Proceedings of the
13th Panhellenic Logic Symposium

Vol.II (2022)

July 6-10, 2022

University of Thessaly, Volos, Greece

FOUNDATION COMPOSITIO MATHEMATICA

University of Cyprus

Association for Symbolic Logic

EUROPEAN MATHEMATICAL SOCIETY

Preface

The 13th Panhellenic Logic Symposium is taking place in Volos, Greece, July 6-10, 2022.

This the second and main part of the proceedings, the first having been published online in July 2021 for a meeting that did not happen due to the COVID epidemic. Here you will find the plenary talks, tutorials, special sessions and the contributed papers of the event physically taking place in Volos.

We would like to emphasize the character of the PLS meeting as an international symposium with local characteristics. There are no parallel sessions, and this year's program is not particularly packed, allowing for increased interaction, discussion and understanding between researchers working different facets of logic. If everything went according to plan, you will probably not find many talks in your particular area of expertise. Take this opportunity to inquire, learn and discuss the concerns of researchers in different facets of logic, perhaps even allowing for some inspiration or ideas in your future work. This applies particularly with regard to our Computer Science special session, which includes topics and researchers that are not traditionally found in logic meetings.

Speakers have been encouraged to give accessible interactive presentations, so the remaining ingredient for a successful meeting is an inquisitive, dynamic and socially interactive open-minded audience.

We thank the committees of the PLS13 and the referees for their work. Particular thanks go to the:

- Local organization by K. Hatzikyriakou and V. Papayiannakopoulou
- Web support by N. Papaspyrou
- Hands-on coordination by A. Kakas, P. Eleftheriou, T. Pheidas, Y. Stephanou

and all the participants for making it to Volos, despite the uncertain and often difficult traveling conditions these days. We also thank our sponsors for making this meeting possible.

It's good to be back together again.

Giorgos Barmpalias & Kostas Tsaprounis
Program Chairs, PLS13
Chinese Academy of Sciences, Beijing
University of Aegean, Greece
July 6, 2022, Volos, Greece.

THE 13th PANHELLENIC LOGIC SYMPOSIUM

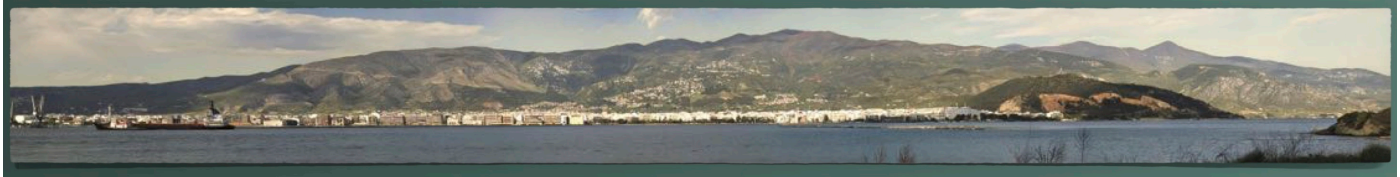
July 06-10, 2022, Volos, Greece

E-MAIL: pls13@softlab.ntua.gr

Organized by the University of Thessaly

WEB: http://panhellenic-logic-symposium.org

* All health protocols in effect will be strictly followed.



INVITED TALKS

Andrew Brooke-Taylor (University of Leeds, UK)

Vassileios Koutavas (Trinity College Dublin, Ireland)

Thanases Pheidas (University of Crete, Greece)

TUTORIALS

Alex Kavvos (University of Bristol, UK)

Nikos Leonardos (University of Athens, Greece)

Stathis Zachos (National Tech. Uni. Athens, Greece)

ORGANIZING COMMITTEE

Kostas Hatzikiriakou, University of Thessaly (chair)

Nikolaos Papaspyrou, National Technical University of Athens

Vasiliki Papayiannakopoulou, University of Thessaly

SCIENTIFIC COMMITTEE

Antonis Achilleos, Reykjavik University

George Barmpalias, Chinese Academy of Sciences (co-chair)

Costas Dimitracopoulos, University of Athens

Pantelis Eleftheriou, University of Leeds

Vassilis Gregoriades, National Technical University of Athens

Kostas Hatzikiriakou, University of Thessaly

Antonis Kakas, University of Cyprus

Alex Kavvos, University of Bristol

Nikolaos Papaspyrou, National Technical University of Athens

Thanases Pheidas, University of Crete

Rizos Sklinos, Chinese Academy of Sciences

Ana Sokolova, University of Salzburg

Alexandra Soskova, Sofia University

Mariya Soskova, University of Wisconsin–Madison

Yannis Stephanou, University of Athens

Konstantinos Tsaprounis, University of the Aegean (co-chair)

Nikos Tzevelekos, Queen Mary University of London

Niki Vazou, IMDEA Institute

Stathis Zachos, National Technical University of Athens

SPECIAL SESSIONS

Computer Science

Bruno Bauwens (HSE University, Russia)

Juan Garay (Texas A&M University, USA) (TBC)

Andrew Lewis-Pye (London School of Economics, UK)

Vassilis Zikas (Purdue Uni. USA & Uni. Edinburgh, UK)

Philosophical Logic

Michael Glanzberg (Rutgers University, USA)

Volker Halbach (University of Oxford, UK)

Elia Zardini (Uni. Lisbon, Portugal & HSE Uni., Russia)

EUROPEAN MATHEMATICAL SOCIETY

Association for Symbolic Logic

FOUNDATION COMPOSITIO MATHEMATICA

University of Cyprus

List of Participants - PLS13

1. Achilleos Antonis- Reykjavík University
2. Anastasiadi Elli - Reykjavík University
3. Baanders Pauline - University of Amsterdam
4. Barmpalias George - Chinese Academy of Science
5. Bauwens Bruno - National Research University & Higher School of Economics
6. Brooke-Taylor Andrew - University of Leeds
7. Chobitaki Dimitra - University of Crete
8. Dimitrakopoulos Costas - University of Athens
9. Eleftheriadis Ioannis - University of Cambridge
10. Eleftheriou Pantelis - University of Leeds
11. Galatos Nick - University of Denver
12. Garay Juan - Texas A&M University
13. Georgiev Ivan - Sofia University
14. Glanzberg Michael - Rutgers University
15. Gregoriades Vassilios - National Technical University of Athens
16. Gronwald Jan - University of Amsterdam
17. Hatzikiriakou Kostas - University of Thessaly
18. Iatrou Evangelos - University of Amsterdam
19. Halbach Volker - University of Oxford
20. Kakas Antonis - University of Cyprus
21. Kartas Konstantinos - University of Oxford
22. Kavvos Alex - University of Bristol
23. Kolev Grigor - Sofia University

24. Konstantakatos Vangelis - National Technical University of Athens

25. Koutavas Vassileios - Trinity College, Dublin

26. Landes Juergen - LMU Munich/Milano

27. Leonardos Nikos - University of Athens

28. Lewis-Pye Andrew - London School of Economics

29. Marangelis Georgios - Aristotle University of Thessaloniki

30. Papadopoulos Aristomenis-Dionysios - University of Leeds

31. Papakonstantinou Dimitrios

32. Papayannopoulos Philippos - Université Paris 1 Panthéon-Sorbonne & CNRS

33. Pawlowski Pawel - Ghent University

34. Petsi Katerina - University of Athens

35. Petrakis Iosif - Ludwig-Maximilians-Universitaet

36. Pheidas Thanases - University of Crete

37. Picenni Simon - University of Bristol

38. Rivello Edoardo - University of Toronto

39. Stepanov Vladimir - CC FIC CSC RAS, Moscow

40. Stephanou Yiannis - University of Athens

41. Titov Ivan - Ruprecht-Karls-Universität

42. Tsakalofas N. Alexandros - University of Athens

43. Tsaprounis Konstantinos - University of the Aegean

44. Zachos Stathis - National Technical University of Athens

45. Zardini Elia - University of Lisbon

46. Zikas Vassilis - Purdue University & University of Edinburgh

Timetable - 13th Panhellenic Logic Symposium

	06.07 WED	07.07 THU	08.07 FRI	09.07 SAT	10.07 SUN
9:00-9:40	Registration	**Plenary** Brooke-Taylor	**Plenary** Koutavas	**Tutorial** Zachos (3)	**Plenary** Pheidas
9:40-10:00					
10:00-10:30		**Contributed** Rivello	**Contributed** Stepanov	**CS Session** Lewis-Pye	**CS Session** Bauwens
10:30-11:00	**Coffee Break**	**Contributed** Papayannopoulos	**Contributed** Iatrou		
11:00-11:30	Opening	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Syskepsis:** Logic MSc (45mins) & Phil. Assoc. (45mins)
11:30-12:30	**Tutorial** Zachos (1)	**Tutorial** Leonardos (1)	**Tutorial** Zachos (2)	**Tutorial** Zachos (3)	
12:30-13:00	**Phil Session** Michael Glanzberg	**Phil Session** Volker Halbach	**Phil Session** Elia Zardini	**Contributed** Pawlowski	**Bus returning to Athens Airport**
13:00-13:30				**Contributed** Petsi	
13:30-14:00	**Contributed** Kartas	**Contributed** Kakas	**Tutorial** Leonardos (2)	**Tutorial** Leonardos (3)	
14:00-17:00	**Break**	**Break**	**Break**	**Break**	
17:00-17:30	**Tutorial** Kavvos (1)	**Tutorial** Kavvos (2)	**Tutorial** Kavvos (3)	**Contributed** Anastasiadi	
17:30-18:00				**Contributed** Picenni	
18:00-18:30	**Coffee Break**	**Coffee Break**	**Coffee Break**	**Coffee Break**	
18:30-19:00	**Contributed** Landes	**CS Session** Garay	**CS Session** Zikas	**Contributed** Papadopoulos	
19:00- 19:30	**Contributed** Chompitaki			**Contributed** Petrakis	
19:30-20:00		**Contributed** Galatos	**Contributed** Titov		
20:00-21:00		**Dinner**	**Syskepsis:** Future of PLS		
21:00-22:00					

Plenary

Speaker: Andrew Brooke-Taylor (University of Leeds, UK)

Title: Categorifying Borel reducibility

Abstract: The framework of Borel reducibility has had great success in showing that various classification programmes cannot be completed, in areas such as ergodic theory and C*-algebras. However, this framework ignores a crucial feature of many classification programmes: namely, that they are expected to be functorial. I will talk about joint work in progress with Filippo Calderoni, introducing a categorified version of Borel reducibility, and noting some differences in consequences from the original framework.

Speaker: Vasileios Koutavas (Trinity College Dublin, Ireland)

Title: Program Equivalence from Theory to Practice

Abstract: Equivalence is a key concept in the area of programming language semantics as it can concisely express concepts such as modularity and abstraction and expose all the intentional and unintentional ways that different parts of a program can affect each other. There is a long history of theoretical techniques for proving equivalence in programming language semantics and related areas. More recently it has found applications in the verification of cryptographic protocols, compiler correctness and regression verification. In this talk we will overview the history of equivalence in programming language semantics, primarily focusing on the techniques of game semantics and bisimulation. We will show a new approach which combines these two theoretical techniques and leads to the development of a practical verification tool of equivalence in a programming language with functions and local state. This is a challenging setting, first studied in the semantics of ALGOL, which is quite common in modern programming languages. In this setting, verifying equivalence of even simple program expressions requires the exploration of infinite behaviours. Being able to reason within a verification tool about many cases of infinite behaviour, while still avoiding reporting any false-positives or false-negatives, has been the main success of this approach.

Speaker: Thanassis Pheidas (University of Crete, Greece)

Title: Definability in Number Theory, Algebra and Geometry and Hilbert's Tenth Problem

Abstract: We survey developments on the questions of decidability (and related definability) of the existential (and first-order) theories of rings of common use, eg. the rational integers and rational numbers and rings of polynomials and rational functions. The starting point of our questions is "Hilbert's Tenth Problem".

Tutorials

Speaker: Stathis Zachos (National Technical University of Athens, Greece)

Title: Introduction to Computational Complexity

Abstract: Complexity classes, Randomness, Interactivity, PCP, Counting. How to deal with NP-completeness, Approximation algorithms, Search Complexity, Parameterized Complexity, Quantum Complexity, Counting revisited (#P, #PE, TotP),Nonuniform Circuit Families, Descriptive Complexity.

Speaker: Alex Kavvos (University of Bristol, UK)

Title: Type Theory and Homotopy

Abstract: This is a whirlwind tour of Martin-Loef's Type Theory. We explain some of the basic ideas underlying type-theoretic approaches to the foundations of mathematics, introduce the rules of MLTT, and present a few examples. We also provide a sketch of the recently discovered connections to homotopy theory, the relationship with computer-assisted formalised mathematics, and point the audience to recent advances in the field.

Speaker: Nikos Leonardos (University of Athens, Greece)

Title: Bitcoin's backbone algorithm and the consensus problem

Abstract: The purpose of the tutorial is to study Bitcoin's protocol as a distributed algorithm. We will begin by discussing the core of the protocol in a simple model that is static and synchronous. We will prove two of its fundamental properties which we call common prefix and chain quality. Subsequently, we will discuss the Consensus problem. We will show that Bitcoin solves an interesting variant of this fundamental problem. Finally, we discuss attributes of the Bitcoin protocol that allow it to be secure in more realistic models (partially synchronous model, variable difficulty, timestamp rules).

