$ in $\mathbf{K}_+$ is captured by the deducibility statements $\Diamond A \wedge \Box B \vdash \Diamond(A \wedge B)$ and $\Box(A \vee B) \vdash \Box A \vee \Diamond B$. Note that if we replace $\vdash$ by $\to$ in these deducibility statements, we obtain $(\Diamond A \wedge \Box B) \to \Diamond(A \wedge B) \in \mathsf{PK}^d$ and $\Box(A \vee B) \to (\Box A \vee \Diamond B) \notin \mathsf{PK}^d$.

1. $\leq$ is reflexive and transitive.
2. $\leq^{-1} \circ R \subseteq R \circ \leq^{-1}$.

A PK^d-model $\mathcal{M} = \langle W, \leq, R, v\rangle$ is a PK^d-frame together with a valuation function $v : Prop \to \langle W, \leq\rangle^+$. The forcing relation is defined as above, but we use the same relation R to define the forcing conditions of $\Box A$ and of $\Diamond A$. In the usual way we define the notions of validity on a model and on a frame and of PK^d-validity.

Lemma 2. *(Persistence) For any PK^d-model $\mathcal{M} = \langle W, \leq, R, v\rangle$, $s, t \in W$, and a formula A, if $s \leq t$, then $\mathcal{M}, s \models A$ implies $\mathcal{M}, t \models A$.*

Theorem 2. *For any A, $A \in \mathsf{PK}^d$ iff $\models_{\mathsf{PK}^d} A$.*

The proof is a modification of the completeness proof for the intuitionistic modal logic IK (alias Fischer Servi logic FS), see [12, pp. 52, 53]). PK^d is a proper sublogic of the positive fragment of IK. Below we give this proof, because it seems that PK^d has not been considered in the literature yet, and because the proof reveals how the interrelations between the modalities arise expressed by Axioms 5 and 6. (We shall omit the index in '$\vdash_{\mathsf{PK}^d}$'.)

Proof (of Theorem 2). The soundness of PK^d wrt to the considered semantics can be verified directly. For the completeness proof we shall use the canonical model method. A set of formulas Γ is called PK^d-*prime* if it contains all PK^d-theorems, is closed under MP and possesses the disjunction property in the form (if $\Gamma \vdash A \vee B$, then $\Gamma \vdash A$ or $\Gamma \vdash B$). For two sets of formulas Γ and Δ, the notation $\Gamma \vdash \Delta$ means that

$$A_1 \wedge \ldots \wedge A_n \vdash B_1 \vee \ldots \vee B_m$$

for some formulas $A_1, \ldots, A_n \in \Gamma$ and $B_1, \ldots, B_m \in \Delta$. As usual, we prove an extension lemma.

Lemma 3. *If $\Gamma \nvdash \Delta$, then there exists a PK^d-prime set Γ' such that $\Gamma \subseteq \Gamma'$ and $\Gamma' \nvdash \Delta$.*

We define the canonical PK^d-model

$$\mathcal{M}^c = \langle W^c, \subseteq, R^c, v^c\rangle$$

as follows. W^c is the family of all PK^d-prime sets and $\subseteq$ is set-inclusion. The relation $\Gamma R^c \Delta$ holds iff $\Gamma_\Box \subseteq \Delta$ and $\Delta^\Diamond \subseteq \Gamma$, where

$$\Gamma_\Box := \{A \mid \Box A \in \Gamma\} \text{ and } \Delta^\Diamond := \{\Diamond A \mid A \in \Delta\}.$$

Finally, $v^c(p) = \{\Gamma \in W^c \mid p \in \Gamma\}$. Note that the relation $\Delta^\Diamond \subseteq \Gamma$ is equivalent to $\Delta \not\vdash \{A \mid \Diamond A \notin \Gamma\}$.

Lemma 4. *$\mathcal{M}^c$ is a PK^d-model.*

Proof. It is enough to check that $\subseteq^{-1} \circ R^c \subseteq R^c \circ \subseteq^{-1}$. Let $\Gamma_0, \Gamma, \Delta \in W^c$ be such that $\Gamma \subseteq \Gamma_0$, $\Gamma_\Box \subseteq \Delta$, and $\Delta^\Diamond \subseteq \Gamma$. We have to construct a set $\Delta_0 \in W^c$ such that $\Gamma_0 R^c \Delta_0 \supseteq \Delta$. Let us consider the set $\Delta_0' := \Delta \cup (\Gamma_0)_\Box$. We claim that

$$\Delta_0' \not\vdash \{A \mid \Diamond A \notin \Gamma_0\}.$$

Indeed, assume that for some formulas $\Box A_1, \ldots, \Box A_n \in \Gamma_0$ and $\Diamond B_1, \ldots, \Diamond B_m \notin \Gamma_0$, we have

$$\Delta \cup \{A_1 \wedge \ldots \wedge A_n\} \vdash B_1 \vee \ldots \vee B_m.$$

In this case, for some $C \in \Delta$,

$$C \to ((A_1 \wedge \ldots \wedge A_n) \to (B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d,$$
$$\Diamond C \to \Diamond((A_1 \wedge \ldots \wedge A_n) \to (B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d \text{ by Rule R}_\Diamond,$$
$$\Diamond C \to (\Box(A_1 \wedge \ldots \wedge A_n) \to \Diamond(B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d \text{ by Axiom 5.}$$

Since $\Delta^\Diamond \subseteq \Gamma$, we have $\Diamond C \in \Gamma \subseteq \Gamma_0$, whence

$$\Box(A_1 \wedge \ldots \wedge A_n) \to \Diamond(B_1 \vee \ldots \vee B_m) \in \Gamma_0,$$
$$(\Box A_1 \wedge \ldots \wedge \Box A_n) \to (\Diamond B_1 \vee \ldots \vee \Diamond B_m) \in \Gamma_0 \text{ by Axioms 2 and 4.}$$

Since $\Box A_1, \ldots, \Box A_n \in \Gamma_0$, also $\Diamond B_1 \vee \ldots \vee \Diamond B_m \in \Gamma_0$, and $\Diamond B_i \in \Gamma_0$ for some i in view of the disjunction property of Γ_0, a contradiction.

We have thus proved $\Delta_0' \not\vdash \{A \mid \Diamond A \notin \Gamma_0\}$. By the previous lemma, there exists a PK^d-prime extension Δ_0 of Δ_0' with $\Delta_0 \not\vdash \{A \mid \Diamond A \notin \Gamma_0\}$. The latter means that $\Delta_0^\Diamond \subseteq \Gamma_0$. Moreover, $(\Gamma_0)_\Box \subseteq \Delta_0$ by definition of Δ_0'. Thus, Δ_0 is the desired element of W^c. ■

Now we prove the canonical model lemma.

Lemma 5. *For any* $\Gamma \in W^c$ *and formula* A,

$$\mathcal{M}^c, \Gamma \models A \textit{ iff } A \in \Gamma.$$

The *proof* is by induction on the structure of formula and we consider only the cases of modal operators.

$\Box$. By definition and induction hypotheses we have the equivalence

$$\mathcal{M}^c, \Gamma \models \Box A \;\Leftrightarrow\; \forall \Gamma_0 \supseteq \Gamma \; \forall \Delta (\Delta^\Diamond \subseteq \Gamma_0 \text{ and } (\Gamma_0)_\Box \subseteq \Delta \text{ imply } A \in \Delta).$$

We have to prove that the right-hand side of this equivalence is equivalent to $\Box A \in \Gamma$.

If $\Box A \in \Gamma$, then for any $\Gamma_0 \supseteq \Gamma$ and Δ with $(\Gamma_0)_\Box \subseteq \Delta$ we shall have $A \in \Delta$.

Conversely, assume $\Box A \notin \Gamma$ and construct $\Gamma_0 \supseteq \Gamma$ and Δ such that $\Gamma_0 R^c \Delta$ and $A \notin \Delta$.

Lemma 6. *The following facts are true for any* $\Gamma \in W^c$ *and formula* A.

1. *If* $\Gamma_\Box \vdash A$, *then* $A \in \Gamma_\Box$.
2. $\Gamma_\Box = (Cl(\Gamma \cup \{\Diamond(B \to B)\}))_\Box$.

Proof. 1. Assume $\Box A \notin \Gamma$ and $\Gamma_\Box \vdash A$. For some $A_1, \ldots, A_n \in \Gamma_\Box$, we have $(A_1 \wedge \ldots \wedge A_n) \to A \in \mathsf{PK}^d$, from which we obtain by Rule $R_\Box$ and Axiom 1

$$(\Box A_1 \wedge \ldots \wedge \Box A_n) \to \Box A \in \mathsf{PK}^d,$$

i.e., $\Box A \in \Gamma$, a contradiction.

2. This item follows from Axiom 6. ■

We turn to the proof of the canonical model lemma. If $\Diamond C \in \Gamma$ for some C, we put $\Gamma_0 = \Gamma$. Otherwise, let us consider $\Gamma' = Cl(\Gamma \cup \{\Diamond(p_0 \to p_0)\})$. By Item 2 of the previous lemma, $\Box A \notin \Gamma'$ and we denote by Γ_0 a PK^d-prime extension of Γ' with $\Gamma_0 \nvdash \Box A$. By Item 1 we have $(\Gamma_0)_\Box \nvdash A$.

We have defined Γ_0 so that $\Diamond C \in \Gamma_0$ for some C. Note that

$$(\Gamma_0)_\Box \nvdash \{B |\; \Diamond B \notin \Gamma_0\}.$$

Otherwise, for some $\Box A_1, \ldots, \Box A_n \in \Gamma_0$ and $\Diamond B_1 \ldots, \Diamond B_m \notin \Gamma_0$, we have

$C \to ((A_1 \wedge \ldots \wedge A_n) \to (B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d$,

$\Diamond C \to \Diamond((A_1 \wedge \ldots \wedge A_n) \to (B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d$ by the rule $R_\Diamond$,

$\Diamond C \to (\Box(A_1 \wedge \ldots \wedge A_n) \to \Diamond(B_1 \vee \ldots \vee B_m)) \in \mathsf{PK}^d$ by Axiom 5.

Thus, $\Box(A_1 \wedge \ldots \wedge A_n) \to \Diamond(B_1 \vee \ldots \vee B_m) \in \Gamma_0$. Applying Axioms 2 and 4 we obtain $\Diamond B_1 \vee \ldots \vee \Diamond B_m \in \Gamma_0$, which means that $\Diamond B_i \in \Gamma_0$ for some i, a contradiction.

By Lemma 3 there exists $\Delta \in W^c$ such that $\Delta \supseteq (\Gamma_0)_\Box$ and

$$\Delta \nvdash \{A\} \cup \{B|\ \Diamond B \notin \Gamma_0\}.$$

It is not difficult to see that this is the desired Δ.

$\Diamond$. By definition and induction hypotheses

$$\mathcal{M}^c, \Gamma \models \Diamond A \ \Leftrightarrow\ \exists \Delta(\Gamma_\Box \subseteq \Delta \text{ and } \Delta^\Diamond \subseteq \Gamma \text{ and } A \in \Delta)$$

We must prove that the latter statement is equivalent to $\Diamond A \in \Gamma$. If the right-hand side of the equivalence holds, from $\Delta^\Diamond \subseteq \Gamma$ and $A \in \Delta$ we immediately obtain $\Diamond A \in \Gamma$.

Assume $\Diamond A \in \Gamma$. Arguing as in the previous item we can show

$$A \cup \Gamma_\Box \nvdash \{B|\ \Diamond B \notin \Gamma\}.$$

Extend $A \cup \Gamma_\Box$ to the set $\Delta \in W^c$ such that $\Delta \nvdash \{B|\ \Diamond B \notin \Gamma\}$. We have obtained Δ with $A \in \Delta$ and $\Gamma R^c \Delta$. ■

The canonical model lemma immediately implies the conclusion of the theorem. ■

The semantical characterization of the logics PK and PK^d allows one to obtain the following

Proposition 1. *The logics* PK *and* PK^d *are conservative extensions of the positive fragment of intuitionistic logic.*

3 Adding strong negation

From now on, we will work with the language $\mathcal{L} := \mathcal{L}^+ \cup \{\sim\}$, where $\sim$ is a symbol for strong negation. Let For denote the set of all formulas of the language $\mathcal{L}$, and let

$$\Box For := \{\Box A|\ A \in For\} \text{ and } \Diamond For := \{\Diamond A|\ A \in For\}.$$

We define four extensions of the logics PK and PK^d with strong negation. Let V be the set of strong negation axioms for non-modal propositional logic, i.e.,

$$\begin{array}{rcl} V & := & \{\sim\sim A \leftrightarrow A, \sim (A \vee B) \leftrightarrow (\sim A \wedge \sim B), \\ & & \sim (A \wedge B) \leftrightarrow (\sim A \vee \sim B), \sim (A \to B) \leftrightarrow (A \wedge \sim B)\} \end{array}$$

and let $D := \{\sim \Box A \leftrightarrow \Diamond \sim A, \sim \Diamond A \leftrightarrow \Box \sim A\}$ be the set of duality axioms for the modalities. We set

$$\begin{array}{lcl} \mathsf{NK}^- & := & \mathsf{PK} + V \\ \mathsf{NK}^{d-} & := & \mathsf{PK}^d + V \\ \mathsf{NK} & := & \mathsf{NK}^- + D \\ \mathsf{NK}^d & := & \mathsf{NK}^{d-} + D \end{array}$$

A formula is in *negation normal form* if it contains $\sim$ only in front of propositional variables. Let $\Box^d = \Diamond$ and $\Diamond^d = \Box$. The following translation $\overline{(\cdot)}$ sends every formula A to a formula in negation normal form, where $p \in Prop$, $\diamond \in \{\vee, \wedge, \to\}$ and $\sharp \in \{\Box, \Diamond\}$:

$$\begin{array}{rclrcl} \overline{p} & = & p & \overline{\sim p} & = & \sim p \\ \overline{\sim\sim A} & = & \overline{A} & \overline{A \diamond B} & = & \overline{A} \diamond \overline{B} \\ \overline{\sim (A \vee B)} & = & \overline{\sim A} \wedge \overline{\sim B} & \overline{\sim (A \wedge B)} & = & \overline{\sim A} \vee \overline{\sim B} \\ \overline{\sim (A \to B)} & = & \overline{A} \wedge \overline{\sim B} & & & \\ \overline{\sharp A} & = & \sharp \overline{A} & \overline{\sim \sharp A} & = & \sharp^d \overline{\sim A} \end{array}$$

Proposition 2. *For any formula A, $A \leftrightarrow \overline{A} \in \mathsf{NK}$ (NK^d).*

Corollary 1. *The logics NK and NK^d are closed under the weak replacement rule*

$$\frac{A \leftrightarrow B \quad \sim A \leftrightarrow \sim B}{C(A) \leftrightarrow C(B)}.$$

The axioms of the logics NK^- and NK^{d-} say nothing about negations of modal formulas $\sim \Box A$ and $\sim \Diamond A$, therefore, analogues of the previous proposition and corollary do not hold for these logics.

Now we turn to the semantics for the introduced modal logics. The notions of an NK^--frame and NK-frame coincide with that of a PK-frame. The notions of an NK^{d-}-frame and NK^d-frame coincide with that of a PK^d-frame.

Models for the logics NK^- and NK^{d-} have the form $\mathcal{M} = \langle \mathcal{W}, v^+, v^- \rangle$, where $\mathcal{W}$ is an NK^--frame, respectively, a NK^{d-}-frame and the valuation functions v^+ and v^- are defined on the following sets:

$$v^+ : Prop \to \langle W, \leq \rangle^+ \text{ and } v^- : Prop \cup \Box For \cup \Diamond For \to \langle W, \leq \rangle^+.$$

Models for the logics NK and NK^d are also obtained by augmenting the respective frames with two valuation functions v^+ and v^-, but in this case, these are functions $v^+, v^- : Prop \to \langle W, \leq \rangle^+$.

For NK^-, the positive and negative forcing relations $\mathcal{M}, t \models^+ A$ and $\mathcal{M}, t \models^- A$ are defined essentially so as for Nelson's logic with strong negation in case of propositional variables and positive connectives . For the other connectives, we set:

$$\begin{array}{lll}
\mathcal{M}, t \models^+ \sim A & \text{iff} & \mathcal{M}, t \models^- A \\
\mathcal{M}, t \models^- \sim A & \text{iff} & \mathcal{M}, t \models^+ A \\
\mathcal{M}, t \models^+ \Box A & \text{iff} & \forall u \geq t \, \forall v (u R_\Box v \text{ implies } \mathcal{M}, v \models^+ A) \\
\mathcal{M}, t \models^- \Box A & \text{iff} & t \in v^-(\Box A) \\
\mathcal{M}, t \models^+ \Diamond A & \text{iff} & \exists u (t R_\Diamond u \text{ and } \mathcal{M}, u \models^+ A) \\
\mathcal{M}, t \models^- \Diamond A & \text{iff} & t \in v^-(\Diamond A)
\end{array}$$

In case of the logic NK, the forcing relations are defined for modalities in a different way:

$$\begin{array}{lll}
\mathcal{M}, t \models^+ \Box A & \text{iff} & \forall u \geq t \, \forall v (u R_\Box v \text{ implies } \mathcal{M}, v \models^+ A) \\
\mathcal{M}, t \models^- \Box A & \text{iff} & \exists u (t R_\Diamond u \text{ and } \mathcal{M}, u \models^- A) \\
\mathcal{M}, t \models^+ \Diamond A & \text{iff} & \exists u (t R_\Diamond u \text{ and } \mathcal{M}, u \models^+ A) \\
\mathcal{M}, t \models^- \Diamond A & \text{iff} & \forall u \geq t \, \forall v (u R_\Box v \text{ implies } \mathcal{M}, v \models^- A)
\end{array}$$

For the logic NK^{d-}, forcing relations are defined essentially so as for NK^-, and for NK^d essentially so as for NK. The only difference is that both relations $R_\Box$ and $R_\Diamond$ are replaced by R.

For each of the four logics defined above, validity on a model, $\mathcal{M} \models A$, means that $\mathcal{M}, t \models^+ A$ for all t. The validity on a frame and $\models_L$-validity, where $L \in \{\mathsf{NK}^-, \mathsf{NK}^{d-}, \mathsf{NK}, \mathsf{NK}^d\}$, are defined in the usual way.

Theorem 3. *$\vdash_L A$ iff $\models_L A$, where $L \in \{\mathsf{NK}^-, \mathsf{NK}^{d-}, \mathsf{NK}, \mathsf{NK}^d\}$.*

All items can be proved via the canonical model method. The main difference in the construction of canonical models is as follows. We put

$$v^{+c}(p) := \{\Gamma \in W^c | \; p \in \Gamma\} \text{ and } v^{-c}(p) := \{\Gamma \in W^c | \; \sim p \in \Gamma\}$$

for the logics NK and NK^d; and for NK^- and NK^{d-}, additionally,

$$v^{-c}(\Box A) := \{\Gamma \in W^c | \ \sim \Box A \in \Gamma\} \text{ and } v^{-c}(\Diamond A) := \{\Gamma \in W^c | \ \sim \Diamond A \in \Gamma\}.$$

The canonical model lemma also has a different formulation.

Lemma 7. *For any $\Gamma \in W^c$ and formula A,*

$$\mathcal{M}^c, \Gamma \models^+ A \textit{ iff } A \in \Gamma, \ \mathcal{M}^c, \Gamma \models^- A \textit{ iff } \sim A \in \Gamma.$$

Comparing the definitions of the positive forcing relations for the logics NK^-, NK^{d-}, NK, and NK^d with the forcing relation defined for PK and PK^d, we easily obtain the following result.

Proposition 3. *1. NK^- and NK are conservative extensions of PK.*

2. NK^{d-} and NK^d are conservative extensions of PK^d.

On the other hand, the definitions of the forcing relations for non-modal connectives coincide with the forcing relations for Nelson's logic $\mathsf{N4}$, from which we obtain

Proposition 4. *NK^-, NK^{d-}, NK, and NK^d conservatively extend $\mathsf{N4}$.*

It is not hard to check also that all the logics introduced above possess the disjunction property and, if strong negation is present, also the constructible falsity property.

Proposition 5. *Let $L \in \{\mathsf{NK}^-, \mathsf{NK}^{d-}, \mathsf{NK}, \mathsf{NK}^d, \mathsf{PK}, \mathsf{PK}^d\}$ and let $L' \in \{\mathsf{NK}^-, \mathsf{NK}^{d-}, \mathsf{NK}, \mathsf{NK}^d\}$.*

1. *If $\vdash_L A \vee B$, then $\vdash_L A$ or $\vdash_L B$.*
2. *If $\vdash_{L'} \sim (A \wedge B)$, then $\vdash_{L'} \sim A$ or $\vdash_{L'} \sim B$.*

4 Extensions via first-order translations

In this section, finally, we make use of constructive predicate logic, $\mathsf{QN4}$, quantified $\mathsf{N4}$. An axiomatization and the relational semantics of $\mathsf{QN4}$ are presented in [10]. We define two translations from $\mathcal{L}$ into the language of $\mathsf{QN4}$ containing a unary predicate P for every propositional variable p and two binary relation symbols $R_\Box$ and $R_\Diamond$. The first translation, tr_x, is defined

$$\begin{array}{rcll} p & \stackrel{tr_x}{\longmapsto} & P(x) & \\ \sim A & \stackrel{tr_x}{\longmapsto} & \sim tr_x(A) & \\ A \bowtie B & \stackrel{tr_x}{\longmapsto} & tr_x(A) \bowtie tr_x(B), & \bowtie \in \{\wedge, \vee, \rightarrow\} \\ \Box A & \stackrel{tr_x}{\longmapsto} & \forall y(R_\Box(x,y) \rightarrow tr_y(A)) & \\ \Diamond A & \stackrel{tr_x}{\longmapsto} & \exists y(R_\Diamond(x,y) \wedge tr_y(A)) & \end{array}$$

Table 13.1: The translation tr_x into constructive predicate logic.

$$\begin{array}{rcl} \sim \Box A & \stackrel{tr'_x}{\longmapsto} & \exists y(R_\Diamond(x,y) \wedge tr'_y(\sim A)) \\ \sim \Diamond A & \stackrel{tr'_x}{\longmapsto} & \forall y(R_\Box(x,y) \rightarrow tr'_y(\sim A)) \end{array}$$

Table 13.2: The translation tr'_x of formulas $\sim \Box A$ and $\sim \Diamond A$.

in Table 13.1. The second translation, tr'_x, is defined like tr_x, except for the translation of formulas $\sim \Diamond A$ and $\sim \Box A$, see Table 13.2. In these definitions, y is a fresh individual variable not used so far in the translation. We also consider a pair of translations from $\mathcal{L}$ into the language of $\mathsf{QN4}$ containing a unary predicate P for every propositional variable p and one binary relation symbol R. The first translation, T_x, is defined like tr_x, and the second translation, T'_x, is defined like tr'_x, except that in both cases $R_\Box$ and $R_\Diamond$ are replaced by R. Whereas translation tr'_x ensures formal duality, translation tr_x does not:

$$\begin{array}{rcl} tr'_x(\sim \Diamond A) & = & \forall y(R_\Box(x,y) \rightarrow tr'_y(\sim A)) \;=\; tr'_x(\Box \sim A) \\ tr_x(\sim \Diamond A) & = & \sim \exists y(R_\Diamond(x,y) \wedge tr_y(A)) \\ tr_x(\Box \sim A) & = & \forall y(R_\Box(x,y) \rightarrow tr_x(\sim A)) \end{array}$$

Although the duality axioms for the quantifiers are provable in $\mathsf{QN4}$, so that $tr_x(\sim \Diamond A)$ is provably equivalent with $\forall y(\sim R_\Diamond(x,y) \vee tr_y(\sim A))$, it is not the case that an implication $(A \rightarrow B)$ is provably equivalent with $(\sim A \vee B)$. Also T'_x guarantees formal duality, whereas T_x does not. Note that the restriction of T'_x to $\mathcal{L}^+$ is the standard translation of $\mathcal{L}^+$ into positive intuitionistic predicate logic QInt^+.

We define three logics $\mathsf{FS}L$, $L \in \{\mathsf{K}^-, \mathsf{K}^{d-}, \mathsf{K}^d\}$. The notation '$\mathsf{FS}L$' points to Fischer Servi's logic FS, the logic faithfully embedded into QInt by the standard translation of the language of intuitionistic modal logic into the language of QInt, see [5], [8], [9]. The interaction between $\Box$ and $\Diamond$ in FS can be captured by Axioms 5 and

7. $(\Diamond A \to \Box B) \to \Box(A \to B)$

Note that Axiom 6 is equivalent to a substitution instance of Axiom 7.

The logics $\mathsf{FS}L$ are defined as follows:

FSK^d := NK^d + Axiom 7 (instead of Axiom 6)

FSK^{d-} := NK^{d-} + Axiom 7 (instead of Axiom 6) +
8. $\sim\Box A \leftrightarrow \Diamond \sim A$ + the impossibility axiom
9. $\sim\Diamond A \wedge \sim\Diamond B \leftrightarrow \sim\Diamond(A \vee B)$
+ the rule $(R^{\sim}_{\Diamond})$ $\sim A \to \sim B / \sim\Diamond A \to \sim\Diamond B$

FSK^{-} := NK^{-} +
5′. $\sim\Box\sim(A \to B) \to (\Box A \to \sim\Box\sim B)$
7′. $(\sim\Box\sim A \to \Box B) \to \Box(A \to B)$
+ Axiom 9 +
10. $\sim\Box\sim(A \vee B) \to (\sim\Box\sim A \vee \sim\Box\sim B)$
+ $(R^{\sim}_{\Diamond})$ + $(R^{\sim}_{\Box})$ $\sim A \to \sim B / \sim\Box A \to \sim\Box B$

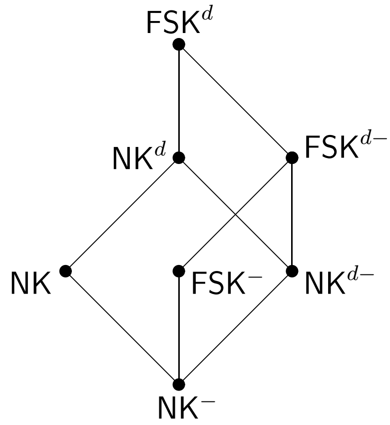
Axiom 7 is added to FSK^d to guarantee that FSK^d is faithfully embedded into QN4 via the translation T'_x. Axiom 9 and the rule $(R^{\sim}_{\Diamond})$ of FSK^{d-} show that despite of the lack of the second duality axiom $\sim \Diamond A \leftrightarrow \Box \sim A$, the strong negation of possibility can be treated as a new impossibility operator. Axioms 5′, 7′, and 10 of FSK^{-} express the fact that $\sim \Box \sim$ is a new possibility operator different from $\Diamond$, which is semantically and formally dual to $\Box$.[2] It can easily be verified that we obtain the following picture:

[2] Note that we could also define logics 'dual' to FSK^{d-} and FSK^{-} by considering $\sim \Box$ as a new unnecessity (non-necessity) operator and replacing Axioms 8 and 9 by the axioms

8.′ $\sim \Diamond A \leftrightarrow \Box \sim A$

9.′ $\sim \Box A \vee \sim \Box B \leftrightarrow \sim \Box(A \wedge B)$

and, in the case of FSK^{d-}, replacing rule $(R^{\sim}_{\Diamond})$ by $(R^{\sim}_{\Box})$.



FSK^d-frames and models are like NK^d-frames and models, except that in addition (as in the case of FS) it is postulated that

$$R \circ \leq \ \subseteq \ \leq \circ R.$$

The relational semantics for FSK^{d-} is inspired by T_x. A FSK^{d-}-frame is a tri-relational structure $\langle W, \leq, R, R^{\sim} \rangle$, such that $\langle W, \leq, R \rangle$ is a FSK^d-frame and $R^{\sim} \subseteq W \times W$. A FSK^{d-}-model $\mathcal{M} = \langle W, \leq, R, R^{\sim}, v^+, v^- \rangle$ is a FSK^{d-}-frame extended by two valuation functions $v^+, v^- : \mathit{Prop} \to \langle W, \leq \rangle^+$. The forcing relations are defined exactly as the forcing relations for NK^{d-}, except that we now have the following falsification clauses for modal formulas:

$$\begin{array}{lll} \mathcal{M}, t \models^- \Diamond A & \text{iff} & \forall s \geq t\ \forall u (sR^{\sim}u \text{ or } \mathcal{M}, u \models^- A) \\ \mathcal{M}, t \models^- \Box A & \text{iff} & \exists s (tRs \text{ and } \mathcal{M}, s \models^- A) \end{array}$$

In the presence of the impossibility relation $R^{\sim}$, the falsification clause for $\Diamond A$ is plausible. A state t supports the falsity of $\Diamond A$ iff for every possible expansion s of t and every state u it holds true that either u falsifies A or u is impossible relative to s.

FSK^{-}-frames and models are essentially the same as FSK^{d-}-frames and models, except that instead of a single relation R there are two relations $R_\Box$ and $R_\Diamond$, the relation $R_\Box$ satisfies the same condition as R in FSK^{d-}-frames, the relation $R_\Diamond$ is as in NK^{-}-frames, $\leq^{-1} \circ R_\Diamond \ \subseteq \ R_\Diamond \circ \leq^{-1}$, and we have the following forcing relations for modal formulas:

$$\begin{array}{lll}\mathcal{M},t \models^+ \Box A & \text{iff} & \forall u \geq t \forall v(uR_\Box v \text{ implies } \mathcal{M},v \models^+ A)\\ \mathcal{M},t \models^- \Box A & \text{iff} & \exists s(tR_\Box s \text{ and } \mathcal{M},s \models^- A)\\ \mathcal{M},t \models^+ \Diamond A & \text{iff} & \exists u(tR_\Diamond u \text{ and } \mathcal{M},u \models^+ A)\\ \mathcal{M},t \models^- \Diamond A & \text{iff} & \forall s \geq t \forall u(sR^{\sim}u \text{ or } \mathcal{M},u \models^- A)\end{array}$$

For each of the logics $\mathsf{FS}L$, validity on a model, $\mathcal{M} \models A$, means that $\mathcal{M},t \models^+ A$ for all t. The validity on a frame and $\models_{\mathsf{FS}L}$-validity are defined in the usual way.

Theorem 4. *For every logic* $\mathsf{FS}L$ *with* $L \in \{\mathsf{K}^-, \mathsf{K}^{d-}, \mathsf{K}^d\}$ *we have* $\vdash_{\mathsf{FS}L} A$ *iff* $\models_{\mathsf{FS}L} A$.

Proof. In each case soundness can be shown directly, and for completeness we can use canonical models. For FSK^d it suffices to show that the canonical model satisfies $(R^c \circ \subseteq) \subseteq (\subseteq \circ R^c)$, see [5], [12]. In the case of FSK^{d-} and FSK^-, the new element in the canonical model is defined as follows:

$$\Gamma R^{\sim c} \Delta \text{ iff } \Gamma_{\sim\Diamond} \not\subseteq \Delta,$$

where $\Gamma_{\sim\Diamond} = \{\sim A \mid \sim\Diamond A \in \Gamma\}$. In the proof of the canonical model lemma, we have: $\mathcal{M}^c, \Gamma \models^- \Diamond A$ iff $\forall \Gamma' \supseteq \Gamma\ \forall \Delta(\Gamma' R^{\sim c} \Delta$ or $\sim A \in \Delta)$. The latter holds iff $\sim\Diamond A \in \Gamma$. If $\sim\Diamond A \in \Gamma$, then for every $\Gamma' \supseteq \Gamma$ and any Δ, if not $\Gamma' R^{\sim c} \Delta$, then $\Gamma'_{\sim\Diamond} \subseteq \Delta$ and hence $\sim A \in \Delta$.

For the other direction suppose that $\sim\Diamond A \notin \Gamma$. Then $\sim A \notin \Gamma_{\sim\Diamond}$ and one can prove $\Gamma_{\sim\Diamond} \not\vdash \sim A$. Indeed, assume $\Gamma_{\sim\Diamond} \vdash \sim A$.

This means that there are $B_1, \dots, B_n$ such that $\sim\Diamond B_1, \dots, \sim\Diamond B_n \in \Gamma$ and $\sim B_1 \wedge \dots \wedge \sim B_n \vdash \sim A$. Then

$$\begin{array}{ll}\sim(B_1 \vee \dots \vee B_n) \to \sim A \in \mathsf{FSK}^{d-}(\mathsf{FSK}^-) & \\ \sim\Diamond(B_1 \vee \dots \vee B_n) \to \sim\Diamond A \in \mathsf{FSK}^{d-}(\mathsf{FSK}^-) & R^{\sim}_{\Diamond}\\ (\sim\Diamond B_1 \wedge \dots \wedge \sim\Diamond B_n) \to \sim\Diamond A \in \mathsf{FSK}^{d-}(\mathsf{FSK}^-) & \text{Axiom 9}\end{array}$$

The latter formula together with the assumption leads to a contradiction, $\sim\Diamond A \in \Gamma$.

By the extension lemma, there exists $\Delta \in W^c$ with $\Gamma_{\sim\Diamond} \subseteq \Delta \not\vdash \sim A$. Thus $\sim A \notin \Delta$ and not $\Gamma R^{\sim c} \Delta$. ■

Corollary 2. *The logics* $\mathsf{FSK}^{\mathsf{d}}$, FSK^{d-}, *and* FSK^- *are closed under the weak replacement rule*

$$\frac{A \leftrightarrow B \quad \sim A \leftrightarrow \sim B}{C(A) \leftrightarrow C(B)}.$$

Proposition 6.

$$\begin{array}{llll} 1. & \mathsf{FSK}^{-} & \subseteq & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} tr_x(A)\} \\ 2. & \mathsf{NK} & \subseteq & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} tr'_x(A)\} \\ 3. & \mathsf{FSK}^{d-} & \subseteq & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} T_x(A)\} \\ 4. & \mathsf{FSK}^{d} & = & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} T'_x(A)\} \end{array}$$

Proof. 1.-3.: By induction on proofs. 4.: It must be shown that T'_x is a faithful embedding. Observe that the translation of $\mathcal{L}$-formulas in negation normal form into $\mathcal{L}^+$ that maps every propositional variable p to itself, sends every negated propositional variable $\sim p$ to a fresh propositional variable p', and commutes with the positive connectives is a faithful embedding of FSK^d into FS^+, positive FS. This suffices to prove the claim, since FS^+ is faithfully embedded by the standard translation, coinciding on negation-free formulas with T'_x, into $\mathsf{QInt}^+ = \mathsf{QN4}^+$, positive $\mathsf{QN4}$. ■

Open questions

$$\begin{array}{llll} \mathsf{FSK}^{-} & = & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} tr_x(A)\} & ? \\ \mathsf{NK} & = & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} tr'_x(A)\} & ? \\ \mathsf{FSK}^{d-} & = & \{A \in \mathcal{L} | \vdash_{\mathsf{QN4}} T_x(A)\} & ? \end{array}$$

The $\subseteq$-directions are proved in the previous proposition. A treatment of the $\supseteq$-directions is left to future research. The second equality explains why we did not introduce the new logic FSK.

5 Bibliography

[1] S. Akama. On Constructive Modality, in: S. Akama (ed.), *Logic, Language and Computation*, Kluwer Academic Publishers, Dordrecht, 1997, 143–158.

[2] J. van Benthem. Modal Foundations for Predicate Logic, *Logic Journal of the IGPL* **5** (1997), 259–286.

[3] M. Božić and K. Došen. Models for Normal Intuitionistic Modal Logics, *Studia Logica* **43** (1984), 217–245.

[4] J.M. Dunn. Positive Modal Logic, *Studia Logica* **55** (1995), 301–317.

[5] G. Fischer Servi. Axiomatizations for some Intuitionistic Modal Logics, *Red. Sem. Mat. Universi. Politec. Torino* **42** (1984), 179–194.

[6] A. Fuhrmann. Models for Relevant Modal Logics, *Studia Logica* **49** (1990), 501–514.