Philosophical Logic Special Session

Speaker: Michael Glanzberg (joint work with Lorenzo Rossi) (Rutgers University, USA)

Title: Truth and Quantification

Abstract: Theories of self-applicable truth have been motivated in two main ways. First, if truth-conditions provide the meaning of (many kinds of) natural language expressions, then self-applicable truth is instrumental to develop the semantics of natural languages. Second, a self-applicable truth predicate is required to express generalizations that would not be otherwise expressible in natural languages. In order to fulfill its semantic and expressive role, we argue, the truth predicate has to be studied in its interaction with constructs that are actually found in natural languages and extend beyond first-order logic---modals, indicative conditionals, arbitrary quantifiers, and more. Here, we focus on truth and quantification. We develop a Kripkean theory of self-applicable truth for the language of Generalized Quantifier Theory. More precisely, we show how to interpret a self-applicable truth predicate for the full class of type <1,1> (and type <1>) quantifiers to be found in natural languages. As a result, we can model sentences which are not expressible in theories of truth for first-order languages (such as `Most of what Jane's said is true', or `infinitely many theorems of T are untrue', and several others), thus expanding the scope of existing approaches to truth, both as a semantic and as an expressive device.

Speaker: Volker Halbach (University of Oxford, UK)

Title: Axiomatic Theories of Truth: A survey

Abstract: Axiomatic theories of truth are obtained by adding a unary predicate T to the language of arithmetic or another language in which syntax theory is usually developed. Then axioms for T are conjoined with the axioms of arithmetic (or another syntax theory). The liar and other paradoxes impose limits on which truth axioms can consistently be added. I survey some systems that have been seen as promising and interesting. There are several motivations for investigating axiomatic theories of truth. They are of philosophical interest not only because they give us a better understanding of the concept of truth, but also because, for instance, they can be used to reduce away commitment to second-order objects and express generalizations over sentences. They are also of purely mathematical interest. First, proof theorists have analyzed their strength and used them as intermediary systems for the analysis of subsystem of second-order arithmetic, but also set-theoretic systems. Some axiomatic theories of truth have also intriguing properties studied by model theorists.

Philosophical Logic Special Session

Speaker: Elia Zardini (University of Lisbon, Portugal & HSE University, Russia)

Title: The Final Cut

Abstract: In a series of works, P. Cobreros, P. Égré, D. Ripley and R. van Rooij have proposed a nontransitive system (call it 'K3LP') as a basis for a solution to the semantic paradoxes. I critically consider that proposal at three levels. At the level of the background logic, I present a conception of classical logic on which K3LP fails to vindicate classical logic not only in terms of structural principles, but also in terms of operational ones. At the level of the theory of truth, I raise a cluster of philosophical difficulties for a K3LP-based system of naive truth, all variously related to the fact that such an extension proves things that would seem already by themselves jointly repugnant, even in the absence of transitivity. At the level of the theory of validity, I consider an extension of the K3LP-based system of naive validity that is supposed to certify that validity in that system does not fall short of naive validity, argue that such an extension is untenable in that its nontriviality depends on the inadmissibility of a certain irresistible instance of transitivity (whence the advertised "final cut") and conclude on this basis that the K3LP-based system of naive validity cannot coherently be accepted either. At all these levels, a crucial role is played by certain metaentailments and by the extra strength they afford over the corresponding entailments: on the one hand, such strength derives from considerations that would seem just as compelling in a general nontransitive framework, but, on the other hand, such strength wreaks havoc in the particular setting of K3LP.

Keywords: classical logic; naive truth; naive validity; nontransitive logics.

CS Special Session

Speaker: Bruno Bauwens (HSE University, Russia)

Title: The algorithmic information distance

Abstract: The algorithmic information distance is a metric on bit-strings. In the machine-learning literature, it is defined as for bit-strings x and y as d(x,y) = max{K(x|y), K(y|x)} + c for some large c, where K(x|y) denotes the minimal length of a program that on input y produces x on an optimal prefix-free machine. In some sense, this measure exploits all types of computable structure that can be found between 2 bit-strings. This optimality property can be formalized in a similar way as Kolmogorov complexity is optimal for measures of information content. The measure has inspired a few interesting machine-learning algorithms and we will briefly discuss some recent ones. The definition given above satisfies the triangle inequality, but it is rather technical. Therefore, the following characterization is usually mentioned: d(x,y) = min {|p| : U(p,x) = y and U(p, y) = x} + O(log |xy|). In fact, the minimum above was the historically first definition. Note that in this minimum, one considers programs p that are {\em bidirectional}: the program should simultaneously map x to y and map y to x. On the other hand, in the definition max{K(x|y), K(y|x)}, the conditional complexities only consider 1-directional programs that either map y to x for K(x|y) or x to y for K(y|x). Can we improve the precision from O(\log |xy|) to O(1)? It is known that for plain complexity this is indeed true. But the triangle inequality does not hold for plain complexity. For a long time, it was an open question whether whether the O(1) precision holds for prefix-free machines. Recently, it has been claimed in several places that the answer is yes, but these proofs are wrong. Recently, I gave a counter example, for the equality with O(1) precision. This is remarkable, because many equalities that hold with O(log |xy|) precision for plain complexity can be transformed to inequalities that hold with O(1) for prefix-free complexity. A famous example is the Kolmogorov-Levin formula K(x,y) = K(x) + K(y | x, K(x)). For the first time, we have found an important (in)equality that holds with better precision for plain complexity. Even more mysteriously, the equality does hold with O(1) precision whenever d(x,y) > 6log |xy| and x and y are large, and the proof of this result is rather complex.

Speaker: Andy Lewis-Pye (London School of Economics, UK)

Title: Chained Fever — achieving optimal synchronisation for Hotstuff

Abstract: This will be a talk of two halves, to accommodate different backgrounds in the audience. First, I'll give an easy introduction to consensus protocols, focussing on the classical 'permissioned' approach rather than Bitcoin, so that there should not be too much overlap with Leonardos' tutorial. Then I'll describe an improvement on the state-of-the-art in consensus protocols, which is a modification of 'Hotstuff' requiring only O(n) messages per `view change', meaning O(n^2) message complexity to produce a confirmed block of transactions in the worst case. This improves on the previous best of O(n^3) message complexity for optimistically responsive protocols.

CS Special Session

Speaker: Juan Garay (Texas A&M University)

Title: Cryptography in the Blockchain Era

Abstract: The advent of blockchain protocols such as Bitcoin has ignited much excitement, not only for their realization of novel financial instruments, but also for offering alternative solutions to classical problems in fault-tolerant distributed computing and cryptographic protocols. Underlying many of such protocols is a primitive known as "proof of work" (PoW), which for over 20 years has been liberally applied in the cryptography and security literature to a variety of settings, including spam mitigation, sybil attacks, and denial of service protection; its role in the design of blockchain-based protocols, however, is arguably its most impactful application. At a high level, the way PoWs enable such protocols is by slowing message passing for all parties indiscriminately, thus generating opportunities for honest parties' views to converge, under the assumption that their aggregate computational power exceeds that of the adversary. This talk comprises two parts. First, despite the evolution of our understanding of the PoW primitive, pinning down the exact properties sufficient to prove the security of Bitcoin and related protocols has been elusive. In this talk we identify such properties, and then construct protocols whose security can be reduced to them in the standard model, assuming a common reference string (CRS -- cf. a "genesis" block). All previous protocols rely on the "random oracle" methodology.Second, regarding the realizability of two important problems in the area of cryptographic protocols -- Byzantine agreement (BA) and secure multiparty computation (MPC) -- we show how PoW-based blockchains allow to realize them even in the presence of a minority of corrupted parties (i.e., $t < n/2$, where t is the number of corrupted parties and n is their total number) , as long as the majority of the computation resources remain in honest hands, while "classically" (i.e., no PoWs), protocols can only tolerate up to $t < n/3$ corruptions in the absence of a private trusted setup, such as a public-key infrastructure. We resolve this apparent contradiction with a new paradigm we call "Resource-Restricted Cryptography."

The bulk of this talk is based on joint work with Marshall Ball, Aggelos Kiayias, Rafail Ostrovsky, Giorgos Panagiotakos and Vassilis Zikas.

Speaker: Vasileios Zikas (Purdue University, USA & University of Edinburgh, UK)

Title: From Blockchain to Global-scale Trustworthy Infrastructure

Abstract: The wide adoption of global computer networks, such as the Internet, creates immense opportunities, and challenges the traditional centralized trust model. The idea of giving control of widely-used critical infrastructure to its users is becoming ever more popular. Blockchain and Distributed Ledger Technology (DLT) promise to bring the decentralization ideas to reality and disrupt traditional strongholds of trust in the financial, digital, biomedical, and manufacturing sectors, as well as in governance. In this talk I will discuss blockchain – from its current structure to its vast potential for future applications. The talk will discuss novel design choices that go into blockchain-based DLT, and how these choices critically impact the security of the solutions and address implementation and deployment challenges. It will also tease the potential of using reputation-based blockchain to enhance trustworthiness of decentralized worldwide systems.

List of Mentors and Topics		
1	Thanasis Pheidas	Decidability and Undecidability of Diophantine Problems
2	Antonis Kakas	Artificial Inteligence
3	Iosif Petrakis	Constructive mathematics
4	Nick Galatos	Universal algebra and algebraic logic
5	Michael Glanzberg	Philosophical Logic
6	Nikos Leonardos	Crypto and computational complexity
7	Andrew Lewis-Pye	Computability, Network Science and Computer Science
8	Juan Garay	Cryptography, Security, Algorithms
9	Alex Kavvos	Logic in Computer Science
10	Costas Dimitrakopoulos	Logic, Fragments of Arithmetic
11	Stathis Zachos	Computational complexity
12	Bruno Bauwens	Algorithmic information and computational complexity
13	George Barmpalias	Algorithmic information, Computability, Dynamics of Networks
14	Andrew Brooke-Taylor	Set theory
15	Pantelis Eleftheriou	Model Theory
16	Yiannis Stephanou	Philosophical Logic

As well as every expert and willing participant that we didn't get the chance to ask.

In the remaining pages you will find:

- A survey article by one of our plenary speakers, Thanases Pheidas
- The contributed papers, ordered as in the table below.

	Contributed talks	
	Author(s)	Title
1	Konstantinos Kartas	Decidability of local fields and their extensions
2	Jürgen Landes	A Triple Uniqueness of the Maximum Entropy Approach
3	Manos Kamarianakis, Dimitra Chompitaki and Thanases Pheidas	Decidability of the theory of addition and the Frobenius map in fields of rational functions
4	Matteo de Ceglie	The V-logic Multiverse and MAXIMIZE
5	Ivan Titov, Wolfgang Merkle	Total variants of Solovay reducibility and speedability
6	Nick Galatos, Peter Jipsen	Generalized bunched implication logic
7	Simone Picenni	Exact truthmaking and self-reference: a model for self-applicable exact truthmaking
8	Ioannis Eleftheriadis and Aristomenis-Dionysios Papadopoulos	NSOP in Classes of Graphs
9	Iosif Petrakis	Proof-relevance in Bishop-style constructive mathematics
10	Fabian Pregel	Frege's Concept of Completeness
11	Vladimir Stepanov	Dynamic Approximation of Self-Referential Sentences
12	Pawel Pawlowski, Elio La Rosa	8-valued non-deterministic semantics for modal logics
13	Katerina Petsi	The Foundations of Arithmetic: Peano vs Dedekind
14	Philippos Papayannopoulos	Some Philosophical Remarks on the Current Definitions of Algorithms
15	Evan Iatrou	Non-monotonic rule-based logical programming for modelling legal reasoning the example of Answer Set Programming (ASP)
16	Antonis Kakas	Argumentation: Reasoning Universalis
17	Edoardo Rivello	The largest intrinsic disquotational definition of truth
18	Elli Anastasiadi, Antonis Achilleos, Adrian Francalanza, Jasmine Xuereb	Epistemic logic for verifying runtime verification communication protocols

Notes on the Theory of Numbers, applied to questions of definability

Thanases Pheidas

May, 2022

Contents

1 Introduction

This is a survey of some questions about the decidability of existential theories of certain commonly used domains in Mathematics, i.e. analogues of Hilbert's Tenth Problem (HTP).The subjects include analogues of HTP for the field of rational numbers, fields of rational functions of a variable z, such as $\mathbb{C}(z)$ and rings of analytic or meromorphic functions of one variable. All these are open problems. In this sense this is an effort for continuing the presentation of [43], partly complementing [22] - where the reader may find, among other things, a nice exposition of the definition of $\mathbb{Z}$ in $\mathbb{Q}$ by Julia Robinson and a comparison with the new improvements of [23]. A presentation of questions in Number Theory that I think that, if answered, will have very important consequences in the subjects we are discussing may be found in [35]. Surveys in closely related fields of research may be found in [3], [53] and [45].

In the last twenty years the subjects around analogues of HTP have attracted a relatively large number of researchers, coming both from a Logic as well as a Number Theory background. This has led to a wealth of new knowledge and results. My intention is to list problems that may interest young logicians and questions in Number Theory which might produce progress towards answering the logical problems. I will assume some familiarity with [43] where the reader may find coordinates of most of the relative literature up to 2000 - and I am including an updated (but certainly incomplete) list of questions and known answers in the last Section.

I have included some questions in Number Theory which are, in my view, interesting to logicians that work on this subject: a discussion of a derivation on the integers and rationals (Section 3.1) and a problem of Grothendieck and Katz for algebraic differential equations (Section 3.2). I am also discussing the question of detecting the varieties that have infinitely many points, rational over some number field, which may be found in [44].

I will leave out of my presentation several important developments, like analogues of HTP for number rings and global fields of positive characteristic - for these see [32], [53] and the references in [22].

Trying to address this article to both logicians and number-theorists I have tried to keep the terminology to a minimum - little more than S. Lang's *Algebra*, the beginning chapters of C. C. Chang and H. J. Keisler's *Model Theory* and elements of the theory of elliptic curves.

The last Section contains a short list of major open questions and bibliography on related areas.

I would like to thank Hector Pasten, Xavier Vidaux and two referees of the Proceedings of the Panhellenic Logic Symposium for contributing information and very useful suggestions during the preparation of this article.

2 Diophantine Equations

A diophantine equation is one of the form

$$f(x_1, \ldots, x_m) = 0$$

where f is a polynomial in the variables $x_1, \ldots, x_m$, with integer coefficients. One wants to find integer and rational solutions.

The history of diophantine equations spans the era from Pyhtagoras, Euclid and Diophantos of Classical Greece, to the Arabs, the Indians, Gauss and Euler and continuously throughout our times. What may be surprising is that many of even the most ancient problems remain unsolved. For a sampler see [57] and [28].

A closely related question is to describe the sets over R, where R is any of $\mathbb{Z}$ and $\mathbb{Q}$, of the form

$$\{\bar{x} \in R^n \mid \exists \bar{y} \in R^m \ f(\bar{x}, \bar{y}) = 0\}$$

where $f(\bar{x}, \bar{y}) = 0$ is a diophantine equation in the n-tuple of variables $\bar{x}$ and the m-tuple $\bar{y}$. These sets, geometrically, are projections (over $\mathbb{Z}$ or $\mathbb{Q}$) of algebraic sets over the ring under consideration. We will call them *diophantine sets* and we will denote the class of these sets by $DIO(\mathbb{Z})$ and $DIO(\mathbb{Q})$, accordingly. The question, *what are the sets in* $DIO(\mathbb{Z})$? has been answered over $\mathbb{Z}$ in a quite satisfactory but also very surprising way. The answer is

$$DIO(\mathbb{Z}) = RE$$

i.e. the class of diophantine sets over $\mathbb{Z}$ coincides with the class RE of *recursively enumerable sets* , which are all sets of tuples of integers which may

be listed - eventually- by some algorithm. And, because there are recursively enumerable sets which are not recursive - *recursive* are the sets for which there is an algorithm which tests membership in the set - we conclude that there is no algorithm which, with input any diophantine equation, replies whether or not the equation has solutions over $\mathbb{Z}$. This is a negative answer to Hilbert's tenth problem (HTP):

Entscheidung der Lösbarkeit einer diophantischen Gleichung. Eine diophantische Gleichung mit irgendwelchen Unbekannten und mit ganzen rationalen Zahlkoefficienten sei vorgelegt: man soll ein Verfahren angeben, nach welchen sich mittels einer endlichen Anzahl von Operationen entscheiden lässt, ob die Gleichung in ganzen rationalen Zahlen lösbar ist.

translated into English
to find a process according to which one can determine in a finite number of steps whether a polynomial equation with integer coefficients has or does not have integer solutions.
The problem was the 10th in a famous list of problems announced by Hilbert at the World Congress of Mathematicians which was held in Sorbonne in 1900. It was solved in the negative in 1970:

Theorem 1 *(Julia Robinson, Martin Davis, Hillary Putnam, Yuri Matijasevich - 1970)*
$DIO(\mathbb{Z}) = RE$, hence there is no such 'process'.

Presentations of this and relevant questions from various points of view may be found in [12], [27], [39], [47], [50] and [25].

The similar questions about $\mathbb{Q}$ are mostly open - but some facts are known. For example the set of squares $\{x \in \mathbb{Q} \mid \exists y \in \mathbb{Q}\ x = y^2\}$ is obviously diophantine over $\mathbb{Q}$. What about its complement, i.e. the set of non-squares? The answer is that this is also diophantine:

Theorem 2 *(Bjorn Poonen) The set of non-squares $\{x \in \mathbb{Q} \mid \forall y \in \mathbb{Q} x \neq y^2\}$ is diophantine over $\mathbb{Q}$.*

The proof uses deep knowledge from *arithmetic algebraic geometry*. It is proved that for certain surfaces, the *Brauer-Manin obstruction* is the only obstruction to the Hasse principle (we will not go into details). The same result is proved in [23] by elementary means and generalised for n-th powers in [6] and further in [16].

Given a language L, a structure (model) an *existential* (resp. *positive-existential*) formula of L is one of the form $\exists \bar{x} \phi$, where ϕ is a boolean combination (respectively positive boolean combination, i.e. with no negation symbols) of atomic formulas of L. Given a structure (model) $\mathcal{A}$ of L, the *existential theory* (resp. *positive existential theory*) of $\mathcal{A}$ is the set of existential (resp. positive existential) sentences of L which are true in $\mathcal{A}$. A subset of a power of the universe of $\mathcal{A}$ is *existential* (resp. *positive existential*) if it has an existential (resp. positive existential) definition.

In general we will consider rings R and a subring R_0. In this setting a *diophantine equation* will be a polynomial equation with coefficients in R_0. A *diophantine set over R with coefficients in R_0* will be a set of the form $\{x \in R^n \mid \exists y \in R^m \ f(x,y) = 0\}$, where $f \in R_0[\bar{x}, \bar{y}]$. For simplicity, unless we state otherwise, we will consider only the cases where R_0 is $\mathbb{Z}$ or $\mathbb{Z}[z]$ - the latter for rings of functions of the variable(s) z - and $\mathbb{F}_p[z]$ for rings of functions of positive characteristic p.

2.1 The set $\mathbb{Q} \setminus \mathbb{Z}$ is diophantine in $\mathbb{Q}$.

A very surprising result, in my opinion, of the last 15 years has been that of [23]:

Theorem 3 *(Jochen Koenigsmann- 2016) The set $\mathbb{Q} \setminus \mathbb{Z}$ is diophantine in $\mathbb{Q}$.*

The proof involves heavily the Algebra of Quaternia. The methods signify a very substantial improvement over the use of quadratic forms that Julia Robinson used in her Thesis in order to define (first-order, with many alterations of quantifiers) $\mathbb{Z}$ in $\mathbb{Q}$.

So, what about the possibility that $\mathbb{Z}$ is diophantine in $\mathbb{Q}$? This is an outstanding question and open problem. There is a conjecture by Barry Mazur, stating:

Conjecture 4 *(Barry Mazur) The real topological closure of an algebraic set over $\mathbb{Q}$ has only finitely many components.*

This, if true, would imply that the projection of an algebraic set over $\mathbb{Q}$ has only finitely many topological $\mathbb{R}$-components (finitely many components project onto finitely many components), hence $\mathbb{Z}$ could not be diophantine over $\mathbb{Q}$. There is also a p-adic version of the conjecture - cf [6].

But be aware: *The function-theoretic version of this conjecture is not true*. In particular the set $\{z^{p^n} \mid n \in \mathbb{N}\}$ is diophantine over $\mathbb{F}_p(z)$ with coefficients in the ring of polynomials $\mathbb{F}_p[z]$ - we will discuss this in Section 6.3 . Therefore, if proved correct, the Conjecture will give a concrete example in which $\mathbb{Q}$ and the $\mathbb{F}_p(z)$ have different behaviour.

In [23] it is proved that the same statement ($\mathbb{Z}$ is not diophantine over $\mathbb{Q}$) follows from some strong version of the *Bombieri-Lang Conjecture*, cf. [22].

2.2 Injective rational polynomials

In order to achieve a structure theory for diophantine sets over the rationals, one would like to have some sort of 'pairing' or 'Goedel Numbering', i.e., in its simplest form, an algebraic way to associate pairs of rational numbers to rational numbers. A natural question along these lines is:

Is there a polynomial f of two variables, over $\mathbb{Q}$, which, as a function, induces an injection from $\mathbb{Q} \times \mathbb{Q}$ into $\mathbb{Q}$? A bijection?

This remains open. It was asked by Harvey Friedman. Don Zagier has speculated that the polynomial

$$x^7 + 3y^7$$

does induce an injection. Bjorn Poonen has shown that the *Bombieri-Lang Conjecture* (an open conjecture in Arithmetic Algebraic Geometry) implies the existence of a polynomial injection : $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$. On the other hand, in a recent paper Giulio Bresciani has proved that the Bombieri-Lang Conjecture implies that there can not be a polynomial bijection from $\mathbb{Q} \times \mathbb{Q}$ onto $\mathbb{Q}$. In [34] Hector Pasten proved that there is an affine surface X over $\mathbb{Q}$ and a polynomial map $f : X \to \mathbb{Q}$, defined over $\mathbb{Q}$, such that the Q-rational points of X are dense in X, and, nonetheless, f induces an injective function $X(\mathbb{Q}) \to \mathbb{Q}$ on $\mathbb{Q}$-rational points (unconditionally, i.e. without using any conjecture). The discussion on this is going on.

On a relevant subject, cf. the results of [48]. They are a generalisation of Lagrange's Theorem, which states that *every non-negative integer (or rational) number is a sum of four squares.*

3 Global Fields and Derivations

The following has been observed throughout time:

When it comes to classes of diophantine equations, such as 'of degree two" or 'elliptic curves'

- Properties of a class of diophantine equations over $\mathbb{Z}$ are similar to the common properties of the analogous class over almost all rings $\mathbb{F}_p[z]$, as p varies.
- Properties of a class of diophantine equations over $\mathbb{Q}$ are similar to the common properties of the analogous class over almost all fields $\mathbb{F}_p(z)$, as p varies.

The statement is vague but it has worked over time. The usual train of arguments is: If some family of diophantine equations is examined for the number and type of their solutions, first check the same equations for function solutions (often for non-constant function solutions) and, according to the answer, try to transfer knowledge to $\mathbb{Z}$ or $\mathbb{Q}$ (and finite extensions of these).

An axiomatisation which is common to number fields (finite extensions of $\mathbb{Q}$) and global function fields (finite extensions of some $\mathbb{F}_p(z)$) was first given in [2] - they are the fields that posses a 'product formula for valuations' and have a valuation with a specific property. The idea has been central to Andre Weil's *Basic Number Theory*. Much of Number Theory today is done along these lines.

Here is an example: An old question, *Fermat's Last Theorem*, asks for the solutions of the equation

$$a^n + b^n = c^n$$

over the integers, where $n \geq 3$ is a natural number. The conjecture ('Theorem') was that these equations have no non-trivial solutions (i.e. with $abc \neq 0$). It was proved by Andrew Wiles in 1995. But the proof of a similar statement over polynomial rings was known long before, it has actually been a standard exercise in advanced algebra courses.

Theorem 5 *Prove that, if F is a field of characteristic co-prime to the natural number $n \geq 3$, then the equation $a^n + b^n = c^n$ has no non-constant polynomial solutions $a, b, c \in F[z]$ (z is a variable) with $abc \neq 0$.*

Proof Say $n \geq 3$ and (a, b, c) is a solution with a, b and c coprime polynomials (if not then cancel common factors so that this hypothesis

is satisfied). Differentiate to obtain $a'a^{n-1} + b'b^{n-1} = c'c^{n-1}$. Consider it as a linear system in the unknowns a^{n-1} and b^{n-1}, solve and obtain $a^{n-1}(a'c - c'a) = b^{n-1}(c'b - b'c)$. Say that b has the maximum degree, h, among a, b, c. Then, by the coprimality of a and b we have that b^{n-1} (which has degree $(n-1)h$) divides $a'c - c'a$, which has degree less than $2h - 1$. So $(n-1)h \leq 2h - 1$, which contradicts the fact that $n \geq 3$. ∎

But the similar problem over $\mathbb{Z}$ took centuries to solve - observe that, here, the problem of the equation having solutions over $\mathbb{Z}$ is considered similar to the same equation having *non-constant* solutions over $F[z]$.

Why can not one find a 'similar' solution over $\mathbb{Z}$? One answer is obvious: One has differentiation over $F[z]$ - but no similar (nontrivial) operation over $\mathbb{Z}$. We will soon see a proposal towards a remedy of this. But before we do so, let us state the

The abc Conjecture: Let a, b and c be coprime natural numbers such that $a + b = c$. Then, for any $\varepsilon > 0$ we have $c \leq Rad(abc)^{1+\varepsilon}$ with only finitely many exceptions. Here the *radical* of x, is $Rad(x) = \Pi_{p|x} p$, the product of all primes that divide the natural number x (each prime taken once). □

The analogue of the abc-Conjecture for polynomials is with the inequality meaning inequality of degrees and is known to be true with $\varepsilon = 0$. The proof uses differentiation. At this point most experts agree that the Conjecture is open, with the Japanese mathematician Shinichi Mochizuki claiming that he has proved it. The abc-Conjecture has many consequences, for example implies that Fermat' Last Theorem is true for some exponent n and higher.

3.1 An arithmetic differential

For $x \in \mathbb{N}$ we write

$$\partial(x) = x \sum_{p|x} \frac{v_p}{p} \xi_p \tag{3.1}$$

where the sum is taken over all primes p which divide x and $v_p = \mathrm{ord}_p(x)$ is the order of x at p. The ξ_p are variables, one for each prime p. We denote by $\xi = (\xi_p)$ the vector of the ξ_p, say ordered by the size of p. We adopt the convention that $\partial(0) = 0$.