[7] D. Gabbay. *Fibring Logics*, Oxford University Press, 1998.

[8] D. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-Dimensional Modal Logics: Theory and Applications*, Elsevier, Amsterdam, 2003.

[9] C. Grefe. Fischer Servi's Intuitionistic Modal Logic has the Finite Model Property, in: M. Kracht et al. (eds.), *Advances in Modal logic. Vol. 1.*, CSLI Lecture Notes 87, 1998, 85–98.

[10] S.P. Odintsov and H. Wansing. Inconsistency-tolerant Description Logic: Motivation and Basic Systems, in: V.F. Hendricks and J. Malinowski (eds.), *Trends in Logic: 50 Years of Studia Logica*, Kluwer Academic Publishers, Dordrecht, 2003, 301–335.

[11] R. Routley. Semantical Analyses of Propositional Systems of Fitch and Nelson, *Studia Logica* **33** (1974), 283–298.

[12] A.K.Simpson. *The Proof Theory and Semantics of Intuitionistic Modal Logics,* PhD Thesis, University of Edinburgh, 1994.

[13] V. Sotirov. Modal Theories With Intuitionistic Logic, in: *Mathematical Logic. Proceedings of the Conference on Mathematical Logic, Dedicated to the Memory of A.A. Markov (1903-1979), Sofia, September 22-23, 1980*, Sofia, 1984, 139–171.

[14] S.K. Thomason. Reduction of Second-order Logic to Modal Logic, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **21** (1975), 107–114.

[15] H. Wansing. Diamonds are a Philosopher's Best Friends, *Journal of Philosophical Logic* **31** (2002), 591–612.

Logic is not the whole story

J. A. Robinson
robinson@equinox.shaysnet.com
Syracuse University

When in 1979 I put out a book [6] about my work in the mechanization of FOL proof procedures, I felt pleased that it had been exactly one hundred years since Frege published the *Begriffschrifft*, essentially inventing FOL and ushering in the modern era of formal logic. His system inspired me to write [6, p. 1]

> ...the correctness of a piece of reasoning ...does not depend on what the reasoning is about ...so much as on how the reasoning is done; on the pattern of relationships between the various constituent ideas rather than on the actual ideas themselves.

Since 1879, logicians have developed many such systems of formalized reasoning. The many versions or presentations of Frege's FOL are in several ways the most attractive of these. FOL is the closest in spirit to the forms of actual reasoning found in real mathematical proofs and its later semantical metatheory (due to Tarski) sheds much light on our informal notion of logical consequence. These formal systems all successfully capture one essential feature of real mathematical proofs, namely their *objective validity* or *logical correctness*. Unfortunately their success has still left uncaptured what may well be an even more important feature of mathematical proofs, namely their *epistemological cogency*.

The mathematician Saunders Mac Lane [4, p. 377] describes the typical professional's view of proof in these words:

> proof in mathematics is both a means to understand why some result holds and a way to achieve precision. ...the test for the

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 287–302

> correctness of a proposed proof is by formal criteria and not by reference to the subject matter at issue.

He goes on to say why actual proofs are rarely dressed up in fully formal clothing (p. 378):

> ... there are good reasons why mathematicians do not usually present their proofs in fully formal style. It is because proofs are not only a means to *certainty*, but also a means to *understanding*. Behind each substantial formal proof there lies an *idea*. The idea ... *explains* why the result holds.

The great mathematician G.H.Hardy in 1929, the year after the publication of [2], set forth his own view of proof in a vivid metaphor:

> I have myself always thought of a mathematician as in the first instance an observer, a man who gazes at a distant range of mountains and notes down his observations. His object is simply to distinguish clearly, and notify to others, as many different peaks as he can. There are some peaks which he can distinguish easily, while others are less clear. He sees A sharply, while of B he can obtain only transitory glimpses. At last he makes out a ridge which leads from A, and following it to its end he discovers that it culminates in B. B is now fixed in his vision, and from this point he can proceed to further discoveries. In other cases perhaps he can distinguish a ridge which vanishes in the distance, and conjectures that it leads to a peak in the clouds or below the horizon. But when he sees a peak he believes that it is there simply because he sees it. If he wishes someone else to see it, he points to it, either directly or though the chain of summits which led him to recognise it himself. When his pupil also sees it, the research, the argument, the proof is finished.

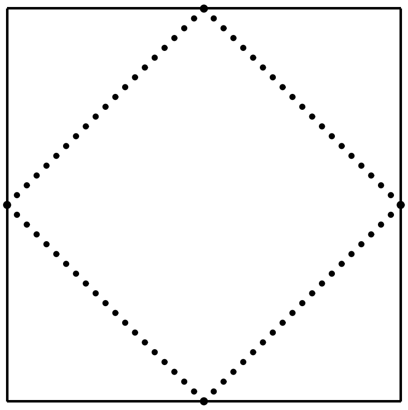
Hardy here is expressing the attitude of most real mathematicians to the nature and function of mathematical proof. He views a proof as a means of enabling the mind to see the *truth* of a proposition, as a device to engineer *understanding* and *conviction* by assembling and advantageously displaying an array of compelling evidence. The act of proving is a process of mind manipulation, which takes time to achieve its purpose. What happens to the assimilator of a successful proof is the expansion of his or her knowledge. A person *knows* something after experiencing a proof which he or she did not know before. If one goes through a proof of something one already knows, the effect is often to reinforce one's knowledge and to help one to understand familiar facts in a new way.

Hardy is aware that his account is somewhat poetic and fanciful:

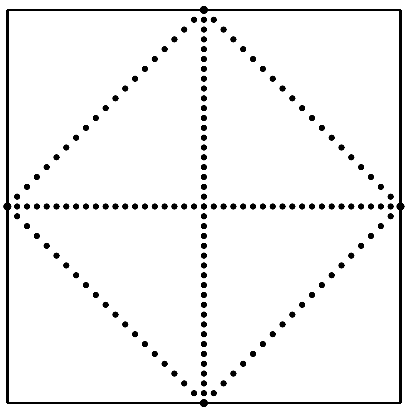
> . . . the analogy is a rough one, but I am sure that it is not altogether misleading. If we were to push it to its extreme we should be led to a rather paradoxical conclusion; that there is, strictly, no such thing as mathematical proof; that we can, in the last analysis, do nothing but point; that proofs are what Littlewood and I call gas, rhetorical flourishes designed to affect psychology, pictures on the board in the lecture, devices to stimulate the imagination of pupils. This is plainly not the whole truth, but there is a good deal in it. The image gives us a genuine approximation to the processes of mathematical pedagogy on the one hand and of mathematical discovery on the other; it is only the very unsophisticated outsider who imagines that mathematicians make discoveries by turning the handle of some miraculous machine. Finally the image gives us at any rate a crude picture of Hilbert's metamathematical proof, the sort of proof which is a ground for its conclusion and whose object is to convince.

The reference here to Hilbert is salutary: for Hilbert's name is somewhat misleadingly associated with the doctrine that formalization of mathematical reasoning is the fundamental technique for achieving certainty and setting mathematics on a sound footing. Hilbert's own predilection in his lectures was for crystal-clear, direct, simple explanation of mathematical propositions, providing so compelling a vision of their truth that a formal proof often seemed like an almost obvious and natural afterthought [5, p. 103].

In Plato's dialogue *Meno*, Socrates anticipates Hardy's method by conducting a pedagogical demonstration with the help of a (we are assured by Plato) uneducated and naive servant boy. He enables the boy to see the truth of the proposition: the lines connecting the midpoints of adjacent sides of any square form another square having exactly half its area.



The key step in this process is to suggest that the boy notice the effect of also connecting the midpoints of opposite sides of the given square.



The large square is thereby seen to be dissected into eight triangles which are all congruent, hence have the same area. The smaller square whose vertices are the midpoints of the sides of the larger square contains exactly four of these triangles. Thus the boy *is enabled to see* that the area of the smaller square is exactly half that of the larger sqare.

The *Meno* example illustrates Hardy's claim – all that Socrates does is to *point* at features of 'what is there' until the boy comes to see that the proposition is true. In the course of his epistmological journey the boy experiences what have been called 'mathematical epiphanies' – Professor Benson's term [1] for the sudden *revelations* by which the mind recognizes a truth. It does not affect the *Meno* demonstration's significance that Socrates himself has a special implausible interpretation of this process, that the boy is simply *recalling* knowledge which he always had. The significance of the demonstration is to show how any mind is capable of experiencing these moments of truth-revelation and how they can be exploited to build knowledge in that mind.

Hardy's point is that all proofs, no matter how long or complex they may be, can be understood as processes similar to the *Meno* proof process. A mind is enabled to experience one or more epiphanies which lead to the final *quod erat demonstrandum* and make it clear and obvious.

One can appreciate the Hardy view directly by going over proofs oneself and experiencing them as an organized web of epiphanies taking place in one's own consciousness. The subject matter need not be geometrical or perceptual. It can be more abstract and immaterial, involving, for example, computational and numeric intuitions, or intuitions involving mappings and transformations, and indeed whatever mental constructions are clearly viewable in the mind's eye.

Consider Euclid's simple but wonderfully intuitive proof of the proposition that there are infinitely many primes. This is not a geometric or spatial truth, but our understanding of its proof is clear and direct. Its central epiphany is the fact that no *finite* set of primes contains them all. To see why, consider any such set. To find a prime which is not in it, all we need to do is compute the number – call it P – which is the *successor* of the *product* of all members of the set. *Every prime factor of* P lies outside the set, since none of the primes inside the set exactly divides P – each of them will always leave a remainder of 1.

This proof does of course require some prior knowledge of the number system: the operations of multiplication, taking the successor, and division; the definition of primeness; and the concept of a finite set. The intuitions available to the mind are not only the *wired-in* perceptual, spatial ones of the elementary geometry example, but also the *learned* symbolic, computational intuitions of arithmetic, of comparing, sorting, and rearranging. Proofs work only on minds which are already suitably equipped with concepts and facts on which the new concepts and facts can be built. Crucially, in this example one has to know that *every number is either itself a prime or else is a product of two or more numbers each of which is prime.*

The *Meno* example is perhaps misleading in that it deals with a mind that is totally naive – quite empty of all *learned* knowledge and furnished only with the intrinsic, hard-wired perceptual machinery provided by the human brain's basic design. Although the primitive mental processes differ in that some are innate and some are learned, the evocation of them in Hardy-like epiphanic proof-engineering is the same. They are awakened when needed.

Formal logic does not try to address this amazing *epistemological* process. How can we go about developing a theory of proof, which does? Clearly, the first step is to study the complete proof process in greater depth and detail, using as data not just tiny, single-epiphany proofs like those earlier discussed, but the more complex, multi-epiphany proofs of the professional mathematician. These are informal, but rigorous nevertheless. They are complex, but with simple individual modules of comprehension. Our minds are limited in their epiphanic capacity. They can deal with insights of only a certain restricted size and scope. Anything larger needs to be broken down (if possible) into some combination of smaller ones. Not all minds are alike in this respect. No doubt those of Ramanujan or Riemann were well above the human norm, but even they needed multi-epiphany proofs to see some of the things that they saw, and in some cases it took them a long time to discover one. Famously, Riemann never did find a proof for what we all now call Riemann's Hypothesis (nor has anyone else) but even so he was convinced of its truth.

In Hardy's view, proofs are not in themselves the static written objects that one finds in books and journals. The book and journals contain only, so to speak, the *scripts* or *recipes* which are to be followed in acting out the proof so that it can produce its effect on the mind. The proofs themselves are the *performances* of these scripts. Studying the *real* proofs therefore means that we must *watch these performances* and try to see what is going on as they take place. It is not enough to study the written scripts. As Wittgenstein put it [7, III, 27]:

> ... pay attention to the patter by means of which we convince someone of the truth of a mathematical proposition. It tells us something about the function of this conviction. I mean the patter by which intuition is awakened.

He later adds;

> Do not look at the proof as a procedure that *compels* you, but as one that *guides* you.

He added [7, III, 30]

> . . . in the course of the proof our way of seeing is changed .. our way of seeing is remodelled.

Wittgenstein is here stressing the Hardy-like pedagogical accompaniments – the 'gas' – which mathematicians use to guide the mind into the successive required epiphanies. Because we have come to think of proofs only as structured static texts we tend to think that nothing *happens* in proofs. We have not been paying attention to that 'patter' or indeed to the effect on the audience for who the performance is intended.

It is as though we think music exists only in the form of the static scores written by composers, whereas in fact the music is *what happens when the score is performed*.

A convenient way to observe proof performances is by *introspective experiments*. Choose a theorem whose meaning is clear but whose truth is not. Then take yourself through the experience of a proof of it, while simultaneously monitoring the process. There is no need to *invent* the proof; just follow and understand it. If all goes well, the truth of the theorem will have become obvious by the end of the proof. *Something happens* in one's head, and *the traversal of the proof is what makes it happen*. Ideally, one should enlist the help of a mathematician who is willing and able to perform the proof with oneself as audience. This way one avoids the complication of being both the performer and the observer.

Either way, we follow the proof and try to watch what happens.

Professor Hardy's metaphor may be too fanciful and flowery for some tastes. The analogy between natural observation and rational conviction is nevertheless worth pursuing. What he is saying is that, however you deliver or communicate a proof, whether it be informally or formally, in writing or in a spoken lecture, with or without visual aids, the effect of the proof must in the last analysis be that its reader, or hearer, *sees the truth* of the proposition which is being proved. The proof must produce conviction in a human mind. The better the proof is, the more it will accomplish beyond merely producing such conviction; it will bring about understanding as well. It will provide not just a *guarantee* (that the conclusion is true), but also an *explanation* why it is true.

There are some similarities between the way the vision system works and the way that the more abstract mental processing works which we associate with mathematical cognition. One of the similarities is its *involuntary* character.

In saying that we see or fail to see that a proposition is true, there lurks a suggestion that the moment of intellectual acquisition is spontaneous and involuntary, just as the event of visual acquisition is involuntary. When we look at a scene, we *cannot help seeing what we see.* Our brains are processing the incoming visual information automatically by cascading recognition and detection signals through neural pathways layer-by-layer back into deeper brain structures, and our mental picture springs into existence rapidly and unbidden. Similarly with our other sensory modalities: hearing, touch, pain, and so on.

It seems clear that physiologically something of the same sort must be taking place when we follow a mathematical proof, digesting its steps and letting them register their effects on our state of knowledge. The assent we give to the successive claims in the proof is not a matter of our *will power.* We do not *decide* to acquiesce in believing what the proof demands we believe. Our subjective experience of the flow of the proof is as a series of *revelations* which *happen to us* whether we like it or not.

The modern attitude to logical rigor is that it should be totally free of psychologism – dependence upon subjective considerations, such as intuition and insight. It was a mistake in earlier times to mix up mental phenomena with objective, combinatorial, structures of derivations in formal languages and their set-theoretic semantical interpretations. From this point of view, people do not or need not exist. Formal proofs are simply certain kinds of mathematical entity, like integers and holomorphic functions. Whether a certain object is or is not a (correct) formal proof depends only on its objective structure, in the same sort of way as whether an entity is a pentagon. Its effect on the observer, whatever that might mean, is completely irrelevant. A formal proof is a formal proof is a formal proof.

That is all very well, but the science of logic loses its primary motivation if we sever all its connections with actual thinking. The point of a formal proof is that it should at least correspond to, by being an abstract representation of, a real proof, which in turn might serve to convince somebody of the truth of its conclusion.

Like the objects in a visual scene, the landmarks of a good, convincing proof spring out and engage the mind in spite of itself. The mind is a kind of spectator or camera. Hardy's metaphor brings out the essential passiveness of the elemental observations making up a proof, the viewings of those intervening peaks. We simply see them, or we don't. The person doing the proving (and not necessarily for another person; one often undertakes to prove something for oneself, silently and in private) needs indeed only to *point*, and the person following the proof needs only to *look*, at the right things, and the rest is amazingly automatic.

At the level of the underlying flow of information between its neurons, the brain is clearly a massively parallel asynchronous computing device. Hundreds of billions of neuron-firings take place simultaneously at any given moment. Each individual neuron can fire – send an output pulse to its immediate successors – hundreds of times per second, asynchronously. It fires when ready without reference to any sort of central clock.

Events at this level are like the lowest-level switching events or pulses in the electronic circuits of a computer. In a computer we need to ascend several conceptual levels above the pulse level to arrive at events and structures which correspond to cognitively meaningful processes and entities. At these higher levels we keep track of the changing values of variables, the calling of subroutines, the evaluation of arguments and the application of functions to them, and so on.

In the brain, similarly, it is only at higher levels of organization, well above the elemental level of neuron-firings, that we find cognitively significant processes and structures. Conscious thought and experience belong to these higher levels.

It is, incidentally, a curious feature of the design of the brain that *conscious* mental processes are serial in character, despite the fact that the underlying neural machine which supports them is highly *parallel.* It may well be therefore that the *unconscious* mental processes, if such there be, are capable of occurring in parallel. Neuroscientists may one day be in a position to discover whether this is so. At present the serial nature of conscious thought is simply an observed fact. Our logics reflect this – many of them define a formal proof as a *linear sequence* of formulas, each successive formula following from earlier (often, *much* earlier) formulas by a rule of inference. The systems which define a formal proof as a tree structure are nearer the mark.

Some intellectual tasks associated with formal logics – such as checking an alleged formal proof to see if is correct – can actually be performed in parallel, but not by individual humans. This task does not correspond to anything we encounter in real proofs. A formal tree-structured proof is made up of one or more inferences. Each inference is a self-contained construct consisting of a conclusion and a set of premises. It is correct if and only if its conclusion follows from its premises by some rule of inference. The proof as a whole is a linked complex of such inference steps. Each inference can be checked for correctness independently of the others, and given enough agents, all of them can be checked at the same time. Some or all of the premises of one inference step may be themselves the conclusions of other inference steps. The entire formal proof has the structure of a tree with sentences as its nodes. Its inferences are then its various subtrees each consisting of a node together with the immediate successors of that node. A node without any immediate successors (a leaf of the tree) is a premise of the proof. It is considered to be a special case of an inference step, one whose conclusion is inferred from the empty set of premises. A node which is not the successor of any node (the root of the tree) is the theorem that is the conclusion of the proof. Note that there are many different ways to visit each node in the tree in order to examine, or experience, the inference whose conclusion is that node. The corresponding linear sequences of formulas are the traces of such tree-traversals, and none of them is particularly superior to any other.

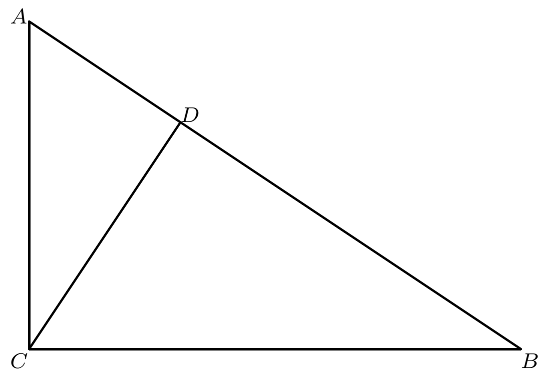
A Hardy-like proof-journey can be thought of as exhibiting this general tree pattern, too. What is interesting is that the mind can cope with an arbitrary ordering of the visits to the inferences: it is not necessary to begin with the premisses and 'follow the flow' of the inferences from premisses to conclusion, layer by layer. Such a 'bottom up' discipline is one option, to be sure, but only one of many. At the other extreme ('top down') one can start with the theorem and work backwards, seeing that *if* we could see that its immediate premisses were true, *then* we could see that it too must be true, and so on, backwards through the tree to the ultimate premisses, which can indeed then be seen to be true directly (assuming the proof works).

Immediacy: a simple example from elementary geometry.
When Wittgenstein said [7, III, 32]

> What interests me is not the immediate realization of a truth, but the phenomenon of immediate realization,

he was seizing on the central element of the proof process, namely its being a complex of *atomic* cognitive acts or events, each an indivisible whole phenomenologically, and consisting of seeing or recognizing a truth.

What kinds of mental event count as atomic cognitive acts? That is, what is the difference between *immediately* recognizing the truth of some proposition and coming to know its truth in a way that is not immediate? Consider the following simple geometrical proof in which the difference between what is and what is not immediate is not always obvious. In the well-known 'similar triangles' proof of the theorem of Pythagoras we are asked to contemplate a plane figure consisting of a right-angled triangle ABC whose right angle is at the vertex C. On its hypotenuse AB there is a point D such that the line CD is perpendicular to AB.



The following facts *are* surely immediate:

1 There are three triangles in the figure: ABC, ADC and DBC.

2 Each of these triangles is right-angled.

3 Triangles ABC and ADC share an angle, namely the one at A.

4 Triangles ABC and DBC share an angle, namely the one at B.

5 length of AB = length of AD + length of DB.

These are directly observable features. Our knowledge of them is essentially *perceptual*. We can hardly avoid (assuming that we actually look) seeing that these features are present in the figure. It is the initial, hard-wired, spatial feature-recognition layer of our neural cognitive net which, so to speak, simply *detects* them.

The following facts are not quite so immediate, but are detected, perhaps, by the *next* cognitive layer:

6 Angles ABC and ACD are equal.

7 Angles BAC and DCB are equal.

To account for our being able to see these two secondary facts we have to suppose that not only the primary facts 1 through 5 are available to the second layer, but that somehow the pair of triangles ABC and ADC, and the pair of triangles ABC and DBC can each be seen as particular cases of the following *general* fact about pairs of triangles:

8 Triangles which agree in two of their three angles also agree in the third.

These facts, 6 through 8, are *immediate conclusions from* what we can directly see, rather than things we can *actually* directly see. They are *secondary* rather than *primary* facts, and they are detected not by the first, but only by the second, cognitive layer of the brain. Here are some tertiary facts, for the detection of which a third cognitive layer imports another general item of knowledge:

9 triangles agreeing in all three of their angles also agree in the ratios of the lengths of each pair of their corresponding sides

and applies it to triangles ABC and ADC. At this depth of perception the mind has to be aware that

10 side AC of triangle ADC corresponds to side AB of triangle ABC (since angle ADC = angle ACB)

11 side AD of triangle ADC corresponds to side AC of triangle ABC (since angle ACD = angle ABC)

so that it can appreciate the equality of the ratios

12 (length of AC) / (length of AD) = (length of AB) / (length of AC).

13 (length of BC) / (length of DB) = (length of AB) / (length of BC).

or in other words (multiplying out the denominators – one more cognitive layer deep?)

14 (length of AC)^2 = (length of AB) $\times$ (length of AD).

15 (length of BC)^2 = (length of AB) $\times$ (length of DB).

Now we are finished with the direct observations of the figure (the data gathering, as it were) and our minds begin a short flurry of algebraic manipulation. Such manipulation is a combinatorial process whose atomic actions are transformations of symbolic expressions purporting to preserve their denotations. That the transformations actually do this must be every bit as convincing as are other kinds of direct confrontation with the facts. First, by *adding equations 14 and 15* we get the equation

16 (length of AC)^2 + (length of BC)^2 = (length of AB) $\times$ (length of AD) + (length of AB) $\times$ (length of DB).

in which we 'undistribute' the factor (length of AB) on the right hand side to obtain the equation:

17 (length of AC)^2 + (length of BC)^2 = (length of AB) $\times$ ((length of AD) + (length of DB)).

In doing this, we are appealing to the general distributivity principle of elementary algabra which allows us to replace a term of the form $x \times y + x \times z$ by a term of the form $x \times (y + z)$ – and *vice versa* – since both denote the same number no matter what numbers x, y and z may be. Finally, we recall what we already have observed, and recorded as fact 5, to conclude that in 17 we can replace the term (length of AD) + (length of DB) by the term (length of AB) to obtain the equation

18 (length of AC)^2 + (length of BC)^2 = (length of AB) $\times$ (length of AB) = (length of AB)^2.

The outcome of this Hardy-like journey is that we have come to see that

19 (length of AC)^2 + (length of BC)^2 = (length of AB)^2

which is the Theorem of Pythagoras.

The short symbolic computation is a series of almost immediate cognitive perceptions. What are involved here are not so much *spatial* intuitions as *computational* intuitions. Here the brain is holistically perceiving, as it were in a single *gestalt*, each of a series of elementary actions and their results. The elementary actions are *redex-replacements* or *rewritings*, in which one expression is substituted for another expression known to be equivalent to it (that is, to have the same denotation). Such rewritings are the very stuff of symbolic computation , and in our culture we are equipped by education early in life with a repertoire of learned mental 'recognition demons' consisting the corresponding rewriting rules. We memorize the multiplication tables, and as a result we have built-in calculators based on table look-up. Seven nines are sixty-three, four twelves are forty-eight, and so on – these are immediate for most educated people. Someone who regularly uses mathematics will have mental immediacies which are not so common: but they are learned in the same way, grooved in by constant repetition and use. Algebraic manipulation is a typical arena for this slightly more specialized mental skill. Increasingly specialized and sophisticated immediacies make up *mathematical expertise*. There is no obvious limit on what the mind can assimilate and package in this way. Learned immediacies go a long way to account for the way that a proof works. The proof needs to have certain immediacies already available in the mind, and it will build others in the course of its unfolding.

Proofs are quite crucial in the *creation of new immediacies*. Our brains are (it seems) hard-wired already at birth or very soon thereafter for recognizing certain features of the visual field, as indeed they also are for the auditory and tactile features. The brain develops in early life by acquiring more and more of these as the child expands its repertoire, by interaction with parents, teachers and others, and by solo exploration of its environment.

In the similar triangles proof the immediacy of the judgment that the area of the large triangle is the sum of the areas of the two smaller triangles springs straight out of the direct perception that the larger triangle consists of the two smaller triangles placed together. This perception/judgement generalizes, though, to situations which might not be so simple to experience directly. We come to see that, in general, if a finite region of the plane is divided into finitely many disjoint subregions which together exhaust the whole region, then the area of the whole region is the sum of the areas of the subregions. We must however be careful. This principle applies to *real space*, but not to the abstract model of space as a three-dimensional mathematical continuum, in view of the so-called Banach-Tarski Theorem (often also called the Banach-Tarski Paradox) which seems like a shocking contradiction to the immediacy of our spatial intuition: a sphere of radius r can be decomposed into a finite number of disjoint pieces in such a way that the pieces can then be reassembled into *two* spheres of radius r. At issue is the precise definition (formalization) of *volume* (measure) and the status of the set-theoretic proposition known as the *Axiom of Choice*. It is not that our intuition about space is at fault. The problem lies in our *model* of space as a continuum and perhaps also in our imperfect understanding of the concept of an infinite set. It is interesting to note that we have no *direct* intuitions concerning infinite sets, and yet we seem to have *strong* intuitions about the concept of infinite sets. For example, it seems extremely obvious that if an infinite set is the union of two disjoint sets, then at least one of these must also be infinite. If, however, we are challenged to *prove* this, it is not clear how to proceed. Is it not simply a primitive intuition which does not admit of proof in the usual sense? If it is, how do we know it with such certainty?

An understanding of real proofs, then, calls for an understanding of the mind's ability to learn, store and deploy the thousands of immediacies which are triggered by the activity of thinking about the various entities we invent and define for whatever reason. Introspective monitoring of proof experiences is one way to make progress towards this goal. With the help of modern computers it has been possible take quite extensive and nontrivial mathematical proofs and *formalize* them within some formal logic (often in FOL, but higher order logics have also been used). This is interesting and important work, but we also need to take the same proofs and *perform* them in a critical, self-aware attempt to find out how and why they work epistemologically. *Formalization* of a proof is very nearly antagonistic to this: the very act of formalization removes from the proof not only the meaning of the constituent nonlogical terms so as to leave exposed only the syntactic structures, but it also obliterates every remnant of the Wittgensteinian 'patter' which largely determines the pragmatic efficacy of the proof.

There is much work to be done. Let us begin!

1 Bibliography

[1] D. C. Benson. *The Moment of Proof: Mathematical Epiphanies.* Oxford, 1999.

[2] G.H. Hardy. Mathematical proof. *Mind*, 38:1–25, 1929.

[3] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik.* Springer, Berlin, 1928.

[4] S. Mac Lane. *Mathematics: Form and Function.* Springer, 1986.

[5] C. Reid. *Hilbert.* Springer, 1970.

[6] J. A. Robinson. *Logic: Form and Function.* North Holland, 1979.

[7] L. Wittgenstein. *Remarks on the Foundations of Mathematics.* Blackwell, 1956.

First-Order Logic, Second-Order Logic, and Completeness

Marcus Rossberg
mr30@st-andrews.ac.uk
Arché - AHRB Research Centre for the Philosophy of
Logic, Language, Mathematics, and Mind
School of Philosophical and Anthropological Studies
University of St Andrews
St Andrews, Fife KY16 9AL
Scotland, U.K.

1 Introduction

Both first- and second-order logic (FOL and SOL, respectively) as we use them today were arguably[1] created by Frege in his *Begriffsschrift* – if we ignore the notational differences. SOL also suggests itself as a natural, and because of its much greater strength, desirable extension of FOL. But at least since W. V. Quine's famous claim that SOL is "set theory in sheep's clothing"[2] it is widely held that SOL is not proper logic – whatever this is taken to be by different authors – but some kind of mathematics. Even contemporary advocates of SOL like Stewart Shapiro point out its mathematical character, albeit without regarding this as problematic.[3] Recent criticisms focus both on the ontological commitment of SOL, which is believed to be to the set-theoretic hierarchy, and on the allegedly problematic epistemic status of the second-order consequence relation.

[1]See William Ewald's contribution to this volume, pp. 89 ff., on this question.
[2][Quine, 1970], pp. 66–68.
[3][Shapiro, 1991], see esp. p. vi–vii and p. 48.

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 303–321

This paper focuses on the second issue, and investigates the claim that the second-order consequence relation is intractable because of the incompleteness result for SOL. The opponents' claim is that SOL cannot be proper logic since it does not have a complete deductive system. I will argue that the lack of a completeness theorem, despite being an interesting result, cannot be held against the status of SOL as a proper logic.

I mainly deal with SOL in this paper. But of course there is an unsurveyable manifold of logics that are not complete – a whole range of first-order modal logics, for instance – to which my argument equally applies (if it holds for SOL). I will also only consider the classical versions of propositional, first- and second-order logic, since most of the present discussion concerning SOL focuses on this case. This is especially true for the incompleteness allegation. It seems to me, though, that the argument is general enough to carry over to non-classical versions of these logics as well.

2 The Complaint

The notion of completeness that is the focus for the present discussion, and that will be my concern here, is sometimes called *strong semantic completeness*. This kind of completeness is a metatheoretic result that holds (or fails to hold) between two different consequence relations. The first of these is *deductive consequence*. It is usually symbolized as '$\Gamma \vdash S$' which can loosely be read as 'S can be deduced from Γ' where 'S' is a sentence of the relevant language and 'Γ' a – possibly empty – class of sentences of the same language. This relation is contrasted with that of *semantic* (or *model-theoretic*) *consequence*. We usually symbolize this as '$\Gamma \models S$' and read it informally as 'All models that make Γ true make S true as well'. If the deductive consequence relation does not hold between any sentences between which the model-theoretic consequence relation fails to hold, we speak of *soundness*. The converse we call *completeness*.

$$\Gamma \vdash S$$

$$\text{Soundness} \Downarrow \qquad\qquad \Uparrow \text{Completeness}$$

$$\Gamma \models S$$

The worry mentioned in the introductory section is directed at the failure of completeness for SOL: there are model-theoretic validities of SOL that cannot be derived in its deductive system. If a consequence relation cannot be tracked by way of the deductive system, the criticism usually continues, then it cannot be the case that we are dealing with a proper logic. Rarely are arguments provided as to why this should be the case.[4] Quine, for example, raises the incompleteness objection against SOL, but interestingly enough not in his *Philosophy of Logic* where he claims SOL is set theory in disguise. He raises the issue of incompleteness in this book, but only in his objections against branching quantifiers.[5] With respect to SOL, it seems, Quine first mentions incompleteness in a response to Hao Wang.[6] In neither case does he provide an argument why the lack of completeness is supposed to show that we are not dealing with a proper logic.

Perhaps the motivation behind the demand for the completeness of a proper logic can be reconstructed as follows: Logical consequence is intuitively taken to be a semantic notion. The logical consequence relation might be said to hold between some premises and a conclusion if, and only if, the truth of the conclusion is guaranteed by the truth of the premises in virtue of their logical form alone. 'Logical form' here can be taken to mean something like the semantic structure of the sentences (or, if you will, the propositions they express). It is therefore the formal semantics, i.e. the model theory, that captures logical consequence. The deductive system merely gives us a bunch of inference rules. If completeness fails this shows that the deductive system does not properly capture logical consequence.

Such a position exhibits a preference of model theory over the deductive system. Let's call a logician who favours model theory as the right way to capture logical consequence the *model-theorist*, and her opponent who favours the deductive system the *proof-theorist*.[7] (Of course I do not suggest that every actual logician can be put in one of these boxes, let alone exactly one of them.) Say, the model-theorist claims that in absence of a model theory it cannot be decided which rules of inference are the logical ones. Furthermore, she might say that model theory provides the proper analysis

---

[4] Jan Woleński's contribution to this volume, pp. 369 ff., is one of the rare exceptions. See also [Wagner, 1987].

[5] [Quine, 1970], pp. 90–91.

[6] [Quine, 1986], p. 646.

[7] [Shapiro, 1991], chapter 2, see esp. p. 35, draws a similar distinction: He discusses the difference between what he calls the *foundational* and the *semantic* conception of logic.

of the concept of logical consequence and therefore gets close to its very nature, while the deductive system at best achieves extensional adequacy. The proof-theorist, on the other hand, could insist that the logical form of a sentence is exhibited by the logical constants that it contains, and that these get their meaning from their introduction- and elimination-rules. She could resist the thought that we need model theory to separate the good, i.e. logical, introduction- and elimination-rules from the bad ones, and might claim that, on the contrary, it is not clear how model theory is supposed to help us in deciding this question. And why shouldn't one be able to give a conceptual analysis using a deductive system? Logic, after all, is about inference, and so are deductive systems.[8]

All these issues are important, and good arguments have been put forward on both sides. The above paragraph certainly does not faithfully represent the complexity of the actual debate, which not only goes beyond the issues I have mentioned but also far beyond the scope of this paper. That many other criteria are suggested by both proof- and model-theorists and adopted to argue for or against SOL and other logics is of course understood, but will not concern us here. For very little in the sketched arguments and the whole debate depends on completeness proofs. Crucially, the question of completeness comes into play only after the other issues have been settled. This is the point I will examine and develop further in the following sections.

3 Logical Consequence

The unqualified claim that SOL is incomplete needs to be made more precise. No deductive system is semantically incomplete in and of itself; rather a deductive system is incomplete with respect to a specified formal semantics. The deductive system of SOL, for instance, is incomplete with respect to the *standard* model-theoretic semantics. In standard model theory a model consists of a set of objects called the *domain* and an interpretation function. This function assigns objects in the domain to names of the language, subsets of the domain to predicate letters, subsets of the Cartesian product of the domain with itself to binary relation symbols and so on. The first-order quantifiers range over the domain, while the second-order quantifiers range over the subsets of the domain in case the quantifier binds a predicate variable, over the subsets of the Cartesian product of the domain with itself in

[8] See for example [Tennant, 1986].

case the quantifier binds a binary relation variable etc.

As is well-known, standard semantics is not the only semantics available. *Henkin semantics*, for example, specifies a second domain of predicates and relations for the upper case constants and variables. The second-order quantifiers binding predicate variables, e.g., can be thought of as ranging over a subset of the powerset of the first-order domain. What is relevant to the present discussion is that the deductive system of SOL is sound and complete with respect to a Henkin semantics.[9] With suitable restrictions on the class of models even compactness, Löwenheim-Skolem, and Löwenheim-Skolem-Tarski hold.[10] To pick up a thought of Shapiro's,[11] one might think that standard (as opposed to Henkin) semantics does not provide *enough* models to invalidate all sentences of the language of SOL that are not theorems. So here we have a completeness theorem for SOL. But it would be bizarre to claim that the incompleteness complaint is thereby refuted.

It is often suggested that, interpreted with a Henkin semantics, SOL is basically a two-sorted first-order logic. This would also explain the apparent tension between the above mentioned results and Lindström's theorem: No logic that goes beyond the expressive power of FOL satisfies both the compactness and the Löwenheim-Skolem theorem.[12] It is also worth noting that a feature of SOL gets lost when a Henkin semantics is adopted; the very feature that attracts many of those who are interested in SOL to it.[13] SOL with standard semantics allows for *categorical* axiomatizations of certain mathematical theories, such as arithmetic or real analysis. A mathematical theory is categorical if, and only if, all of its models are isomorphic. Such a theory then essentially has just one model, the standard one. First-order axiomatizations of, for instance, real analysis cannot be categorical since the Löwenheim-Skolem theorem holds, which directly contradicts categoricity: If there is an uncountable model, there will be a countable one as well. For the same reason SOL with Henkin semantics cannot deliver categoricity results. It seems that the desire to have a complete logic leaves us with one that deprives us of the possibility of having categorical characterizations of infinite structures. Completeness and categoricity apparently pull in opposite

[9] [Henkin, 1950].

[10] [Shapiro, 1991], pp. 88–95.

[11] [Shapiro, 1998], p. 141, dismisses this view, but also see his fn. 10 on the same page.

[12] [Lindström, 1969].

[13] See for example [Shapiro, 1991] and [Shapiro, 1997].

directions. As desirable as it might seem, one cannot have both.[14] (Other projects in the philosophy and foundation of mathematics which require SOL, e.g. the Neo-Fregean programme,[15] might not depend on standard semantics, of course.)

I will not take sides here in the debate regarding whether standard or Henkin semantics is the "right" semantics. Nevertheless, it is important to realise that Henkin and standard semantics are not the only options available. Substitutional, game-theoretical, or topological semantics, Boolos' plural interpretation,[16] or even a semantics inspired by Leśniewski recently suggested by Peter Simons[17] surely do not exhaust the alternatives. Further, I contend that the issues concerning completeness and logical consequence are more likely to be obscured than elucidated by a discussion about "the right semantics". (What are the criteria for the "right" semantics? On which independent, i.e. non-question-begging grounds can we decide? Is the "right" semantics "right" *tout court* or is it the "right" one with respect to some purpose? Is there only one "right" semantics?) A more direct approach is called for.

So let's take a step back and reflect on key features we want a proper logic to have and what its relation to logical consequence is. We want our logic to be *formal*. I will henceforth use '*formal system*' (or sometimes even just '*system*') in a somewhat unorthodox way. We often use 'formal system' as short for 'formal *deductive* system', i.e. a formal language together with some rules of inference. The meaning I intend is broader: It includes any axiomatic system as long as it is based on a formal language; in particular it includes model theory. Whenever I refer to *deductive* systems in particular I will explicitly use 'deductive system'. So we want a logic to be a formal system in the above sense. The reason for the possibly non-standard usage will become clear in a moment. It is to be found in a Fregean thought, made explicit in the short 1882 paper *On the Scientific Justification of a*

[14]The relation between and historical significance of categoricity and completeness is investigated in [Corcoran, 1980] and [Read, 1997].

[15]See [Wright, 1983], [Hale and Wright, 2001] and [MacBride, 2003].

[16]Cf. [Boolos, 1984], [Boolos, 1985], [Boolos, 1994]; for ways to account for relations along the lines that Boolos suggest for the predicates see [Burgess et al., 1991], [Hazen, 1997a] and [Hazen, 1997b], and for an alternative way in Boolosian spirit [Rayo and Yablo, 2001]; investigations into the strength of Boolos' plural semantics are [Rayo and Uzquiano, 1999] and [Uzquiano, 2003].

[17]Cf. [Simons, 1985] and [Simons, 1993], see also [Simons, 1997].

Begriffsschrift. Frege published this paper in response to criticism directed at his *Begriffsschrift*:[18] We need a formal symbolic system such that the content cannot escape the rigorous logical form.[19] This thought can certainly be found in the *Begriffsschrift* itself already.

I take this thought to be of general importance: We are looking for formal systems which axiomatize, characterize, or formalize in some other way some notion, or notions, in such a way as to secure the intended content. These notions may themselves be pre-theoretic. Our particular concern here are *logical* systems, which are those formal systems that capture and formalize the notion of *logical consequence*. In this light, one might be tempted to read the soundness result as: "We will not deduce a sentence from a class of premises that is not a logical consequence of them" (we will come back to this later), and the completeness result accordingly as: "We can deduce every sentence from a class of premises that is a logical consequence of them". But this should give us pause. First, logical truths are (by definition) logical consequences of the empty class of premises and hence, by monotonicity, logical consequences of every class of premises. But certainly a logical truth of FOL cannot be deduced in the deductive system of propositional logic. Propositional logic is complete, and is a proper logic if anything is. Yet a logical truth like '$(\forall x(Fx \supset Gx) \land Fa) \supset Ga$' escapes its consequence relation. The solution to this "puzzle" is of course simple[20] and leads us to the second point: As noted above, completeness is a metatheoretic relation that holds between a deductive system and a formal semantics. So the two consequence relations we are dealing with are the deductive and the formal-semantic ones. Let's assume for the sake of simplicity that the formal semantics is model theory. So completeness shows that every deductive consequence of a class of sentences is, in addition, a model-theoretic consequence of those sentences. But how does *logical* consequence get into the picture?

Given the broad notion of a formal system as described above it now seems clear that we are dealing with *two* formal systems here: the deductive system and the model theory. If there is a soundness and a completeness proof

[18] [Frege, 1879].

[19] "Wir bedürfen eines Ganzen von Zeichen, aus dem jede Vieldeutigkeit verbannt ist, dessen strenger logischer Form der Inhalt nicht entschlüpfen kann." [Frege, 1882], p. 52.

[20] Another easy solution that comes to mind is that propositional logic simply lacks the expressive power that these sentences require. It can therefore not be accused of not capturing logical truths of predicate logic. That is of course correct. I will say more about systems with increasing expressive resources in the beginning of section 4.

we know that this duplication does not matter: Both systems capture the same consequence relation. If we can derive nothing but logical consequences of given premises in one system, then this will be the case for the other system as well. But if completeness fails, we will have to decide which of the two systems (if any) is the one that captures logical consequence. If one decides that it is the deductive system, one will hold that the model theory is defective in the sense that it produces a surfeit of consequences of a set of sentences which are not actually logical consequences of it. One might then say that the model theory does not provide an appropriate model of the logical consequence relation that is specified by the deductive system, and consequently reject the semantics.[21] The failure of completeness, therefore, does not disqualify the deductive system from capturing logical consequence and therefore being a proper logic (other features might still have this effect, of course). If, on the other hand, one has convinced oneself that model theory is the system that properly codifies logical consequence, one will presumably think that the right thing to do, when one wants to do logic, is just that, viz. model theory.[22] It is hard to see why the lack of a complete deductive system should cast doubt on the model-theoretic system as a logic if one has independent reasons to believe the model theory to properly capture logical consequence.

So let's re-assess what the significance of soundness and completeness proofs is. They show us very important features of the two systems. For example, they show that results from the one system can be carried over to the other. If I prove a theorem of first-order logic, by soundness I know that this sentence is valid, i.e. it is true in all models. If I provide a model that makes a sentence false, again by soundness I know that I won't be able to prove it. Completeness allows us to make these transitions in the reverse direction. A model-theoretic argument can establish that a sentence is a consequence of some other sentences. If completeness holds one knows that there is also a

[21]John Etchemendy seems to hold such a view with respect to SOL. In his terminology, the standard semantics for SOL "overgenerates". See [Etchemendy, 1990], esp. pp. 158–159.

[22]Jon Barwise might have had something like this in mind when he wrote: "Mathematicians often lose patience with logic simply because so many notions from mathematics lie outside the scope of first-order logic, and they have been told that that *is* logic. The study of model-theoretic logics should change that, by finding applications, and by the isolation of still new concepts that enrich mathematics and logic. [...] There is no going back to the view that logic is first-order logic." [Barwise, 1985], p. 23.

derivation in the deductive system. The reason to prefer a proof to be carried out in one or the other system might have to do with the time it takes to carry out the proof or other matters of convenience. George Boolos provides us with the example of an inference that can be shown to be valid with a relatively short model-theoretic argument.[23] Since all the premises and the conclusion are first-order, by completeness we know that there has to be a derivation of it. As it turns out, this derivation would contain more symbols than particles exist in the known universe. Our incapability to carry out such a proof seems to be a pretty good reason to prefer the model-theoretic argument over the derivation, no matter whether one thinks that the true way to formalize logical consequence lies in a deductive system or in model theory. Very often it will be more convenient, though, to derive consequences in the deductive system than to provide a model-theoretic argument. Take the case of a second-order inference and imagine a logician who thinks that the standard model theory properly captures the logical consequence relation. A derivation in the deductive system is, by soundness, as good as a model-theoretic argument as a means of showing that the conclusion of the deduction is a logical consequence of its premises. The mere fact that there are semantic consequences for which there is no derivation in the deductive system does not throw any doubt on it being a logical consequence (again, other considerations might well do so).[24]

I said above that one might be tempted to think that a soundness proof shows us that one can only derive logical consequences in a deductive system. It seems quite obvious now that this can only be the case if one already has established independently that the model-theoretic consequence relation

[23]For details see [Boolos, 1987]. Boolos' main interest in this paper, though, is that SOL allows us to carry out the proof on two pages while the first-order proof appears to be physically impossible. His observation, however, holds for the model-theoretic proof, too.

[24]Another way in which the existence of a soundness and completeness proof can be interesting is pointed out by [Kreisel, 1967], pp. 152–157: We want to capture the pre-theoretic notion of logical consequence. Now we convince ourselves that any derivation in the deductive system is licensed by our pre-theoretical notion, and any pre-theoretically valid inference is valid in the model theory as well. A soundness and completeness proof then ties all three notions together and shows that they are all equivalent. This is consistent with what I said above: It first needs to be shown that the pre-theoretic notion is indeed sandwiched in this way between the deductive system and model theory, and it does not follow from the impossibility of applying Kreisel's strategy, e.g. in the case of SOL, that either system fails to capture logical consequence. The refinements in section 4 will make this point even clearer.

embraces only logical consequences. But it provides us with another very important insight: A soundness proof assures us that the deductive system in question is *consistent*, i.e. that we will never be able to prove a contradiction in it (from no premises) – provided the model theory is consistent. This certainly is a desirable feature of logic, to say the least.

4 Refining the Picture

The considerations above require refinements in two respects. First, I have been speaking of the logical consequence relation that is captured by a formal system. It seems better to speak of a *part* of logical consequence here. Propositional logic, for example, is a proper subsystem of FOL, yet even if we had reason to believe that FOL captured all of logical consequence we would not want to deny the status of a proper logic to propositional logic just because it fails to capture all of logical consequence. The situation is more dramatic for someone who does not think that there is no proper logic stronger than FOL. Second-, third-, fourth-order logic and so on might be thought to be proper logics. If we thought we could chase all the way up to ω (and beyond?) there might not even be an all-inclusive system that captures the whole of logical consequence. Modal logics might be considered proper logics and the various systems might not mix well, especially when the conception of logic is a semantical one. Since my aim here is to stay as neutral as possible on the question "What is proper logic?" it seems advisable rather to speak in general of systems that partially formalize logical consequence.

The second refinement that is needed is due to the fact that the actual process of coming up with a formal system that captures logical consequence is a bit more complicated than the above picture suggests. For example, it might well be the case that the attempt to provide a formal system as an alternative to one that is in use already leads to the discovery of a surprising mismatch between the original system and the new one. Let's say we have two deductive systems, X and Y. X is traditionally thought to capture a certain class of logical inferences. The new system Y is suggested because it contains some features which are thought to be advantageous in application; it might be that it allows for shorter proofs, for example. When the logician constructs Y she wants to come up with a system that is equivalent to X in the sense that any sentence is amongst the theorems of X if, and only if, it is amongst those of Y.[25] Now let's imagine it turns out that the two formal systems do not match in that sense. I will consider three ways in which this could be the case:[26] (1) Y is a proper extension of X, i.e. all theorems of X are theorems of Y, but not the other way around, (2) X is a proper extension of Y, or (3) the systems contradict each other, i.e. the union of the class of theorems of X and the class of theorems of Y is inconsistent.[27] The reaction to this as described above is to decide on one of the systems, the one that properly captures the pre-theoretic notion. Let's say that in cases (1) and (2) one would convince oneself that the respective stronger system does not generate theorems that lie outside the pre-theoretic notion. Now, in the first case it seems that the natural response would be henceforth to stick to the new system Y. It formalizes a part of the pre-theoretic consequence relation that was hitherto unaccounted for. Since X is a subsystem of Y we know that all the sentences we have proved in the past in X will be theorems of Y as well. In the second case one would probably try to fix Y such that it captures the missing part of the consequence relation as well.

If we, on the other hand, come to believe that the "extra theorems" of the respective stronger deductive system are illegitimate in the sense that they lie outside of what is warranted by the pre-theoretic notion, in case (1) we might try to weaken Y until it matches X and so does not generate these "extra theorems" anymore. But (2), here, will be the really interesting case. We are in the position of having discovered that the old system X allowed us to prove sentences which on reflection upon the pre-theoretic notion should not be provable. A likely response is to stick to the new system, Y, in the future and weaken the old one, X, to see which axioms or rules are responsible for the "false" theorems. Depending on how well entrenched X was, some revision of previous results derived in X might be necessary.

Case (3) is equally interesting. One has to find out in this case which of the two systems really does capture the pre-theoretic notion. And it might

[25] I restrict myself here and in the following to the theorems of the systems for the sake of the simplicity of the exposition. The picture extends in the usual way to the consequence relations.

[26] One can of course distinguish at least six relations in which two such systems can stand to each other: The three I discuss in the following, the case where the systems are disjoint, and two cases of overlap: in one the union of the systems is consistent, in the other one it is not. In any case, the resulting situations would be similar in the relevant respects to the ones I describe.

[27] We might differentiate between considering the union and the closure of the union. This would give us even more cases than mentioned in the previous footnote.

turn out that it is neither. It seems that a case of type (3) is most likely to occur when the logician develops the new system because she is unsatisfied with the old one. She is then most likely to design a formal system directly due to reflection on the pre-theoretic notions and to be little, if at all, guided by the old system. In any of these cases, though, the mere mismatch of the systems is not going to provide one with an answer to the question: Which of the systems is to be modified or rejected? But it is a valuable indicator that one has to look out for arguments and that some adjustments are needed in at least one of the systems.

It is worth emphasising that fit between two systems does not guarantee that the intended notion is properly captured. The original system might fail to provide theorems it should have, or have too many. If the new system is then designed to match the old one, this failure will carry over. The mere proof that two systems agree on their theorems cannot show that the pre-theoretic notion is captured unless one has independent reasons already to believe that one does. Equally, in the case of a mismatch there is certainly no guarantee that one of the systems is the correct one. Both might be wrong.

I doubt that anyone would have deep objections to my case descriptions. And if I'm right about the relation between formal semantics and deductive systems with respect to pre-theoretic notions then the same picture must hold true in this case as well. Let the X and Y above be a deductive system and a formal semantics, respectively. (1) then corresponds to a situation in which we have soundness and incompleteness, as is the case for SOL with standard semantics, (2) would be completeness, but failure of soundness, (3) the failure of both. Let's consider (1), and the case of SOL with standard semantics, where we decide to favour the model theory. In light of Gödel's incompleteness theorem there is no hope of coming up with some additional axiom that will provide us with a deductive system that will allow us to deduce everything that is given by the standard model-theoretic consequence relation. But where it works out, we can rely on the standard deductive system anyhow, since the soundness theorem holds. For the rest we have to do model theory which – after all – we have assumed is the right place to look for logical consequence anyway. If, on the other hand, we convince ourselves that the deductive system properly captures logical consequence, but for some purposes we require a complete model theory (rather than rejecting model-theory as a whole), we would go for a Henkin semantics.[28] (2) corresponds

[28][Shapiro, 1999], p. 51, suggests that Henkin semantics is the right tool to study the

to SOL with entirely unrestricted Henkin models: completeness holds, but soundness fails.[29] If we decide to favour the deductive system we will restrict the class of Henkin models in a way that puts us into the position to prove soundness. Should we decide that the Henkin semantics captures what we are after we have to weaken our deductive system. It is, however, hard to imagine cogent reasons for adopting this last option. Concerning (3) we again find ourselves in the position where both systems are most clearly up for discussion; although it has to be said that this, of course, is the case for all three scenarios.

Failures of match between formal systems which are meant to capture the same pre-theoretic notion show us that some reconsideration needs to be undertaken. The failure of soundness or completeness is a special and important case of this, and there are some well-known examples of such investigations which have given rise to fruitful developments in formal logic. Tarski's original model-theoretic account, for example, arose from his dissatisfaction with the previous characterizations of the concept of logical consequence.[30] This first account operated with a domain that does not change in cardinality.[31] Hence for an infinite domain, for any $n \in \mathbb{N}$ any sentence of the language that expresses $\ulcorner$there exist at least n objects$\urcorner$ comes out as a logical truth – a feature usually deemed undesirable. Even worse, for finite domains, which sentences of this form are logical truths depends on the size of the domain. Up to the n that is the cardinality of the domain the sentences come out as logical truths, for greater n they come out as logical falsehoods. This is fixed in the model theory that we use today, and model theory has grown to be a very powerful and fruitful area of mathematical logic.

Something similarly exciting happened in the area of modal logic. Saul Kripke gave possible worlds semantics to a whole range of modal logics, and C.I. Lewis' original systems S4 and S5 turned out to be both sound and complete with respect to their respective possible worlds semantics. By having this semantical tool to hand, many more axiomatic systems could be designed (K, D, T, B, etc.). On the other side of the story, the soundness proof for Lewis' other systems failed, leading to the development of a semantics that includes so-called "impossible worlds". Again, more axiomatic systems were developed and "impossible worlds semantics" plays an important role,

deductive system of SOL, for example.

[29]Cf. [Shapiro, 1991], p. 88.

[30][Tarski, 1936].

[31]Cf. [McGee, 1992].

e.g., in epistemic logics, some relevant and paraconsistent logics, and formal semantic theories of information.

Before we conclude, note that it might be the case that the study of formal systems effects our so-called pre-theoretic intuitions.[32] Nothing in the discussion above hangs on whether pre-theoretic notions are fixed. Should a pre-theoretic notion change over time, for whatever reason, a new system is called for to capture the changed content. The old systems might survive and still codify technical concepts which are not to be identified with the changed pre-theoretic notions. The point remains the same: Should metatheoretical studies yield a mismatch between two systems – and one of those possible mismatches is what we call semantic incompleteness – independent arguments have to be provided concerning which of the two systems is the one that properly captures the pre-theoretic notion, if any. When the pre-theoretical notion that is to be formalized is that of logical consequence, incompleteness alone cannot serve as an *argument* to disqualify a system as a proper logic, since it does not provide us with a criterion to decide whether to reject the deductive system of the model theory. Much less is it acceptable to dismiss both of a *pair* of formal systems, say the deductive system of SOL and its standard model theory, on the grounds that they don't match.

5 Conclusion

I argue above that completeness results can provide insights into properties of formal systems. This is because soundness and completeness really constitute equivalence proofs. The failure of equivalence of two systems which are meant to formalize the same pre-theoretic notion suggests that these systems have to be investigated again to see what is wrong with at least one of them. But one will always have to provide independent arguments why a formal system under discussion does not properly capture (part of) logical consequence. The mere perfect match of a model theory to a sound deductive system, i.e. completeness, *cannot* provide such an argument unless one already has provided *other* arguments that show that one of the systems does in fact properly capture this part of logical consequence.

Imagine, for example, that the standard model theory is not only the

[32]There is a debate on this under the heading of "normative vs descriptive aspects of logic". [Resnik, 1985] argues that pre-theoretic and formal notions eventually reach a reflective equilibrium.

right tool to capture logical consequence, but also the only way to give a conceptual analysis of logical consequence – as briefly mentioned as an option in the end of section 2. It should therefore be the case that the best shot a deductive system had was to be extensionally adequate with respect to the right model theory. Now, if the completeness proof fails, this surely does not show that the model theory did not properly capture logical consequence. The deductive system on this view is an *addition* to the conceptual analysis of logical consequence given by the model theory. The latter is not to be measured in light of the former. The possible failure of the conceptual analysis would have to be shown in another way. The same holds *mutatis mutandis* if the deductive system is given precedence.

I'm under no illusion of having shown that SOL is proper logic. What I have argued for is merely that one cannot draw on the incompleteness of deductive systems of SOL with respect to standard semantics to show that either of the systems is *not* a proper logic. The arguments provided should be sufficiently unspecific to SOL to hold generally for logical systems which lack a completeness result. Traditionally, most logics started their formal career as a deductive or axiomatic system. All that should be required to get a semantics relative to which a given deductive system is complete is a sufficiently cunning model-theorist. Whether this semantics answers to the pre-theoretic notion that was meant to be captured is an entirely different issue and must be argued for independently. Of course, should it turn out that it does not capture logical consequence, and the deductive system is sound and complete with respect to it, so much the worse for the deductive system. But note that in such a case a *completeness* theorem, rather than an incompleteness theorem seals the fate of this system. If the logic starts out as a model-theoretic one, it is much less clear whether we can find a complete deductive system for it or a complete axiomatization of it. But the mere lack of such a system cannot count as showing that the model-theoretic logic is not a proper logic, especially not if we independently decide that the model theory properly captures the (part of the) logical consequence relation we were after.[33]

[33] I am indebted to Ross Cameron, Daniel Cohnitz, Roy T. Cook, Philip Ebert, Amy Hughes, Nikolaj Pedersen, Agustín Rayo, Stephen Read, Stewart Shapiro, Robbie Williams, Crispin Wright, and the participants of the FOL75 conference in Berlin for many helpful comments and suggestions.

6 Bibliography

[Barwise, 1985] Barwise, Jon (1985). Model-theoretic Logics: Background and Aims. In Barwise, Jon and Feferman, Solomon, editors (1985). *Model-Theoretic Logics*. Springer-Verlag, New York, pages 3–23.

[Boolos, 1984] Boolos, George (1984). To Be is to Be the Value of a Variable (or to Be Some Values of Some Variables). *Journal of Philosophy*, 81:430–450.

[Boolos, 1985] Boolos, George (1985). Nominalist Platonism. *Philosophical Review*, 94:327–344.

[Boolos, 1987] Boolos, George (1987). A Curious Inference. *Journal of Philosophical Logic*, 16:1–12.

[Boolos, 1994] Boolos, George (1994). 1879? In Clark, Peter and Hale, Bob, editors, *Reading Putnam*. Blackwell, Oxford, pages 31–48.

[Burgess et al., 1991] Burgess, John P., Hazen, A.P., and Lewis, David (1991). Appendix on Pairing. In Lewis, David, *Parts of Classes*. Oxford University Press, Oxford, pages 121–149.

[Corcoran, 1980] Corcoran, John (1980). Categoricity. *History and Philosophy of Logic*, 1:187–207.

[Etchemendy, 1990] Etchemendy, John (1990). *The Concept of Logical Consequence*. Harvard University Press, Cambridge, MA.

[Frege, 1879] Frege, Gottlob (1879). *Begriffsschrift. Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Verlag L. Nebert, Halle/Saale. Reprinted in [Frege, 1964a]. English translation in van Heijenoort, Jean, editor (1967), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, Cambridge, MA, pp. 1–82.

[Frege, 1882] Frege, Gottlob (1882). Ueber die wissenschaftliche Berechtigung einer Begriffsschrift. *Zeitschrift für Philosophie und Philosophische Kritik*, 81:48–56. Reprinted in [Frege, 1964a], pp. 106–114. English translation: [Frege, 1964b].

[Frege, 1964a] Frege, Gottlob (1964a). *Begriffsschrift und andere Aufsätze.* Mit E. Husserls und H. Scholz' Anmerkungen. Herausgegeben von Ignacio Angelelli. Olms, Hildesheim.

[Frege, 1964b] Frege, Gottlob (1964b). On the Scientific Justification of a Concept-Script. Translated by J.M. Bartlett. *Mind*, 73:155–160.

[Hale and Wright, 2001] Hale, Bob and Wright, Crispin (2001). *The Reason's Proper Study.* Clarendon Press, Oxford.

[Hazen, 1997a] Hazen, A.P. (1997a). Relations in Lewis's Framework Without Atoms. *Analysis*, 57:243–248.

[Hazen, 1997b] Hazen, A.P. (1997b). Relations in Monadic Third-Order Logic. *Journal of Philosophical Logic*, 26:619–628.

[Henkin, 1950] Henkin, Leon (1950). Completeness in the Theory of Types. *Journal of Symbolic Logic*, 15:159–171.

[Kreisel, 1967] Kreisel, Georg (1967). Informal Rigour and Completeness Proofs. In Lakatos, Imre, editor, *Problems in the Philosophy of Mathematics.* North-Holland, Amsterdam, pages 138–171.

[Lindström, 1969] Lindström, Per (1969). On Extensions of Elementary Logic. *Theoria*, 35:1–11.

[MacBride, 2003] MacBride, Fraser (2003). Speaking with Shadows: A Study of Neo-Logicism. *British Journal for the Philosophy of Science*, 54(1):103–163.

[McGee, 1992] McGee, Vann (1992). Two Problems with Tarski's Theory of Consequence. *Proceedings of the Aristotelian Society*, 92:273–292.

[Quine, 1970] Quine, W.V. (1970). *Philosophy of Logic.* 2nd edition. Oxford University Press, Oxford.

[Quine, 1986] Quine, W.V. (1986). Reply to Hao Wang. In Hahn, Lewis Edwin and Schilpp, Paul Arthur, editors, *The Philosophy of W.V. Quine.* Open Court, La Salle, Ill., pages 644–648.

[Rayo and Uzquiano, 1999] Rayo, Agustín and Uzquiano, Gabriel (1999). Toward a Theory of Second-Order Consequence. *Notre Dame Journal of Formal Logic*, 40:315–325.

[Rayo and Yablo, 2001] Rayo, Agustín and Yablo, Stephen (2001). Nominalism Through De-Nominalization. *Noûs*, 35(1):74–92.

[Read, 1997] Read, Stephen (1997). Completeness and Categoricity: Frege, Gödel and Model Theory. *History and Philosophy of Logic*, 18:79–93.

[Resnik, 1985] Resnik, Michael (1985). Logic: Normative or Descriptive? The Ethics of Believe or a Branch of Psychology? *Philosophy of Science*, 52:221–238.

[Shapiro, 1991] Shapiro, Stewart (1991). *Foundations Without Foundationalism: A Case For Second-Order Logic*. Oxford University Press, Oxford.

[Shapiro, 1997] Shapiro, Stewart (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, Oxford.

[Shapiro, 1998] Shapiro, Stewart (1998). Logical Consequence: Models and Modality. In Schirn, Matthias, editor, *Philosophy of Mathematics Today*. Oxford University Press, Oxford, pages 131–156.

[Shapiro, 1999] Shapiro, Stewart (1999). Do Not Claim Too Much: Second-Order and First-Order Logic. *Philosophia Mathematica*, 7:42–64.

[Simons, 1985] Simons, Peter (1985). A Semantics for Ontology. *Dialectica*, 39:193–216.

[Simons, 1993] Simons, Peter (1993). Who's Afraid of Higher-Order Logic? *Grazer Philosophische Studien*, 44:253–264.

[Simons, 1997] Simons, Peter (1997). Higher-order Quantification and Ontological Commitment. *Dialectica*, 51:255–271.

[Tarski, 1936] Tarski, Alfred (1936). Über den Begriff der logischen Folgerung. *Actes du Congrès International de Philosophie Scientifique*, 7:1–11. English translation: [Tarski, 1983], new English translation: [Tarski, 2003].

[Tarski, 1983] Tarski, Alfred (1983). On the Concept of Logical Consequence. In Tarski, Alfred, *Logic, Semantics, Metamathematics.* Edited by John Corcoran. Hackett, Indianapolis, pages 409–420.

[Tarski, 2003] Tarski, Alfred (2003). On the Concept of Following Logically. Translated from the Polish and German by Magda Stroińska and David Hitchcock. *History and Philosophy of Logic*, 23:155–196.

[Tennant, 1986] Tennant, Neil (1986). The Withering Away of Formal Semantics. *Mind and Language*, 1:302–318.

[Uzquiano, 2003] Uzquiano, Gabriel (2003). Plural Quantification and Classes. *Philosophia Mathematica*, 3:67–81.

[Wagner, 1987] Wagner, Stephen J. (1987). The Rationalist Conception of Logic. *Notre Dame Journal of Formal Logic*, 28:3–35.

[Wright, 1983] Wright, Crispin (1983). *Frege's Conception of Numbers as Objects*. Aberdeen University Press, Aberdeen.

Logic-Based Agents and the Frame Problem: A Case for Progression

Michael Thielscher
mit@inf.tu-dresden.de
Department of Computer Science, Dresden University of Technology

Intelligent agents that reason logically about their actions have to cope with the classical Frame Problem. We argue that a progression-based solution is necessary for agent programs to run efficiently over extended periods of time. We support this claim by comparing the computational behavior of two popular logic programming systems for reasoning agents: Regression-based GOLOG and progression-based FLUX.

1 Introduction

An intriguing application of logic as a formal model of rational thought is to endow artificial systems with the ability to reason. Software agents and autonomous robots exhibit rational behavior as a result of reasoning about the effects of their actions based on an abstract, symbolic model of their environment. This approach to Artificial Intelligence is inherently connected with the famous Frame Problem of how to axiomatize the effects of actions in a concise way so as to enable an automated agent to infer what has and what has not changed after a sequence of actions [6, 7].

Throughout its history, the Frame Problem has initiated many important developments—a prominent example is nonmonotonic logic [2]—but satisfactory solutions did not emerge until the past decade. These solutions have recently evolved into declarative, high-level programming languages and systems which can be used to create reasoning agents and robots. The core of

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 323–336

each such system is its underlying inference schema for solving the Frame Problem. These inference schemata come in two different flavors.

In a *regression-based* solution to the Frame Problem, the question whether a property φ holds after the agent has performed a sequence of actions, is reduced to the question whether another property $\mathcal{R}[\varphi]$ (the *regression* of φ) holds after the last but one action. This reduction is applied recursively through the whole sequence, so that in the end the fully regressed formula can be checked against what was initially true.

In a *progression-based* solution to the Frame Problem, a (possibly incomplete) initial world model is updated upon the performance of an action. In this way, the model is progressed through an action sequence executed by the agent, and the current model is used directly to decide whether a property φ holds in the current situation. We argue that this principle is mandatory for the efficient control of agents over extended periods of time. To support this claim, we analyze and compare the computational behavior of the regression-based logic programming system GOLOG [4] with progression-based FLUX [13]. Our analysis shows that when the former is used, the computational effort continually increases as a program proceeds, whereas the latter system scales up effortlessly to long-term control.

The remainder of this paper is organized as follows. In the next section, we compare the two principles of regression and progression in the context of logic-based agents. In Section 3 we present and analyze experimental results with GOLOG and FLUX applied to a mail delivery problem which requires to reason about action sequences of non-trivial length. We conclude in Section 4. We assume that the reader is familiar with basic notations of logic programming and Prolog (as can be found, e.g., in [1]). Lack of space does also not permit to give a full explanation of syntax and semantics of GOLOG and FLUX; we refer to [4, 9] and [13], respectively.

2 Progression vs. Regression

Consider a robot whose task is to pick up and deliver mail packages exchanged among a number of offices. The robot is equipped with several slots, a kind of mail bag, each of which can be filled with one such package. Figure 1 depicts a sample scenario in an environment consisting of six offices and a robot with three mail bags. A simple, general strategy for the robot is to deliver packages whenever it finds itself at some office for which it carries

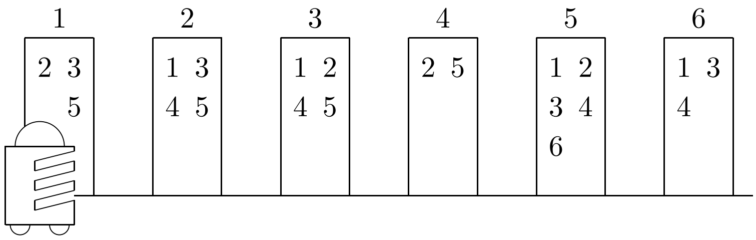


Figure 1: The initial state of a sample mail delivery problem, with a total of 21 delivery requests.

mail, then pick up packages whenever it happens to be at some place where items are still waiting to be collected, and finally move either up or down the hallway toward an office where a package can be picked up or delivered. This strategy is implemented by the following semi-formal algorithm:

loop
  if possible to deliver a package
      then do it
  else if possible to pick up a package
      then do it
  else if can pick up or deliver a package up (resp. down) the hallway
      then go up (resp. down)
  else stop
end loop

This algorithm obviously requires the robot to evaluate conditions which depend on the current state of the environment. For in order to decide on its next action, the robot always needs to know the current contents of its mail bags, the requests that are still open, and its current location. Since these properties constantly change as the program proceeds, the robot has to keep track of what it does as it moves along. For this purpose, it needs an internal representation of the environment, which throughout the execution of the program conveys the necessary information about the current location of all packages that have not yet been delivered. Logical reasoning on the basis of this model allows the robot to decide which actions are possible and how the model needs to be updated after each action in accordance with the effects of the action. With regard to the scenario in Figure 1, for instance,

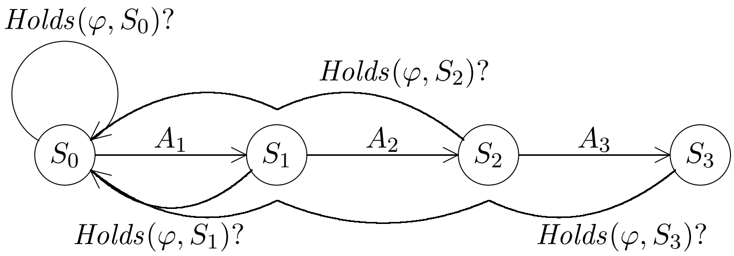


Figure 2: In regression-based solutions to the Frame Problem, the question whether a property φ holds in a situation S_i is decided by regressing φ through the actions that lead from the initial situation S_0 to S_i.

---

the robot needs to be able to conclude that it can start with putting one of the three available packages into one of its mail bags. Furthermore, the robot needs to infer that after this action, the package is in one of the mail bags while the other two bags are still empty. Hence, the robot has to cope with the Frame Problem [6].

In a *regression-based* inference schema for solving the Frame Problem [8], the question whether a property φ holds after a particular action, is reduced to the question whether another property $\mathcal{R}[\varphi]$ (the *regression* of φ) holds before the action. This reduction is applied recursively through all actions the agent has performed thus far, so that in the end the fully regressed formula can be checked against the initial world model. Figure 2 gives a schematic illustration of this principle. The graph shows that in general the effort of examining the validity of a property depends on the length of the history. As a consequence, the computational behavior of a regression-based agent program can be expected to worsen the longer the program runs.

The family of GOLOG dialects rooted in [4] is an example of regression-based implementations. The effects of actions are encoded by *successor state axioms* [8], which are of the form

$$Holds(f, Do(a, s)) \leftrightarrow \Phi_f(a, s) \tag{1}$$

Here, f is an atomic property, a so-called *fluent*, and $Do(a, s)$ denotes the *situation*, i.e., sequence of actions, reached by performing action a in situation s. Formula Φ_f describes the conditions on action a and situation s under which f can be concluded to hold in the successor situation $Do(a, s)$.