Observe that, if Ω is the free $\mathbb{Z}$-module generated by the variables ξ_p, then $\partial : \mathbb{Z} \to \Omega \otimes \mathbb{Q}$.

It is easy to see that

$$\partial(xy) = x\partial(y) + \partial(x)y \ .$$

(the Leibnitz multiplication rule for derivatives holds)

and if $g = (x, y)$ (the greatest common divisor) then

$$\partial(gx + gy) - \partial(gx) - \partial(gy) = g \cdot \{\partial(x + y) - \partial(x) - \partial(y)\} \ .$$

We define the *differential* of a natural number x to be $\partial(x)$, - which is a functional. For example $\partial(27) = 27\xi_3$ and $\partial(12) = 12\xi_2 + 4\xi_3$. Notice that if one applies the definition of the operation ∂ to polynomials in, say, $\mathbb{F}_p[z]$ - instead of integers, where the 'primes' p are irreducible and monic polynomials - then for the values $\xi_p = \frac{dp}{dz}$ the value of $\partial(x)$ is $\frac{dx}{dz}$ (the derivative of x). We intend to use the above as shown in the following example:

Say natural numbers a, b, c and $n \geq 3$ are given, with $a^n + b^n = c^n$. Look at the above proof of 'Fermat's Last Theorem for polynomials' and replace every occurrence of a derivative by ∂ (so replace a' by $\partial(a)$ etc.). If one knows that the resulting equation (in the unknowns ξ_p, one for each prime that divides a or b or c) has *small* and *suitable* solutions $\xi_p = \tilde{\xi}_p$, then we might be able to reproduce the proof for polynomials. This has not been done so far (without resorting to conjectures) but in [36] the following is proved:

Theorem 6 *(Hector Pasten) The abc-Conjecture is equivalent to the following statement:*

There is an absolute constant η with $0 < \eta < 1$ such that, for any triple (a, b, c) of co-prime natural numbers, not of the form $(1, N, q)$ with q a prime (up to permutation) and such that $a + b = c$, the following holds:
There are integer values $\tilde{\xi}_p$ of the variables ξ_p, so that

$$\partial(a) + \partial(b) = \partial(c)$$

with $\partial(x)$ as in (3.1) and such that $a\partial(b) - \partial(a)b \neq 0$ ('the Wronskian of a and b is non-zero') and $\sup_p\{|\tilde{\xi}_p|\} < c^\eta$ ($\sup_p$ is the supremum norm).

Moreover it is proved that the operation ∂ is a *universal object* in a natural category and this may be considered as an indication that ∂ is a 'natural' - rather than 'artificial' - construction.

Despite the fact that all this has not resulted so far in new mathematical results - other than expressing old conjectures in new ways - I am quite optimistic that it may present a new way to look into old problems, a way that is closer to objects that we have encountered before successfully, such as derivatives and solutions of linear systems of equations over $\mathbb{Z}$ (see for instance *Siegel's Lemma*). Note that the definition of the operation ∂ gives no obvious clue for what a *second differential* may be.

Before we go on we will state one more open problem:

Problem 7 *(Buchi's Problem)*

Is the following true?

There is an absolute constant M such that any sequence of natural numbers $(\bar{x}) = (x_0, \ldots, x_{M-1})$ (with M terms), which has the property that the second difference of the squares of the $\bar{x}_n$ is the constant sequence $(2)_{0,\ldots,M-1}$, i.e. for $n = 1, \ldots, M-1$ we have

$$(x_{n+1}^2 - x_n^2) - (x_n^2 - x_{n-1}^2) = 2 \ , \tag{3.2}$$

is a sequence of successive squares (i.e. $x_{n+1} = \pm x_n \pm 1$). Similarly for rational - rather than natural numbers.

Numerical experimentation indicates that the Problem may have a positive answer for $M = 5$ for the natural numbers and $M = 6$ for the rationals. Pasten has proved that a positive answer follows from the abc-Conjecture.

Analogues of this for rings and fields of functions, and for higher order differences have been established by, among others, Vojta, Vidaux, Pasten, Wang (see [38] and the bibliography therein), but the original problem remains open.

An application of a positive answer to the Problem, known to Buchi, is that, if true, it has as a consequence the following: in the structure of the integers with addition, constant symbols for 0 and 1 and a predicate symbol which is interpreted as 'x is a square ' multiplication is diophantine and therefore the positive existential theory of this structure is undecidable (an improvement over the negative answer to Hilbert's Tenth Problem). A relevant question is

Problem 8 *(asked by Leonard Lipshitz) Is the positive existential theory of the structure $(\mathbb{Z}, +, $ 'x is a square $'$, 'x is a cube $', 0, 1)$ undecidable?*

There is an extensive bibliography on extensions of the group-theory of $\mathbb{Z}$ by various predicates, in many directions, for example compare [33] and the references in [7] - the intuition here is that 'unstable structures are very likely to have undecidable theories'. The decidability result for the theory of the structures $(\mathbb{N}, +, \{2^n \mid n \in \mathbb{N}\}, 0, 1)$, by Semenov, has been extended to $(\mathbb{F}_p[z], +, \{z^n \mid n \in \mathbb{N}\}, 0, 1, z)$ in [54] but the - much stronger - similar result for $(\mathbb{N}, +, n \mapsto 2^n, 0, 1)$ has no clear analogue in the case of polynomials. We ask:

Problem 9 *Is the theory of the structure*
$(\mathbb{F}_p[z], +, \{z^n \mid n \in \mathbb{N}\}, (z^n, x) \mapsto z^n x, 0, 1, z)$ *decidable?*

3.2 A problem of Grothendieck

Grothendieck's Problem for linear and homogeneous algebraic differential equations, which is also known as the *Grothendieck-Katz p-curvature conjecture* asks whether the following is true:

Conjecture 10 *(Grothendieck-Katz)*
Consider a linear and homogeneous algebraic differential equation

$$a_n x^{(n)} + \cdots + a_1 x^{(1)} + a_0 x = 0 \tag{3.3}$$

with all $a_k \in \mathbb{Z}[z]$ *(*z *is a variable) and with* $x^{(k)}$ *being the* k*-th derivative of* x *with respect to* z*.*

Assume that the reduction of (3.2) modulo almost any (any but a finite number) prime number, has n *solutions in* $\mathbb{F}_p((z))$*, linearly independent over the field of constants of differentiation (i.e.* $\mathbb{F}_p((z^p))$*). Then (3.3) has* n *many solutions over the Puiseu series over* $\mathbb{C}((z))$*, linearly independent over* $\mathbb{C}$ *and they are all algebraic over* $\mathbb{C}(z)$*.*

It has come from Geometry. It has been proved for large classes of equations but the general problem remains open. As an indicative example, the equation $2(1+z)x^{(1)} = x$, with p an odd prime, has the polynomial solution $x = (1+z)^{\frac{p+1}{2}}$ over $\mathbb{F}_p(z)$ and over $\mathbb{C}(z)$ the algebraic solution $x = (1+z)^{\frac{1}{2}}$. The case of equations of order one is shown in [20] to be equivalent to *Chebotarev's Density Theorem* and be true. The 'Puiseu series' may be substituted by 'power series' if one considers only differential equations in which $z = 0$ is not a zero of the highest-order coefficient a_n.

I think that, if it is correct, it would be a very good example of a "Transfer Principle" among characteristics, i.e. a statement that, if it holds true for almost all positive characteristics then it remains true in characteristic zero. An example of an application of this principle is the *Ax-Grothendieck Thorem*, which states that a polynomial map $f : \mathbb{C}^n \to \mathbb{C}^n$ which is injective is also surjective; its proof by Ax proves it over finite fields (instead of $\mathbb{C}$) and transfers it over $\mathbb{C}$ by methods of Model Theory. From Logic's point of view, and asked when a_n in (3.3) is not divisible by z, Problem 3.2 seems to require the study of an ultra-power of $\mathbb{F}_p[[z]]$ over a non-principal ultrafilter, in a language that should contain, at the least, symbols for addition, $+$, differentiation, D, the function $x \mapsto zx$ and the operation of multiplication among constants of differentiation (the elements x for which $D(x) = 0$) - in order to be able to state that the constants of differentiation constitute a field.

Problem 11 *Study the ultrapowers $\Pi_U \mathbb{F}_p[[z]]$ over non-principal ultrafilters U on the set of prime integers, as models of the language described above and extensions of it.*

4 Diophantine Equations with infinitely many solutions over some number field

We outline the presentation of Part B of [44].

A question similar to Hilbert's Tenth Problem is whether there is an algorithm to determine whether any given diophantine equation over $\mathbb{Z}$ has or does not have *infinitely many solutions* over any of $\mathbb{Z}$ and $\mathbb{Q}$. On this Martin Davis has proved:

The problem of whether a diophantine equation has infinitely many solutions over $\mathbb{Z}$ or not is undecidable, even given an oracle which determines the solvability of any given diophantine equation over $\mathbb{Z}$.

The similar problem over $\mathbb{Q}$ is an open problem.

A question similar to Hilbert's Tenth Problem but in a qualitative sense would be:

Problem 12 *a) Is there an algorithm which, given any variety over $\mathbb{Q}$, the algorithm determines whether it has an infinite number of points over some number field?*

b) Similarly, over some number ring (i.e. the ring of elements of a number field, integral over $\mathbb{Z}$)?

The following is a conjecture of Serge Lang

Conjecture 13 *(Serge Lang) For a variety V, defined over the rationals, the following are equivalent:*

- *V has infinitely many points in some number field.*
- *There is a non-constant analytic function $f : \mathbb{C} \to V$.*

Applied to curves the Conjecture is known to hold true, by Falting's proof of the 'Mordell Conjecture' (generalised to global fields of positive characteristic by Ehud Hrushovski using model theoretic methods). For example the fact that the circle $X^2 + Y^2 = Q$, with Q a rational number, has infinitely many points over some number field is associated to the fact that there is a non-constant analytic function $f(z) = (\frac{1}{\sqrt{Q}} \cos z, \frac{1}{\sqrt{Q}} \sin z)$ (for $Q \neq 0$) from $\mathbb{C}$. Similarly, any elliptic curve is parametrised by its Weierstrass $\wp$-function and its derivative. But irreducible curves of genus ≥ 2 admit no analytic parametrisation - this has been known long ago - and, due to Falting's Theorem, they may not have an infinite number of points over any number field. The Conjecture of Lang has been generalised and made much more accurate by Paul Vojta - but the experts think that for the moment it seems inaccessible.

If Lang's Conjecture is true then Problem (12) is reduced to the following question:

Problem 14 *Is there an algorithm which, for any variety V defined over $\mathbb{Q}$, determines whether there is a non-constant analytic function $f : \mathbb{C} \to V$?*

This sort of an analogue of Hilbert's Tenth Problem may be expressed for the structure of meromorphic functions with addition, multiplication, a predicate symbol meaning *'x is a non-constant meromorphic function'* and symbols for 0 and 1. At this point the problem seems inaccessible.

5 A program for proving that the existential theory of $\mathbb{Q}$ is undecidable

Regarding the decidability properties of the existential theory of $\mathbb{Q}$, over time there have been expressed both possibilities - decidable and undecidable. The main argument for decidability is that, over the rationals, varieties may be embedded in varieties which have additional structure, such as the abelian varieties. For the possibility of undecidability there are some more concrete plans. I will present a sort of 'program' which seems to me to have some promise towards proving that the existential theory of $\mathbb{Q}$ is undecidable (if it is).

The beginning of it is in [40] (some familiarity with the basic theory of elliptic curves is assumed). Consider an elliptic curve E over $\mathbb{Q}$ of rank one. One may change coordinates so that the curve has the form

$$(\tilde{x}_1^3 + \alpha\tilde{x}_1 + \beta)Y^2 = X^3 + \alpha X + \beta \tag{5.1}$$

with $(\tilde{x}_1, 1)$ being a base point, i.e. the points of E are torsion points plus some $P_n = n \cdot (\tilde{x}_1, 1)$. It is known that, for $n \in \mathbb{Z} \setminus \{0\}$, there are rational functions of one variable x_n and y_n so that for any point (x, y) of E, rational over $\mathbb{Q}$, we have $n \cdot (x, y) = (x_n(x), yy_n(x))$, where $\cdot$ is meant with respect to addition on E. So $P_n = (x_n(\tilde{x}_1), y_n(\tilde{x}_1))$. It is well known how, from this, to produce a diophantine definition of the the set of points P_n (for elliptic curves *without complex multiplication*). Elliptic curves satisfying all the above are known to exist - and be 'many'. The points P_n are taken to be a *model* of the rational integers - see [43] for a concrete discussion of this). Addition of points of E is quantifier-free, therefore the graph of addition of points of E is diophantine. If one achieves giving a diophantine (over $\mathbb{Q}$) definition of the graph of multiplication, i.e. of the set $\{(P_m, P_n, P_{mn}) \mid m, n \in \mathbb{Z} \setminus \{0\}\}$ then one will have proved undecidability of the existential theory of $\mathbb{Q}$. But no one knows how to perform the last step. Instead, one may try to define (in a diophantine way) divisibility among points of E, i.e. the set $\{(P_m, P_n) \mid m \text{ divides } n\}$. If one succeeds one will have a model of addition and divisibility over $\mathbb{Z}$ encoded in a diophantine way into the diophantine theory of $\mathbb{Q}$. But this will not suffice (in order to prove undecidability of the diophantine theory of $\mathbb{Q}$), because Leonard Lipshitz has proved that addition and divisibility over $\mathbb{Z}$ has a decidable existential theory. The problem of defining positive-existentially divisibility of points of E remains unsolved.