As an example, consider the following successor state axiom, given in Prolog notation, for the fluent *Empty*(b), that is, the property of mail bag b to be empty:

```
holds(empty(B),do(A,S)) :- A=deliver(B)
                           ;
                           holds(empty(B),S),
                              not A=pickup(B,R).
```

This axiom says that mail bag b is empty after performing an action a in a situation s just in case the action was to deliver the contents of bag b, or mail bag b happened to be empty in situation s and the action was not to pick up into b a package for some room r. The atom $\mathit{Holds}(\mathit{Empty}(b), s)$ in the right hand side is solved recursively until the situation argument s is reduced to the initial situation S_0. In this way, the computational effort for deciding whether *Empty*(b) holds depends on the number of actions performed thus far. As a consequence, the time it takes for a GOLOG agent to make a decision can be expected to increase with every action the agent takes.

In a *progression-based* inference schema for solving the Frame Problem [5, 12], a (possibly incomplete) initial world model is updated upon the performance of an action. In this way, the model is progressed through the action sequence performed by the agent, and the updated model is used directly to decide whether a property holds in the current situation. Figure 3 gives a schematic illustration of this principle. The graph shows that the effort of examining the validity of a property is independent of the length of the history. As a consequence, the computational behavior of a progression-based agent program should be expected to remain the same throughout the execution so that this principle has the potential to scale up to long-term control.

FLUX [13] is an example of a progression-based implementation. World models, so-called *states*, are encoded as lists of fluent terms, possibly accompanied by constraints for negative and disjunctive state knowledge. The effects of actions are encoded by *state update axioms* [12], which are of the form

$$\mathit{StateUpdate}(z_1, a, z_2) \;\leftarrow\; \Phi_a(z_1, z_2)$$

Here, formula Φ_a describes the conditions under which z_2 is the state reached by performing action a in state z_1. As an example, consider the

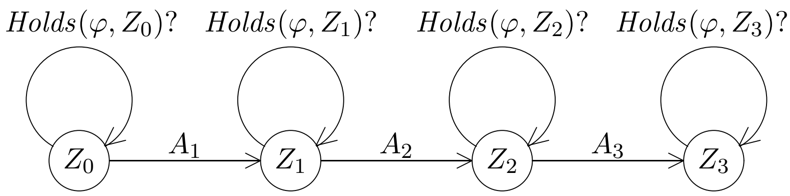


Figure 3: In progression-based solutions to the Frame Problem, the world model Z_i is progressed through the next action in every situation. A property φ can then be decided directly wrt. the current world model.

---

following state update axiom[1] for the action *Deliver*(b) of delivering the contents of mail bag b:

```
state_update(Z1,deliver(B),Z2) :-
  holds(at(R),Z1), update(Z1,[empty(B)],[carries(B,R)],Z2).
```

This axiom says that state z_2 is the result of performing a *Deliver*(b) action in state z_1 if the robot is at room r in z_1, and z_2 is the result of updating z_1 by the positive effect that bag b becomes empty and the negative effect that the robot no longer carries in bag b a package for room r. When executing a FLUX program, conditions of the form *Holds*(φ, z) are always evaluated against the current world model. Since the computational effort for this evaluation is independent of the actions that have been performed thus far, the time it takes for a FLUX agent to make a decision is expected to remain the same as the program proceeds.

3 Progressive FLUX vs. Regressive GOLOG

In order to see how the theoretical differences between regression-based and progression-based implementations manifest in practice, we have applied both GOLOG and FLUX to mail delivery problems which require to reason about action sequences of non-trivial length. We use four fluents to describe a state in the mail delivery world: *At*(r) to represent that the robot is at room r; *Empty*(b) to represent that the robot's mail bag b is

[1] The standard FLUX predicate `update(Z1,P,N,Z2)` used below represents the update of state z_1 to state z_2 by positive effects p and negative effects n.

empty; *Carries*(b, r) to represent that the robot carries in bag b a package for room r; and *Request*(r, r') to indicate a delivery request from room r to room r'. The following logic programming clauses, for example, constitute a GOLOG specification of the initial situation depicted in Figure 1:

```
holds(at(1),s0).
holds(empty(bag1),s0).
holds(empty(bag2),s0).
holds(empty(bag3),s0).
holds(request(1,2),s0).
...
holds(request(6,4),s0).
```

The three elementary actions of the mail agent are: *Pickup*(b, r) to pick up into bag b a package for room r; *Deliver*(b) to deliver the contents of bag b at the current location; and *Go*(d) to move $d = Up$ or $d = Down$ the hallway to the next room. Using GOLOG syntax, where *Poss*(a, s) means that action a is possible in situation s, the following is a suitable definition of the action preconditions in the mail delivery world:

```
poss(pickup(B,R),S) :- holds(empty(B),S), holds(at(R1),S),
                       holds(request(R1,R),S).

poss(deliver(B),S)  :- holds(at(R),S), holds(carries(B,R),S).

poss(go(D),S)       :- holds(at(R),S),
                       ( D=up, R<6 ; D=down, R>1 ).
```

Verifying the executability of an action is a vital aspect of executing the agent program for the mail delivery robot. The effects of the actions are encoded by the successor state axioms given in Appendix A.

With the help of this background theory, our strategy for the mail delivery robot given at the beginning of Section 2 translates into the following recursive GOLOG procedure:[2]

```
proc(main_loop, [deliver(B),main_loop]  #
                [pickup(B,R),main_loop] #
                [continue,main_loop]    # []).
```

[2]For details regarding syntax and semantics of GOLOG, we refer to [4, 9].

```
proc(continue,
 [ [?(empty(B)),?(request(R1,R2))] # ?(carries(B,R1)),
   ?(at(R)), [?(less(R,R1)),go(up)] # go(down) ]).

holds(less(R1,R2),S) :- R1<R2.
```

The auxiliary procedure *Continue* succeeds if there is the possibility for the robot to pick up or deliver mail somewhere up or down the hallway. If neither a *Deliver*(b) nor a *Pickup*(b, r) action is possible, and if the robot needs not continue to another office, then the program terminates.

In FLUX, the initial state of Figure 1 is encoded by this clause:

```
init(Z0) :- Z0 = [at(1),empty(bag1),empty(bag2),empty(bag3),
                  request(1,2),...,request(6,4)].
```

The specification of the precondition axioms is the same as in GOLOG while the effects of the three actions are encoded by the state update axioms given in Appendix B.

The following FLUX program implements the same algorithm as the GOLOG procedure for the mail robot:

```
main :- init(Z), main_loop(Z).

main_loop(Z) :- poss(deliver(B),Z)
                   -> execute(deliver(B),Z,Z1), main_loop(Z1)
              ; poss(pickup(B,R),Z)
                   -> execute(pickup(B,R),Z,Z1), main_loop(Z1)
              ; continue(Z,Z1)
                   -> main_loop(Z1)
              ; true.

continue(Z,Z1) :- ( holds(empty(B),Z), holds(request(R1,R2),Z)
                    ; holds(carries(B,R1),Z) ),
                  holds(at(R),Z),
                  ( R<R1 -> execute(go(up),Z,Z1)
                          ; execute(go(down),Z,Z1) ).
```

Both the GOLOG and the FLUX program are available for download from our web page `www.fluxagent.org`. We ran a series of experiments

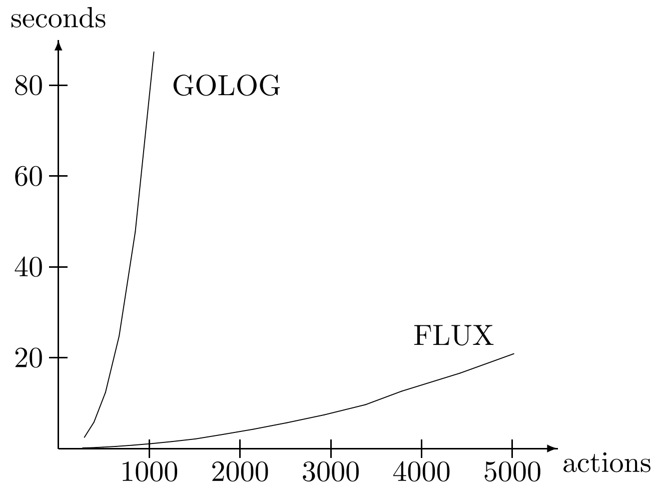


Figure 4: Overall runtime of the mail delivery program in GOLOG and FLUX (vertical axis) depending on the solution length (horizontal axis).

with maximal delivery problems, that is, with initial requests from every office to every other. The following table shows the resulting lengths of the action sequences for all problem sizes from $n = 10$ offices up to $n = 30$ and with a robot with three mail bags:[3]

n	# act	n	# act	n	# act
10	492	17	2144	24	5658
11	640	18	2516	25	6352
12	814	19	2928	26	7100
13	1016	20	3382	27	7904
14	1248	21	3880	28	8766
15	1512	22	4424	29	9688
16	1810	23	5016	30	10672

Figure 4 shows the runtime of the two programs in relation to the length of the solution. The experiments were carried out on a standard PC with

[3]We have kept the value for k constant because while it influences the overall number of actions needed to carry out all requests, this parameter turned out to have negligible influence on the computational effort needed for action selection and effect computation.

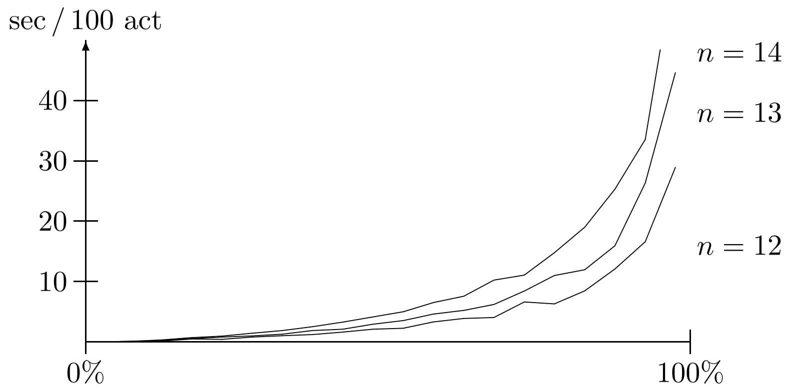


Figure 5: The computational behavior of the GOLOG program for the mail delivery problem in the course of its execution. The horizontal axis depicts the degree to which the run is completed while the vertical scale is in seconds per 100 actions.

---

a 500 MHz processor. A detailed analysis of the computational behavior as the two programs proceed shows that the superiority of FLUX is mainly due to its progressive solution to the Frame Problem: Figure 5 depicts, for three selected problem sizes, the average action selection time in the course of the execution of the GOLOG program. The curves show that the computational effort increases polynomially as the program runs, which is a consequence of the regression-based solution to the Frame Problem. Figure 6 depicts the average time for action selection and state update computation in the course of the execution of the FLUX program, again for three selected problem sizes. The curves show that the computational effort remains essentially constant throughout, thanks to the progression-based solution to the Frame Problem. The slight general descent can be explained by the decreasing state size due to fewer remaining requests.

4 Discussion

We have argued that progression-based solutions to the Frame Problem are necessary for logic-based agents that need to reason about action sequences of non-trivial length: By continually updating their internal model of the environment, agents can evaluate properties directly at every stage. In contrast, regression-based solutions to the Frame Problem give rise to a computational

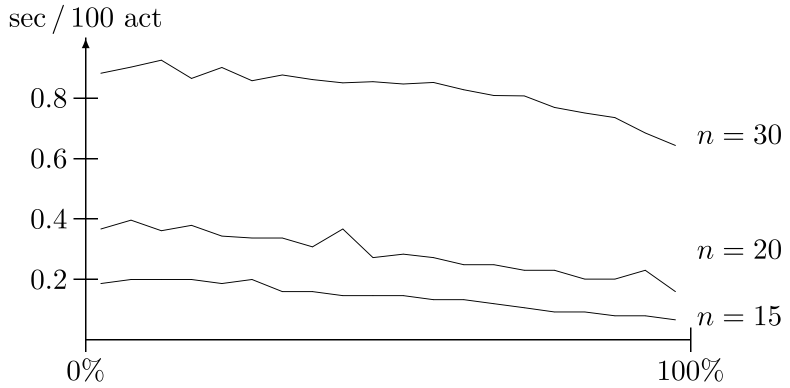


Figure 6: The computational behavior of the FLUX program for the mail delivery problem in the course of its execution. The horizontal axis depicts the degree to which the run is completed while the vertical scale is in seconds per 100 actions.

---

effort for evaluating properties which increases with every action taken by the agent. In the long run, the polynomial effort for regression worsens the complexity of any polynomial algorithm for agent control. We have shown how this difference manifests in practice by comparing regression-based GOLOG with progression-based FLUX on a problem which requires to reason about several hundreds or thousands of actions.

A prominent alternative to GOLOG, the implementation [11] of the event calculus [10] is essentially regression-based just as well: In order to verify that a property holds at some time t, it must be proved that this property was initiated by some previous event and that no event in between terminated this property. This, too, requires to take into account the history of events (i.e., actions) when examining the validity of a property, so that again the computational behavior of a control program can be expected to worsen with every action taken by the agent.

In FLUX, the notion of a history of actions serves different purposes: It is used to give semantics to program execution and to endow agents with the ability of planning. As argued in [3], since planning is a computationally demanding problem, it should be restrictively employed in agent programs and interleaved with action execution. By combining progression with much of GOLOG's powerful concept for plan search control, FLUX combines the best of both worlds.

5 Bibliography

[1] Krzysztof Apt. *From Logic Programming to Prolog.* Prentice-Hall, 1997.

[2] Daniel G. Bobrow, editor. *Artificial Intelligence 13: Special Issue on Non-Monotonic Reasoning.* Elsevier, 1980.

[3] Giuseppe De Giacomo and Hector J. Levesque. An incremental interpreter for high-level programs with sensing. In H. Levesque and F. Pirri, editors, *Logical Foundations for Cognitive Agents*, pages 86–102. Springer, 1999.

[4] Hector J. Levesque, Raymond Reiter, Yves Lespérance, Fangzhen Lin, and Richard B. Scherl. GOLOG: A logic programming language for dynamic domains. *Journal of Logic Programming*, 31(1–3):59–83, 1997.

[5] Fangzhen Lin and Ray Reiter. How to progress a database. *Artificial Intelligence*, 92:131–167, 1997.

[6] John McCarthy and Patrick J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502, 1969.

[7] Zenon W. Pylyshyn, editor. *The Robot's Dilemma: The Frame Problem in Artificial Intelligence.* Ablex, Norwood, New Jersey, 1987.

[8] Ray Reiter. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation*, pages 359–380. Academic Press, 1991.

[9] Raymond Reiter. *Logic in Action.* MIT Press, 2001.

[10] Murray Shanahan. *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia.* MIT Press, 1997.

[11] Murray Shanahan and Mark Witkowski. High-level robot control through logic. In C. Castelfranchi and Y. Lespérance, editors, *Proceedings of the International Workshop on Agent Theories Architectures and Languages (ATAL)*, volume 1986 of *LNCS*, pages 104–121, Boston, MA, July 2000. Springer.

[12] Michael Thielscher. From situation calculus to fluent calculus: State update axioms as a solution to the inferential frame problem. *Artificial Intelligence*, 111(1–2):277–299, 1999.

[13] Michael Thielscher. FLUX: A logic programming method for reasoning agents. *Theory and Practice of Logic Programming*, 2004.

A Successor State Axioms in GOLOG

```
holds(at(R),do(A,S)) :- A=go(up),   holds(at(R1),S),
                                    R is R1+1
                      ; A=go(down), holds(at(R1),S),
                                    R is R1-1
                      ; not A=go(D), holds(at(R),S).

holds(empty(B),do(A,S)) :- A=deliver(B)
                           ;
                           holds(empty(B),S),
                              not A=pickup(B,R).

holds(carries(B,R),do(A,S)) :- A=pickup(B,R)
                               ;
                               holds(carries(B,R),S),
                                  not A=deliver(B).

holds(request(R,R1),do(A,S)) :- holds(request(R,R1),S),
                                ( A=pickup(B,R1)
                                     -> holds(at(R2),S),
                                        R2\=R
                                  ; true ).
```

B State Update Axioms in FLUX

For the sake of simplicity and because our example domain does not involve any sensing actions, we have omitted the argument for sensory input, which is required for general update axioms [13].

```
state_update(Z1,pickup(B,R),Z2) :-
   holds(at(R1),Z1),
   update(Z1,[carries(B,R)],[empty(B),request(R1,R)],Z2).

state_update(Z1,deliver(B),Z2) :-
  holds(at(R),Z1), update(Z1,[empty(B)],[carries(B,R)],Z2).

state_update(Z1,go(D),Z2) :-
  holds(at(R),Z1), ( D=up -> R1 is R+1 ; R1 is R-1 ),
  update(Z1,[at(R1)],[at(R)],Z2).
```

A Version of the Second Incompleteness Theorem For Axiom Systems that Recognize Addition But Not Multiplication as a Total Function

Dan E. Willard
dew@cs.albany.edu
SUNY Albany Computer Science Department

ABSTRACT: Let $A(x,y,z)$ and $M(x,y,z)$ denote predicates indicating $x+y=z$ and $x*y=z$ respectively. Let us say an axiom system α *recognizes* Addition and Multiplication both as **Total Functions** iff it can prove:

$$\forall x \forall y \exists z \; A(x,y,z) \quad \textbf{AND} \quad \forall x \forall y \exists z \; M(x,y,z) \tag{1}$$

We will introduce some new variations of the Second Incompleteness Theorem for axiom systems which recognize Addition as a "total" function but which treat Multiplication as only a 3-way relation. These generalizations of the Second Incompleteness Theorem are interesting because our prior work [30, 32, 34] has explored several types of boundary-case exceptions to the Second Incompleteness Theorem that occur when one weakens the the hypothesis for our main theorems only slightly further.

1 Introduction

The Second Incompleteness Theorem states that sufficiently strong axiom systems are unable to formally verify their own consistency. There has been

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 337–368

an extensive amount of research about how the Second Incompleteness Theorem can be generalized for weak axiom systems in the context of Frege and Hilbert deduction. For instance, Bezboruah-Shepherdson [3] showed how a version of the Second Incompleteness Theorem is valid for an axiom system that Tarski-Mostowski-Robinson [24] called Q. Pudlák [16] proved a more robust version of the Second Incompleteness Theorem, that applied to all extensions of Q and to all versions of its Gödel encoding (including very importantly localized versions on Definable Cuts) involving Frege-Hilbert style deduction. Wilkie-Paris [29] developed several examples of other versions of the Second Incompleteness Theorems involving for example the inability of $I\Sigma_0 + Exp$ to prove the consistency of Q. Solovay [21] observed how one could modify the formalism of Pudlák [16] with some techniques used by Nelson and Wilkie-Paris [11, 29] to obtain the following result:

> $*$ *(Solovay's Extension of Pudlák's Generalization of the Second Incompleteness Theorem [16, 21])* No reasonable consistent axiom system α treating *merely Successor* as a total function (and viewing Addition and Multiplication as 3-way relations) can recognize the assured non-existence of a Frege-Hilbert style proof of 0=1 employing α's axioms.

Our research was greatly stimulated by the Theorem $*$ resulting from the joint work of Nelson, Pudlák, Solovay and Wilkie-Paris. It differs from this theorem's formalism mainly by examining a Semantic Tableaux style of a proof rather than a Frege-Hilbert methodology.

It turns out that some results from Theorem $*$ concerning Frege-Hilbert deduction generalize for Semantic Tableaux, but other aspects of it do not generalize. Our research is also relevant to some open questions raised by Paris and Wilkie [14] concerning the characterization of the exact circumstances where the Semantic Tableaux version of the Second Incompleteness Theorem applies to *weak axiom systems.*

There have been a substantial amount of research in recent years [1, 2, 4, 7, 8, 9, 10, 13, 14, 16, 17, 18, 19, 23, 25, 26, 27, 29, 33] into this topic. Our contribution in [30, 32, 34] was the focus on axiom systems α which *drop* Equation (1)'s assumption that Multiplication is a total function. (Most of [30, 32, 34]'s systems regarded only Addition as total and viewed Multiplication as a 3-way relation.) This topic is interesting because significant boundary case exceptions for the Semantic Tableaux version of the Second

Incompleteness Theorem can be found when an axiom system recognizes merely Addition as *formally total.*

One of the strongest types of boundary-case exceptions to the Second Incompleteness was presented in [34]. It introduced a hierarchy of several increasingly strong definitions of Semantic Tableaux consistency and demonstrated that the prior systems of [30, 32] could be improved so that they had a capacity to recognize their self-consistency under a "Level(1)" definition of Semantic Tableaux consistency (rather than using [30, 32]'s weaker "Level-zero" type definition). The theme of [34] was essentially that all the different levels of definition of Semantic Tableaux consistency were logically equivalent to each other from the standpoint of a strong enough axiom systems. However, a weak axiom system is typically unable to recognize the equivalence of these different levels of definition. Using this fact, [34] demonstrated that some axiom systems were able to evade the force of the Second Incompleteness Theorem when the level-number of their associated definition of Semantic Tableaux consistency was made sufficiently small. In essence, our goal in this paper and in our prior work is to attempt to characterize in as much detail as possible the maximum level number where such an evasion of the Second Incompleteness Theorem is feasible.

Section 2 of this paper will review our definition of different levels of Semantic Tableaux consistency. Our chief goal will be to show that the Semantic Tableaux version of the Second Incompleteness Theorem becomes valid at what is called "Level(2+)" for essentially all axiom systems that recognize merely Addition as a total function. This result is significant because there is a very narrow gap between the Level(1) — where [34] showed the Second Incompleteness Theorem can be evaded when Multiplication is treated as a 3-way relation — and the Level(2+) where the Second Incompleteness Theorem takes force.

There is also a second variant of our results (in both their positive and complementary negative forms) that involves a what we call a TabList style of deduction, described later at the end of the next section.

2 Formal Statement of Main Theorems

Some added notation is needed to define our leveled hierarchy and the various new theorems we will present. Define the mapping $F(a_1, a_2 ... a_j)$ to be a **Non-Growth** function iff $F(a_1, a_2, ...a_j) \leq Maximum(a_1, a_2, ...a_j)$ for all values of $a_1, a_2, ...a_j$. Six examples of non-growth functions are *Integer Subtraction* (where $x-y$ is defined to equal zero when $x \leq y$), *Integer Division* (where $x \div y$ is defined to equal x when $y = 0$, and it equals $\lfloor x/y \rfloor$ otherwise), $Maximum(x, y)$, $Logarithm(x)$, $Root(x, y) = \lceil x^{1/y} \rceil$ and $Count(x, j)$ designating the number of "1" bits among x's rightmost j bits. The term **U-Grounding Function** will refer to this set of six non-growth functions plus the Growth operations of Addition and $Double(x) = x + x$.

All our results in this paper will technically be couched in terms of a language that houses function symbols for the eight operations defined above. This notation is technically unnecessary — because a system that uses the combination of Equation (2) (which implies Addition and Doubling are total functions) along with only our first six non-growth U-Grounding operations would have properties similar to the U-Grounding language.

$$\forall x \, \forall y \, \exists z \quad x = z - y \tag{2}$$

However, it is much easier to present a short proof of our results if our language does employ two additional function symbols for the operations of Addition and $Double(x) = x + x$. The virtue of this notation convention is that it allows us to formally employ constant symbols only for the numbers of 0 and 1 and to encode every other integer $N \geq 2$ using no more than $2 \cdot \text{Log}_2 N$ appearances of the function symbols for Addition and Doubling applied to these two input symbols. Such a mathematical term will be henceforth called the **U-Grounded Binary Representation of N** and be denoted as $\overline{N}$. For instance, 25 can be encoded in a "binary-like" form as: $1 + \text{Double(Double(Double(} 1 + \text{Double(1))))}$.

The use of logic's conventional notation about Π_n and Σ_m sentences is technically inappropriate in this paper because the latter notation (with its Multiplication Function symbol) is suitable only for axiom systems which recognize Multiplication as a total function. Instead, our analogs for Π_n and Σ_m in the U-Grounding Function language are called Π_n^* and Σ_m^*. Here, a *term* t is defined to be a constant, variable or a U-Grounding function symbol (whose input arguments are recursively defined terms). Also, the quantifiers in the wffs $\forall \, v \leq t \;\; \Psi(v)$ and $\exists \, v \leq t \;\; \Psi(v)$ are called *bounded quantifiers*. If Φ is a formula that uses the U-Grounding primitives as its function symbols and the two relation symbols of " $=$ " and " $\leq$ ", then this formula will be called both Π_0^* and Σ_0^* whenever all its quantifiers are bounded. For $n \geq 1$, a formula Υ shall be called Π_n^* iff it is written in the form $\forall v_1 \, \forall v_2 \, ... \, \forall v_k \quad \Phi,$ where Φ is Σ_{n-1}^* . Likewise, Υ is called Σ_n^* iff it is written in the form $\exists v_1 \, \exists v_2 \, ... \, \exists v_k \quad \Phi,$ where Φ is Π_{n-1}^* .

Let us call Υ a Q_n^* sentence iff it is one of a Σ_n^* sentence, a Π_n^* sentence or a Boolean combination of several Σ_n^* and Π_n^* sentences (using the standard connective symbols of $\wedge$, $\vee$, $\neg$ and $\rightarrow$). There will be three types of definitions of Semantic Tableaux consistency that we will examine in this paper. They are defined below:

1. A **Level(n) Definition** of an axiom system α's Tableaux consistency is the declaration that there exists no Π_n^* sentence Υ supporting simultaneous Semantic Tableaux proofs from α of both Υ and its negation.

2. A **Level(n+) Definition** of an axiom system α's Tableaux consistency is the statement that no Q_n^* sentence Υ supports simultaneous Semantic Tableaux proofs for both Υ and its negation.

3. A **Level(0-) Definition** of a system α's Tableaux consistency is the statement that there exists no proof of 0=1 from α.

All definitions of consistency, from Level(0-) up to Level(n+) *for any* n , are equivalent to each other under strong enough models of Arithmetic. However, many weak axiom systems do not have a mathematical strength to formally prove and recognize this equivalence.

Translated into this Π_1^* styled notation convention, the core result in [34] was the construction of a consistent axiom system α which had the following properties:

1. α was capable of recognizing its Level(1) Tableaux consistency.

2. α was capable of recognizing Addition as a total function.

3. α was capable of proving all of Peano Arithmetic's Π_1^* theorems.

The above result beckons one to consider whether or not it would be possible to develop stronger versions of this effect where α could recognize higher levels of its own Tableaux consistency. Theorem 1 and Remark 1 show that most such generalizations are infeasible.

Theorem 1. *Let* α *denote an axiom system that uses the language of the U-Grounding functions (and thus recognizes Addition as a total function). There exists a* Π_1^* *theorem* W *of Peano Arithmetic such that no consistent* $\alpha \supset W$ *of finite cardinality can recognize its own Level(2+) Tableaux consistency.*

Remark 1. It is also possible to generalize Theorem 1 for essentially all axiom system $\alpha \supset W$ of infinite cardinality. In particular, let us say an axiom system α satisfies the **Conventional Deciphering Property** iff there exists a Σ_0^* sentence Test(n) such that n represents the Gödel number of an axiom of α iff and only iff Test(n) is true. We will not have the page space to prove this stronger result here, but Theorem 1 can be strengthened to indicate that no consistent axiom system $\alpha \supset W$ satisfying the Conventional Deciphering Property can prove a theorem affirming its own Level(2+) Tableaux consistency.

Remark 2. The Incompleteness effect described by Theorem 1 and Remark 1 should not be confused with a prior result published by us in [33]. The latter's version of the Incompleteness Theorem showed that there were essentially no interesting axiom systems that could simultaneously recognize their Level(0-) Tableaux consistency and also recognize both *Addition and Multiplication* as total functions. This alternate effect is quite different from our current result, which will apply to axiom systems that recognize *solely Addition* as a total function.

In particular, this distinction is non-trivial essentially because [34] showed that axiom systems could recognize their Level(1) consistency when they treated Multiplication as a 3-way relation (rather than as a total function). Hence, there arises the question concerning at what Level of Tableaux consistency does the Semantic Tableaux version of the Second Incompleteness become valid for this latter class of axiom systems which do not employ [33]'s assumption that Multiplication is a total function? The purpose of Theorem 1's generalization of the Second Incompleteness Theorem is to at least partially answer this question. It shows that while [33]'s Tableaux generalization of the Second Incompleteness Theorem is known from [34] to become false at all levels between $0-$ and 1 when a system fails to recognize Multiplication as total, there is nevertheless available a Level(2+) Tableaux generalization of the Second Incompleteness Theorem for such systems.

We will also discuss in this paper a cousin of Theorem 1's Incompleteness result, involving an alternative rule of inference, called TabList deduction. It is helpful to review Smullyan's formal definition of a Semantic Tableaux proof before formally defining TabList deduction.

Following roughly Fitting's or Smullyan's notation [5, 20], let us define a Φ-**Based Candidate Tree** for the axiom system α to be a tree whose root corresponds to the sentence $\neg\,\Phi$ and whose all other nodes are either axioms of α or deductions from higher nodes of the tree. Let the notation " $\mathcal{A} \implies \mathcal{B}$ " indicate that $\mathcal{B}$ is a valid deduction when $\mathcal{A}$ is an ancestor of $\mathcal{B}$. In this notation, the Tableaux-Deduction rules are:

1. $\Upsilon \wedge \Gamma \implies \Upsilon$ and $\Upsilon \wedge \Gamma \implies \Gamma$.
2. $\neg\neg\,\Upsilon \implies \Upsilon$. Other deduction rules for the $\neg$ symbol include: $\neg(\Upsilon \vee \Gamma) \implies \neg\Upsilon \wedge \neg\Gamma$, $\neg(\Upsilon \rightarrow \Gamma) \implies \Upsilon \wedge \neg\Gamma$, $\neg(\Upsilon \wedge \Gamma) \implies \neg\Upsilon \vee \neg\Gamma$, $\neg\,\exists v\,\Upsilon(v) \implies \forall v \neg\,\Upsilon(v)$ and $\neg\,\forall v\,\Upsilon(v) \implies \exists v\,\neg\Upsilon(v)$
3. A pair of sibling nodes Υ and Γ is allowed when their ancestor is $\Upsilon \vee \Gamma$.
4. A pair of sibling nodes $\neg\Upsilon$ and Γ is allowed when their ancestor is $\Upsilon \rightarrow \Gamma$.
5. $\exists v\,\Upsilon(v) \implies \Upsilon(u)$ where u is a newly introduced Parameter Symbol.
6. $\forall v\,\Upsilon(v) \implies \Upsilon(t)$ where t denotes a "Function Term". These terms are U-Grounding Function objects, whose inputs are any set of constant symbols, parameter symbols or other function-objects.

Define a particular leaf-to-root branch in a candidate tree T to be **Closed** iff it contains both some sentence Υ and its negation $\neg\,\Upsilon$. A **Semantic Tableaux** proof of Φ is then defined [5, 20] to be a candidate tree whose root stores the sentence $\neg\Phi$ and all of whose root-to-leaf branches are closed.

One further definition is needed before we can describe the second variation of Theorem 1's Incompleteness Result explored in this paper. Let H denote a sequence of ordered pairs (t_1, p_1), (t_2, p_2), ... (t_n, p_n) , where p_i is a Semantic Tableaux proof of the theorem t_i , and let $\Re$ denote an arbitrary class of sentences. Define H to be a Tab$-\Re-$List proof of a theorem T from the axiom system α iff $T = t_n$ and also:

1. Each axiom in p_i's proof is either one of $t_1, t_2, ...t_{i-1}$ or comes from α.

2. Each of the "intermediately derived theorems" $t_1, t_2, ...t_{n-1}$ must lie within the "prespecified class" $\Re$ of sentences.

If $\Re$ denotes the set of Q_K^* sentences, the notion of an Tab$-Q_k^*-$List proof is quite similar (although not fully identical) to constructs that have been called R-proofs and Q_k style proofs by Hájek, Paris, Pudlák and Wilkie in [7, 16, 29]. One minor difference between these definitions is that the TabList notion contains some added flexibility because it allows one to set $\Re$ equal to any of the classes of Π_k^* sentences, Σ_k^* sentences, Q_K^* sentences, or for example the union of the sets of Π_k^* and Σ_k^* sentences. Another difference is that the R-proofs and Q_k style proofs of [7, 16, 29] are based on partially limiting the power of Hilbert-style deduction, whereas our dual form of this construct proceeds in the opposite direction — where we seek to progressively expand the logical power of Semantic Tableaux style deduction instead.

For any class $\Re$ of sentences, each of our prior definitions of Level(0), Level(N) and Level(N+) consistency can be generalized for Tab$-\Re-$List deduction. For instance, an axiom system α's Level(0$-$) consistency under Tab$-\Re-$List deduction is the statement that every Tab$-\Re-$List proof from α's axioms fails to prove 0=1.

Below are our main theorems about the generality and limitations of the Second Incompleteness Theorem under TabList deduction.

Theorem 2. *Let α denote an arbitrary axiom system that uses the language of the U-Grounding functions (and thus recognizes Addition as a total function). It is not necessary, but for the sake of simplifying our proof of Theorem 2 we will also assume that the axiom system α has finite cardinality. Then there exists two Π_1^* theorems of Peano Arithmetic, V_A and V_B such that*

A. No consistent $\alpha \supset V_A$ can prove a theorem affirming its own Level(0-) consistency under Tab$-\Pi_2^*-$List deduction.

B. No consistent $\alpha \supset V_B$ can prove a theorem affirming its own Level(0-) consistency under Tab$-\Sigma_2^*-$List deduction.

Theorem 3. *Let Tab$_1$List be an abbreviation for the variant of Tab$-\Re-$List deduction where $\Re$ denotes the union of the set of Π_1^* and Σ_1^* sentences. Then for each consistent axiom system A that is an extension of Peano Arithmetic, there exists a consistent axiom system α that can*

1. *recognize its own Level(1) consistency under* Tab_1List *deduction.*

2. *recognize the validity of all A's* Π_1^* *theorems, and*

3. *recognize Addition as a total function*

Part of the reason Theorems 2 and 3 are interesting is due to the close match between their complementary positive and negative results. Thus, Theorem 3 established that there exists a Boundary-Case exception to the Second Incompleteness Theorem when $\Re$ represents the union of the set of Π_1^* and Σ_1^* sentences, while Theorem 2 shows the Second Incompleteness Theorem comes to force when $\Re$ represents instead either the class of Π_2^* sentences or the class of Σ_2^* sentences. Moreover, Theorem 3 indicates that its Boundary-Case exception rises up to Level(1) definitions of consistency, while Theorem 2 shows that even the lower Level(0−) is problematic under Tab−Π_2^*−List and Tab−Σ_2^*−List deduction.

Most of our discussion in this paper will focus on proving Theorems 1 and 2. Theorem 3's result was technically announced on the last page of our Tableaux-2002 conference paper [34]. However, the latter conference paper was written in a too abbreviated style for it to also include a proof of Theorem 3. Instead, its formal proof examined a slightly more specialized variation of Theorem 3 where Semantic Tableaux deduction replaced Tab_1List deduction (in Clause 1 of Theorem 3). We have therefore also inserted a 3-page appendix into the current article, which roughly outlines how [34]'s proof formalism can be slightly strengthened to obtain Theorem 3's more general result. This appendix is helpful because there is a pleasantly tight and sharp match between Theorem 3's positive result and Theorem 2's complementary negative result, as was explained in the prior paragraph.

3 Overall structure of Theorem 1's Proof

Our method for encoding a Semantic Tableaux proof p is described on page 581 of our article [32]. This encoding is maximally compressed in that it will encode p as an integer whose **"Bit-Length"** is approximately proportional to the length of such a Semantic Tableaux proof when it is written down by hand. Several other authors [7, 29] have also employed roughly similar types of maximally compressed encoding methods. It is therefore probably unnecessary for a reader to examine our exact encoding method in [32].