But here is something that may help in this: In [40] is proved an *existential divisibility lemma*, which gives a diophantine (over $\mathbb{Q}$) definition of a relation between two rational numbers, u and v, that, roughly, states

any prime number $p \equiv -1$ (mod 4) which divides the denominator of u to an odd order, divides also the numerator or denominator of v to an odd order.

So one is able to express the following property $|^*$ between points $P_n = (\tilde{x}_n, \tilde{y}_n)$, above:

(Existential Divisibility) $P_m|^* P_n$ if and only if every prime $p \equiv -1$ mod (4) that divides the denominator of $\tilde{x}_m$ divides also the numerator or denominator of $\tilde{x}_n$.

And, for suitable elliptic curves E - having all the above properties - this may be equivalent to

$$m \text{ divides } n.$$

And for our purpose it would suffice that the last equivalence holds only for m and n in an infinite ideal of $\mathbb{Z}$ (e.g. the even integers).

Problem 15 *Are there any elliptic curves having the properties assumed of E so far (in particular, satisfying the last condition)?*

From discussions with experts in Number Theory I have gotten the impression that this is very likely, and there should be many such curves. One heuristic reason is that *the denominator of the rational function x_m divides teh denominator of x_{mn}* and the fact that the rational functions x_n and y_n have only irreducible factors (of their numerators and denominators) with multiplicities bounded by 2, so one might expect that something similar may happen for their values at the rational $\tilde{x}_1$ - in a way reminiscent of Hilbert's Irreducibility Theorem. Note that the 'Existential Divisibility Lemma' has been generalised for many sets of primes, instead of the primes $p \equiv -1$ (mod 4) in [14].

But, as explained above, this will not suffice for our task (proving undecidability of the existential theory of $\mathbb{Q}$). But the following twist of the above train of thought may have some chance of success. It is due to Gunther Cornelissen: There are *abelian varieties with real multiplication*, i.e whose points (apart from a finite set of torsion points) are given as $P_n = n \cdot P_1$, where n ranges in a ring of real quadratic integers, e.g. in $\mathbb{Z}[\sqrt{5}]$. What if one may do the above - give a definition of divisibility of the indices n in a way which is diophantine over $\mathbb{Q}$? Then one would code in the diophantine theory

of $\mathbb{Q}$ a model of, say, $\mathbb{Z}[\sqrt{5}]$, with addition and divisibility (both meant in $\mathbb{Z}[\sqrt{5}]$). And Leonard Lipshitz has proved that the positive existential theory of addition and divisibility in $\mathbb{Z}[\sqrt{5}]$ (and any real quadratic number ring) is undecidable. At the time that Cornelissen expressed this idea these abelian varieties were a new finding. And the problem of giving a diophantine definition of 'a point dividing another' seemed to be much more complicated than for elliptic curves. But maybe new knowledge for these varieties is enough to change this. So I ask:

Problem 16 *Consider an abelian variety V whose $\mathbb{Q}$-rational points are torsion plus a group of points isomorphic to an order of $\mathbb{Q}[\sqrt{\ell}]$ where ℓ is a positive integer. Is the relation 'the point P divides the point Q' diophantine over $\mathbb{Q}$?*

Of course a far aim in this investigation would be an analogue of $DIO = RE$ (over the integers) over $\mathbb{Q}$:

Problem 17 *Is there a purely algorithmic characterisation of the diophantine subsets of $\mathbb{Q}$ and finite powers of it?*

Despite the fact that the above are mostly open problems, one might try to get some ideas of how to proceed in order to answer them, trying first to produce a proof of undecidability of the analogue of HTP for the fields of rational functions $\mathbb{F}_p(z)$ with coefficients in $\mathbb{F}_p[z]$, in a manner as similar to the above as possible. Note that the existing proof of undecidability of the existential theory of $\mathbb{F}_p(z)$ (given by the author for odd p and by Carlos Videla for $p = 2$) is quite different from the above (see the presentation in [43] for the case of odd p). So I will suggest a likely way to do this.

Say that we work over some $\mathbb{F}_p(z)$ with coefficients in $\mathbb{F}_p[z]$. One may produce elliptic curves E over $\mathbb{F}_p(z)$ whose $\mathbb{F}_p(z)$-rational points (even the $\tilde{\mathbb{F}}_p(z)$-rational points) may be inrdexed as $n \cdot P_1$ plus (a finite) torsion (see how in cf. [41]). One problem is that the indices n range not in $\mathbb{Z}$ but in an order of a non-real quadratic field (e.g. $\mathbb{Z}[\sqrt{-p}]$); but this does not seem to me as a critical problem. A difference of the case of positive characteristic (over $\mathbb{F}_p(z)$) from the case of characteristic zero ($\mathbb{Q}$) is that in the case of characteristic $p > 0$ one may produce a diophantine definition of the relation

$$\{(u, v) \in (\mathbb{F}_p(z))^2 \ \exists s \in \mathbb{Z} \ v = u^{p^s}\} \ . \tag{5.2}$$

(see a discussion of new ways to produce such definitions for $p \neq 2$ in Section 6.3). Then, if one had a diophantine definition of divisibility of indices of the points P_n, the structure of the points P_n, together with addition on E, divisibility among the indices n and the relation $|_p^Z$ among indices, defined by '$m|_p^Z n$ if and only if $\exists h \in \mathbb{Z}\ n = p^s \cdot m$' would give a model of addition, divisibility (denoted by $|$) and $|_p^Z$ on the set of indices. It has been proved by Denef that in the structure $(\mathbb{Z}, +, |, |_p^Z, 0, 1)$, one may define positive-existentially multiplication, hence the positive existential theory of it is undecidable. So, in order to produce an undecidability proof of HTP for $\mathbb{F}_p(z)$ as above it would suffice to define, positive existentially, a definition of divisibility $|$ of indices of the points P_n. Maybe such a definition results from some sort of analogue of the Existential Divisibility Lemma for the fields $\mathbb{F}_p(z)$. I consider it very likely that analogues of the methods of [23] for the function-field case would be useful in this.

Problem 18 *Give a proof of undecidability of the existential theory of any $\mathbb{F}_p(z)$, with coefficients in $\mathbb{F}_p[z]$, along the lines of the last paragraph. See how much of the methods may be generalised to the case of $\tilde{\mathbb{F}}_p(z)$ (the undecidability of the existential theory of which is an open problem).*

6 The question of decidability of the theory of $\mathbb{C}(z)$.

We continue the presentation of Section 2 of [43].

The following is an old unanswered question:

Problem 19 *Is the first order theory of $\mathbb{C}(z)$ decidable in the language L_z, which extends the language of rings by a constant-symbol for the variable z?*

The similar problems in positive characteristic have been answered negatively: The theory of a field $\tilde{\mathbb{F}}_p(z)$ in the language L_z is undecidable, due to Gregory Cherlin (cf. [41]).

The similar question for the diophantine theory seems much harder.

Problem 20 *Is the diophantine theory of $\mathbb{C}(z)$ in the language L_z decidable?*

In fact there is no known example of an algebraically closed field F for which the diophantine theory of $F(z)$ (analogue of HTP) in L_z is undecidable (or decidable).

A problem, seemingly similar to the above, but probably much harder is the following: Consider the language L_T which extends the language of rings by a predicate symbol T, which is interpreted as

$T(x)$ *if and only if the function* x *is not a constant function.*

One may consider the theory and the diophantine theory of a ring of functions of the variable z in the language L_T. Notice that, given a variety V over $\mathbb{Q}$, the question of whether V contains a non-trivial rational curve may be expressed as a positive-existential sentence of L_T. As we saw in Section 4, similar problems in rings of analytic or meromorphic functions of one variable relate to a qualitative version to Hilbert's tenth problem, cf. [44]. Here we ask:

Problem 21 *a) Is the theory of* $\mathbb{C}(z)$ *in the language* L_T *decidable?*
b) Is the diophantine theory of $\mathbb{C}(z)$ *in the language* L_T *decidable?*

There are only few results on undecidability of L_T-theories of rings of functions - even fewer for existential theories. One of them is for rings of polynomials (in any number of variables) over an integral domain - due to Zahidi and the aurhor, cf. [37].

It is known that a proof of undecidability of the theory (respectively, existential theory) of $\mathbb{C}(z)$ in L_z would result from a definition (resp. existential definition) of the property *'the function* x *has no pole at* $z = 0$*'* - along the lines of Denef's similar proof for $\mathbb{R}(z)$. The next subsection constitutes an effort towards producing such a definition. Another effort has been through using elliptic curves over $\mathbb{C}(z)$ of rank equal to two - along the lines of the proof of undecidability of the existential theory of $\mathbb{C}(z_1, z_2)$ (z_1 and z_2 are independent variables, with constant symbols to represent them in the language) of Kim and Roush; it has been mostly unsuccessful so far. The last subsection contains a definition of a countable subset of the integral algebraic closure of $\mathbb{Z}$ in L_z over $\mathbb{C}(z)$- the first of this kind, to the best of my knowledge.

6.1 Defining order at a fixed point in fields of functions

Let K be a field of functions of the (one) variable z which is one of a) rational functions, b) the field $\mathcal{M}_z(\mathbb{C})$ of meromorphic functions of z as this ranges on$\mathbb{C}$, c) the field $\mathcal{M}_z(\mathbb{C}_p)$ of meromorphic functions of z as this ranges on $\mathbb{C}_p$ (the field of p-adic complex numbers). By $\mathrm{ord}_z(x)$ we denote the order of the function x at $z = 0$, i.e. the multiplicity of the divisor z in x.

Problem 22 *Let K be as above. For any $M \in \mathbb{N}$, consider the set K_M of functions of K*

$$K_M = \{x \mid \exists a, a_1, \dots a_M \in K[x = a^2 \cdot \Pi_{k=1}^{M}(a_k^4 - z^2)]\} \ . \tag{6.1}$$

Is there an M such that $\{x \in \mathbb{Q}(z) \mid \mathrm{ord}_z(x)$ is even $\} \subseteq K_M$?

It remains open . Notice that for a meromorphic function x in $\mathcal{M}_z(\mathbb{C})$ or $\mathcal{M}_z(\mathbb{C}_p)$ with $p \neq 2$ the following are equivalent:

- $\mathrm{ord}_z(1 + zx^2)$ is even.
- $\mathrm{ord}_z(x) \geq 0$.

Therefore a positive answer to Problem 22 would result in a diophantine definition in L_z of a subset of functions $K^{\square}$ of any K as above, such that

$$\{x \in \mathbb{Q}(z) \mid \mathrm{ord}_z(x) \geq 0\} \subseteq K^{\square} \subseteq \{x \in K \mid ord_z(x) \geq 0\} \ .$$

It is known that a positive answer to Problem 22 in the case $K = \mathbb{C}(z)$ will imply that the analogue of HTP for $\mathbb{C}(z)$ with coefficients in $\mathbb{Z}[z]$ has a negative answer (cf. the discussion of the similar subject in the Thesis of Karim Zahidi).

6.2 A set of algebraic integers, definable over $\mathbb{C}(z)$.

The following is due to an unknown, to me, referee, around 1990. It gives a set J_2, definable over $\mathbb{C}(z)$ in L_z, which is guaranteed to be a subset of the set of algebraic integers.

Consider an elliptic curve E over $\mathbb{C}$, given by some equation on the (X, Y)-plane, say $Y^2 = X^3 + \beta X^2 + \gamma X + \delta$, with $\beta, \gamma, \delta \in \mathbb{C}$ (the right-hand side must be a polynomial of X with only simple zeros) and with identity the (uniques) point at infinity. Its *j-invariant* is a rational function of the coefficients β, γ and δ . Consider an element s, algebraic over $\mathbb{C}(z)$, such that $s^2 = z^3 + \beta z^2 + \gamma z + \delta$,

The non-constant algebraic endomorphisms $e : E \to E$ (i.e. maps from E to E which preserve addition on the curve and are algebraic over the inputs) may be described as pairs of functions, $e = (x(z), sy(z))$ of the variables z and s, where $x, y \in \mathbb{C}(z)$. This set of endomorphisms, which we will denote by $End(E)$, is a ring under the operations of addition on E and composition. It is known that (in characteristic zero) $End(E)$ is isomorphic to either $\mathbb{Z}$ or to some order in an imaginary quadratic extension of $\mathbb{Q}$. In the second case

we say that E has *'complex multiplication'*. It is known that the j-invariant of a curve with complex multiplication is an algebraic integer. Look at the ring $End(E)/(2End(E)$, i.e. $End(E)$ modulo 2 (with addition meant on E). If this ring contains elements other than 0 and 1 then E has complex multiplication, hence its j-invariant is an algebraic integer. So define J_2 to be the set of j-invariants of elliptic curves E (with the coefficients β, γ and δ as parameters), for which the above holds (i.e. $End(E)/(2End(E))$ has more than two elements). It is easy to see that J_2 is definable over $\mathbb{C}(z)$ in the language L_z. Technically, the, so called, *Manin-Denef curve* E_z, associated to E, given by the equation

$$(z^3 + \beta z^2 + \gamma z + \delta)y^2 = x^3 + \beta x^2 + \gamma x + \delta \tag{6.2}$$

in the variables x and y, which defines an elliptic curve over $\mathbb{C}(z)$. For each solution (x, y) of it over $\mathbb{C}(z)$ the function defined by $(z, s) \mapsto (x, sy)$ is a function from E into E. By a theorem of Weil any such function (from E to E and with an algebraic graph over $\mathbb{C}(z, s)$) is of the form $e \oplus P$ where $\oplus$ denotes addition on E, e is an endomorphism of E and P is a point of E rational over $\mathbb{C}$. Looking at the possible values of (x, y) at the points of E of order two, i.e. for $s = 0$, one sees that the point P is, if not the identity (the point at infinity) necessarily of order 2 on E - and there are three such points. We conclude that the affine points (x, y) of E_z, rational over $\mathbb{C}(z)$, are precisely the points for which $(x, sy) = e \oplus P_0$ for some endomorphism e of E and P_0 is either the identity or one of the three points of E of order 2. The rest (for defining J_2) is easy and left to the reader.