This section will sketch the overall structure of Theorem 1's proof. Let $\mathrm{Prf}_\alpha(x,y)$ denote a Σ_0^* formula indicating y is a semantic tableaux proof of the theorem x from the axiom system α. Also, let $\mathrm{Log}(z)$ denote Base-2 Logarithm, with *downwards rounding* to the lowest integer, and $\mathrm{Log}^\lambda(z)$ denote the operation $\mathrm{Log}(\mathrm{Log}(\mathrm{Log}...(\mathrm{Log}(z))))$ — where λ designates an integer indicating the number of iterations of Log here. It is useful to employ the notation $\mathrm{ShortPrf}_\alpha^\lambda(x,y,z)$ to denote a Σ_0^* formula indicating that y represents a Semantic Tableaux proof of the theorem x from the axiom system α **and** that $y = \mathrm{Log}^\lambda(z)$.

Takeuti [23] introduced a form of the λ–Short-Proof concept for studying integers y satisfying the condition $\exists\, z \ \mathrm{Log}^\lambda(z) = y$. His goal was to use this construct to help explicate the relationship between Buss's Bounded Arithmetic, Gentzen's sequent calculus and some of NP's properties [23]. An entirely different type of application of the λ–Shortness concept was subsequently observed by Adamowicz, Salehi, Willard and Zbierski [1, 2, 19, 31, 33] (largely independently of Takeuti's research). This second line of research used the λ–Shortness concept as an intermediate step to help answer some open questions about $I\Sigma_0$'s Incompleteness properties raised by Paris and Wilkie in [14]. Thus in approximate chronological order, the latter research included Adamowicz-Zbierski's observation [1, 2] that a cut-free version of the Second Incompleteness Theorem was valid at the level of $I\Sigma_0 + \Omega_1$, Willard's strengthening of this result so that the threshold for the Cut-Free Second Incompleteness effect would be lowered so that it would include all extensions of $I\Sigma_0$ and most extensions of Q [31, 33], and Salehi's more recent second type of proof [19] of Willard's $I\Sigma_0$ Incompleteness Theorem.

The λ-Short concept will also help us prove Theorem 1 in this paper. Thus, let $D(\alpha)$ denote the following Gödel sentence:

> "There is no Semantic Tableaux proof of **this sentence** from α's set of proper axioms"

Let $D^\lambda(\alpha)$ denote the "$\mathrm{ShortPrf}_\alpha^\lambda(x,y,z)$" analog of this diagonalization sentence, defined below:

> "In a context of the $\mathit{ShortPrf}_\alpha^\lambda(x,y,z)$ notation convention, there exists no code (y,z) that *proves* **this sentence** from α's axioms"

It is easy to give $D^\lambda(\alpha)$ a Π_1^* encoding. Thus, let Subst(g,h) denote the following Σ_0^* formula:

Subst(g,h) = The integer g is an encoding of a formula, and h encodes a sentence identical to g, except all g's free variables are now replaced by a term equal to the constant g *itself.*

Then following Gödel's example, $D^\lambda(\alpha)$ is formally defined to be the sentence $\Gamma(\bar{n})$, where Equation (3) defines the formula $\Gamma(g)$ and $\bar{n}$ is a term whose numerical value represents $\Gamma(g)$'s Gödel number.

$$\forall y\ \forall z\ \forall h < y \quad \{ \quad \text{Subst}(g,h) \quad \rightarrow \quad \neg\ \text{ShortPrf}_\alpha^\lambda\ (h\ ,\ y\ ,\ z\ ) \quad \} \tag{3}$$

Our proof of Theorem 1 will use the $D^\lambda(\alpha)$ Diagonalization sentence as an intermediate step to help corroborate Theorem 1. In particular, let **Pair(s,t)** denote a Σ_0^* formula indicating that s is the Gödel number of a Q_2^* sentence and that t is the Gödel number of a second Q_2^* sentence which is the negation of s . Also, let $\lceil D^\lambda(\alpha) \rceil$ denote $D^\lambda(\alpha)$'s Gödel number. Then the Theorem 4 (below) will help prove Theorem 1.

Theorem 4. *Suppose α is a consistent axiom system capable of proving all the Σ_0^* sentences that are valid in the Standard Model. Suppose there exists two constants, λ and L, such that α can also prove:*

A) $\quad \forall g\ \forall h\ \forall h^* \ \{ \ [\ Subst(\,g\,,\,h\,) \wedge Subst(\,g\,,\,h^*\,)\ ] \ \rightarrow \ h\ =\ h^* \ \}$

B) $\quad \forall\, z > L \quad \forall\, y \quad \{ \quad ShortPrf_\alpha^\lambda(\ \lceil D^\lambda(\alpha) \rceil\ ,\ y,\ z\ ) \quad \rightarrow$

$\exists\ p < z\ \ \exists\ q < z\ \ \exists\ s < z\ \ \exists\ t < z\ \ [\ Pair(s,t)\ \wedge\ Prf_\alpha(p,s)\ \wedge\ Prf_\alpha(q,t)\ \ ]\}$

Then α must be incapable of proving:

$$\forall p\ \forall q\ \forall s\ \forall t\ \ \neg\ \ [\ Pair(s,t)\ \wedge\ Prf_\alpha(p,s)\ \wedge\ Prf_\alpha(q,t)\ ].$$

We will not prove Theorem 4 here because its justification is similar to the Theorem 2.3 from [33]. The remainder of this article will use Theorem 4 as an intermediate step to help prove Theorems 1 and 2.

4 The Formal Proof of Theorem 1.

Let α denote an axiom system and $\varphi(x)$ denote a formula free in only x. The formula $\varphi(x)$ is called [7] a **Definable Cut relative to** α iff α can prove the theorem:

$$\varphi(0) \quad \text{AND} \quad \forall\, x \;\; \varphi(x) \rightarrow \varphi(x+1) \quad \text{AND} \quad \forall\, x\, \forall\, y < x \;\; \varphi(x) \rightarrow \varphi(y) \tag{4}$$

Definable Cuts have been studied by a very extensive literature [1, 2, 4, 6, 7, 8, 9, 10, 11, 13, 14, 16, 17, 22, 25, 26, 27, 28, 29]. They are unrelated to Gentzen's notion of a Sequent Calculus "Deductive Cut Rule".

It is convenient to call a Definable Cut **trivial relative** to α iff α *can formally prove* " $\forall\, x \;\; \varphi(x)$ ". For example since Peano Arithmetic recognizes the validity of the Principle of Induction, all its Definable Cut Formulae are trivial. On the other hand, every arithmetical logical system strictly weaker than Peano Arithmetic contains some non-trivial Definable Cut.

One theme of the literature about Definable Cuts is that they are very helpful for developing new versions of the Second Incompleteness Theorem, as well as for devising new uses of it. For instance, the Theorem $*$, attributed in Section 1 to the joint work of Nelson, Pudlák, Solovay and Wilkie-Paris, was derived by using Definable Cuts as a crucial intermediate step. As we pointed out in Section 1, our new Theorem 1 is related to this literature, and it was greatly stimulated by its over-all perspective. However, one aspect of our main proof will veer in a slightly different direction.

The distinction arises because the Second Incompleteness Theorem is much more difficult to prove and generalize for cut-free proof methods, such as Semantic Tableaux or the Cut-Free variant of the Sequent Calculus, than it is for Cut-Permissive formalisms, such as the Hilbert-style methodology or the Sequent Calculus with a Cut Rule. The intuitive reason for this distinction is that most generalizations of the Second Incompleteness Theorem using Equation (4)'s formalism require a Gentzen-style Deductive Cut Rule (or equivalently some Hilbert-style modus ponens deductions) as an intermediate step. Since we are not allowed to use these methodologies under Semantic Tableaux deductive calculi, our strategy for proving Theorem 1 will instead use Theorem 4's formalism as an alternate interim step to facilitate the proof.

We will now summarize the notation used in Theorem 1's proof. For any fixed integer i , let $G_i(x)$ denote the scalar-multiplication operation that maps the integer x onto the quantity $2^{2^i} \cdot x$. Let $\Lambda(i,x,y)$ denote a Σ_0^* formula which indicates that $y = G_i(x)$. There are many possible Σ_0^* encodings for $\Lambda(i,x,y)$'s graph. Equation (5) provides one example:

$$\{i \neq 0 \wedge x \neq 0 \rightarrow \exists v \leq y \ [\text{LogLog}(v) = i \wedge \text{LogLog}(v-1) < i \wedge \frac{y}{v} = x \wedge \frac{y-1}{v} < x]\}$$

$$\text{AND} \qquad \{ \ [\, i = 0 \ \vee \ x = 0 \,] \ \rightarrow \ y = \ x + x \ \} \tag{5}$$

Let Υ_i denote Equation (6)'s Π_2^* sentence, which states that the operation that maps x onto the integer $G_i(x)$ is a total function.

$$\Upsilon_i \qquad =_{\text{def}} \qquad [\ \forall x \ \exists y \quad \Lambda(\, \bar{i} \,, \ x \ , \ y \ ) \quad] \tag{6}$$

G_i's definition clearly implies $\forall i \, \forall x \ G_{i+1}(x) \ = \ G_i(G_i(x))$. Equation (7) encodes this fact *as a* Π_1^* *sentence*.

$$\forall i \ \forall x \ \forall y \ \forall z \quad [\ \Lambda(i,x,y) \ \wedge \ \Lambda(i,y,z) \ ] \quad \rightarrow \quad \Lambda(i+1,x,z) \tag{7}$$

Also, for any $m > 0$ and $n > 0$, let (8) and (9) define the sentences Θ_m and $\mho_n$ below. (Equation (7) implies their validity.)

$$\Theta_m \qquad =_{\text{def}} \qquad [\ \Upsilon_{m-1} \rightarrow \Upsilon_m \ ] \tag{8}$$

$$\mho_n \qquad =_{\text{def}} \qquad [\ \Upsilon_0 \ \wedge \ \Theta_1 \ \wedge \ \Theta_2 \ \wedge \ ... \ \wedge \Theta_n \ ] \tag{9}$$

The intuitive reason why the Second Incompleteness Theorem is harder to prove for the Semantic Tableaux deductive calculus than for the Hilbert method of deduction can be appreciated by examining an axiom system β whose only *non-trivial* axiom about the U-Grounding functions corresponds to (7)'s axiom. This axiom, combined with the built-in assumption that Addition is a total function, allows one to construct a Hilbert-style proof of the theorem Υ_n from β whose length is no longer than $c \cdot n^c$ for some constant c , using a methodology that many logicians [4, 7, 8, 10, 11, 13, 16, 17, 21, 29, 32] have attributed to unpublished private communications from Robert Solovay. However, it oddly turns out that while the Semantic Tableaux calculus supports equally short proofs of similar approximate length $O(\, c \cdot n^c \,)$ for the sentences Θ_n and $\mho_n$ from β , there is no analogously short *Semantic Tableaux* proof of Υ_n from β !

Our proof of Theorem 1 is related to this fact. It will formalize a Level(2+) Tableaux-style generalization of the Second Incompleteness Theorem that has no analog at Level(1) essentially because $\mho_n$ has a much shorter proof than Υ_n from β (under Semantic Tableaux). No analog of this Tableaux-type separation characterizes the several Hilbert-style versions of the Second Incompleteness Theorem, discussed in say [4, 7, 8, 9, 10, 13, 14, 16, 17, 25, 26, 27, 29], where the Second Incompleteness Theorem is equally valid at all levels L . The intuitive reason the Semantic Tableaux will be shown in the discussion (below) to have contrasting properties for the Levels 1 and 2+ , unlike Hilbert deduction, will ultimately be because the difference between the proof-lengths of Υ_n and $\mho_n$ is much greater under Semantic Tableaux deduction than under Hilbert deduction.

In essence, the above observations combined with Theorem 4's role as a very helpful intermediate step explain the two main underlying intuitions behind Theorem 1's proof.

Summary of Main Proof: In the context of our proof of Theorem 1, the symbol "α" in the predicates $\text{ShortPrf}^{\,2}_{\alpha}\,(x,y,z)$ and $\text{Prf}_{\,\alpha}\,(x,y)$ will have a slightly unconventional interpretation. Rather than treat "α" as a fixed constant that denotes a finite-sized axiom system, this section will view it as an integer that designates a Gödel number that represents a finite sequence of sentences $S_1, S_2 \,...\, S_n$, listing α's axioms. (Under our method for encoding an axiom system α , its Gödel number will have a bit-length proportional to the sum of the lengths of $S_1, S_2 \,...\, S_n$.)

Our formal analysis will begin by defining the Π_1^* sentence W mentioned in Theorem 1's hypothesis. It is defined as a conjunction of nine Π_1^* clauses $W_0, W_1 \,...\, W_8$ — where many of these clauses W_i are in turn conjunctions of several further Π_1^* sub-clauses. These nine clauses are defined below:

Definition of W_0: This axiom will be a multi-clause Π_1^* sentence which provides sufficient information about the eight U-Grounding functions so that W_0 has the capacity to formally prove every Σ_0^* sentence that is logically valid. (It is unimportant which particular finite set of Π_1^* clauses is used to formulate W_0 , as long as one of its clauses is the explicit statement "$0 \neq 1$".)

Definitions of W_1 through W_5: The definition of the Π_1^* sentence W_1 was given in Equation (7). In a context where $\text{Prf}_\alpha(x,y)$, Subst(g,h) and $\Lambda(i,x,y)$ were already given Σ_0^* defining formulae, the further Π_1^* definitions of W_2 through W_5 are given by Equations 10 through 13.

$$\forall g \ \forall h \ \forall h^* \ \{ \ [\ \text{Subst}(\, g \, , \, h \,) \wedge \text{Subst}(\, g \, , \, h^* \,) \,] \ \rightarrow \ h \, = \, h^* \ \} \tag{10}$$

$$\forall \, i \quad \forall z \quad [\quad \Lambda(i, \overline{1} \, , z) \quad \rightarrow \quad \text{LogLog}(z) = i \ \ ] \tag{11}$$

$$\forall \alpha \quad \forall t \quad \forall n \quad [\quad \text{Prf}_{\,\alpha} \, (\, t \, , \, n \,) \quad \vee \quad \neg \ \text{Prf}_{\,\alpha} \, (\, t \, , \, n \,) \quad] \tag{12}$$

$$\forall g \quad \forall h \quad [\quad \text{Subst} \, (g, h) \quad \vee \quad \neg \, \text{Subst} \, (g, h) \quad] \tag{13}$$

Definitions of W_6 and W_7: We will not provide a formal equational description of these two Π_1^* sentences, similar to the prior Equations (10) through (13), because their formal structures are a bit tedious to encode. Instead, we will provide a functional description of them:

1. The Π_1^* sentence W_6 will contain sufficient information about the U-Grounding functions so that for each ordered triple $(\, \overline{\alpha} \, , \, \overline{h} \, , \, \overline{y} \,)$ where $\overline{h}$ denotes the Gödel number of $D^2(\alpha)$ and $\text{Prf}_{\,\alpha} \, (\, h \, , \, y \,)$ is true, the formal proof of $\text{Prf}_{\,\alpha} \, (\, \overline{h} \, , \, \overline{y} \,) \,)$ from W_6 will have a bit-length no larger than $[\text{Log}(y)]^{C_6}$ for some constant C_6 . Likewise, the Π_1^* sentence W_6 will contain sufficient information about the U-Grounding functions so that for each ordered pair $(\, \overline{g} \, , \, \overline{h} \, ,)$ where $\text{Subst}(\overline{g} \, , \, \overline{h} \,)$ is true, the formal proof of this fact will have a bit-length no larger than $[\text{Log}(h)]^{C_6}$ (It is easy to construct a Π_1^* axiom W_6 and accompanying constant C_6 with these properties.)

2. Let us recall that "U-Grounded Binary-encoded Representations" $\overline{n}$ of integers n were defined in Section 2. The Π_1^* sentence W_7 will contain sufficient information about the U-Grounding functions so that for any integer $n > 1$, the proof from W_7 of " $\overline{n-1} \, + \, \overline{1} \, = \, \overline{n}$ " has a bit-length no larger than n^{C_7} , for some sufficiently large fixed constant C_7 .

Definition of W_8: The Π_1^* definition of W_8 will appear in Equation (26) later in this section. Its presentation is postponed because some preliminary lemmas need to first help motivate it.

Lemma 1. *Each of* W_0 *through* W_7 *are* Π_1^* *theorems of Peano Arithmetic.*

Proof: It is obvious that W_1 through W_5 are Π_1^* theorems of Peano Arithmetic. Also, it is trivial to construct Π_1^* sentences W_0, W_6 and W_7 that satisfy their functional requirements. □

Lemma 2. *Let us recall that* Υ_i, Θ_m *and* $\mho_n$ *were defined by Equations (6), (8) and (9). Then there exists three constants* K_0, K_1 *and* K_2 *such that:*

i) A semantic tableaux proof of Θ_m *from W requires a bit-length no greater than* $K_1 \cdot m^{K_1}$.

ii) A semantic tableaux proof of $\mho_n$ *from W requires a bit-length no greater than* $K_2 \cdot n^{K_2}$.

iii) The semantic tableaux proofs of Υ_i *from the union of* W *and* Υ_{i-1} *, and of* $\neg\,\Upsilon_{i-1}$ *from the union of* W *and* $\neg\,\Upsilon_i$ *, each require a bit-length no greater than* $K_0 \cdot i^{K_0}$.

Proof:

It is immediate from the definition of the sentences W_1 and W_7 that these two parts of W are sufficient to assure that Θ_m's proof will have a length satisfying constraint (i). The assertion (ii) follows from (i) because the proof of $\mho_n$ has a length essentially no greater than the sum of the proof lengths for Υ_0 and for Θ_1 , Θ_2 ... Θ_n . The assertion (iii) follows from (i) because the definition of Θ_m makes it obvious that (iii)'s two proofs have a sufficiently similar structure to (i)'s proof for there to be no meaningful difference between their proof lengths.

□

Our proof of Theorem 1 will be centered around showing that the axiom system α satisfies the requirement (B) of Theorem 4 when $\lambda = 2$. After establishing this fact, the remainder of Theorem 1's proof will be quite easy. Paraphrased into the English Language, Theorem 4's Part-B requirement, with $\lambda = 2$ and L representing a fixed constant, is the statement:

> $+$ If an ordered pair (y, z) (with $z > L$) encodes a "2-short proof" (from α) of the Gödel-like diagonalization sentence $D^2(\alpha)$, then there will exist some corresponding Q_2^* sentence (called say S) where both S and $\neg\,S$ have Semantic Tableaux proofs from α whose Gödel numbers are smaller than z.

Our proof of this statement will rest on showing that for some integer n (whose exact value will depend only on the ordered pair (y, z)), one adequate sentence S satisfying the above assertion is Equation (9)'s sentence $\mho_n$. In order to complete Theorem 1's proof, we will need to show that any axiom system satisfying Theorem 1's hypothesis *will both* satisfy $+$'s requirements *and recognize this fact about itself.*

Some added notation will be employed by our next two lemmas. The symbol W^* will denote a multi-clause axiom, similar to W , except that W^*'s clauses will consist of W_0 through W_7 (and thus omit the W_8 condition). Also, p will be called a **Partial Proof** of the theorem Φ from α iff its structure is identical to a Semantic Tableaux proof tree **except** that one of its branches is released from the requirement of containing a pair of contradictory nodes. This unique branch will be called p's **Open Branch**. Its lowest node will be called p's **Bottom Node.** Our proof of $+$ will have a nicely compartmentalized modular nature, where it develops a sequence of increasing complex partial proof trees P_1 , P_2 , ... P_m , where each tree P_{i+1} is an extension of the prior tree P_i and the final object P_m is the well-defined Semantic Tableaux proof of the sentence $\neg\, S$ required by Statement $+$.

Lemma 3. *There exists a constant* $K_3 > 0$ *(whose exact numeric value will be unimportant to our main theorem) such that for each* $n \geq 1$, *it is possible to construct a Partial Proof* P_1 *of* $\neg\, \mho_n$ *from the axiom system* W^* *whose bottom node stores* Υ_n *and where* P_1*'s length is bounded by* $K_3 \cdot n^{K_3}$.

Proof: Since P_1 represents a proof of $\neg\, \mho_n$, its root will consist of the sentence $\neg\, \neg\, \mho_n$. The root's child will be the sentence $\mho_n$ (which is formally derived via Section 2's $\neg$ Elimination rule for Semantic Tableaux proofs). Then via several applications of the $\wedge$ Elimination rule, Equation (9)'s $\mho_n$ sentence will be broken repeatedly into smaller and smaller components until each of the formal Q_2^* sentences of Υ_0, Θ_1 , Θ_2 , ... Θ_n is enumerated along P_1's open branch. The last n steps of P_1's proof will consist in chronological order of n repeated applications of the $\rightarrow$ Elimination Rule, whose $i-$th iteration splits Equation (8)'s Θ_i sentence into a left sibling nodes of the form $\neg\, \Upsilon_{i-1}$ and a right sibling of Υ_i . (These splits will be performed so that the sentences Υ_1 , Υ_2 , ... Υ_n are enumerated in chronological order along P_1's open branch.)

To verify the above tree is a "Partial Proof" whose "Bottom Node" is Υ_n , we need to confirm each of the preceding paragraph's "left sibling nodes" of the form $\neg\ \Upsilon_{i-1}$ are indeed leaves lying at the bottom of closed branches. This fact is a consequence of our inductive construction because the parent of each leaf storing a sentence $\neg\ \Upsilon_{i-1}$ will store its negation Υ_{i-1}.

Hence, P_1 certainly represents a Partial Proof of $\neg\ \mho_n$. It is trivial that for some constant $K_3 > 0$, its proof length is bounded by $K_3 \cdot n^{K_3}$. □

Lemma 4. *There exists a constant* $K_4 > 0$ *such that for every* $n \geq 1$ *it is possible to construct a Partial Proof* P_2 *of the sentence* $\neg\ \mho_n$ *from the axiom system* W^* *— where this proof's bit-length is bounded by* $K_4 \cdot n^{K_4}$ *and where for some parameter* u *(created during existential quantifier elimination) the Bottom Node of this Partial Proof is the sentence "* $LogLog(u) = \overline{n}$ *" .*

Proof: The proof P_2 is constructed by taking Lemma 3's partial proof P_1 (whose Bottom node had stored the sentence Υ_n) and adding seven nodes below this bottom node. The first two of these seven nodes will store the sentences indicated by Equations (14) and (15). In particular, (14) is deduced from Equation (6)'s sentence Υ_n via the $\forall$ Elimination Rule, and (15) is deduced from (14) via the $\exists$ Elimination Rule.

$$\exists y \quad \Lambda(\overline{n},\ \overline{1}\ ,\ y\ ) \tag{14}$$

$$\Lambda(\ \overline{n}\ ,\ \overline{1}\ ,\ u\ ) \tag{15}$$

The next three sentences in P_2's proof tree will consist of the axiom W_3 (whose formal statement was listed in Equation (11)) and then two deductions from this axiom based on applying the $\forall$ Elimination Rule so that Equation (11)'s universally quantified variables of i and z are replaced by $\overline{n}$ and u. The final sentence at the bottom of these three steps is:

$$\Lambda(\ \overline{n}\ ,\ \overline{1}\ ,\ u\ ) \ \rightarrow \quad \text{LogLog}(u)\ =\ \overline{n} \tag{16}$$

The last two nodes of P_2's proof tree will be deduced from (16) via the $\rightarrow$ Elimination rule. These two nodes will thus be sibling nodes storing the sentences of " $\neg\ \Lambda(\ \overline{n}\ ,\ \overline{1}\ ,\ u\ )$ " and " $\text{LogLog}(u)\ =\ \overline{n}$ " .

The tree P_2 (above) is a Partial Proof because " $\text{LogLog}(u)\ =\ \overline{n}$ " is its Bottom Node and the vertex storing " $\neg\ \Lambda(\ \overline{n}\ ,\ \overline{1}\ ,\ u\ )$ " is contradicted by Equation (15)'s sentence (and hence represents the needed leaf closing a branch). Since P_2 differs from P_1 by only having seven additional nodes, both trees have the same approximate bit-length (thereby establishing that P_2's length is sufficiently small to satisfy Lemma 4's claim). □

Lemma 5. *Let* α *denote any axiom system of finite cardinality that is an extension of* W^**. Let us recall the Gödel Sentence* $D^2(\ \alpha\ )$ *was defined in Section 3. Suppose* N *denotes the Gödel number of a semantic tableaux proof of* $D^2(\ \alpha\ )$ *from* α *. Then there will exist a semantic tableaux proof* P *of the sentence* $\neg\ \mho_N$ *from* α *whose bit-length is bounded by* $K_5 \cdot N^{K_5}$ *for some fixed constant* $K_5 > 0$ *.*

Proof Sketch. The proof P is built by taking Lemma 4's partial proof P_2 and inserting directly below P_2's Bottom Sentence nine additional nodes and three further subtrees. The first of these nine nodes will be Equation (12)'s W_4 axiom sentence. The next three nodes will represent reductions from this axiom, using the $\forall$ Elimination Rule to replace (12)'s universally quantified variables t, α and n with $D^2(\alpha)$'s Gödel number, $\overline{\alpha}$ and $\overline{N}$. The resulting sentence at the end of these reductions is:

$$\text{Prf}_{\ \overline{\alpha}}\ (\ \lceil D^2(\alpha)\rceil\ ,\ \overline{N}\ )\ \vee\quad \neg\ \text{Prf}_{\ \overline{\alpha}}\ (\ \lceil D^2(\alpha)\rceil\ ,\ \overline{N}\ ) \tag{17}$$

Below node (17), P's proof will apply the $\vee$ Elimination rule to produce the following two sibling nodes

$$\neg\ \text{Prf}_{\ \overline{\alpha}}\ (\ \lceil D^2(\alpha)\rceil\ ,\ \overline{N}\ ) \tag{18}$$

$$\text{Prf}_{\ \overline{\alpha}}\ (\ \lceil D^2(\alpha)\rceil\ ,\ \overline{N}\ ) \tag{19}$$

It is easy to inject a closed subtree T_1 below node (18) which meets Lemma 5's requirements. This is because the lemma's hypothesis indicates N is a proof of $D^2(\alpha)$ and the axiom W_6 then implies the subtree T_1 can have a small enough bit-length to satisfy Lemma 5's requirements.

In order to construct an analogous second suitably small subtree below (19)'s sentence, let G denote the particular Gödel number satisfying the identity Subst (G , $\lceil D^2(\alpha)\rceil$) . In this context, we will insert below (19)'s sentence, a 3-node sequence beginning with Equation (13)'s W_5 axiom followed by two iterations of the $\forall$ Elimination rule to obtain the deduction:

$$\text{Subst}\ (\ \overline{G}\ ,\ \lceil D^2(\alpha)\rceil\ )\quad \vee\quad \neg\ \text{Subst}\ (\ \overline{G}\ ,\ \lceil D^2(\alpha)\rceil\ ,) \tag{20}$$

Next, let us apply the $\vee$ Elimination Rule to split (20)'s node into (21) and (22)'s pair of sibling nodes.

$$\neg\text{Subst}\ (\ \overline{G}\ ,\ \lceil D^2(\alpha)\rceil\ ,)\tag{21}$$

$$\text{Subst}\ (\ \overline{G}\ ,\ \lceil D^2(\alpha)\rceil\ ,)\tag{22}$$

Using axiom W_6, it is easy to insert below node (21) a closed subtree T_{2A} whose bit-length is sufficiently small to satisfy Lemma 5's requirements. Thus to complete Lemma 5's proof, we must merely show that it is also possible to insert below node (22) a second adequately small subtree T_{2B} .

In order to construct T_{2B}, we will use again the fact that N represents a proof of the theorem $D^2(\alpha)$. Since N proves a theorem $D^2(\alpha)$ (whose formal statement in Equation (3) begins with three universally quantified variables y, z and h) , one can apply standard methods from Proof Theory to deduce the existence of a proof T^* of 0=1 from the union of the axiom system α with the added axioms given in Equations (23) through (25) such that $T^* \leq N^2$ when these two proofs are viewed as Gödel numbers. (The footnote [1] explains the intuition behind T^*'s construction, and a longer version of this paper will give a more formal proof that $T^* \leq N^2$ exists.)

$$\text{Subst}\ (\ \overline{G}\ ,\ u_1\ )\tag{23}$$

$$\text{Prf}_{\ \overline{\alpha}}\ (\ u_1\ ,\ u_2\ )\tag{24}$$

$$\text{LogLog}(u)\ =\ u_2\tag{25}$$

In the above context, T_{2B} will be defined as a tree identical to T^* except that each appearance of u_1 and u_2 in T^* is replaced by respectively $\lceil D^2(\,\overline{\alpha}\,)\rceil$

[1]The intuitive reason that $T^* \leq N^2$ exists is that a proof N of $D^2(\alpha)$ will store $\neg\, D^2(\alpha)$ in N's root, and the main non-degenerate versions of such proofs N will next apply the $\neg$ Elimination Rule to transform $D^2(\alpha)$'s universally quantified variables y, z and h into existentially quantified variables that are subsequently replaced by the parameter symbols u_1, u_2 and u satisfying Equations (23) through (25) in a context where these three nodes appear in a straight-line path in N's proof-tree directly below the root. This footnote should not be considered a formal proof that there exists the required $T^* \leq N^2$ because a formal proof, appearing in a longer version of this paper, must also consider certain degenerate cases, in addition to the main non-degenerate case outlined here.

and $\overline{N}$ in T_{2B}. This subtree must certainly close the portion of P's proof tree that descends from node (22) because the sentences in Equations (23) through (25) are identical to the three respective sentences in (22), (19) and P_2's Bottom Node with $\lceil D^2(\overline{\alpha})\rceil$ and $\overline{N}$ now replacing u_1 and u_2. Essentially because the replacement of u_2 with the longer expression of $\overline{N}$ will cause T_{2B}'s bit-length to have an $O[\ (\text{Log}\ N\ )^2\ ]$ magnitude, the resulting constructed T_{2B} tree will be short enough to satisfy Lemma 5's claim. □

Lemma 6. *Let MinAx(α) denote a Σ_0^* formula that indicates α is an axiom system of finite cardinality which includes the Π_1^* sentences of W_0, W_1 ... W_7 . Then there exists a suitably large constant $L > 0$ such that Equation (26)'s Π_1^* sentence is both valid and a theorem of Peano Arithmetic:*

$$\forall z > L\ \forall y\ \forall r\quad \{\quad [\ \text{ShortPrf}_r^2(\lceil D^2(r)\rceil, y, z)\quad \wedge\quad \text{MinAx}(r)\ ]\quad \rightarrow$$

$$\exists\, p < z\ \exists\, q < z\ \exists\, s < z\ \exists\, t < z\ [\,\text{Pair}(s,t) \wedge \text{Prf}_r(p,s) \wedge \text{Prf}_r(q,t)\ ]\} \quad (26)$$

Proof: The combination of Lemma 2 (part ii) and Lemma 5 easily implies that Equation (26) is valid for sufficiently large enough L . This is because these two lemmas imply that if (z, y, r) satisfies the left side of (26)'s implication clause, then (for suitably large enough L) the Q_2^* sentence $\mho_y$ will have short enough proofs of both itself and its negation to automatically satisfy the right side of (26)'s implication clause. (Moreover, (26) is a Π_1^* theorem of Peano Arithmetic because all the lemmas presented in this chapter can be formally proven by Peano Arithmetic.) □

Finishing the Definition of W . The final clause W_8 of the system W will be defined as Equation (26)'s Π_1^* sentence with a large enough value assigned to the constant L to make this sentence valid. (We had not provided W_8's definition earlier in this section because it seemed more appropriate to first introduce the Lemma 6 — indicating the correctness of W_8 — before defining it.)

Finishing the Proof of Theorem 1. The combination of Lemmas 1 through 6 indicates that all the clause of W are valid Π_1^* theorems of Peano Arithmetic (as the hypothesis of Theorem 1 had required). To finish the proof of Theorem 1, we must show that every finite and consistent axiom system $\alpha \supset W$ is unable to prove its Level(2+) consistency.

This fact is an easy consequence of Theorem 4. In particular, any $\alpha \supset W$ must be able to prove the two sentences (A) and (B) required by Theorem 4 because the sentence (A) is identical to α's W_2 axiom clause and because α can verify (B) by taking Equation (26)'s W_8 axiom-sentence and observing that it reduces to (B) when one sets $r = \overline{\alpha}$ (and uses the fact that MinAx($\overline{\alpha}$) =True). Hence since α can prove (A) and (B), Theorem 4 indicates α is unable to verify its Level(2+) Tableaux consistency. □

Clarification Concerning Theorem 1's Proof and Meaning. The preceding paragraph showed how it was relatively easy to derive from Theorem 4 the conclusion that no consistent $\alpha \supset W$ can verify its own Level(2+) Tableaux consistency. This result would actually *be totally meaningless* if W was an invalid statement because there would then be no example of a consistent $\alpha \supset W$ *actually existing* ! The more subtle aspect of Theorem 1's proof was thus not its second paragraph (which explored the properties of systems satisfying $\alpha \supset W$) but rather it was its first paragraph (which noted Lemmas 1 through 6 imply W is a logically valid Π_1^* sentence). Without this latter crucial fact, Theorem 1 would be devoid of meaningful, non-trivial structural implications. (A similar distinction about the importance of W being a valid Π_1^* sentence will also apply to the generalizations of Theorem 1 explored in the next section.)

5 Sketch of Theorem 2's Proof

Throughout this section, Tab$-\Re-$List$_\alpha^\lambda(t,p)$ will denote a Σ_0^* sentence asserting p is a Tab$-\Re-$List proof of the theorem t from the axiom system α . Also, Short$-\Re-$List$_\alpha^\lambda(x,y,z)$ will denote the TabList analog of Section 3's ShortPrf(x,y,z) predicate. It will thus denote a Σ_0^* sentence that asserts that the two conditions of $y =$Log$^\lambda(z)$ and Tab$-\Re-$List$_\alpha(x,y)$ both hold. Also, $D_\Re^\lambda(\ \alpha\ )$ will denote the analog of Section 3's $D^\lambda(\ \alpha\ )$ Gödel sentence. Thus, $D_\Re^\lambda(\ \alpha\ )$ is defined to be the following diagonalization sentence:

> There exists no code (y,z) that represents a ShortList proof (with exponent λ and intermediate set $\Re$) of **this sentence** from α's axioms.

Throughout this section, the symbol $\perp$ will denote the Gödel number of the sentence $0=1$. Our general technique for proving Theorem 2 will employ a methodology very similar to Theorem 1's proof. It will thus employ an intermediate step, analogous to the Theorem 4, used in Sections 3 and 4. The formal statement of this revised form of Theorem 4 is given below:

Theorem 5. . *Let $\Re$ denote an arbitrary class of sentences. (In our particular applications of Theorem 5, $\Re$ will represent either the set of Π_2^* formulae or the set of Σ_2^* formulae). Suppose α is a consistent axiom system capable of proving all the Σ_0^* sentences that are valid in the Standard Model. Suppose there exists two constants, λ and L, such that α can also prove:*

$$A)\quad \forall g\ \forall h\ \forall h^*\ \{\ [\ Subst(g\,,\,h\,) \wedge Subst(g\,,\,h^*\,)\ ]\ \rightarrow\ h\,=\,h^*\ \}$$

$$B)\quad \forall\, z > L\quad \forall\, y\quad \{\quad Short{-}\Re{-}List_{\alpha}^{\lambda}\,(\,\lceil\, D_{\Re}^{\lambda}(\alpha)\,\rceil\ ,\ y,\ z\ )\quad \rightarrow$$

$$\exists\, p < z\qquad Tab{-}\Re{-}List_{\alpha}(\ \perp\ ,\ p\ )\quad \}$$

Then α must be incapable of proving Equation (27)'s theorem statement (which intuitively indicates that Tab−$\Re$−List proofs using α's set of proper axioms are Level(0–) consistent).

$$\forall p\quad \neg\ \ Tab{-}\Re{-}List_{\alpha}(\ \perp\ ,\ p\ ) \tag{27}$$

In addition to using Theorem 5 to prove Theorem 2, we will need an approximate analog of the prior section's Lemma 6. Below is Lemma 6's analog for TabList deduction:

Lemma 7. *Let $MinW(\ \alpha\ )$ denote a Σ_0^* formula that indicates α is an axiom system of finite cardinality which includes Section 4's Π_1^* sentence W. Then there exists two suitably large constants $L_1 > 0$ and $L_2 > 0$ such that Equations (28) and (29) are both Π_1^* theorems of Peano Arithmetic:*

$$\forall\, z > L_1\ \ \forall\, y\ \ \forall\, r\quad \{\quad [\ \ Short{-}\Pi_2^*{-}List_r^2\,(\lceil\, D_{\Pi_2^*}^2(r)\,\rceil, y, z)\quad \wedge\quad MinW(r)\ \ ]\quad \rightarrow$$

$$\exists\ p_1\ <\ z\qquad Tab{-}\Pi_2^*{-}List_{\alpha}(\ \perp\ ,\ p_1\ )\quad \} \tag{28}$$

$$\forall\, z > L_2\ \ \forall\, y\ \ \forall\, r\quad \{\quad [\ \ Short{-}\Sigma_2^*{-}List_r^2\,(\lceil\, D_{\Sigma_2^*}^2(r)\,\rceil, y, z)\quad \wedge\quad MinW(r)\ \ ]\quad \rightarrow$$

$$\exists\ p_2\ <\ z\qquad Tab{-}\Sigma_2^*{-}List_{\alpha}(\ \perp\ ,\ p_1\ )\quad \} \tag{29}$$

Proof Sketch for Lemma 7. In a very abbreviated form, the proof element p_1 needed to make Equation (28) valid can be summarized as having the following 4-part structure:

1. Its first fragment will be the Tab$-\Pi_2^*-$List proof y , which intuitively represents a proof of the Π_1^* theorem $D^2_{\Pi_2^*}(\alpha)$.
2. Its second fragment will prove the Π_0^* theorem which states that $\overline{y}$, viewed as a Gödel number, proves the theorem $D^2_{\Pi_2^*}(\alpha)$. (This theorem is formally encoded as: Tab$-\Pi_2^*-$List$\,^2_\alpha$ ($\lceil D^2_{\Pi_2^*}(r) \rceil$, $\overline{y}$).)
3. Its third fragment will prove the precise sequence of Π_2^* theorems of Υ_1 , Υ_2 , ... Υ_y in exactly the chronological order just specified. (The reason it is necessary to prove these theorems in increasing chronological order is that it will turn out that Part-iii of Lemma 2 can then assure that each proof is sufficiently compact for p_1 to satisfy Equation (28)'s severe size constraint of $p_1 < z$.)
4. The last fragment of p_1 will use the combination of the preceding three results to prove 0=1, via a variant of Gödel's traditional diagonalization contradiction argument. (The exactly helpful role of the theorem Υ_y in this contradiction proof is that it will guarantee the existence of the integer $z = 2^{2^y}$ — *whose formally proven existence* is required to finish our proof of 0=1.)

Likewise summarized again in an abbreviated form, the proof element p_2 satisfying Equation (29) is the following 4-part structure:

1. Its first fragment will be the Tab$-\Sigma_2^*-$List proof y , which intuitively represents a proof of the Π_1^* theorem $D^2_{\Sigma_2^*}(\alpha)$.
2. Its second fragment will prove the Σ_0^* theorem which states that $\overline{y}$, viewed as a Gödel number, proves the theorem $D^2_{\Sigma_2^*}(\alpha)$. (This Σ_0^* theorem is formally encoded as Tab$-\Sigma_2^*-$List$\,^2_\alpha$ ($\lceil D^2_{\Sigma_2^*}(r) \rceil$, $\overline{y}$).)
3. Its third fragment will prove the precise sequence of Σ_2^* theorems of $\neg\,\Upsilon_y$, $\neg\,\Upsilon_{y-1}$, ... $\neg\,\Upsilon_1$ in exactly the descending chronological order just specified. (The reason it is necessary to prove these theorems in decreasing chronological order is that the theorem $\neg\,\Upsilon_y$ can be derived

easily as a combined consequence of Item 1 and 2's intermediate results, and for each integer i we can then use Part-iii of Lemma 2 to obtain a suitably short proof of $\neg\Upsilon_{i-1}$ from the preceding intermediate result of $\neg\Upsilon_i$ to satisfy Equation (29)'s severe size constraint of $p_2 < z$.)

4. The last fragment of p_2 will use Item 3's intermediate result of $\neg\Upsilon_1$ to prove the desired contradiction theorem of 0=1.

The proof that the objects p_1 and p_2 satisfy the size constraints in Lemma 7's two claims can be summarized as being analogous to Section 4's multi-step proof of Lemma 6 — except that roughly Part-iii of Lemma 2 will now replace its prior Part-ii in certain intermediate steps of Lemma 7'a proof. (A more formal proof of Lemma 7 will appear in a longer version of this article.)

Sketch of the Remainder of Theorem 2's proof: To complete Theorem 5's proof, we will set V_A equal to the conjunction of Section 4's axiom W with Equation (28)'s Π_1^* sentence, and set V_B equal to the conjunction of W with Equation (29)'s Π_1^* sentence. Then a similar type of diagonalization proof, as was used in Section 4, can finish Theorem 2's proof. In particular, one can combine the intermediate results of Theorem 5 and Lemma 7 to derive Theorem 2 in the same manner that the prior section combined Theorem 4 and Lemma 6 to derive Theorem 1. (Several additional details concerning exactly how Section 4's proof of Theorem 1 can generalize to also verify Theorem 2 shall appear in a longer version of this paper.)

Significance of Theorem 2's Result. Part of the reason that Theorem 2 is significant is that Theorem 3 shows that if one were to weaken only slightly either Part-A or Part-B of Theorem 2's hypothesis then significant Boundary-Case Exceptions to the Second Incompleteness Theorem will arise. It should also be stated that Theorem 2 generalizes to all axioms systems of infinite cardinality satisfying Remark 1's Conventional Deciphering Property.

It is also apparent that the Incompleteness effects described by Theorems 1 and 2 can generalize from Semantic Tableaux deduction to any other cut-free rule of inference, such as for example Herbrand deduction or the cut-free variants of the Sequent calculus. (In the particular case of Theorem 2, the analogs of TabList deduction obviously will have their subcomponent proofs $p_1, p_2, ...p_n$ consist of Herbrand or Cut-Free sequent calculus proofs.)

6 Appendix: Sketch of Theorem 3's Proof

This appendix will sketch a proof of Theorem 3. As Section 2 had explained, Theorem 3's result was technically announced on the last page of our Tableaux-2002 paper [34]. However, the latter conference paper only formally proved the correctness of a weaker version of this result — where Semantic Tableaux deduction replaced Tab_1List deduction in Clause 1 of Theorem 3's formal statement. Our goal in this Appendix is to briefly summarize the added functionality needed to prove the stronger and more general version of [34]'s theorem. Throughout this appendix, we will assume that the reader has already examined our prior paper [34] and has a copy of it on his desk. Thus, we will not review most of the definitions from [34]. Also, if a Lemma number ends with "t" (as in say "Lemma 2-t") then it refers to a result from our Tableaux–2002 Conference paper [34].

In all candor, we first hesitated to include a very abbreviated proof of Theorem 3 as an appendix to this article. However, it seemed desirable to include some type of justification of Theorem 3 here because there is such a tight match between Theorem 3's positive result and Theorem 2's complementing negative result for this topic to be worth mentioning.

Notation Conventions: The deductive rule of inference that had been called "R(1,1) Hierarchy Deduction" in our prior paper [34] has now been renamed, and it is called instead "Tab_1List" deduction in the current paper. Thus $\text{Tab}_1\text{List}\;_{\alpha}\;(t,p)$ will denote a Σ_0^* formula indicating that p is a Tab_1List proof of the theorem t from the axiom system α

As in our earlier article [34], $\text{Pair*}(x,y)$ will denote a Σ_0^* formula indicating that x is a Π_1^* sentence and y is its negation. The last page of [34] had defined IS-1*(A) as an axiom system identical to [34]'s IS-1(A) system, except that Tab_1List deduction had replaced Semantic Tableaux deduction in the statement of IS-1*(A)'s self-justifying Group-3 axiom. Thus, the Group-3 axiom of IS-1*(A) is a self-referencing axiom of the form:

$$\forall x\,\forall y\,\forall p\,\forall q\quad \neg\;[\;\text{Pair*}(x,y)\wedge\;\text{Tab}_1\text{List}_{\;\text{IS-1*}(A)}(x,p)\wedge\;\text{Tab}_1\text{List}_{\;\text{IS-1*}(A)}(y,q)\,]\quad (30)$$

Our proof of Theorem 3 will rest on showing that one can generalize the Theorem 2-t from [34] to establish that:

\+ If A is consistent then IS-1*(A) is automatically also consistent.

Henceforth, $\omega(x,y,p,q)$ will denote the Σ_0^* formula enclosed within (30)'s square bracket expression. In order to establish $+$, it is sufficient to prove the following alternate form of this assertion:

$++$ If A is consistent then $\forall\, x\, \forall\, y\, \forall\, p\, \forall\, q \quad \neg \quad \omega(x,y,p,q)$

Outline of the Proof of ++ . For the sake of constructing a proof-by-contradiction, let us assume $++$ was false and that (X,Y,P,Q) denotes the particular tuple satisfying $\omega(X,Y,P,Q)$ that has minimal value for the quantity $G = \text{Max}(P,Q)$. Then Equation (31) is valid:

$$\forall\, x\, \forall\, y\, \forall\, p\, \forall\, q \quad \{\, p < G \,\wedge\, q < G \,\} \;\rightarrow\; \neg\, \omega(p,q,x,y) \tag{31}$$

The procedure PROBE, the notion of a $(L,M)-$Conservative Branch of a Semantic Tableaux proof and the condition called Constraint(p,β) were each defined in [34]. Our goals will be 1) to use Equation (31) and these concepts to construct an ordered pair satisfying Constraint(t,β), and then 2) to combine this fact with [34]'s Lemma 1-t to finish our proof-by-contradiction.

Some added notation is now needed to formalize this proof of ++. Let H again denote a Tab$_1$List proof, comprised of the sequence of ordered pairs of (t_1,p_1) , $(t_2,p_2)...(t_n,p_n)$, where p_i is a semantic Tableaux proof of the intermediate result t_i. Let us call Υ the **Conclusion** of the proof H iff it is the last theorem that H proves (i.e. this means that $t_n = \Upsilon$). Define $\chi(p_i)$ to be the number of logical symbols appearing in p_i's proof. Also, let $\Im(H)$ denote the quantity $\sum_{i=1}^{n} \chi(p_i)$. Let us say that:

1. The Σ_1^* sentence " $\exists\, v_1\, \exists\, v_2\, ...\, \exists\, v_m \;\; \psi(v_1,v_2,...v_m)$ " is $G-$good iff there exists a Tab$_1$List proof $H \leq G$ from IS-1*(A) of this sentence, and it is accompanied with a valid Equation (32).

$$\exists\, v_1 \leq 2^{\Im(H)}\, \exists\, v_2 \leq 2^{\Im(H)}\, ...\, \exists\, v_m \leq 2^{\Im(H)} \quad \psi(v_1,v_2,...v_m) \tag{32}$$

2. The Π_1^* sentence " $\forall\, v_1\, \forall\, v_2\, ...\, \forall\, v_m \;\; \psi(v_1,v_2,...v_m)$ " is $G-$good iff there exists a Tab$_1$List proof $H \leq G$ from IS-1*(A) of this sentence, and it is accompanied with a valid Equation (33).

$$\forall\, v_1 \leq G\cdot 2^{-\,\Im(H)}\, \forall\, v_2 \leq G\cdot 2^{-\,\Im(H)}\, ...\, \forall\, v_m \leq G\cdot 2^{-\,\Im(H)} \quad \phi(v_1,v_2,...v_m) \tag{33}$$

The opening paragraph of this proof had assumed that $\omega(X, Y, P, Q)$ was true, and it had defined the integer G to equal Max(P, Q). In this context, all the standard encoding conventions imply that the two formal inequalities of $\Im(P) < \frac{1}{3}\,\mathrm{Log}_2(\mathrm{G})$ and $\Im(Q) < \frac{1}{3}\,\mathrm{Log}_2\,(G)$ both hold when the Π_1^* sentence X and the Σ_1^* sentence Y are both $G-$good. Since it is impossible for both X and Y to have $G-$good proofs under the preceding circumstances, we are forced to conclude that there exists some sentence Υ^* which fails to be $G-$good and whose proof $H^* \leq G$ has the smallest Gödel number among the set of proofs that fail the $G-$good criteria.

Let p^* denote the particular tableaux proof belonging to H^* that proves the theorem Υ^* . It turns out that [34]'s procedure PROBE can construct under these assumptions a branch β^* of p^* satisfying Constraint(p^*, β^*). In particular, this procedure PROBE will produce such an output when it is given the input arguments of

1. $M = G \cdot 2^{-\Im(H^*)} - 1$, $L = 2^{\Im(H^*) \,-\, \chi(p^*)}$ and $T = p^*$ when Υ^* is a Π_1^* sentence.

2. $M = G \cdot 2^{\chi(p^*) \,-\, \Im(H^*)} - 1$, $L = 2^{\Im(H^*)}$ and $T = p^*$ when Υ^* is a Σ_1^* sentence.

The formal proof that these input values for L, M and T will enable the procedure PROBE to produce an $(L, M)-$Conservative branch is roughly similar to the 8-step construction used in [34] to justify its Lemmas 2-t and 3-t. The added details needed to now justify our stronger effect appear in a longer version of this paper. It differs from the analogous results from [34] by essentially using Equation (31) and the certifiable fact that H^* is the smallest Tab$_1$List style proof failing $G-$goodness to enable us to strengthen Lemmas 2-t and 3-t.

Our justification of ++ can now be finished with the same type of proof-by-contradiction that was used in [34]. The preceding four paragraphs have shown that if ++ was false, then we could construct some β satisfying Constraint(p^*, β^*). However such a construction is actually impossible because the axioms in p^*'s proof would then form a system satisfying the hypothesis of Lemma 1-t. In this context, Lemma 1-t *certainly forbids* Constraint(p^*, β^*) from being true. Hence, ++ must be true to avoid this inherent contradiction. □

Acknowledgment: I thank Robert Solovay for several stimulating telephone conversations that we had during 1994. In essence, the Theorems 1 and 2 of the current article were designed to strengthen the formalism from the Theorem $*$ that Solovay communicated to me during 1994 (see [21]).

This paper's research was funded partially by NSF Grant 99-02726.

7 Bibliography

[1] Z. Adamowicz, "Herbrand Consistency and Bounded Arithmetic", *Fundamenta Mathematica* 171 (2002) pp. 279-292.

[2] Z. Adamowicz and P. Zbierski, "On Herbrand consistency in weak theories", *Archive for Mathematical Logic* 40 (2001) pp. 399-413.

[3] A. Bezboruah and J. C. Shepherdson, "Gödel's Second Incompleteness Theorem for Q", *Journal of Symbolic Logic* 41 (1976) pp. 503-512.

[4] S. R. Buss and A. Ignjatovic, "Unprovability of Consistency Statements in Fragments of Bounded Arithmetic", *Annals of Pure and Applied Logic* 74 (1995) pp. 221-244.

[5] M. Fitting, *First Order Logic & Automated Theorem Proving,* Springer 1996.

[6] P. Hájek, "On Intrepretability in Set Theory", Part I in *Comm, Math, Univ. Carol* Vol 12 (1971) pp. 73–79 and Part II in *Comm, Math, Univ. Carol* Vol 13 (1972) pp. 445-455.

[7] P. Hájek and P. Pudlák, *Metamathematics of First Order Arithmetic,* Springer Verlag 1991.

[8] R. Kaye, *Models of Peano Arithmetic*, Oxford University Press 1991.

[9] J. Krajícek, "A Note on Proofs of Falsehood", *Archives on Math Logic* 26 (1987) pp. 169-176.

[10] J. Krajícek, *Bounded Propositional Logic and Complexity Theory,* Cambridge University Press, 1995.

[11] E. Nelson, *Predicative Arithmetic,* Math Notes, Princeton Press, 1986.

[12] R. Parikh, "Existence and Feasibility in Arithmetic", *Journal of Symbolic Logic* 36 (1971), pp.494-508.

[13] J. B. Paris and C. Dimitracopoulos, "A Note on the Undefinability of Cuts", *Journal of Symbolic Logic* 48 (1983) pp. 564-569.

[14] J. B. Paris and A. J. Wilkie, "Δ_0 Sets and Induction", Proceedings of the Jadswin Logic Conference (Poland), Leeds Univ. Press (1981) pp. 237-248.

[15] P. Pudlák, "On the Lengths of Finistic Consistency Statements in First Order Theories", *Logic Colloquium 1984*, North Holland Press, pp. 165-196.

[16] P. Pudlák, "Cuts, Consistency Statements and Interpretations", *Journal of Symbolic Logic* 50 (1985) pp.423-442.

[17] P. Pudlák, "On the Lengths of Proofs of Consistency", in *Colleqium Logicum: Annals of the Kurt Gödel Society Volume 2*, Springer-Wien-NewYork (1996) pp 65-86.

[18] Z. Ratajczyk, "Subsystems of True Arithmetic and Hierarchies of Functions", *Annals or Pure and Applied Logic* 64 (1993) pp. 95–152.

[19] S. Salehi, Herbrand Consistency in Arithmetics with Bounded Induction, Polish Academy of Sciences Ph D thesis, October 2001.

[20] R. Smullyan, *First Order Logic,* Springer Verlag, 1968.

[21] R. Solovay, Some detailed telephone conversations between Robert Solovay and Dan Willard (in April of 1994), subsequently summarized and published in Appendix A of [32]. During these conversations, Solovay described his generalization of one of Pudlák's theorems from [16]. Solovay never published any of his observations about "Definable Cuts" that numerous logicians [4, 7, 8, 10, 11, 13, 16, 17, 29] have attributed to his private communications. A 1-sentence summary of Solovay's idea, communicated to us, is described by Theorem $*$, whose formal statement can be found in the first paragraph of Section 1. See the Appendix A

of [32] for a much longer 4-page summary of these telephone conversations with Robert Solovay and an interpretation of how the joint work of Nelson, Pudlák, Solovay and Wilkie-Paris had collectively produced the basic result of Theorem $*$.

[22] V. Švejdar, "Modal Analysis of Generalized Rosser Sentences", *Journal of Symbolic Logic* 48 (1983) pp. 986-999.

[23] G. Taekuti, "Gödel Sentences of Bounded Arithmetic", *Journal of Symbolic Logic* 65 (2000) pp. 1338-1346 (A preliminary version of this result appeared in the *Proceedings of the Logic 1996 conference,* pp. 247–261.)

[24] A. Tarski, A. Mostowski and R. Robinson, *Undecidable Theories*, N. Holland 1953.

[25] A. Visser, "An Inside View of Exp", *Journal of Symbolic Logic* 57 (1992) 131–165

[26] A. Visser, "The Unprovability of Small Inconsistency", *Archive for Mathematical Logic* 32 (1993) pp. 275-298.

[27] A. Visser, "Faith and Falsity", Utrecht Univ. Tech. Report (October 2002).

[28] P. Vopěnka and P. Hájek, "Existence of a Generalized Semantic Model of Gödel-Bernays Set Theory", *Bulletin de l'Academie Polonaise des Sciences, Mathmatiques, Astromiques et Physiques* 12 (1973) pp.1079-1086.

[29] A. J. Wilkie and J. B. Paris, "On the Scheme of Induction for Bounded Arithmetic", *Annals of Pure and Applied Logic* (35) 1987, 261-302

[30] D. Willard, "Self-Verifying Axiom Systems", *Third Kurt Gödel Colloquium* (1993), Springer-Verlag LNCS#713, pp. 325-336.

[31] D. Willard, "The Semantic Tableaux Version of the Second Incompleteness Theorem Extends Almost to Robinson's Arithmetic Q", in *Automated Reasoning with Analytic Tableaux and Related Methods* (2000), SpringerVerlag LNAI#1847, pp. 415-430.

[32] D. Willard, "Self-Verifying Systems, the Incompleteness Theorem and the Tangibility Reflection Principle", *Journal Symbolic Logic* 66 (2001) 536-596.

[33] D. Willard, "How to Extend The Semantic Tableaux And Cut-Free Versions of the Second Incompleteness Theorem Almost to Robinson's Arithmetic Q", *Journal of Symbolic Logic* 67 (2002) pp. 465-496.

[34] D. Willard, "Some New Exceptions for the Semantic Tableaux Version of the Second Incompleteness Theorem", in *Automated Reasoning with Semantic Tableaux and Related Methods*, Springer–Verlag LNAI#2381, (2002) 281-297.

First-Order Logic: (Philosophical) Pro and Contra

Jan Woleński
wolenski@if.uj.edu.pl
Jagiellonian University, Kraków, Poland

The nature of logic has been debated ever since logic was born in ancient Greece. These discussions focus on a considerable number of issues. What are the functions of logic? How is logic related to other disciplines? What is the scope of logic? Is logic inborn in our minds or perhaps learned and mastered through experience? Is it possible to reason without logic? The answers to these and other questions determine a variety of different, sometimes conflicting philosophies of logic. Logic in its mathematical form provokes the same problems as logic in its more traditional dressing. What is new in the present discussions is that we can use firm metalogical results in order to elaborate old philosophical questions about logic. This is what I will try to do with respect to the so-called first-order thesis (FOT) (this paper follows my earlier works [48, 49]). The "metalogical methodology" adopted in this paper is neutral toward ontology in this sense that it makes possible to neglect the ontological commitment, if any, of logic to abstract entities. Thus, I will not discuss Quine"s objection against higher-order logic (since higher-order logic is reduced to second-order logic, I will speak about the latter), namely that it is committed to universals. Generally, speaking, FOT says that first-order logic (FOL) is *the* logic (extensive discussions are to be find in [33]; the author rejects FOT, but the book reviews the main arguments pro and contra; [35, 36]; the second book is an anthology including many relevant papers pro and contra FOT). I consider the problem in question not as a

Hendricks et al. (eds.):
First-Order Logic Revisited
Logos Verlag Berlin (2004), 369–398

contribution to terminology, but as a conceptual analysis (see [10] for various dimensions of the controversy over FOT). This means that the acceptance or rejection FOT is not taken here as deciding how the word "logic" should be employed. There is nothing new in that. Incidentally, one distinguishes logic in the narrow (strict) sense, and logic in the broad sense. The first reduces logic to the formal one, but the second sees logic as consisting of semiotic (the logical theory of language), formal logic and methodology of science; it would be unwise to demand that the term "logic" should (or should not) be restricted to formal logic as its reference. Returning to my main task, I will, roughly speaking, defend FOT by attempting to show that it fits better the most fundamental conceptual intuition concerning logic than its rivals. In particular, it precisely displays the intuition that logic is universal.

Still one preliminary remark should be made in advance. Since we have a variety of items to be covered by the rubric "logic", the main question of this paper can be worded as "Which logic is the right logic?" There are two possible lines of understanding this issue. The first one, and older, concerns the choice between classical logic and a non-classical logic (many-valued, intuitionistic, paraconsistent, relevant, etc.). I will not discuss (except for incidental remarks) this problem, although FOT could (and even should) be considered also in this context. The second path does not lead us outside classical logic. More specifically, it consists in arguing whether FOL or second-order logic or infinitary logic, etc. is the right logic (compare the title of [41]). Thus, the rivals all identified on this path remain in the same classical camp. This concerns not only classical logic seen from the point of view of possible levels (first-order vs. second-order) or the length of its formulas (the finite length vs. infinite strings of signs), but also modal extensions (alethic modal logic, deontic logic, etc.) or modifications as non-monotonic logic, for example. Although these issues are important as far as the matter concerns the scope of classical logic, they will be touched only parenthetically in what follows.

Since I will appeal to various intuitions, a historical account of various understanding of logic is in order. The most extensive historical vocabulary (see [27]) lists (it omits more or less exotic items, like "dialectical logic", "hermeneutical logic", etc. or metaphors as "logic of history", "logic of love", etc.; I keep Risse's order and labels, but note that the last entry should be named *l'art de penser*, if French authors are pointed out) nine meanings: (1) dialectic (analysis and synthesis of concepts; Plato), (2) analytic (deduction;

Aristoteles), (3) *organon* (methods of reasoning; Aristoteles), (4) canonic (norms of knowledge; Epicurus), (5) *medicina mentis* (descriptive and normative account of mental capacities; Cicero), (6) *Vernunftslehre* (rules of pure reason; the tradition of *philosophia rationalis*), (7) *Kunstlehre* (the art of arguing; Husserl); (8) Wissenschaftslehre (the theory of science; Petrus Hispanus: *ars artium scientia scientiarum ad omniam aliarum scientiarum methodorum principiam viam habent*); (9) *Denklehre* (the theory of thinking; Arnauld, Nicole).

The particular items occurring in this list send us various messages. Logic is theoretical or descriptive on some accounts (e. g., as dialectics or analytic), but practical or normative in other views (e. g., as *organon* or *Kunstlehre*). This important difference was also captured by the medieval distinction (see [9]) of *logica docens* (logic as a theory) and *logica utens* (applied logic). Petrus Hispanus stressed an important feature of logic, namely that *dialectica* (that is, logic) *est art atrium et scientia scientiarum ad omnium aliarum scientiarum methodorum principia viam habent.* Since Leibniz, projects of logic as *characteristica universalis* or *logica magna* became fairly popular. Logic in this role played the role of a general scheme providing a methodological and linguistic framework for the whole of science or for mathematics at least. Logic as *characteristica universalis* was usually contrasted with *calculus ratiocinator* conceived as a battery of the rules of inference (it was refreshed by the distinction of logic as calculus and logic as language; see [42]). Although formal logic played the central part in the whole of logic almost always (I say "almost", because sometimes, particularly in the Enlightenment, due to the general philosophical environment of this period formal logic was treated as secondary and unimportant), this attitude became explicit in mathematical logic. Let me quote a sample of preliminary explanations of what logic is (from classical sources and standard textbooks):

> "Symbolic or formal logic [...] is the study of the various general types of deduction." [31, p. 10]

> "Logic is concerned only with those grounds of judgment which are truths. To make a judgment because we are cognizant of other truths as providing a justification for it is known as *inferring.* There are laws governing this kind of justification, and to set up these laws of valid inference is the goal of logic." [13, p. 3]

> "[...] [formal] logic is concerned with the analysis of sentences

> or of propositions [...] and of proofs [...] with attention to the *form* in abstraction from the *matter*." [7, p. 1]

> "The aim of constructing our symbolic logic is that is shall serve as a precise criterion for determining whether or not a given instance of [...] reasoning is correct." [29, p. 5]

> "The primary subject-matter of logic is the structure pattern of demonstrative inference." [20, p. 10]

> "[...] logic is [...] the analysis of methods of reasoning [...] logic is interested in the form rather than the content of argument" [22, p. 1]

> "Logic is the study of reasoning" [36, p. 1]

> "Symbolic logic is a mathematical model of deductive thought." [11, p. 1]

> "Logic is the study of correct reasoning." [4, p. 3]

> "Formal logic is [...] a theory of logical consequence" [24, p. 19]

All quoted explanations (we could continue this list *ad infinitum*) are approximately the same and explain that logic studies deductive, correct or demonstrative inferences and does that by reference to the form of arguments with abstraction from their content. It is clear that the terms "valid argument", "reasoning", "correct reasoning", "deductive reasoning" or "inference" are considered as synonyms by the quoted authors or, at least, as co-extensive labels. This allows skipping the question, otherwise very interesting, whether inductive reasoning is logical or not, or what is the status of reasonings performed in ordinary life.

Yet some traditional problems borrowed from the historical analysis of the development of logic and its philosophy are still alive. The essence of Petrus Hispanus' *dictum* is clearly expressed in the views of Gödel, Tarski and Quine:

> "[...] [logic] is a science prior to all others, which contains the ideas and principles underlying all sciences." [16, p. 125]

> "[...] the word 'logic' is used [...] in the present book [...] as the name of the discipline which analyses the meaning of the concepts common to all sciences, and establishes general laws governing these concepts." [40, p. XII]

> "The lexicon is what caters distinctively to special tastes and interests. Grammar and logic are the central facilities, serving all comers." [25, p. 102]

Let me refer to this intuition as suggesting that logic is universally applicable. Still another aspect of logic found its impressive expression in Ryle:

> "[Logical principles] are perfectly general, anyhow in this sense, that differences of concrete subject-matters make no difference to the validity or fallaciousness of inferences [...] 'logical constants' are indifferent to subject-matter or are topic-neutral." [32, p. 115]

Roughly the same idea is expressed if one says (see [45, p. 149]; page-references to the reprint) that logic is fundamental, because "its principles and content cannot depend on non-logical ones" or points out (see [19, p. 177]; page-reference to the reprint) that logic is, so to speak, content-free, because "no substantial content is coded in it".

The collected material suggests a distinction of three understandings of the universality of logic:

(a) logic is universal, because it is universally applicable;

(b) logic is universal, because it is topic-neutral;

(c) logic is universal, because its principles are universally valid.

Although (a) – (c) can be attributed to *logica utens* as well as to *logica docens*, (a) seems to be primarily addressed rather to the former, but (b) and (c) as applied to the latter. Since *logica utens* acts as an applied science, its essence consists in formulating rules of performing inferences. On the other hand, *logica docens* has fairly descriptive tasks. It aims at a theoretical description of the world of logic, whatever this reality seems to be.

The contrast between normative (*logica utens*) and descriptive (*logica docens*) aspects of logic finds its illustration in Frege and Russell:

> "Like ethics, logic can also be called a normative science. How must I think in order to reach the goal, truth? We expect logic to give us the answer to this question, but we do not demand that it should go into what is peculiar to each branch of knowledge and its subject-matter. On the contrary, the task we assign logic is only that of saying what holds with the utmost generality, whatever its subject-mater. We must assume that the rules for our thinking and for our holding something to be true are prescribed by the laws of truth. The former are given along with the latter. Consequently we can also say: logic is the science of the most general laws of truth." [14, p. 128]

> "Logic is concerned with the real world just as truly as zoology, though with its more abstract and general features." [30, p. 169]

According to Frege, the normative aspect of logic consists in prescribing how to reach truth. I will not enter into various possible interpretations of Frege's related view but, in his view, the main property of the norms of logic is that they preserve truth (or they are truth-preserving), that is, they lead from true premises to true conclusions by necessity. Frege links the normative function of logic with its descriptive task, that is, the account of the most general laws of truth. Let us assume as given that the same understanding can be ascribed to Russell's words that logic is concerned with "more abstract and general features" of the real world. Thus, if we take Frege's and Russell's views together, we have a link between *logica utens* as the set of truth-preserving rules and *logica docens* as dealing with the most general laws of thought. Any reasonable philosophy of logic should explain this regularity in a way.

What about *logica magna*? Well, the Frege-Russell logicism or the Leśniewski foundational project are good examples of *logica magna* in its maximal proposal. Today, we have rather more moderate accounts, like this:

> "[...] logic consists of a collection of mathematical structures, a collection of formal expressions, and a relation of satisfaction between the two [...]. We can say, then, that a logic is something we construct to study the logic of some parts of mathematics." [2, pp. 4–5]

Logic in this view offers suitable tools envisaged as possible descriptions of various mathematical structures. Doubtless, any good logic must be strongly expressive in order to capture as much mathematical content as possible. The requirement of a great expressive power of logic leads to the fourth understanding of logical universality (I will refer to it by (d)). According to this meaning of universality, logic is universal in this sense if its content is rich: its universality is, so to speak, proportional to its content.

How are the understandings of the universality of logic displayed by (a) – (d) mutually related? A tentative answer is that (a) – (c) are equivalent, but (d) captures another understanding of universality. Consider the following characterization of logic ([47, pp. 7–8]; I omit the issue of extensionality, because it is not relevant in the present context):

(1) The study of logic is the study of a certain type of concepts, most important of which are the concept of logical consequence and logical truth. [. . .]. Put differently, it is the study of theories or instrument of *deduction*. [. . .].

(2) Logical truth is truth due (only) to logical form. [. . .].

(3) Truth is a relation between sentences on the one hand and the structures on the other [. . .]. [. . .]

(4) In logic there are not privileged objects.

Further (p. 16), Westerståhl defines logic as an ordered pair $\langle S_L, \models\rangle$, where S_L is the class of sentences of a language L and $\models$ is the truth relation (I omit additional constraints concerning morphisms between structures interpreting L).

It is clear that Westerståhl combines various accounts of universality. The point (4) gives a version of the thesis that logic is topic-neutral, the points (1) and (2) are familiar from the previous remarks. The view expressed in (3) stresses the semantic nature of the concept of truth. Now, Westerståhl's definition of logic is related (I am not sure whether consciously or not) to the mentioned idea of *characteristica universalis* (*logica magna*) in its moderate version. In fact, Westarståhl develops the idea of abstract logic as a collection of formal schemes constructed in order to investigate various mathematical structures. Thus, we obtain a new characterization of the old idea, which consists in the semantic explication of the nature of *logica magna* as language *cum* the satisfaction relation. Although this explanation of the nature of logic is interesting in itself, it does not contribute very much to the problem of how (a) – (c) and (d) are related. We can go a step further, when we try to characterize *calculus rationicator* in the manner similar to that used on the occasion of *characteristica universalis*. *Calculus rationicator* appears as a syntactic object. Generally speaking, we have two objects (I slightly change Westarståhl's notation): (I) $\langle L, \models \rangle$ and (II) $\langle L, \vdash \rangle$ (the second is principally used in Rasiowa and Sikorski [26, p. 187], although not for the discussion of problems advanced in this paper). The question concerning the relation of (I) and (II) is good, because, as we will see, the answer to it opens the way to a promising account of the universality property of logic. FOT claims that both accounts of logic are equivalent. It means, the rejection of this thesis means that the equivalence in question does not hold generally.

Let me stress here that the distinction of (I) and (II) does not mean the same as that of *logica utens* and *logica docens*. On the contrary, the people who reject FOT insist that we need a powerful expressive scheme just because logic as a codification of deductive means (usually identified with FOL) has a very limited application. A message of this kind is clearly indicated by the following words:

> "As logicians we do our subject a disservice by convincing others that logic is first-order logic and then convincing them that almost none of the concepts of modern mathematics can really be captured in first-order logic. Paging through any modern mathematics book, one comes across concept after concept that cannot be expressed in first-order logic. Concepts from set theory (like *infinite set*, *countable set*), from analysis (like *set of measure* 0 or *having the Baire property*), from topology (like *open set* and *continuous function*), and from probability theory (like *random variable* and *having probability greater than some number* r), are central notions in mathematics which, on the mathematician-in-the street view, have their own logic. Yet none of them fit within the domain of first-order logic. In some cases the basic presuppositions of first-order logic about the kinds of mathematical structures one is studying are inappropriate (as the examples from topology or analysis show). In other cases, the structures dealt with are of the sort studied in first-order logic, but the concepts themselves cannot be defined in terms of 'logical constants'. [...]. Extended model theory adds a new dimension and new tools to

> the study of the logic of mathematics. The first-order thesis, by contrast, confuses the subject matter of logic with one of its tools. First- order logic is just an artificial language constructed to help investigate logic, much as the telescope is a tool constructed to help study heavenly bodies. From the perspective of the mathematician in the street, the first-order thesis is like the claim that astronomy is the study of the telescope. Extended model theory attempts to take the experience gained in first-order model theory and apply it in ever broader contexts, by allowing richer structures and richer ways of building expressions. It attempts to build languages similar to the first-order predicate calculus to study concepts that are banned from logic by the first-order thesis. [...] Mathematicians often lose patience with logic simply because so many notions from mathematics lie outside the scope of first-order logic, and they have been told that is logic. The study of model-theoretic logics should change that, by getting at the logic of the concepts mathematicians actually use, by finding applications, and by the isolation of still new concepts that enrich mathematics and logic. [...] one thing is certain. There is no going back to the view that logic is first-order logic." [2, pp. 5–6, p. 23]

Barwise's rejection of FOT is explicit and radical. His main argument appeals to a very poor applicability of FOL in mathematics. His arguments are pragmatic, because they point out that FOL is not suitable for defining and analysing mathematical structures and mathematical concepts. In particular, extended model theory (other labels: abstract model theory, abstract logic) is of the utmost significance for mathematics, because it considerably increases the expressive power of logic. In fact, Barwise does not claim that FOL is to be rejected, but argues that it is not sufficient "from the perspective of the mathematician in the street" and must be enriched by devices offered by extended model theory. "To put FOL into its right place" can serve as a concise summary of Barwise's position toward FOT and first-order logic. On the other side, the defenders of FOT argue that FOL has various elegant and nice properties. For example, it is semantically complete and has an effective (recursive) proof-procedure, contrary to second-order logic. We have also the Lindström characterization theorem (LI) saying that if a logic LOG has only Boolean connectives, countable language, is either (semantically) complete or compact (if every finite subset of a given set has a model, this set has a model)) and satisfies the Löwenheim property (if a set of sentences has an infinite model, it has a countable model), then LOG is equivalent with FOL with respect to expressive power. On the other hand, one should also note a hot controversy, whether these properties are natural or not. The opponents of FOT say that compactness limits the expressive power of first-order theories, because it is responsible for them being insufficient to define important mathematical concepts (for example, "to be a finite set"). Further, the Löwenheim property leads to a well-known puzzle (the Skolem paradox): set theory with uncountable sets has countable models. Generally speaking, FOL has a small expressive power and is not categorical, that is, it does not provide the unique (up to isomorphism) of defined concepts, for example, the concept of natural number. Thus, although the completeness theorem (every tautology of FOL is provable, every consistent set of first-order sentences has a model) and the Lindström theorem provide an elegant formal account of essential properties of FOL, it is far from being obvious that "essential" means "logically natural".