Of course J_2 may not contain all j-invariants of curves with complex multiplication, because in some cases it might be that all endomorphisms are conguent to 0 or 1 modulo 2. But I would guess that J_2 is an infinite set. Notice that, given any positive integer $d > 1$ one may work similarly with $End(E)/(dEnd(E))$ in order to define a set J_d, likewise. So, in trying to determine the decidability properties of the theory of $\mathbb{C}(z)$ in L_z, one may use for free a predicate for any set J_d. Also notice that the definition of J_d, apparently is not existential - the statement "$End(E)/(2End(E))$ has more than two elements", which translates to '$End(E)$ has at least three elements, which are pairwise in-equivalent modulo 2', needs a universal quantifier. So I ask:

Problem 23 *Are the sets J_d, defined above, existential over $\mathbb{C}(z)$ in L_z?.*

I can not see an immediate application of the above. But the sets J_d, if they are large enough as I expect, are non-trivial sets of algebraic integers, such sets are often quite complicated and this seems to indicate a step towards undecidability.

6.3 The case of characteristic $p > 0$

In the case of global fields of characteristic $p > 0$ the Frobenius map $x \mapsto x^p$ may be used in order to produce a positive-existential definition of the relation $x|_p y$, defined by $\exists s \in \mathbb{Z}\ y = x^{p^s}$. This was established first for fields of rational functions (by the author for odd p and Carlos Videla for $p = 2$). A uniform way to define the same relation in a positive existential way but, also, uniformly throughout the positive characteristics $p > 0$ is given in [37]. So one may produce, over $\tilde{\mathbb{F}}_p(z)$ in L_z, a 'model' $\mathbb{Z}$ with the structure of addition and the relation $|_p^Z$ in $\mathbb{Z}$, where, for $m, n \in \mathbb{Z}$, $|_p$ is defined by $m|_p n$ if and only if $\exists s \in \mathbb{Z}\ n = p^s m$.

Problem 24 *Is the positive-existential theory of the structure* $(\mathbb{Z}, +, |_p^Z, 0, 1)$ *undecidable?*

If the answer is positive then the analogue of HTP for any $\tilde{\mathbb{F}}_p(z)$ in L_z is undecidable.

It is interesting to see the latest sort of definitions of the relation $|_p$ over $\tilde{\mathbb{F}}_p(z)$ (for sufficiently large p):

Theorem 25 *Let* $p \geq 23$ *be a prime number. Then, for any* $x, y \in \tilde{\mathbb{F}}_p(z) \setminus \tilde{\mathbb{F}}_p$ *the relation* $x|_p y$ *holds (i.e.*$\exists s \in \mathbb{Z}\ y = x^{p^s}$*) if and only if there is a sequence* $x_1, \ldots, x_{20}$ *such that* $x_2 \neq \pm x_1 \pm 1$ *and any three successive* x_{n-1}, x_n, x_{n+1} *of it satisfy the conditions (3.2) of Buchi's Problem and* $xy = x_1^2$ *and* $x + y = x_2^2 - x_1^2 - 1$.

Notice that this gives an existential definition of the relation $|_p$ over $\tilde{F}_p(z)$ even in the language L_T. For a proof see Lemma 6.1 and the proof of Proposition 4.1 of [37]. Observe that, for any non-negative integer s, the squares of successive terms of any sequence of the form $\left((x+n)^{\frac{p^s+1}{2}}\right)_n$ satisfy the relations (3.2).

More definitions of the same relation $|_p$ have been given, among others by Pasten.

7 Analogues of HTP for rings of analytic and meromorphic functions.

The ring of functions of one variable, analytic on a domain of $\mathbb{C}$ with non-empty interior has an undecidable theory in the language L_z but even in the language of rings (due to Raphael Robinson). For analogues of HTP the following problems, due to Lee Rubel, are open:

Problem 26 *Let $\mathcal{H}_z(\mathbb{C})$ be the ring of entire functions of the single variable z and let $\mathcal{M}_z(\mathbb{C})$ be the field of fractions of $\mathcal{H}_z(\mathbb{C})$.*

a) Is the positive existential (respectively, existential) theory of $\mathcal{H}_z(\mathbb{C})$ in L_z decidable?

b) Let $L_{z,ord} = L_z \cup \{\text{ord}\}$ be the extension of the language L_z by the predicate-symbol ord*, which, in $\mathcal{M}_z(\mathbb{C})$ is interpreted as the set of functions which are analytic at $z = 0$. What about the similar problem for the field $\mathcal{M}_z(U)$ of meromorphic functions on the open unit disk U? (or the closed unit disk $\bar{U}$, meaning meromorphic on some open set containing $\bar{U}$?)*

c) Is the existential theory of $\mathcal{M}_z(\mathbb{C})$ in $L_{z,ord}$ decidable?

d) As in c) for $\mathcal{M}_z(U)$ or $\mathcal{M}_z(\bar{U})$?

The analogue of a) of the Problem for the ring $\mathcal{H}_z(\mathbb{C}_p)$ of entire functions on a p-adic analogue $\mathbb{C}_p$ of $\mathbb{C}$ has a negative answer (due to Leonard Lipshitz and the author) - $\mathbb{Z}$ is diophantine in L_z. Xavier Vidaux proved in [56] a negative answer for the field $\mathcal{M}_z(\mathbb{C}_p)$ of p-adic meromorphic functions on $\mathbb{C}_p$ - $\mathbb{Z}$ is diophantine- but only in the language $L_z \cup \{ord\}$ (the question raised in Subsection 6.1 may be asked of $\mathcal{M}_z(\mathbb{C}_p)$ and is open).

The first of these results uses the fact that p-adic entire functions which are not polynomials have an infinite number of zeros. In particular there is no such function without zeros (such as the exponential in the complex case). Out of this (and with a lot of work) comes the fact that there is no non-polynomial map of the form (x, sy) from the curve $s^2 + z^2 = 1$ (the unit circle over $\mathbb{C}_p$) into itself, i.e. the equation

$$(1 - z^2)y^2 = 1 - x^2 \ , \tag{7.1}$$

in the unknowns (x, y), and with $x, y \in \mathcal{H}_z(\mathbb{C}_p)$ has only polynomial solutions. These are known: they are given by

$(\pm x, y) = (x_n, y_n) = (\cos(n \arccos z), \frac{1}{\sqrt{1-z^2}} \sin(n \arccos z))$

(the $\pm$ sign is read as 'plus or minus').
For each integer n - the x_n are the *Chebishev polynomials*. These solutions were first used by J. Robinson, then by M. Davis and by J. Denef in order to prove undecidability of the analogue of HTP for polynomial rings such as $\mathbb{C}[z]$ in L_z. Here one may repeat Denef's proof and give a diophantine definition of the rational integers over $\mathcal{H}_z(\mathbb{C}_p)$ in L_z.

For the meromorphic functions $\mathcal{M}_z(\mathbb{C}_p)$ a similar phenomenon holds: Any function from an elliptic curve E with affine equation in (z,s) coordinates $s^2 = z^3 + \beta z^2 + \gamma z + \delta$ (with $\beta, \gamma, \delta \in \mathbb{Q}$) into itself is a rational function of z and s- but the proof here is much more difficult than in the case of the circle. In consequence Equation (6.2):

$$(z^3 + \beta z^2 + \gamma z + \delta)y^2 = x^3 + \beta x^2 + \gamma x + \delta$$

in the unknowns (x,y), with $x, y \in \mathcal{M}_z(\mathbb{C}_p)$ has only rational function solutions (those presented in Subsection 6.2). Choosing the coefficients (i.e. the curve E of Subsection 6.2) so that it does not have complex multiplication, a variant of the proof of Denef for the case of $\mathbb{R}(z)$ works in order to obtain a definition of the rational integers, diophantine over $\mathcal{M}_z(\mathbb{C}_p)$ - but one needs the predicate interpreting ord.

Since those results Natalia Garcia-Fritz and Hector Pasten have produced new similar (undecidability) results for various non-archimedean rings and fields - see the introduction of [42]. The latest result, in [42], proves the following:

Consider the ring $\mathcal{H}_{z_1,z_2}(\mathbb{C})$ of functions of the two variables z_1 and z_2, analytic as (z_1, z_2) ranges over $\mathbb{C}^2$ and let $\mathcal{M}_{z_1,z_2}(\mathbb{C})$ be the field of fractions of $\mathcal{H}_{z_1,z_2}(\mathbb{C})$. Let $L_{z_1,z_2,eval}$ be the language of rings, extended by symbols for z_1, z_2 and a predicate symbol for the property *Eval* of elements of $\mathcal{M}_{z_1,z_2}(\mathbb{C})$, defined by

Eval is the set of meromorphic functions which, evaluated at $z_1 = 0$ give functions of z_2, analytic around $z_2 = 0$ and with a zero at $z_2 = 0$.

Then

Theorem 27 *(Xavier Vidaux and the author)*

The set of rational integers $\mathbb{Z}$ is diophantine over $\mathcal{M}_{z_1,z_2}(\mathbb{C})$ in $L_{z_1,z_2,eval}$, therefore the positive-existential $L_{z_1,z_2,eval}$-theory of $\mathcal{M}_{z_1,z_2}(\mathbb{C})$ is undecidable.

The proof uses the following: Set $z_1 = z-1$ and $z_2 = \delta+2$, so $\mathcal{M}_{z_1,z_2}(\mathbb{C}) =$

$\mathcal{M}_{z,\delta}(\mathbb{C})$. Consider the solutions $(x,y) \in \mathcal{M}_{z,\delta}(\mathbb{C})$ of

$$(z^3 + \delta z^2 + z)y^2 = x^3 + \delta x^2 + x. \tag{7.2}$$

The solutions $(x,y) = (x_n, y_n)$, which were presented in Subsection 6.1 for given any fixed δ, are actually rational functions of the pair of variables (z,δ). We do not know whether there are any solutions of (7.2) other than those. But yet the following is proved:

Consider a pair of solutions (x,y) of (7.2), where y is not the zero function. Write

$$A_{xy} = \frac{x_z}{y}$$

where x_z is the partial derivative of x with respect to z and, for any $w \in \mathcal{M}_{z,\delta}(\mathbb{C})$, $w|_{z=1}|_{\delta.=-2}$ is the value of the meromorphic function w, evaluated first at $z=1$ and then at $\delta = -2$ - and the notation implies that $w|_{z=1}$ is a function of δ, analytic around $\delta = -2$). The main point is to prove that

• *The function A_{xy} is an analytic function of the pair of variables (z,δ) over $\mathbb{C}^2$ and the value $A_{xy}|_{z=1}|_{\delta.=-2}$ is a rational integer.*

Observe that for each $\delta \neq \pm 2$ Equation (7.2) defines an elliptic curve E_δ. It is shown that in the limit case $\delta \to -2$ the value of A_{xy} is a rational integer. The proof is too complicated to be presented here - for an outline see Section 2 of [42], Here we ask:

Problem 28 *Does Equation (7.2) have any non-rational solutions (x,y), over $\mathcal{M}_{z,\delta}(\mathbb{C})$?*

Notice that questions that may be expressed in the language $L_{z_1,z_2,eval}$ (i.e. equations which are polynomials of the unknowns, together with conditions of the form $Eval(x)$) include many questions in the Sciences that come from solutions of differential equations together with initial or boundary conditions. A slightly different (but probably quite more difficult) problem would be to determine the question of decidability in the following situation, also coming from natural problems:

Problem 29 *Let L_{z_1,z_2} be the extension of the language of rings by symbols for the variables z_1 and z_2.*

a) Is the positive existential theory of $\mathcal{H}_{z_1,z_2}(\mathbb{C})$ decidable?

b) Let $L_{z_1,z_2,ord}$.be the extension of L_{z_1,z_2} by a symbol for the property ord, *defined by*

'ord *is the set of elements of* $\mathcal{M}_{z_1,z_2}(\mathbb{C})$ *which are analytic around* $(z_1, z_2) = (0, 0)$ *and obtain the value zero there'.*

Is the positive existential theory of $\mathcal{M}_{z_1,z_2}(\mathbb{C})$ *in* $L_{z_1,z_2,ord}$ *decidable?*

There are many more questions that come naturally from the proof of Theorem 27, especially in Analytic Geometry - but asking some of them would take us out of the aim of the present article.