The last sentence brings us back to philosophical issues, at least if we assume that philosophy of logic should investigate the essence of logic just as a conceptual matter. As I already noticed, the universality property of logic is taken in this paper as its most essential attribute. Although none of the mentioned formal metalogical results directly says anything about the universality property and its aspects defined by (a) – (c) above, metalogic can, however, be used for illuminating some points. Thus, let us investigate what can be derived for the universality property of logic from metalogical results. Anticipating the further discussion, if this work should be done, we will have material allowing us to make a contrast of (a) – (c) with (d), selected as the philosophical choice by Barwise. On the way we will touch some corollary problems, namely the relation of *logica docens* and *logica utens* or the relation of (I) and (II), that is, $\langle L, \models \rangle$ and $\langle L, \vdash \rangle$.

The first thing to do is to define logic by metalogical devices. I will begin with *calculus rationicator*, that is, logic conceived as a deductive manual. Intuitively speaking, such manuals provide instructions on how to prove some propositions on the basis of other propositions adopted as premises. It is done in virtue of inference rules; for example, modus ponens informs us that it is logically allowed to pass A and $A \longrightarrow B$ as premises to B as the conclusion. Assume that $\mathbf{R}$ is a set of inference rules. The notation $X \vdash^{\mathbf{R}} A$ expresses that a formula A is provable (derivable) from the set X of assumptions, relative to rules of inference from **R** (further I will omit the superscript indexing the provability sign). We define

Definition 1 (*Cn*). $A \in \mathrm{Cn}(X) \Longleftrightarrow X \vdash A$.

Cn (the consequence operation) and $\vdash$ (the consequence relation) are mutually interdefinable, but a categorial difference between them should be noted. Let L be a language understood as a set of formulas. *Cn* maps 2^L into 2^L, that is, sets of formulas into sets of formulas, but the consequence relation act on the subsets of $2^L \times L$, that is, links sets of formulas with single formulas.

Although analysis of logic *via* the consequence operator appears more often than the approach *via Cn*, I take the latter approach (I follow [38]). How many consequence operations have we? The answer is that there are infinitely (even, uncountably) many of them. In this situation, we need to fix some constraints selecting a "reasonable" consequence operation (or operations). Tarski (see [39]) characterized the classical *Cn* axiomatically (in fact, Tarski axiomatized the consequence operation associated with the propositional calculus; the axioms given below concern the consequence operation suitable for first-order logic). The axioms are these (FIN – "finite"):

(C1) $\emptyset \leq L \leq \aleph_0$;

(C2) $X \subseteq CnX$;

(C3) $X \subseteq Y \longrightarrow CnX \subseteq CnY$;

(C4) $CnCnX = CnX$;

(C5) $A \in CnX \longrightarrow \exists Y \subseteq X \wedge Y \in \mathsf{FIN} \wedge (A \in CnY)$;

(C6) $(A \longrightarrow B) \in CnX \longrightarrow B \in Cn(X \cup \{A\})$;

(C7) $B \in Cn(X \cup \{A\}) \longrightarrow (A \longrightarrow B) \in CnX$;

(C8) $Cn\{A, \neg A\} = L$;

(C9) $Cn\{A\} \cap Cn\{\neg A\} = \emptyset$;

(C10) $A(\nu/t) \in Cn\{\forall \nu A(\nu)\}$, if the term t is substitutable for ν;

(C11) $A \in CnX \longrightarrow \forall \nu A(\nu) \in CnX$, if ν is not free in B, for any $B \in X$;

(C12) $t_i = t_i \in Cn\emptyset$;

(C13) $(\ldots(s_1 = t_1) \wedge \ldots \wedge (s_n = t_n)\ldots)$
$$\longrightarrow A(t_1, \ldots, t_n) \in Cn\{A(s_1, \ldots, s_n)\}.$$

The set (C1) – (C13) can be divided into two groups. The first group includes (C1) – (C5) as general axioms for *Cn*. (C1) says that the cardinality of L is at most denumerably infinite, (C2) that any set is a subset of the set of its consequences, (C3) established the monotonicity of *Cn*, (C4) its idempotency, (C5) states the finiteness condition, which means that if something belongs to $Cn(X)$, it belongs to the set of consequences of a finite subset of X. In other words: every inference is finitary, that is, performable on the base of a finite set of premises and, according to the character of rules, finitely long. (C1) – (C5) do not provide any logic in its usual sense, because they do not generate any rules of inference. The logical machinery is encapsulated by the rest of the axioms (related to logic based on negation, implication and the universal quantifier): (C6) the deduction theorem (if it is to be applied to predicate logic, we must assume that A and B are closed formulas, that is, sentences), (C7) formulates modus ponens, (C8) – (C9) characterize negation, (C10) – (C11) characterize the universal quantifier, and (C12) – (C13) deal with identity. The status of identity leads to controversies. The main reason to include it in the list of logical constants is that first-order logic with identity satisfies the main metalogical theorems concerning elementary logic: the completeness, the compactness, the Löwenheim-Skolem(LS) and the Lindström (LI). This suggests that identity behaves like propositional connectives and quantifiers. On the other hand, since identity makes it possible to define numerical quantifiers, like "there are exactly two", "there are exactly three", etc. (for arbitrary natural n), it seems to introduce extralogical contents to logic and, thereby, should not be considered as a logical notion. This controversy can be settled only by a decision, and my choice is to include identity among logical constants (see also Appendix). Thus, FOL refers to first-order (classical) logic with identity.

How to define logic *via Cn*? Having the deduction theorem we say that LOG is identified as $Cn\emptyset$. More formally we have:

(D1) $A \in \mathsf{LOG} \Longleftrightarrow A \in Cn\emptyset$, or, equivalently $\mathsf{LOG} = Cn\emptyset$.

(D1) looks artificial at first sight; clearly, and the empty set looks here like a convenient metaphor. In particular, one might argue, we can derive something from the empty set only because of the logical machinery is already incorporated into Cn. Otherwise speaking, we tacitly assumed that axioms for Cn have a certain logical content. Hence, the question arises how to justify that stipulations (C1) – (C13) about the consequence operation are proper for logic. As far as the general axioms are concerned, we can for instance drop the requirement of monotonicity (it leads to non-monotonic logics used in computer science) or finiteness in order to obtain infinitary logics. Hence, any definition of logic *via* the consequence operation needs additional justification. (C1) and (C5) are related to human faculties in performing inferences. A possible defence of these axioms consists in pointing out that our inferential performances have the finitary character, because we always employ finite sets of premises of a finite length. This is not at odds with (C1), which admits that the set of sentences can be countably infinite, because it means that this set is simply inductively extendible; even if we admit that $\aleph_0$ represents the actual infinity, it is a fairly moderate ontological presupposition. (C2) is obvious as including axioms as well as other asserted assumptions among theorems. (C3) says that Cn acting more than once on the given set, produces nothing more. The problem of monotonicity ((C4)) is more complicated and I restrict myself only to one remark in favour of this property, namely that it is plausible to say that if we can derive something from the empty set, it is also derivable from any other set. Let us take for granted that (C1) – (C5) are justified. The rest of the axioms characterize classical logic. If they are changed, for example, by weakening the force of negation, we will obtain a non-classical logic, for example, intuitionistic. The deduction theorem is, of course, very desirable. In particular, it is essential for obtaining (D1).

However, (D1) applies not only to FOL. Leaving aside non-classical cases, (D1) is equivalent, modulo (C1) – (C13) (in fact, (C1) – (C9) are enough; moreover, we have intuitionistic counterparts of (C8) and (C9)) to two other statements, namely:

(D2) $A \in \mathsf{LOG}$ if and only if $\neg A$ is inconsistent.

(D3) LOG is the only non-empty product of all deductive systems (theories).

(D2) and (D3) surely define the properties, which we expected to be possessed by any reasonable logic (paraconsistency is to be separately discussed at this point). We agree that negations of logical principles are contradictory and that logic is the common part of all, even mutually, inconsistent theories. Additionally, (D3) entails that logical laws are derivable from arbitrary premises. Thus, we immediately obtain the equivalence: $A \in Cn\emptyset$ if and only if $A \in CnX$, for any X, and the equality $\mathsf{LOG} = Cn\emptyset = CnX$, for any X.

Yet one can point out objections to the above explanations. In particular, that they seem to play with FOL and LOG, sometimes regarding them as interchangeable, sometimes not. Further, every formal system can be defined as $Cn\emptyset$, if axioms for axioms for Cn are modified. Let T is a theory axiomatized by a set A of axioms or axiom schemes. Assume that the symbol CA denotes the conjunction of the axioms of T (in general, we do not need to claim that CA is finite; see below). Assume further that $t \in CnA$. By the deduction theorem we have $(\mathsf{CA} \longrightarrow t) \in Cn\emptyset$. This is all right. However, if we add the formula

($*$) $\mathsf{CA} \in Cn\emptyset$

as a new axiom for Cn, we immediately obtain that $t \in Cn\emptyset$. On the syntactic level, nothing prevents such moves. In fact, the axiom (C12) is of this kind. It was added, because there are reasons for considering identity as a logical concept. However, it is difficult to agree that the axioms of the type ($*$) are sound in every case. In most cases they are not.

The last section suggests that it is significant to have another account of logic, which would be independent of the way *via Cn*. It can be achieved with the help of semantics, which motivates

(D4) $A \in \mathsf{LOG}$ if and only if for every model M, A is true in M.

This definition describes logic as consisting of laws true in every model (domain, possible world, interpretation, etc).

Now we can return to the universality property of logic. I distinguished four ways of how this property could be understood. To repeat: (a) logic is universal, because it is universally applicable; (b) logic is universal, because it is topic-neutral; (c) logic is universal, because its principles are universally valid; (d) logic is universal, because it has a great expressive power. I also suggested that that (a) – (c) are equivalent. Now we can precisely state these intuitions. If LOG is a part of every theory, it means that it is universally applicable. Exactly the same follows from (D4), because logic as true in every model is applicable in every concrete deductive inference. Further, since LOG belongs to every theory T, independently of T-content, LOG is true in every model and LOG does not depends on specific assumptions, it is also topic-neutral. Thus, starting from (a) or (b) or (c), we intuitively obtain other points. This is only an informal reasoning. Formally speaking, (a) – (c) are equivalent, it we can accept that (D1) and (D4) are equivalent as well, that is,

(#) $A \in Cn\emptyset$ if and only if for every model M, A is true in L.

The justification for (#) comes from the completeness theorem. It has two versions: strong and weak. The former says (I use the consequence relation in this case)

(SV) $X \models A$ iff $X \vdash A$,

and the latter states

(WV) $\emptyset \models A$ iff $\emptyset \vdash A$.

(SV) is much more attractive, because it concerns all theories, not only strictly logical systems, that is, exclusively consisting of tautologies. The strong version requires the following definition of logic

(D5) $\mathsf{LOG} = \langle L, Cn \rangle$,

where L is an arbitrary first-order language. (D5) suggests that its right side should be replaced by $\langle L, \models \rangle$ on the semantic level.

There is, of course, nothing wrong with looking at logic as an arbitrary first-order language together with a consequence operation, but it does not deal directly with the universality of logic. Assume that a LOG satisfies (SV), (C5) and (C7). Thus, every derivation in LOG is reducible to a derivation from a finite set of premises and the right side of (SV) can be replaced by $\mathsf{CX} \vdash A$. By (C7), that is, the deduction theorem, we obtain $\emptyset \vdash \mathsf{CX} \longrightarrow A$ and, further, by (WV), that the implication (i) $\mathsf{CX} \longrightarrow A$ is universally valid. Therefore, (i) is a tautology. In fact, (SV) says that a derivation represented by $\mathsf{CX} \vdash A$ proceeds *via* a rule of logic, which is represented by a logical theorem (i). The universality property of (i) is directly established by (WV). Although (WV) is obtainable from (SV) by putting $\emptyset$ in the place of X, the former still says something non-trivial, namely

(##) for any X, $A \in CnX$ if and only if A is true in all models.

I am inclined to say that (SV) is about *logica utens*, but (WV) about *logica docens* (see below). It seems (WV) is philosophically more important for logic conceived as the stock of tautologies. Having (D4) justified, it is easy to show that (a) – (c) express the same property. (D1) says that logic is independent of any specific assumptions. It is formally displayed by (D1) and its corollary, which says that logic is a part of every theory. (D4) indicates that logical laws are universally valid and topic neutral. Now (##) establishes that (a) – (c) are equivalent, not only by convention, by due to the completeness theorem, that is, a firm metalogical result.

Now I can return to *logica utens* and *logica docens*. The former consists of rules of inferences, the latter from theorems. Let $\mathsf{LOG}^{\mathbf{R}}$ consists of a stock of rules and LOG^{T} covers a stock of theorems. Assume that $\mathbf{R} = \langle\{A_1, \ldots, A_n\}, A\rangle$ is a rule of inference with the premises $A_1, \ldots, A_n$ and the conclusion A. The deduction theorem and (SV) justifies

(###) $\langle A_1, \ldots, A_n, A\rangle \in \mathsf{LOG}^{\mathbf{R}}$ iff $(A_1 \longrightarrow (\ldots \longrightarrow (A_n \longrightarrow A)\ldots) \in \mathsf{LOG}^{T}$.

This establishes the parity of $\mathsf{LOG}^{\mathrm{T}}$ and $\mathsf{LOG}^{\mathbf{R}}$ and thereby, also the parity of *logica utens* and *logica docens*. It means that the definitions (D1) – (D4) can be applied to logical theorems and logical rules as well. The normativity of logic, as related to (D4), has an interesting feature, which is related to Frege's point that logic tells us how to think in order to reach truth. Since logic does not favour any possible world (model), every world is logically accessible from any other. The standard definition of obligation tells us that $\mathbf{O}A$ is true in our world M^* if and only if A is true in all possible world accessible from M^*. If A is a tautology, it is true in all worlds, including M^*. Thus $\mathbf{O}1$ (where the symbol 1 denotes an arbitrary tautology) is true in M^* (in any other world as well). Thus, tautologies generate the realm of logical oughtness (we do not worry about the ontological status of this realm). Further, the relation of logical accessibility is reflexive. It means that **O**1 implies 1. We have also the reverse dependence. Briefly, ought and is are not distinguishable in logic. This can be interpreted as the exception to the Hume thesis that ought is logically separate from is, but it appears to be the only exception. I guess that it is a proper interpretation of Frege's idea that logic is normative. Moreover, the most general laws of truths in Frege's sense or more abstract and general features of the real world can be identified with tautologies or all models (also at this point, Frege was more subtle than Russell). Assume further, together with Frege, that if A is true, we should assert A. Hence, since tautologies are true, we should assert them unconditionally. Logic in itself does not force anybody to assert it, but if it comes to the cognitive game, the situation changes, because the obligation to assert something appears. On the other hand, truth and assertion are not the same. Contrary to Frege, if A is asserted, it does not need to be true, unless we have to do with tautologies. Therefore, Frege's opinion that logic is normative, because it says how we must think in order to reach truth, is to be corrected. The normativity of logic as far as matter concerns reaching truth is restricted to inferences. Assume that Cn closes assertion. This simply means that if A is asserted (the grounds of extralogical assertions are not relevant here), $B \in CnA$, then B is asserted. In order to be more realistic, we can add that it is known to the inferring person that $B \in CnA$. Assume that A is asserted, it is known that $B \in CnA$ and B is not asserted. By the deduction theorem, we obtain that $(A \longrightarrow B) \in Cn\emptyset$. Thus, the formula $(A \longrightarrow B)$ is a tautology. Applying the principle of the assertion of tautologies, we obtain that the formula $(A \longrightarrow B)$ is asserted. Since assertion is distributive over implication, we get $(Ass\ A \longrightarrow\ Ass\ B)$. However, if B is not asserted, A also is not asserted, contrary to the first assumption. An easy reasoning shows that if A is asserted conditionally as a conclusion of a correct inference with asserted premises, it ought be asserted as well.

Let me return to FOL and the claim that it is *the* logic. It consists, as *logica docens*, of tautologies and, since it satisfies (WV), (#) is justified for it. It shows that FOL possesses the universality property. There is something more to be said about FOL in the light of (WV). At first, let me note that (D5) is also applicable in this case, but with the proviso that L has a purely logical vocabulary, that is, individual variables, propositional connectives, quantifiers, identity and predicate letters understood as non-specified parameters. We assume that a logical theorem is a formula consisted of the above building blocks and derivable from the empty set of premises; in particular, we assume that instantiations of logical theorems do not belong to logic in this understanding. Clearly, logic as the system theorems is never systematized by listing all logical tautologies; it would even be impossible for there are infinitely many logical truths. Hence, logic is codified by a suitable axiomatic system. In particular, FOL has a complete axiomatization. Assume that $\mathsf{AX}^{\mathsf{FOL}}$ is a set of axioms for FOL, that is $\mathsf{FOL} = Cn\mathsf{AX}^{\mathsf{FOL}} = Cn\emptyset$. This means that *logica docens* is generated from the axioms. Hence, the rules leading from axioms are theorems must preserve tautologicity, although rules associated with logic as $\langle L, \vdash\rangle$ preserve "only" truth. It means that the former are stronger, than the latter. Of course, every tautologicity-preserving rule is also truth-preserving, but one can observe that the rule of substitution for predicate letters (as applied in $\langle L, \vdash\rangle$) does not generally lead from tautologies to tautologies (see [23, p. 90]); independently of that, this rule transforms abstract theorems into their instantiations. Although truth-preserving is sufficient for logic as $\langle L, \vdash\rangle$, where L is arbitrary, the difference touched in this section is essential, at least from the philosophical point of view, because it points out a certain peculiarity of the universality of *logica docens*. It is important to see that $\mathsf{LOG} = Cn\emptyset$ is also a language with the consequence relation. The completeness theorem allows to see it as $\langle L, \models\rangle$ as well. This shows that Shapiro's (see [34, p. XV]) qualification of first-order logic as only a calculus is not well-founded, because it is also a language with the satisfaction relation, although very limited.

There is still one result, which contributes to the interpretation of the universality property. FOL satisfies the following neutrality theorem (c_i, c_j are individual constants, P_k, P_n are predicate parameters, the notation $A(c)$ and $A(P)$ means that a constant c (predicate parameter P) occurs in A):

(ND) (a) $A(c_i) \in \mathsf{LOG} \longrightarrow A(c_j/c_i \in \mathsf{LOG}$;
(b) $A(P_k) \in \mathsf{LOG} \longrightarrow A(P_n/P_k) \in \mathsf{LOG}$.

This theorem says that if something is provable in logic about an object or its property, the same can be also proved about any other object or property. Otherwise speaking, FOL does no distinguish any extralogical item, that is, an individual constant or a predicate parameter. The semantic proof of (ND) uses (WV), but one can also prove this theorem syntactically without any reference to semantics (see [17, Ch. I §9], [23, Ch. 13]). The use of the completeness theorem in the semantic proof of (ND) indicates its link with the universality property. In fact, (ND) displays in a way that FOL is topically neutral. Incidentally, what Westerståhl proposed as an intuitive mark of logic, finds, *via* (ND), its precise wording in metalogic.

Is (#) as a criterion of logic a sufficient and necessary condition? Certainly, it is a necessary condition. As such it excludes second-order logic, because its completeness theorem (Henkin) does not treat all models al pari. More specifically, second order logic with full models is incomplete, but it becomes complete if its model are stratified in a way. However, second order logic in the later case is equivalent to many-sorted FOL. Thus, second-order logic with standard semantics (no model is distinguished) is not universal for its incompleteness, but it is also not universal, when non-standard (Henkin) semantics is admitted for the stratification of models (it should be considered as a first-order extralogical theory). Boolos (see [5, p. 77], page-reference to the reprint) tries to overcome this argument. He says:

> "I know of no perfectly effective reply to this view [that logic is topic-neutral – J. W.]. But, in the first place, one should perhaps be suspicious of the identification of subject matter and range. (Is elementary arithmetic really not *about* addition, but only *about* numbers?) And then it might be said that logic is not so "topic-neutral" as it is often made out to be: it can easily be said to be about the notions of negation, conjunction, identity, and the notions expressed by "all" and "some", among others (even though these notions are almost never quantified over). In the second place, unlike *planet* or *field*, the notion as of *set*, *class*, *property*, *concept*, and *relation*, etc. have often been considered to be distinctively logical notions, probably for some such very simple reason that anything whatsoever may belong to a set, have a property, or bear a relation. That some set- or relation-existence assertion are counted as logical truths in second-order systems does not, it seems to me, suffice to disqualify them as system of logic, as a system would be disqualified if it classified as a truth of logic the existence of planet with at least two satellites."

First of all, logic is not about the logical concepts. They are studied in metalogic. Take the notion of conjunction. It can be construed as a function from L to the set $\{\mathsf{True}, \mathsf{False}\}$, that is, the set of truth-values. If $A, B \in L$, then $v(A \wedge B) = \mathsf{True}$, when $v(A) = v(B) = \mathsf{True}$; otherwise, $v(A \wedge B) = \mathsf{False}$. Yet no theorem of propositional logic asserts that conjunction behaves in that way. We rather should say that the formula $A \wedge B \longrightarrow B$ becomes universally valid according to the above definition of conjunction. The second argument has also a very weak force, because just its content is the subject of the controversy in question. Following Boolos, one could say that second-order logic is logic for astronomy is not. Similarly, the argument (see [8]) pointing out the indispensable role of second-order sentences (for example, "true sentences logically imply true sentences") in elaborating properties of first-order tautologies seems to confuse logic and metalogic.

On the other hand, (WV) in its general form does not provide a necessary condition, because there are logics other than FOL, which are semantically complete, for example, some infinitary logics or logics with infinitary rules, for example, the ω-rule. However, if we say that (C5) is a natural property of logic, then only FOL remains. Thus, *the* logic, on the proposed views, has two marks, namely the universality property and the finitary (in the sense of (C5)) character of inference rule. Note, however, that cancelling (C5) as a source of logical properties still gives a definition, which ascribes the universality property in the considered sense to related. But if some generalized quantifiers (for instance, "there exists countably many" or "there exists uncountably many") are added, (ND) does not hold and the universality property is broken. This suggests, otherwise than many contemporary proposals do (see [47]) that generalized quantifiers, as favouring some cardinalities, are not logical constants, contrary to the usual quantifiers, that is, "for every" and "there is". Perhaps another argument (due to Alexander M. Levin and pointed to me by Valentin Shehtman) additionally enlightens this point. Logic should take into account the absolute properties. Now, due to the (LS), the notion of cardinality is not absolute. Hence, any theory, which distinguishes cardinalities is not a logic. As far as the matter concerns the quantifiers, only "for every" and "there is" (in particular, the former) appear as purely logical.

One can ask what (LI) tells us about the universality property. First of all, the Lindström theorem concerns rather *logica utens* (in the semantic version), that is, $\langle L, \models \rangle$, than *logica docens.* Secondly, (LI) stresses rather the expressive power of logic than its universality property. It is of course very interesting that completeness and compactness act to the same effect, when they occur together with the Löwenheim property. As far as the matter concerns logic as $Cn\emptyset$, its compactness is a trivial property. Applying it to the universality property, we obtain that a set of sentences is universally valid if and only if its every finite subset is universally valid, but this is nothing surprising (see also Appendix). The Löwenheim property displays an aspect of the universality, namely that FOL makes no difference between models, according to their cardinality. Also this feature of FOL is not surprising, because it treats all models al pari. Thus, compactness and the Löwenheim property are fairly natural from the point of view of *logica docens*, if it is identified with FOL. Its expressive power is very poor, almost null in the case of first-order logic without identity ("almost", because we assume that M's are not empty) and somehow richer (numerical quantifiers), when identity is added. However, although numerical quantifiers are definable in FOL (with identity), no cardinality is distinguished (by metalogical results, that is, from the point of view of metalogic) with one exception, namely countability. It is really a very surprising fact (due to (LS)) that if something is first-order satisfiable at all, it is satisfiable in a denumerable domain, even if we explain this by the cardinality of L. The most essential observation is perhaps this: a small expressive power of FOL as set of tautologies is a cost of its universality property.

These considerations should be supplemented by the remark that our semantics is based on the standard set theory. It is not without importance. One can ask how reliable is set theory as the base for semantics or metalogic for FOL. Of course, its reliability does not exceed that witnessed in the case of other parts of mathematics. However, using standard set theory in semantics and metalogic of FOL, we need to employ only a part of the set theoretical universe (in fact, the weak second order arithmetic with the arithmetical comprehension axiom is enough for first-order model theory; see [18, 37]). This circumstance is related to the absoluteness of FOL (see [43, 44]; this second paper shows how the absoluteness of FOL is related to (LI)). Roughly speaking, a logic (in the sense of $\langle L, \models \rangle$ is absolute if the truth value of the expression "$M \models A$" depends on the existence of some chosen sets (the existence of such sets is just guaranteed by the arithmetical comprehension axiom). On the other, hand, second-order logic is not absolute in this sense, because generates problems connected with the continuum hypothesis and other independent set-theoretical statements (see [44, 1]). If we go to metalogic of second-order logic, we cannot neglect the differences between various possible extensions of ZFC. In particular, Väänänen argues that the 1$^{\text{st}}$ order ZFC is equally good as second-order logic. Since the latter operates a very relative notion of set, this makes impossible to decide which set-theoretical universe is really "good" on clear logical grounds. Thus, according to Vänäänen, it is quite illusory to maintain that second-order logic gives us the proper characterization of the set-theoretical universe. Certainly, it is possible to appeal to other metalogical schemes, for example, to category theory or topoi, but I do not think that it would change the situation in a radical way.

If we pass to *logica utens*, that is $\langle L, \vdash \rangle$ and $\langle L, \models \rangle$, what is natural from the perspective *logica docens*, might be seen otherwise from the point of view of applications. Although I have no ambition to introduce terminological innovations, let me temporary speak about first-order formalizations (FOF) of theories, instead first-order logic. In order to display this idea in a explicit way, let $\langle L, \vdash \rangle$ and $\langle L, \models \rangle$ be replaced by $\langle L^1, \vdash \rangle$ and $\langle L^1, \models \rangle$, where the superscripts refer to the order of language. Logic is therefore hidden in $\vdash$, and semantic in $\models$. Now, we see that one should not speak about expressive power of $\vdash$ or $\models$, but refer this capacity to L^1. There is no doubt that expressive power of this language is very limited, but it fairly independent of the properties of the consequence relation. The same concerns non-categoricity of FOF. The matter whether FOF are good (or, how far good) for mathematics and science is still a controversial question (see [44, 1] for defence of FOF and the quoted works of Shapiro for the opposite view) and must be here omitted except invoking one of Vänäänen's remarks. He argues that second-order logic with Henkin's model cannot be distinguished from the full second-order logic. However, it seems clear that if logicians want to have powerful languages, they have to abandon FOF in favour of second-order formalisms. This move leads to system with the universality property in the sense (d). Now, it is also clear that universality in this sense is at odds with the universality property as determined by (a) – (c) and formally displayed by the metalogical characterization of FOL. Observe also that $\langle L \models \rangle$ is simply not comparable with $\langle L, \vdash \rangle$ without appealing to metalogical properties. Thus, we have a kind of dialectic between the universality property as validity (and its cognates) and the universality as expressive power. Some logicians want to have both universalities without any price. However, this task seems to be fairly impossible. Speaking of logic, one should choose the position: either the universality property or the great expressive power. Incidentally, the opinion that second-order theories are categorical is simply misleading. They are incomplete by the first Gödel theorem, because if arithmetic is consistent its extensions obtained by adding undecidable sentences are also consistent and have models. However, these models are radically different, although can be equicardinal. It means that also second-order theories have non-standard models, and this fact seems to be a derivative of powerful expressive devices of second-order languages. I see, then, the objection pointing out that first-order theories do not distinguish standard and non-standard models as simply unfair. It reminds Descartes' famous argument that if senses deceive us sometimes, they can deceive us in every case. This argument is strange, because if Descartes know that senses deceive us sometimes, he should also know, when they provide reliable information. There is no doubt which models are standard from the first-order perspective. The mistake consists here in an unfounded belief that we have purely logical criteria of the standardness of models. In fact, these criteria are extralogical also in the case of second-order theories.

An explicitly emotional tone, present in the above quotation from Barwise (about the point of view of "the mathematician in the street") is not a proper background for philosophical discussions about logic. Similarly, I regard as misleading the following words:

> "When we are interested in set theory or classical analysis, the Löwenheim theorem is usually taken as a sort of defect (often thought to be inevitable) of the first-order logic. Therefore, what is established (by Lindström theorems) is not that first-order logic is the only possible but rather that it is the only possible logic when we in a sense deny reality to the concept uncountability [...]." [46, p. 154]

There are two thoughts in Wang. Firstly, that the Löwenheim property is defective. As I tried to explain, it depends on the point of view or expectations concerning logic. Secondly, Wang connects the view that FOL is *the* logic with rejecting the concept of uncountability as referring to something real. I think that it is a wrong diagnosis. FOL as *logica docens* considers all cardinalities al pari (although the peculiar role of countability is obvious) and, thereby, has no ontological consequences stated by Wang. The only ontological assumption made by FOL is that something exists. Other ontological features of this logic are to be discussed independently, but (see introductory remarks) this issue goes beyond this paper.

A more promising view is offered by something like that:

> "Lindström's results show that it makes no sense to classify logic as either good or bad, depending on whether they are complete (compact) and have the Löwenheim-Skolem property of not. On the contrary, Lindström's result gave special emphasis to the proposal – already expressed by Kreisel [...] – that there must be balance between syntax and semantics of a logic and that the semantic properties we consider must be adapted to the expressive power and the special features of the given logic." [12, p. 78]

> "[...] we can use the Compactness Theorem to get a better understanding of the limitations of first-order logic – or, to put a more positive spin on it, a better understanding of the richness of mathematics!" [21, p. 103].

And what I have tried to do consisted in investigating that the balance is going on, when the universality property is taken as basic property of logic. Contrary to Barwise there is "going back to the view that [*the*] logic is first-order logic" under quite natural provisos. On the other hand, returning to terminological questions, it would be unwise to fight against the extended usage of the word 'logic' or constructing and studying various model-theoretic logics and languages, although it seems to me that the distinction between first-order logic and n-order formalizations is important. Perhaps a compromise, combining FOF as languages and FOL as the universal deductive code for mathematical proofs, is a tenable solution. However, one should always remember that terminology does not decide about properties of sets sentences considered as logical theories.

Appendix FOL has three segments (I omit general axioms for Cn): (A) propositional calculus (the axioms (C6) – (C9)); (B) first-order predicate logic without identity (the axioms: (C6) – (C111); (C) FOL itself (the axioms (C6) – (C13)). There are properties possessed by some parts, which are not attributable to others. (A) is i. a. semantically complete, Post-complete and decidable. (A) + (B) and (A) + (B) + (C) are neither Post-complete nor decidable. (A) + (B) obeys the inflation theorem (IT) (if a formula is satisfiable in a model of the cardinality n, it is also satisfiable in models of any greater cardinality) and the deflation theorem (DT) (if a formula is true in a model of a cardinality n, it is also satisfiable in models of any smaller cardinality), although these theorems do not hold for (A) + (B) +

(C). If some will insist that decidability is a natural property, only propositional logic (eventually, plus some fragments of predicate calculus) remains as *the* logic. At the first sight, (IT) and (DF) theorem seem to be arguments against including identity into the purely logical vocabulary. However, this argument can be met. The axioms (C12) and (C13) keep their validity in all models. Hence, the problem arises only for formulas, which are true in one-element models, two-element models, etc. This follows from the fact that identity defines numerical quantifiers, what is impossible in (B); logical devices available in (B) do not suffice to determine concrete models. Certainly, the expressive power of (C) is greater than that of (B) (similarly, the expressive power of (B) is greater than the content expressible in (A)), but the identity-tautologies still behave as any other tautologies. The systems (A), (A) + (B) and (A) + (B) + (C) are semantically complete and it is perhaps their most important logical property (they are, of course, also consistent, but this attribute has no use in determining what is *the* logic). Returning to identity, we must choose between the universality property and "having an amount of non-tautological content". This situation displays traditional disputes on the status of identity. I choose the universality property (together with (C5)) as determining *the* logic.

Although the set of tautologies of FOL is not maximally (Post) consistent, we can show that it is maximal in a sense (it is also absolute, because assumes that there is a non-empty set of objects). First, note that the consequences of tautologies should be tautologies too (recall that proofs inside $Cn\emptyset$ preserve tautologicity). We need to fix the semantical status of the empty set of sentences. This set is finite. Every finite set X of sentences is representable by a finite conjunction $\mathsf{C}X$. In general, $\mathsf{C}X \in CnX$. Thus, $C\emptyset \in Cn\emptyset$. This, assuming that we are working in logic satisfying the completeness theorem, implies that $C\emptyset$ is a tautology. Since $C\emptyset$ represents the set $\emptyset$, the latter has to be considered as the tautological set. In other words, we have the property TAUT associated with the set $Cn\emptyset$ such that TAUT (A) if and only if for any model M, A is true in M. It is clear that: (i) $\mathsf{TAUT}(\emptyset)$, and (ii) TAUT $(Cn\emptyset)$ if and only if $\mathsf{TAUT}(X)$, for every finite set X such that $X \subseteq Cn\emptyset$. Thus, TAUT is a property of the finite character. If we will identify the universality property with TAUT, this property is also of the finite character. Moreover (the Tukey Lemma), if any other set Y of sentences true in all models and closed by Cn satisfies (i) and (ii), then $Y = Cn\emptyset$. Although adding non-tautologies to $Cn\emptyset$ does not produce inconsistency in general, TAUT and the universality property are maximal in a well-defined sense. Observe that this reasoning does not go with respect to $\langle L, \vdash \rangle$ and (SV) of the completeness theorem. We can naturally proof that all tautologies are universally valid and derivable from the empty set of premises, but the consequence relation preserves truth, not tautologicity. However, truth is not a property of the finite character. Manipulating constraints for Cn can change this situation, but it brings us back to the issue of what is natural in metalogic and what should be captured by logic; for instance, we can postulate that every truth is a consequence of the empty set. It is also possible to employ some new metalogical concepts and results in the discussion about the essence of logic. FOL without identity is structurally complete, but the full FOL does not possess this property (see [23, Ch. 10]; [24, pp. 434–436]). Hence, if the structurality of the rules of inference expresses a natural logical property, identity loses its status as a logical constant. However, it seems to me that the structurality is rather a property of some formal codifications of logic, than of logic itself.