7.1 Exponential polynomials

It has been a question of Lee Rubel to examine the analogue of HTP for rings that result from extending $\mathbb{C}[z]$ (and $\mathbb{C}(z)$) by as many functions of elementary Calculus as possible and closing under composition. One of the simplest rings of this form is the *ring of exponential polynomials*, often written as $\mathbb{C}[z]^E$ i,e, the result of closing $\mathbb{C}[z]$ under the operations of addition, multiplication and composition. It is not known whether the theory of $\mathbb{C}[z]^E$ in L_z (or L_T) is decidable or not, much more for the existential or positive existential theory.

Problem 30 *Which of the following is decidable?*

- *The theory of* $\mathbb{C}[z]^E$ *in* L_z.
- *The existential (respectively, positive existential theory) of* $\mathbb{C}[z]^E$ *in* L_z.
- *The theory of* $\mathbb{C}[z]^E$ *in* L_T.
- *The existential (respectively, positive existential theory) of* $\mathbb{C}[z]^E$ *in* L_T.

In [4] a negative answer to the analogue of HTP in L_z is given for the ring of exponential polynomials 'of finite order' (in the sense of Complex Analysis). It is shown that in this ring the solutions of (7.1) are only the polynomial ones and $\mathbb{Z}$ is diophantine.

8 Characterisation of diophantine sets

How much may one extend the answer $DIO = RE$ over the integers to other domains? Do analogous facts hold true for rings such as $\mathbb{F}_p[z]$ (with coefficients in the natural image of $\mathbb{Z}[z]$)? The answer is essentially *yes*, see [15]. What about the fields $\mathbb{Q}(z)$ and $\mathbb{F}_p(z)$ - where negative answers to the

analogue of HTP have been given (with coefficients in $\mathbb{Z}[z]$)? The answer is not known.

Problem 31 *a) Give characterisations of the diophantine subsets, in L_z, of $\mathbb{Q}(z)$ and $\mathbb{F}_p(z)$, if possible in algorithmic terms. In particular answer to the question whether $\mathbb{F}_p[z]$ is diophantine in $\mathbb{F}_p(z)$.*

b) Is $\mathbb{Q}$ diophantine over $\mathbb{Q}(z)$ with coefficients in $\mathbb{Z}[z]$?

For a) notice that it is known that $\mathbb{C}[z]$ *is not diophantine* over $\mathbb{C}(z)$ in L_z - see [24].

For b), notice that the proof by Denef of the undecidability of $\mathbb{Q}(z)$ in L_z does not give a definition of $\mathbb{Q}$ (even more so for an *existential definition*).

9 A short list of results

(complementing those presented in [43])

1. Analogue of HTP for rational solutions (HTP($\mathbb{Q}$)): **An open problem see Section 1**.

2. Is $\mathbb{Z}$ existentially definable in $\mathbb{Q}$ (in the language of arithmetic)? **An open problem; see Section 1**.

3. Analogue of HTP for 'diophantine equations' with coefficients in the natural image of $\mathbb{Z}[z]$ and solutions in

   (a) Any of $\mathbb{F}_p[z]$, $\mathbb{Q}[z]$, $\mathbb{C}[z]$ (in general: $F[z]$, F a field or even an integral domain) (HTP($F[z]$): **Undecidable, due to J. Denef**.

   (b) Any of $\mathbb{R}(z)$, $\mathbb{F}_p(z)$: **Undecidable, due to J. Denef, the author and C. Videla; see Section 6.3**.

   (c) Any of $\tilde{\mathbb{F}}_p(z)$ (where $\tilde{\mathbb{F}}_p$ is an algebraic closure of $\mathbb{F}_p$), $\mathbb{C}(z)$: **An open problem, see Section 6**.

   (d) Any of $\mathbb{R}[[z]]$, $\mathbb{R}((z))$, $\mathbb{C}[[z]]$, $\mathbb{C}((z))$: **Decidable, due to J. Ax and S. Kochen**.

   (e) The ring of entire (i.e. global analytic) functions of the variable z: **An open problem; see Section 7**.

4. Is $\mathbb{C}[z]$ existentially definable in $\mathbb{C}(z)$ in the language L_z? **No, due to J. Kollar.**

5. Is $\mathbb{Q}$ existentially definable in $\mathbb{Q}(z)$ in the language L_z? **An open problem; see Section 8.**

6. Is the existential theory of any of $\mathbb{F}_p[[z]]$, $\mathbb{F}_p((z))$, $\tilde{\mathbb{F}}_p[[z]]$, $\tilde{\mathbb{F}}_p((z))$ in the language $L_z$decidable? **An open problem**.

7. Given a diophantine equation with coefficients in $\mathbb{Z}[z]$, does it have a power series solution with radius of convergence > 1? **An open problem**.

8. Analogues of HTP for rings of transseries (see definitions in [17]): **Open problems**.

Some comments:

For rings and fields of power series (Item 3): The diophantine theory of $\mathbb{F}_p[[z]]$ and $\tilde{\mathbb{F}}_p[[z]]$ with coefficients in $\mathbb{F}_p[z]$ is decidable by the analogue of the *Artin-Greenberg approximation*, due to J. Denef and L. Lipshitz (see [10] and for recent results and references [31]); in other words we have a way to effectively solve systems of equations over these rings but not for systems of equations and in-equations.

For rings and fields of power series in positive characteristic (Item 6) J. Denef and Hans Schoutens have proved in [11] that if there is resolution of singularities in positive characteristic then the existential theory under consideration is decidable. There are new (decidability) results in extensions of the language of rings (but not in L_z) due to Sylvy Anscombe and Arno Fehm in [1]. Notice that for fields F of characteristic other than 2, $\exists y\ y^2 - zx^2 = 1$ is equivalent over $F((z))$ to $x \in F[[z]]$ (an application of Hensel's Lemma, due to J. Ax, see generalisations in [19]). For connections with the fields of p-adic numbers see [21] and its references. For power series of more than one variables see [13]. For a rather complete bibliography see he page *The valuation theory home page* of Franz-Victor Kuhlmann. For some very interesting relative applications and additional bibliography see [9].

For rings of rational or algebraic functions of one or more variables there is a number of results - all negative - by A. Shlapentokh, Ki Hang Kim and Fred Roush (a characteristic result with method different from most of the bibliography: undecidabity of the existential theory of $\mathbb{C}(z_1, z_2)$, with coefficients in $\mathbb{Z}[z_1, z_2]$) and relevant results by K. Zahidi, Kirsten Eisentraeger, G. Demeyer and Claudia Degroote. Many analogues of HTP for global fields rely on knowing that, for any given such field, there is an elliptic curve of

rank one, with generator the point $(z, 1)$, given by Equation (6.2); this holds true by [30].

For analogues of HTP for rings between $\mathbb{Z}$ and $\mathbb{Q}$ see the survey [53] and the references therein.

For the status of the following

Conjecture 32 *(Jan Denef and Leonard Lipshitz) Let K be a number field and $\mathcal{O}_K$ the ring of integers of K. Then the positive-existential theory of $\mathcal{O}_K$ in the language of rings is undecidable (i.e. the analogue of HTP for $\mathcal{O}_K$ has a negative answer.*

all existing results on this area negative answers to HTP (by J. Denef, L. Lipshitz, A. Shlapentokh, the author and recently by N. Garcia-Fritz and H. Pasten - using Iwasawa Theory). An analogue for global fields of characteristic > 0 is known, see [51]. In [46] and [8] it is proved that each of two different conjectures in Number Theory would imply Conjecture 32. B. Mazur and K. Rabin have shown that the Shafarevich-Tate Conjecture (in Number Theory) implies both the conjectures.

Another direction is to examine *questions of uniformity*, i.e. a common definition of similar sets in each of a class of structures, cf. [5], or (un)decidability of the problem whether a formula (or an existential formula) is true in almost all structures in a class (e.g. almost all $\mathbb{F}_p[z]$, as p varies), cf. [37].

For more interesting relevant results see [29], [18], [26] and work by - except for the authors of papers in the references - Lou Van den Dries (see especially questions for rings of transseries), David Marker, and Maxim Vsemirnov. Also by Mihai Prunescu and Leonidas Cerda-Romero (about the structure of addition and divisibility).

References

[1] Sylvy Anscombe and Arno Fehm, *Algebra and Number Theory*, **10-3** (2012), 665-683.

[2] Emil Artin and George Whaples, *Axiomatic characterization of fileds by the product formula for valuations*, Bulletin of the American Mathematical Society, **51-7** (1945), 469-492.

[3] Alexis Bes, *A Survey of Arithmetical Definability*, Bulletin of the Belgian Mathematical Society, Simon Stevin, (2013), pp.1-54.

[4] D. Chompitaki, N. Garcia-Fritz, H. Pasten, T. Pheidas, X. Vidaux, *The diophantine problem for rings of exponential polynomials*, Annali della Scuola Normale Superiore di Pisa, Classe di Scienze (accepted).

[5] Raf Cluckers, Jamshid Derakhshanc, Eva Leenknegt, Angus Macintyre, *Uniformly defining valuation rings in Henselian valued fields with finite or pseudo-finite residue fields*, Annals of Pure and Applied Logic, —bf 164-12 (2013),

[6] Jean-Louis. Colliot-Thélène and Jan van Geel, *Le complémentaire des puissances nièmes dans un corps de nombres est un ensemble diophantien*, Compos. Math. **151-10** (2015), 1965-1980 . 1236-1246.

[7] Gabriel Conant and Christian d'Elbée and Yatir Halevi and Léo Jimenez and Silvain Rideau-Kikuchi, *Enriching a predicate and tame expansions of the integers*, arXiv:2203.07226 [math.LO].

[8] Gunther Cornelissen, Thanases Pheidas, and Karim Zahidi, *Division-ample sets and the Diophantine problem for rings of integers*, Journal de Théorie des Nombres de Bordeaux, **17-3** (2005), 727–735.

[9] Jan Denef, *Arithmetic and Geometric Applications of Quantifier Elimination for Valued Fields*, Model Theory, Algebra and Geometry, MSRI Publications, **39** (2000).

[10] Jan Denef and Leonard Lipshitz, *Ultraproducts and approximation in local rings II*, Mathematische Annalen, **253-1** (1980), 1–28.

[11] Jan Denef and Hans Schoutens *On the decidability of the existential theory of Fp[[t]]*, Fields Institute Communications (2003).

[12] Martin Davis, Yuri Matiyasevich, Julia Robinson, *Hilbert's tenth problem. Diophantine equations: positive aspects of a negative solution*, *Mathematical developments arising from Hilbert problems* (ed. F. E. Browder), Proc. Sympos. Pure Math., vol. 28, Part 2, Amer. Math. Soc., Providence, RI, (1976), 323–378.

[13] Francoise Delon, *Formal power series*, Annals of Mathematics and Artificial Inteligence, —bf 16 (1996), 59-73.

[14] Jeroen Demeyer and Jan Van Geel, *An Existential Divisibility Lemma for global fields*, Monatsefte fur Mathematik, **147-4** (2006), 293-308.

[15] Jeroen Demeyer, *Recursively enumerable sets of polynomials over a finite field are diophantine*, Inventiones Mathematicae, **170-3** (2007). 655-670.

[16] Philip Dittmann, *Irreducibility of polynomials over global fields is diophantine*, Compos. Math. **154** (2018), 761-772.

[17] G. A. Edgar, *Transseries for beginners*, Real ANalysis Exchenge, **35-2**, , obtainable from https://people.math.osu.edu/edgar.2/preprints/trans_begin/beginners.pdf

[18] Kirsten Eisentraeger, Russell Miller, Caleb Springer, Linda Westrick, *A topological approach to undefinability in algebraic extensions of* $\mathbb{Q}$, arXiv:2010.09551 .

[19] Arno Fehm, *Existential-definability of henselian valuation rings*, The Journal of Symbolic Logic, **80-1** (2015), 301-307.

[20] , Taira Honda, *Algebraic Differential Equations*, Symposia Mathematica XXIV, (1981).

[21] Konstantinos Kartas, *Decidability of local fields and their extensions*, Proceedings of the PLS13 (2021), obtainable from http://panhellenic-logic-symposium.org/13/proceedings-2021.pdf.

[22] Jochen Koenigsmann, *Undecidability in Number Theory*, arXiv:1309.0441 [math.NT], (2013).

[23] ———, *Defining* $\mathbb{Z}$ *in* $\mathbb{Q}$, Annals of Mathematics, **183** (2016), 73-93.

[24] Janos Kollar, *Diophantine subsets of function fields of curves*, Algebra and Number Theory, **2-3** (2008), 299-311.

[25] Angus Macintyre, *The history of interactions between logic and number theory 2-4* https://www.youtube.com/watch?v=P7i-qmsXHFk

[26] Carlos Martinez-Ranero and Javier Utreras and Carlos R. Videla, *Undecidability of $Q^{(2)}$*, Proceedings of the . American Mathematical Society, **148** (2020), 961-964.

[27] Barry Mazur. *The topology of rational points*, Experiment. Math. **1-1**,(1992), 35 - 45.

[28] ———, *Questions about Number*, in the volume: New Directions in Mathematics.

[29] Alice Medvedev and Thomas Scanlon, *Invariant varieties for polynomial dynamical systems*, Annals of Mathematics (to appear).

[30] Laurent Moret-Bailly, *Elliptic curves and Hilbert's tenth problem for algebraic function fields over real and p-adic fields*, Journal fur die reine und Angewandte Mathematik, **587** (2005), 77-143.