And still a word about modal logic. Their possible-world semantics is based on various properties of the accessibility relation, like reflexivity, symmetry, etc. However, they are not sources of universality (some semantics have these properties, others not). Thus, they introduce some extralogical element to the logical behaviour of modalities. In fact, only the system called K treats all possible words equally. Its modalities, that is, necessity and possibility, behave exactly as quantifiers. Perhaps this system represents the pure modal logic. However, and it is not surprising, K is not sufficient to cover all intuitions connected with modalities, because its expressive power is relatively small.

Acknowledgements. I am indebted to István Németi and Jerzy Pogonowski for valuable comments and advices.

1 Bibliography

[1] H. Andréka, J. Madárasz, and I. Németi. Why first-order logic? Appendix of On the logical structure of relativity theories. http://www.math-inst.hu/pub/algebraic -logic/Contents.html, 2003.

[2] J. Barwise. Model-theoretic logics: Background and aims. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 3–23. Springer, Berlin, 1985.

[3] J. Barwise and S. Feferman, editors. *Model-Theoretic Logics*. Springer, Berlin, 1985.

[4] D. Bonevac. *Deduction – Introductory Symbolic Logic*. Mayfield Publishing Company, Pao Alto, 1987.

[5] G. Boolos. On second-order logic. *Journal of Philosophy*, 72:509–527, 1975. Repr. in [35, pp. 70-87].

[6] L. Borkowski. Characterization of quantifiers in the axiomatic theory of consequence. In L. Borkowski and A. B. Stepien, editors, *Studies in Logic and the Theory of Knowledge*, volume 2, pages 37–39. Towarzystwo Naukowe KUL, Lublin, 1991.

[7] A. Church. *Introduction to Mathematical Logic*. Princeton University Press, Princeton, 1956.

[8] J. Corcoran. Second-order logic. In C. A. Anderson and M. Zel editors, *Logic, Meaning and Computation*. Kluwer Academic Publishers, 2001.

[9] S. Ebbesen. Logica docens/logica utens. In *[28]*, pages 353–355. 1980.

[10] M. Eklund. On how logic became first-order. *Nordic Journal of Philosophical Logic*, 1:147–167, 1996.

[11] H. Enderton. *A Mathematical Introduction to Logic*. Academic Press, New York, 1972.

[12] J. Flum. Characterizing logic. In *[3]*, pages 77–120. 1985.

[13] G. Frege. Logic. In *[15]*, pages 1–9, 1979.

[14] G. Frege. Logic. In *[15]*, pages 126–151. 1979.

[15] G. Frege. *Posthumus Writings*. Basil Blackwell, Oxford, 1979. Ed. by H. Hermes, K. Kambartel and F. Kaulbach, tr. By P. Long and R. White.

[16] K. Gödel. Russell's mathematical logic. In P. Schillpp, editor, *The Philosophy of Bertrand Russell*, pages 125–153. Open Court, La Salle, 1944.

[17] A. Grzegorczyk. *An Outline of Mathematical Logic. Fundamental results and Notions Explained with All Details.* Reidel, Dordrecht, 1974.

[18] V. Halbach. *Axiomatische Wahrheitstheorien.* Akademie-Verlag, Berlin, 1999.

[19] I. Jané. A critical appraisal of first-order logic. *History and Philosophy of Logic*, 14:67–86, 1993. Repr. in [35, pp. 177-196].

[20] G. T. Kneebone. *Mathematical Logic and the Foundations of Mathematics.* Van Nostrand, London, 1963.

[21] Ch. Leary. *A Friendly Introduction to Mathematical Logic.* Prentice-Hall, Upper Saddle River, 2000.

[22] E. Mendelson. *Introduction to Mathematical Logic.* Van Nostrand, New York, 1964.

[23] W. Pogorzelski. *Klasyczny rachunek kwantyfikatorów. Zarys teorii.* Pañstwowe Wydawnictwo Naukowe, Warszawa, 1981. Classical Quantification Theory. An Outline of Theory.

[24] W. Pogorzelski. *Notions and Theories of Elementary Formal Logic.* Warsaw University – Bialystok Branch, Bialystok, 1994.

[25] W. v. O. Quine. *Philosophy of Logic.* Prentice Hall, Englewood Clifts, 1970.

[26] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics.* PWN – Polish Scientific Publishers, Warszawa, 1970.

[27] W. Risse. Logik. In *[28]*, pages 357–362. 1980.

[28] J. Ritter and K. Gründer, editors. *Historisches Wörterbuch der Philosophie*, volume 5, Basel, 1980. Schwabe.

[29] J. B. Rosser. *Logic for Mathematicians.* McGraw Hill, New York, 1953.

[30] B. Russell. *Introduction to Mathematical Philosophy.* Georg Allen and Unwin, London, 1919.

[31] B. Russell. *The Principles of Mathematics.* Georg Allen and Unwin, London, 1903.

[32] G. Ryle. *Dillemmas.* Cambridge University Press, Cambridge, 1954.

[33] S. Shapiro. *Foundations without Foundationalism.* Oxford University Press, Oxford, 1991.

[34] S. Shapiro. Introduction. In *[35]*, pages XI–XXII. 1996.

[35] S. Shapiro, editor. *The Limits of Logic. Higher-Order Logic and the Löwenheim-Skolem Theorem.* Dartmouth, Aldershot, 1996.

[36] J. Shoenfield. *Mathematical Logic.* Addison-Wesley, London, 1967.

[37] S. G. Simpson. *Subsystems of Second-Order Arithmetic.* Springer-Verlag, Berlin, 1999.

[38] S. Surma. The growth of logic out of the foundational research in the foundations of mathematics. In E. Agazzi, editor, *Modern Logic*, pages 15–33. Reidel, Dordrecht, 1981.

[39] A. Tarski. Über einige fundamentale Begriffe der Metamathematik. *Comptes Rendus des séances de la Société des Sciences et de Lettres de Varsovie*, 23:22–39, 1930. Eng. tr. in A. Tarski, Logic, Semantics, Metamathematics. Papers from 1923 to 1938, Clarendon Press, Oxford, pp. 30–37.

[40] A. Tarski. *Introduction to Logic and the Methodology of Deductive Sciences.* Clarendon Press, Oxford, 1941.

[41] L. Tharp. Which logic is the right logic. *Synthese*, 31:11–16, 1975. Repr. in [35, pp. 47–67].

[42] J. Van Heijenoort. *Selected essays*, chapter Logic as Calculus and Logic as Language, pages 11–16. Bibliopolis, 1985.

[43] J. Vänäänen. Set-theoretical definability of logics. In *[3]*, pages 599–643. 1985.

[44] J. Vänäänen. Second-order logic and the foundations of mathematics. *Bulletin of Symbolic Logic*, 7:504–520, 2001.

[45] S. Wagner. The rationalist conception of logic. *Notre Dame Journal of Formal Logic*, 28:3–35, 1987. Repr. in [35, pp. 143–175].

[46] H. Wang. *From Mathematics to Philosophy.* Routledge and Kegan Paul, London, 1974.

[47] D. Westerståhl. *Some Philosophical Aspects of Abstract Model Theory.* Institutionen för Filosofi, Göteborgs Universitet, Göteborg, 1976.

[48] J. Woleński. In defence of the first-order thesis. In P. Kolar and V. Svoboda, editors, *Logica '93*, pages 1–11. Filosofia, Praha, 1994.

[49] J. Woleński. Logic from a metalogical point of view. In E. Orlowska, editor, *Logic at Work. Essays Dedicated to the Memory of Helena Rasiowa*, pages 25–35. Physica-Verlag A Springer-Verlag Company, Berlin, 1998.

Bereits erschienene und geplante Bände der Reihe

Logische Philosophie

Hrsg.: H. Wessel, U. Scheffler, Y. Shramko, M. Urchs

ISSN: 1435–3415

In der Reihe „Logische Philosophie“ werden philosophisch relevante Ergebnisse der Logik vorgestellt. Dazu gehören insbesondere Arbeiten, in denen philosophische Probleme mit logischen Methoden gelöst werden.

Uwe Scheffler/Klaus Wuttich (Hrsg.)

Terminigebrauch und Folgebeziehung

ISBN: 978-3-89722-050-8 Preis: 30,- €

Regeln für den Gebrauch von Termini und Regeln für das logische Schließen sind traditionell der Gegenstand der Logik. Ein zentrales Thema der vorliegenden Arbeiten ist die umstrittene Forderung nach speziellen Logiken für bestimmte Aufgabengebiete - etwa für Folgern aus widersprüchlichen Satzmengen, für Ersetzen in gewissen Wahrnehmungs- oder Behauptungssätzen, für die Analyse von epistemischen, kausalen oder mehrdeutigen Termini. Es zeigt sich in mehreren Arbeiten, daß die nichttraditionelle Prädikationstheorie eine verläßliche und fruchtbare Basis für die Bearbeitung solcher Probleme bietet. Den Beiträgen zu diesem Problemkreis folgen vier diese Thematik erweiternde Beiträge. Der dritte Abschnitt beschäftigt sich mit der Theorie der logischen Folgebeziehungen. Die meisten der diesem Themenkreis zugehörenden Arbeiten sind explizit den Systemen F^S bzw. S^S gewidmet.

Horst Wessel

Logik

ISBN: 978-3-89722-057-7 Preis: 37,- €

Das Buch ist eine philosophisch orientierte Einführung in die Logik. Ihm liegt eine Konzeption zugrunde, die sich von mathematischen Einführungen in die Logik unterscheidet, logische Regeln als universelle Sprachregeln versteht und sich bemüht, die Logik den Bedürfnissen der empirischen Wissenschaften besser anzupassen.

Ausführlich wird die klassische Aussagen- und Quantorenlogik behandelt. Philosophische Probleme der Logik, die Problematik der logischen Folgebeziehung, eine nichttraditionelle Prädikationstheorie, die intuitionistische Logik, die Konditionallogik, Grundlagen der Terminitheorie, die Behandlung modaler Prädikate und ausgewählte Probleme der Wissenschaftslogik gehen über die üblichen Einführungen in die Logik hinaus.

Das Buch setzt keine mathematischen Vorkenntnisse voraus, kann als Grundlage für einen einjährigen Logikkurs, aber auch zum Selbststudium genutzt werden.

Yaroslav Shramko

Intuitionismus und Relevanz

ISBN: 978-3-89722-205-2 Preis: 25,- €

Die intuitionistische Logik und die Relevanzlogik gehören zu den bedeutendsten Rivalen der klassischen Logik. Der Verfasser unternimmt den Versuch, die jeweiligen Grundideen der Konstruktivität und der Paradoxienfreiheit durch eine „Relevantisierung der intuitionistischen Logik“ zusammenzuführen. Die auf diesem Weg erreichten Ergebnisse sind auf hohem technischen Niveau und werden über die gesamte Abhandlung hinweg sachkundig philosophisch diskutiert. Das Buch wendet sich an einen logisch gebildeten philosophisch interessierten Leserkreis.

Horst Wessel

Logik und Philosophie

ISBN: 978-3-89722-249-6 Preis: 15,30 €

Nach einer Skizze der Logik wird ihr Nutzen für andere philosophische Disziplinen herausgearbeitet. Mit minimalen logisch-technischen Mitteln werden philosophische Termini, Theoreme und Konzeptionen analysiert. Insbesondere bei der Untersuchung von philosophischer Terminologie zeigt sich, daß logische Standards für jede wissenschaftliche Philosophie unabdingbar sind. Das Buch wendet sich an einen breiten philosophisch interessierten Leserkreis und setzt keine logischen Kenntnisse voraus.

S. Wölfl

Kombinierte Zeit- und Modallogik. Vollständigkeitsresultate für prädikatenlogische Sprachen

ISBN: 978-3-89722-310-3 Preis: 40,- €

Zeitlogiken thematisieren „nicht-ewige“ Sätze, d. h. Sätze, deren Wahrheitswert sich in der Zeit verändern kann. Modallogiken (im engeren Sinne des Wortes) zielen auf eine Logik alethischer Modalbegriffe ab. Kombinierte Zeit- und Modallogiken verknüpfen nun Zeit- mit Modallogik, in ihnen geht es also um eine Analyse und logische Theorie zeitabhängiger Modalaussagen.

Kombinierte Zeit- und Modallogiken stellen eine ausgezeichnete Basistheorie für Konditionallogiken, Handlungs- und Bewirkenstheorien sowie Kausalanalysen dar. Hinsichtlich dieser Anwendungsgebiete sind vor allem prädikatenlogische Sprachen aufgrund ihrer Ausdrucksstärke von Interesse. Die vorliegende Arbeit entwickelt nun kombinierte Zeit- und Modallogiken für prädikatenlogische Sprachen und erörtert die solchen logischen Systemen eigentümlichen Problemstellungen. Dazu werden im ersten Teil ganz allgemein multimodale Logiken für prädikatenlogische Sprachen diskutiert, im zweiten dann Kalküle der kombinierten Zeit- und Modallogik vorgestellt und deren semantische Vollständigkeit bewiesen.

Das Buch richtet sich an Leser, die mit den Methoden der Modal- und Zeitlogik bereits etwas vertraut sind.

H. Franzen, U. Scheffler

Logik. Kommentierte Aufgaben und Lösungen

ISBN: 978-3-89722-400-1 Preis: 15,- €

Üblicherweise wird in der Logik-Ausbildung viel Zeit auf die Vermittlung metatheoretischer Zusammenhänge verwendet. Das Lösen von Übungsaufgaben — unerläßlich für das Verständnis der Theorie — ist zumeist Teil der erwarteten selbständigen Arbeit der Studierenden. Insbesondere Logik-Lehrbücher für Philosophen bieten jedoch häufig wenige oder keine Aufgaben. Wenn Aufgaben vorhanden sind, fehlen oft die Lösungen oder sind schwer nachzuvollziehen.

Das vorliegende Trainingsbuch enthält Aufgaben mit Lösungen, die aus Klausur- und Tutoriumsaufgaben in einem 2-semestrigen Grundkurs Logik für Philosophen entstanden sind. Ausführliche Kommentare machen die Lösungswege leicht verständlich. So übt der Leser, Entscheidungsverfahren anzuwenden, Theoreme zu beweisen u. ä., und erwirbt damit elementare logische Fertigkeiten. Erwartungsgemäß beziehen sich die meisten Aufgaben auf die Aussagen- und Quantorenlogik, aber auch andere logische Gebiete werden in kurzen Abschnitten behandelt.

Diese Aufgabensammlung ist kein weiteres Lehrbuch, sondern soll die vielen vorhandenen Logik-Lehrbücher ergänzen.

U. Scheffler

Ereignis und Zeit. Ontologische Grundlagen der Kausalrelationen

ISBN: 978-3-89722-657-9 Preis: 40,50 €

Das Hauptergebnis der vorliegenden Abhandlung ist eine philosophische Ereignistheorie, die Ereignisse über konstituierende Sätze einführt. In ihrem Rahmen sind die wesentlichen in der Literatur diskutierten Fragen (nach der Existenz und der Individuation von Ereignissen, nach dem Verhältnis von Token und Typen, nach der Struktur von Ereignissen und andere) lösbar. In weiteren Kapiteln werden das Verhältnis von kausaler und temporaler Ordnung sowie die Existenz von Ereignissen in der Zeit besprochen und es wird auf der Grundlage der Token-Typ-Unterscheidung für die Priorität der singulären Kausalität gegenüber genereller Verursachung argumentiert.

Horst Wessel

Antiirrationalismus

Logisch-philosophische Aufsätze

ISBN: 978-3-8325-0266-9 Preis: 45,- €

Horst Wessel ist seit 1976 Professor für Logik am Institut für Philosophie der Humboldt-Universität zu Berlin. Nach seiner Promotion in Moskau 1967 arbeitete er eng mit seinem Doktorvater, dem russischen Logiker A. A. Sinowjew, zusammen. Wessel hat großen Anteil daran, daß am Berliner Institut für Philosophie in der Logik auf beachtlichem Niveau gelehrt und geforscht wurde.

Im vorliegenden Band hat er Artikel aus einer 30-jährigen Publikationstätigkeit ausgewählt, die zum Teil nur noch schwer zugänglich sind. Es handelt sich dabei um logische, philosophische und logisch-philosophische Arbeiten. Von Kants Antinomien der reinen Vernunft bis zur logischen Terminitheorie, von Modalitäten bis zur logischen Folgebeziehung, von Entwicklungstermini bis zu intensionalen Kontexten reicht das Themenspektrum.

Antiirrationalismus ist der einzige -ismus, dem Wessel zustimmen kann.

Horst Wessel, Klaus Wuttich

daß-Termini

Intensionaliät und Ersetzbarkeit

ISBN: 978-3-89722-754-5 Preis: 34,- €

Von vielen Autoren werden solche Kontexte als intensional angesehen, in denen die üblichen Ersetzbarkeitsregeln der Logik nicht gelten. Eine besondere Rolle spielen dabei *daß*-Konstruktionen.

Im vorliegenden Buch wird gezeigt, daß diese Auffassungen fehlerhaft sind. Nach einer kritischen Sichtung der Arbeiten anderer Logiker zu der Problematik von *daß*-Termini wird ein logischer Apparat bereitgestellt, der es ermöglicht, *daß*-Konstruktionen ohne Einschränkungen von Ersetzbarkeitsregeln und ohne Zuflucht zu Intensionalitäten logisch korrekt zu behandeln.

Fabian Neuhaus

Naive Prädikatenlogik

Eine logische Theorie der Prädikation

ISBN: 978-3-8325-0556-1 Preis: 41,- €

Die logischen Regeln, die unseren naiven Redeweisen über Eigenschaften zugrunde liegen, scheinen evident und sind für sich alleine betrachtet völlig harmlos - zusammen sind sie jedoch widersprüchlich. Das entstehende Paradox, das Russell-Paradox, löste die sogenannte Grundlagenkrise der Mathematik zu Beginn des 20. Jahrhunderts aus. Der klassische Weg, mit dem Russell-Paradox umzugehen, ist eine Vermeidungsstrategie: Die logische Analysesprache wird so beschränkt, daß das Russell-Paradox nicht formulierbar ist.

In der vorliegenden Arbeit wird ein anderer Weg aufgezeigt, wie man das Russell-Paradox und das verwandte Grelling-Paradox lösen kann. Dazu werden die relevanten linguistischen Daten anhand von Beispielen analysiert und ein angemessenes formales System aufgebaut, die Naive Prädikatenlogik.

Bente Christiansen, Uwe Scheffler (Hrsg.)

Was folgt

Themen zu Wessel

ISBN: 978-3-8325-0500-4 Preis: 42,- €

Die vorliegenden Arbeiten sind Beiträge zu aktuellen philosophischen Diskussionen – zu Themen wie Existenz und Referenz, Paradoxien, Prädikation und dem Funktionieren von Sprache überhaupt. Gemeinsam ist ihnen der Bezug auf das philosophische Denken Horst Wessels, ein Vierteljahrhundert Logikprofessor an der Humboldt-Universität zu Berlin, und der Anspruch, mit formalen Mitteln nachvollziehbare Ergebnisse zu erzielen.

Vincent Hendricks, Fabian Neuhaus, Stig Andur Pedersen, Uwe Scheffler, Heinrich Wansing (Eds.)

First-Order Logic Revisited

ISBN: 978-3-8325-0475-5 Preis: 75,- €

Die vorliegenden Beiträge sind für die Tagung „75 Jahre Prädikatenlogik erster Stufe“ im Herbst 2003 in Berlin geschrieben worden. Mit der Tagung wurde der 75. Jahrestag des Erscheinens von Hilberts und Ackermanns wegweisendem Werk „Grundzüge der theoretischen Logik“ begangen.

Im Ergebnis entstand ein Band, der eine Reflexion über die klassische Logik, eine Diskussion ihrer Grundlagen und Geschichte, ihrer vielfältigen Anwendungen, Erweiterungen und Alternativen enthält.

Der Band enthalt Beiträge von Andréka, Avron, Ben-Yami, Brünnler, Englebretsen, Ewald, Guglielmi, Hajek, Hintikka, Hodges, Kracht, Lanzet, Madarasz, Nemeti, Odintsov, Robinson, Rossberg, Thielscher, Toke, Wansing, Willard, Wolenski

Pavel Materna

Conceptual Systems

ISBN: 978-3-8325-0636-0 Preis: 34,- €

We all frequently use the word “concept”. Yet do we know what we mean using this word in sundry contexts? Can we say, for example, that there can be several concepts of an object? Or: can we state that some concepts develop? What relation connects concepts with expressions of a natural language? What is the meaning of an expression? Is Quine’s ‘stimulus meaning’ the only possibility of defining meaning? The author of the present publication (and of “Concepts and Objects”, 1998) offers some answers to these (and many other) questions from the viewpoint of transparent intensional logic founded by the late Czech logician Pavel Tichý (†1994 Dunedin).

Johannes Emrich

Die Logik des Unendlichen

Rechtfertigungsversuche des *tertium non datur* in der Theorie des mathematischen Kontinuums

ISBN: 978-3-8325-0747-3 Preis: 39,- €

Im Grundlagenstreit der Mathematik geht es um die Frage, ob gewisse in der modernen Mathematik gängige Beweismethoden zulässig sind oder nicht. Der Verlauf der Debatte – von den 1920er Jahren bis heute – zeigt, dass die Argumente auf verschiedenen Ebenen gelagert sind: die der meist konstruktivistisch eingestellten Kritiker sind erkenntnistheoretischer oder logischer Natur, die der Verteidiger ontologisch oder pragmatisch. Die Einschätzung liegt nahe, der Streit sei gar nicht beizulegen, es handele sich um grundlegend unterschiedliche Auffassungen von Mathematik. Angesichts der immer wieder auftretenden Erfahrung ihrer Unverträglichkeit wäre es aber praktisch wie philosophisch unbefriedigend, schlicht zur Toleranz aufzurufen. Streiten heißt nach Einigung streben. In der Philosophie manifestiert sich dieses Streben in der Überzeugung einer objektiven Einheit oder Einheitlichkeit, insbesondere geistiger Sphären. Im Sinne dieser Überzeugung unternimmt die vorliegende Arbeit einen Vermittlungsversuch, der sich auf den logischen Kern der Debatte konzentriert.

Christopher von Bülow

Beweisbarkeitslogik

– Gödel, Rosser, Solovay –

ISBN: 978-3-8325-1295-8 Preis: 29,- €

Kurt Gödel erschütterte 1931 die mathematische Welt mit seinem Unvollständigkeitssatz. Gödel zeigte, wie für jedes noch so starke formale System der Arithmetik ein Satz konstruiert werden kann, der besagt: „Ich bin nicht beweisbar.“ Würde das System diesen Satz beweisen, so würde es sich damit selbst Lügen strafen. Also ist dies ein wahrer Satz, den es nicht beweisen kann: Es ist unvollständig. John Barkley Rosser verstärkte später Gödels Ergebnisse, wobei er die Reihenfolge miteinbezog, in der Sätze bewiesen werden, gegeben irgendeine Auffassung von „Beweis“. In der Beweisbarkeitslogik werden die formalen Eigenschaften der Begriffe „beweisbar“ und „wird früher bewiesen als“ mit modallogischen Mitteln untersucht: Man liest den notwendig - Operator als beweisbar und gibt formale Systeme an, die die Modallogik der Beweisbarkeit erfassen.

Diese Arbeit richtet sich sowohl an Logik-Experten wie an durchschnittlich vorgebildete Leser. Ihr Ziel ist es, in die Beweisbarkeitslogik einzuführen und deren wesentliche Resultate, insbesondere die Solovayschen Vollständigkeitssätze, präzise, aber leicht zugänglich zu präsentieren.

Niko Strobach

Alternativen in der Raumzeit

Eine Studie zur philosophischen Anwendung multidimensionaler Aussagenlogiken

ISBN: 978-3-8325-1400-6 Preis: 46.50 €

Ist der Indeterminismus mit der Relativitätstheorie und ihrer Konzeption der Gegenwart vereinbar? Diese Frage lässt sich beantworten, indem man die für das alte Problem der futura contingentia entwickelten Ansätze auf Aussagen über das Raumartige überträgt. Die dazu hier Schritt für Schritt aufgebaute relativistische indeterministische Raumzeitlogik ist eine erste philosophische Anwendung der multidimensionalen Modallogiken.

Neben den üblichen Zeitoperatoren kommen dabei die Operatoren „überall“ und „irgendwo“ sowie „für jedes Bezugssystem“ und „für manches Bezugssystem“ zum Einsatz. Der aus der kombinierten Zeit- und Modallogik bekannte Operator für die historische Notwendigkeit wird in drei verschiedene Operatoren („wissbar“, „feststehend“, „beeinflussbar“) ausdifferenziert. Sie unterscheiden sich bezüglich des Gebiets, in dem mögliche Raumzeiten inhaltlich koinzidieren müssen, um als Alternativen zueinander gelten zu können. Die Interaktion zwischen den verschiedenen Operatoren wird umfassend untersucht.

Die Ergebnisse erlauben es erstmals, die Standpunkt-gebundene Notwendigkeit konsequent auf Raumzeitpunkte zu relativieren. Dies lässt auf einen metaphysisch bedeutsamen Unterschied zwischen deiktischer und narrativer Determiniertheit aufmerksam werden. Dieses Buch ergänzt das viel diskutierte Paradigma der verzweigten Raumzeit („branching spacetime“) um eine neue These: Der Raum ist eine Erzählform der Entscheidungen der Natur.

Erich Herrmann Rast

Reference and Indexicality

ISBN: 978-3-8325-1724-3 Preis: 43.00 €

Reference and indexicality are two central topics in the Philosophy of Language that are closely tied together. In the first part of this book, a description theory of reference is developed and contrasted with the prevailing direct reference view with the goal of laying out their advantages and disadvantages. The author defends his version of indirect reference against well-known objections raised by Kripke in Naming and Necessity and his successors, and also addresses linguistic aspects like compositionality. In the second part, a detailed survey on indexical expressions is given based on a variety of typological data. Topics addressed are, among others: Kaplan's logic of demonstratives, conversational versus utterance context, context-shifting indexicals, the deictic center, token-reflexivity, vagueness of spatial and temporal indexicals, reference rules, and the epistemic and cognitive role of indexicals. From a descriptivist perspective on reference, various examples of simple and complex indexicals are analyzed in first-order predicate logic with reified contexts. A critical discussion of essential indexicality, de se readings of attitudes and accompanying puzzles rounds up the investigation.

Magdalena Roguska

Exklamation und Negation

ISBN: 978-3-8325-1917-9 Preis: 39.00 €

Im Deutschen, aber auch in vielen anderen Sprachen gibt es umstrittene Negationsausdrücke, die keine negierende Kraft haben, wenn sie in bestimmten Satztypen vorkommen. Für das Deutsche handelt sich u.a. um die exklamativ interpretierten Sätze vom Typ:

Was macht sie nicht alles! Was der nicht schafft!

Die Arbeit fokussiert sich auf solchen Exklamationen. Ihre wichtigsten Thesen lauten:

- Es gibt keine Exklamativsätze aber es gibt Exklamationen.
- *Alles* und *nicht alles* in solchen Sätzen, haben semantische und nicht pragmatische Funktionen.
- Das „nicht-negierende“ *nicht* ohne *alles* in einer Exklamation ist doch eine Negation. Die Exklamation bezieht sich aber trotzdem auf denselben Wert, wie die entsprechende Exklamation ohne Negation.
- In skalaren Exklamationen besteht der Unterschied zwischen Standard- und „nicht-negierenden“ Negation im Skopus von *nicht.*

Die Analyse erfolgt auf der Schnittstelle zwischen Semantik und Pragmatik.

August W. Sladek

Aus Sand bauen. Tropentheorie auf schmaler relationaler Basis

Ontologische, epistemologische, darstellungstechnische Möglichkeiten und Grenzen der Tropenanalyse

ISBN: 978-3-8325-2506-4 (4 Bände) Preis: 198.00 €

Warum braucht eine Tropentheorie zweieinhalbtausend Seiten Text, wenn zweieinhalb Seiten ausreichen, um ihre Grundidee vorzustellen? Weil der Verfasser zuerst sich und dann seine Leser, auf deren Geduld er baut, überzeugen will, dass die ontologische Grundidee von Tropen als den Bausteinen der Welt wirklich trägt und sich mit ihnen die Gegenstände nachbilden lassen, die der eine oder andere glaubt haben zu müssen. Um metaphysischen, epistemologischen Dilemmata zu entgehen, sie wenigstens einigermaßen zu meistern, preisen viele Philosophen Tropen als „Patentbausteine“ an. Die vorliegende Arbeit will Tropen weniger empfehlen als zeigen, wie sie sich anwenden lassen. Dies ist weit mühseliger als sich mit Andeutungen zu begnügen, wie brauchbar sich doch Tropen erweisen werden, machte man sich die Mühe sie einzusetzen. Lohnt sich die Mühe wirklich? Der Verfasser wollte zunächst nachweisen, dass sie sich nicht lohnt. Das Gegenteil ist ihm gelungen. Zwar sind Tropen wie Sandkörner. Was lässt sich schon aus Sand bauen, das Bestand hat? Wenn man nur genug „Zement“ nimmt, gelingen gewiss stabile Bauten, doch wie viel und welcher „Zement“ ist erlaubt? Nur schwache Bindemittel dürfen es sein; sonst gibt man sich mit einer hybriden Tropenontologie zufrieden, die Bausteine aus fremden, konkurrierenden Ontologien hinzunimmt. Die vier Bände bieten eine schwächstmögliche und damit unvermischte, allerdings mit Varianten und Alternativen behaftete Tropentheorie an samt ihren Wegen, Nebenwegen, Anwendungstests.

Mireille Staschok

Existenz und die Folgen

Logische Konzeptionen von Quantifikation und Prädikation

ISBN: 978-3-8325-2191-2 Preis: 39.00 €

Existenz hat einen eigenwilligen Sonderstatus in der Philosophie und der modernen Logik. Dieser Sonderstatus erscheint in der klassischen Prädikatenlogik – übereinstimmend mit Kants Diktum, dass Existenz kein Prädikat sei – darin, dass „Existenz“ nicht als Prädikat erster Stufe, sondern als Quantor behandelt wird. In der natürlichen Sprache wird „existieren“ dagegen prädikativ verwendet.

Diese andauernde und philosophisch fruchtbare Diskrepanz von Existenz bietet einen guten Zugang, um die Funktionsweisen von Prädikation und Quantifikation zu beleuchten. Ausgangspunkt der Untersuchungen und Bezugssystem aller Vergleiche ist die klassische Prädikatenlogik erster Stufe. Als Alternativen zur klassischen Prädikatenlogik werden logische Systeme, die sich an den Ansichten Meinongs orientieren, logische Systeme, die in der Tradition der aristotelischen Termlogik stehen und eine nichttraditionelle Prädikationstheorie untersucht.

Sebastian Bab, Klaus Robering (Eds.)

Judgements and Propositions

Logical, Linguistic, and Cognitive Issues

ISBN: 978-3-8325-2370-1 Preis: 39.00 €

Frege and Russell in their logico-semantic theories distinguished between a proposition, the judgement that it is true, and the assertion of this judgement. Their distinction, however, fell into oblivion in the course of later developments and was replaced by the formalistic notion of an expression derivable by means of pureley syntactical rules of inference. Recently, however, Frege and Russell's original distinction has received renewed interest due to the work of logicians and philosophers such as, for example, Michael Dummett, Per Martin-Lf, and Dag Prawitz, who have pointed to the central importance of o both the act of assertion and its justification to logic itself as well as to an adequate theory of meaning and understanding.

The contributions to the present volume deal with central issues raised by these authors and their classical predecessors: What kind of propositions are there and how do they relate to truth? How are propositions grasped by human subjects? And how do these subjects judge those propositions according to various dimensions (such as that of truth and falsehood)? How are those judgements encoded into natural language, communicated to other subjects, and decoded by them? What does it mean to procede by inference from premiss assertions to a new judgement?

Marius Thomann

Die Logik des Könnens

ISBN: 978-3-8325-2672-6 Preis: 41.50 €

Was bedeutet es, einer Person eine praktische Fähigkeit zu attestieren? Und unter welchen Umständen sind derartige Fähigkeitszuschreibungen wahr, etwa die Behauptung, Max könne Gitarre spielen? Diese Fragen stehen im Zentrum der vorliegenden Untersuchung. Ihr Gegenstand ist die philosophisch-logische Analyse des Fähigkeitsbegriffs. Als Leitfaden dient eine Analyse normalsprachlicher Fähigkeitszuschreibungen, gemäß der Max genau dann Gitarre spielen kann, wenn er dies unter dafür angemessenen Bedingungen normalerweise erfolgreich tut. Drei in der Forschungsliteratur vorgeschlagene Systeme werden diskutiert, die zwar wertvolle Impulse für die formale Modellierung geben, als Vertreter des so genannten modalen Ansatzes aber von der Diagnose ontologischer Inadäquatheit betroffen sind: Die Entitäten, die als Fähigkeiten attribuiert werden, lassen sich nicht über Propositionen individuieren; ohne die explizite Referenz auf Handlungstypen, die eben gekonnt oder nicht gekonnt werden, bleibt Max' Fähigkeit, Gitarre zu spielen, unterbestimmt. Um diesen Einwand zu vermeiden, liegt demgemäß der hier vorgestellten Logik des Könnens ein Gegenstandsbereich zugrunde, dessen Struktur an der Ontologie von Handlungen orientiert ist.

Christof Dobieß

Kausale Relata

Eine Untersuchung zur Wechselbeziehung zwischen der Beschaffenheit kausaler Relata und der Natur der Kausalbeziehung

ISBN: 978-3-8325-5083-7 Preis: 57.00 €

Dieses Buch macht nachdrücklich klar, daß die Thematik „Kausale Relata“ kein Nebenschauplatz der Kausalitätsdiskussion ist und sich die Analyse von Kausalität nicht auf die bloße Betrachtung der Kausalrelation selbst beschränken darf. Zwischen der Metaphysik der kausalen Relata und der Natur der Kausalbeziehung, so die Hauptthese dieses Werks, besteht eine enge theoretische Wechselbeziehung.

Untersucht wird diese These anhand zentraler kausaler Problembereiche: (1) der kausalen Präemption, (2) der Transitivität der Kausalität, (3) der dispositionalen Verursachung, (4) der negativen Verursachung und (5) der Konzeption von Verursachung als „qualitativem Fortbestand“ („qualitative persistence“).

Während die Probleme der Präemption und des qualitativen Fortbestands in der Auseinandersetzung zwischen kontrafaktischen Kausalkonzeptionen und Transfertheorien Bedeutung entfalten, betreffen die Transitivität der Kausalität sowie negative und dispositionale Verursachung nahezu alle Kausaltheorien. Der Forderung nach der Transitivität der Kausalität kann nur durch eine hinreichend präzise und eindeutig gefaßte Konzeption der kausalen Beziehungsträger entsprochen werden. Ob Dispositionen oder Negativereignisse in kausale Beziehungen treten können, hängt entscheidend davon ab, inwiefern Entitäten dieser Art ein ontologisches Bleiberecht zugestanden wird.

Logische Philosophie

Eds.: H. Wessel, U. Scheffler, Y. Shramko and M. Urchs

The series "Logische Philosophie" introduces philosophically relevant results of logical research. In particular, treatises are issued in which logical means are employed to solve philosophical problems.

The volume is the proceedings from the conference *FOL75 – 75 Years of First-Order Logic* held at Humboldt University, Berlin, Germany, September 18 - 21, 2003 on the occasion of the anniversary of the publication of Hilbert's and Ackermann's *Grundzüge der theoretischen Logik*. The papers provide analyses of the historical conditions of shaping of FOL, they discuss several modern rivals to FOL and show the importance of FOL for interdisciplinary research. Although the celebrated book marks a most important step in the development of logic, the volume in hand proves the actuality of the question "Which logic is the right logic."

Logos Verlag Berlin

ISBN 3–8325–0475–3