[31] ———, *An extension of Greenberg's theorem to general valuation rings*, Manuscripta Mathematica, **139** (2012), 153-166.

[32] M. Ram Murty, *Hilbert's Tenth Problem: An Introduction to Logic, Number Theory, and Computability*, American Mathematical Society, 2019.

[33] Daniel Palacin and Rizos Sklinos, *On superstable expansions of free abelian groups*, Notre Dame Journal of Formal Logic **59-2** (2018), 157-169.

[34] Hector Pasten, *Bivariate polynomial injections and elliptic curves*, Selecta Mathematica, **26-2** (2020), , 1-13.

[35] , ———, *Notes on the DPRM property for listable structures*, The journal of Symbolic Logic, **87-1** (2021), 273-312.

[36] ———, *Arithmetic derivatives through geometry of numbers*, Canadian Mathematical Bulletin. To appear (2021).

[37] Hector Pasten and Thanases Pheidas and Xavier Vidaux, *Uniform existential interpretation of arithmetic in rings of functions of positive characteristic*, Inventiones mathematicae **196-2** (2014), 453-484.

[38] Hector Pasten, Xavier Vidaux, *Positive existential definability of multiplication from addition and the range of a polynomial*, Israel Journal of Mathematics, **216** (2016), no. 1, 273-306.

[39] ———, *Extensions of Hilbert's Tenth Problem*, The Journal of Symbolic Logic **59-2** (1994), 372-397.

[40] ———, *An effort to prove that the existential theory of Q is undecidable*, Contemporary Mathematics 270 (2000), 237-252.

[41] ———, *Endomorphisms of elliptic curves and undecidability in function fields of positive characteristic*, Journal of Algebra, **273-1** (2004), 395-411.

[42] Thanases Pheidas and Xavier Vidaux *Hilbert's tenth problem for complex meromorphic functions in several variables with a place*, International Mathematics Research Notices (to appear) [arXiv:1711.09412].

[43] Thanases Pheidas and Karim Zahidi, *Undecidability of existential theories of rings and fields: A survey*, Contemporary Mathematics, **270** (2000), 49-106.

[44] ———, *Analogues of Hilbert's tenth problem*, Model theory with Applications to Algebra and Analysis Vol. 2 (Eds. Zoe Chatzidakis, Dugald Macpherson, Anand Pillay, Alex Wilkie), London Math Soc. Lecture Note Series Nr 250, Cambridge Univ Press, 2008.

Obtainable at https://www.newton.ac.uk/files/seminar/20050406100011001-149003.pdf

[45] Francoise Point and Michel Rigo and Laurent Waxweiler *Defining Multiplication in Some Additive Expansions of Polynomial Rings*, Communications in Algebra, **44:5** (2016), 2075-2099.

[46] Bjorn Poonen, *Using Elliptic Curves of Rank One towards the Undecidability of Hilbert's Tenth Problem over Rings of Algebraic Integers*, Proceedings of the 5th International Symposium on Algorithmic Number Theory (2002), 33-41.

[47] ———, *Hilbert's Tenth Problem over rings of number-theoretic interest*, obtainable from https://math.mit.edu/~poonen/papers/aws2003.pdf.

[48] ———, *Sums of values of a rational function*, Acta Arithmetica **112** (2004), 333-343.

[49] ———, *The set of nonsquares in a number field is diophantine*, Math. Res. Lett. **16-1** (2009), 165-170.

[50] Bjorn Poonen, *Undecidable problems: A sampler*, obtainable from `http://www-math.mit.edu/~poonen/papers/sampler.pdf`

[51] Alexandra Shlapentokh, *Diophantine relations between rings of S-integers of fields of algebraic functions in one variable over constant fields of positive characteristic*, The Journal of Symbolic Logic **58-1** (1993), 158–192

[52] ———, *Hilbert's Tenth Problem: Diophantine Classes and Extensions to Global Fields*, Cambridge University Press, (2007).

[53] ———, *Defining integers*, The Bulletin of Symbolic Logic, **17-2** (2011), 230-251.

[54] Alla Sirokofskich, *On an exponential predicate in polynomials over finite fields*, Proceedings of the American Mathematical Society, **138-7** 2010), 2569-2583, .

[55] ——— *Decidability questions for a ring of Laurent polynomials*, Annals of Pure and Applied Logic **163-5** (2012), 615–619.

[56] Xavier Vidaux, *An analogue of Hilbert's 10th problem for fields of meromorphic functions over non-Archimedean valued fields*, Journal of Number Theory, **101-1** (2003), 48-73.

[57] Andre Weil, *Number theory: an approach through history; from Hammurapi to Legendre*, Modern Birkhauser Classics Series (1984).

Decidability of local fields and their extensions

Konstantinos Kartas

University of Oxford, Mathematical Institute, Woodstock Road, Oxford
kartas@maths.ox.ac.uk

Abstract

We provide a survey of classical decidability results for local fields and then present some new results for various infinite extensions of local fields which are of arithmetic interest.

1 Introduction

The decidability of the p-adic numbers $\mathbb{Q}_p$, established by Ax-Kochen [AK65] and Ershov [Ers65], still remains one of the highlights of model theory. It motivated several decidability results both in mixed and positive characteristic:

- In mixed characteristic, Kochen [Koc74] showed that $\mathbb{Q}_p^{ur}$, the maximal unramified extension of $\mathbb{Q}_p$, is decidable. More generally, by work of [Zie72], [Ers65], [Bas78], [Bél99] and more recently [AJ19], [Lee20] and [LL21], we have a good understanding of the model theory of unramified and finitely ramified mixed characteristic henselian fields.

- In positive characteristic, our understanding is much more limited. Nevertheless, by work of Denef-Schoutens [DS03], we know that $\mathbb{F}_p((t))$ is *existentially* decidable in $L_t = \{+, \cdot, -, 0, 1, t\}$, modulo resolution of singularities. In fact, Theorem 4.3 [DS03] applies to show that any finitely ramified extension of $\mathbb{F}_p((t))$ is existentially decidable relative to its residue field.

Note that all of the above results are restricted to *finitely* ramified extensions of $\mathbb{Q}_p$ and $\mathbb{F}_p((t))$. The situation is less clear for *infinitely* ramified fields and there are many such algebraic extensions of $\mathbb{Q}_p$, whose decidability problem is still open. This is the content of the following sections. Our results are divided into two categories, the wildly ramified extensions and the tamely ramified extensions.

2 Wildly ramified extensions

Recall the definition:

Definition 2.0.1. *A finite extension $(L, w)/(K, v)$ of valued fields is said to be wildly ramified if the ramification degree $e(L/K)$ is p-divisible, where p is the residue characteristic of (K, v). An algebraic extension is said to be wildly ramified if any finite subextension is wildly ramified.*

In practice, one refers to wildly ramified extensions when the ramification degree is highly p-divisible. Important wildly ramified extensions of $\mathbb{Q}_p$ include:

Example. *(a) $\mathbb{Q}_p^{ab}$, the maximal abelian extension of $\mathbb{Q}_p$.*
(b) $\mathbb{Q}_p(\zeta_{p^\infty})$, the totally ramified extension obtained by adjoining all p^n-th roots of unity.

These extensions have been discussed in Macintyre's survey on pg.140 [Mac86] and a conjectural axiomatization of $\mathbb{Q}_p^{ab}$ was given by Koenigsmann on pg.55 in [Koe18]. Another interesting extension is $\mathbb{Q}_p(p^{1/p^\infty})$, a totally ramified extension of $\mathbb{Q}_p$ obtained by adjoining a compatible system of p-power roots of p.
The p-adic completions of the above fields are typical examples of perfectoid fields (see [Sch12]). For any such field K, one can define its tilt, which intuitively is its *local* function field analogue and serves as a characteristic p approximation of K. For our fields of interest, one has that $\mathbb{Q}_p(p^{1/p^\infty})$ and $\mathbb{Q}_p(\zeta_{p^\infty})$ are approximated by $\mathbb{F}_p((t))^{1/p^\infty}$, the perfect hull of $\mathbb{F}_p((t))$, while $\mathbb{Q}_p^{ab}$ is approximated by $\bar{\mathbb{F}}_p((t))^{1/p^\infty}$, the perfect hull of $\bar{\mathbb{F}}_p((t))$. The fields $\mathbb{F}_p((t))^{1/p^\infty}$ and $\bar{\mathbb{F}}_p((t))^{1/p^\infty}$ are typical examples of wildly ramified extensions of $\mathbb{F}_p((t))$.
In [Kar20], the following is established:

Theorem A (Corollary A [Kar20])**.** *(a) Assume $\mathbb{F}_p((t))^{1/p^\infty}$ is decidable (resp. $\exists$-decidable) in L_t. Then $\mathbb{Q}_p(p^{1/p^\infty})$ and $\mathbb{Q}_p(\zeta_{p^\infty})$ are decidable (resp. $\exists$-decidable) in L_{val}.*
(b) Assume $\bar{\mathbb{F}}_p((t))^{1/p^\infty}$ is decidable (resp. $\exists$-decidable) in L_t. Then $\mathbb{Q}_p^{ab}$ is decidable (resp. $\exists$-decidable) in L_{val}.

In the above result, the language L_t is the language of valued fields together with a constant symbol for t. This is essentially a special case of the main result of [Kar20], which is a relative decidability result for perfectoid fields. The proof uses Fontaine's period rings, which are relevant in the construction of the Fargues-Fontaine curve.
One may also prove the following unconditional decidability result:

Theorem B. *There is an algorithm that decides whether a system of polynomial equations and inequations, defined over $\mathbb{Z}$, has a solution modulo p over each of the valuation rings of $\mathbb{Q}_p(p^{1/p^\infty}), \mathbb{Q}_p(\zeta_{p^\infty})$ and $\mathbb{Q}_p^{ab}$.*

The proof of Theorem B goes via reduction to characteristic p, but unlike Theorem A only existential decidability in L_{val} is needed on the characteristic p side. The latter is known by work of Anscombe-Fehm [AF16].
Relative decidability results in the reverse direction are also established in [Kar20]. For example:

Proposition. *If $\mathbb{Q}_p(p^{1/p^\infty})$ is $\forall^1\exists$-decidable in L_{val}, then $\mathbb{F}_p[[t]]^{1/p^\infty}$ is $\exists^+$-decidable in L_t.*

The above Proposition is not exactly a converse of the existential version of Theorem A but still suggests that if we eventually want to understand the theories of $\mathbb{Q}_p(p^{1/p^\infty})$, $\mathbb{Q}_p(\zeta_{p^\infty})$ and $\mathbb{Q}_p^{ab}$ (even modest parts of their theories), we have to face the diophantine problem over the perfect hull of $\mathbb{F}_p((t))$ and $\bar{\mathbb{F}}_p((t))$.

3 Tamely ramified extensions

We now discuss some new results for tamely ramified extensions that are established in [Kar21], where details and proofs may be found. Theorem C below is a general *existential* Ax-Kochen-Ershov principle for tamely ramified fields, with *no restriction* on the characteristic, but which is conditional on a certain form of resolution of singularities.
For our model-theoretic purposes, we need to extend the usual notion of a tamely ramified field extension to the context of transcendental valued field extensions:

Definition 3.0.1. *A valued field extension $(L,w)/(K,v)$ is said to be tamely ramified if l/k is separable*[1]*, the quotient group Δ/Γ has no p-torsion, where $p = char(k)$, and every finite subextension is defectless.*

Example. *(a) Every valued field extension is tamely ramified when the residue characteristic is zero.*
(b) The valued field extension $(\mathbb{Q}_p(p^{1/n}),v_p)/(\mathbb{Q},v_p)$ is tamely ramified if and only if $p \nmid n$.
(c) Let $\mathbb{F}_p((t^\Gamma))$ be the Hahn series field with residue field $\mathbb{F}_p$ and value group Γ. The valued field extension $\mathbb{F}_p((t^\Gamma))/\mathbb{F}_p(t)$ is tamely ramified if and only if 1 is not p-divisible in Γ.

Our results in this section depend on a certain form of resolution of singularities. In very simple terms, resolution of singularities allows us to transform a given variety, which may have lots of singularities, to one which is non-singular. Moreover, the latter variety is in some sense close to the former, so that anything useful that can be said about the latter variety can often be translated into something useful about the former. The advantage of resolving the singularities of a variety lies in the fact that it is usually much easier to deal with non-singular varieties for all sorts of problems.
We now state the precise form that is assumed in [Kar21]:

Conjecture R (Log-Resolution)**.** *Let X be a reduced, flat scheme of finite type over an excellent discrete valuation ring R. Then there exists a blow-up morphism $f : \tilde{X} \to X$ in a nowhere dense center $Z \subset X$ such that*

1. *$\tilde{X}$ is a regular scheme.*
2. *$\tilde{X}_s = \tilde{X} \times_{SpecR} Spec(R/\mathfrak{m}_R)$ is a strict normal crossings divisor.*

The notion of an excellent ring, introduced by Grothendieck(see §7.9 [Gro65]), is quite technical to define here. However, for the case of discrete valuation rings, this simply means that $\widehat{K}/K$ is a separable (not necessarily algebraic), where $K = Frac(R)$ and $\widehat{K}$ denotes the completion of K. A divisor is said to be strict normal crossings if its reduced underlying scheme locally looks like a union of smooth varieties crossing transversely. In [Kar21], the following general existential Ax-Kochen-Ershov is obtained:

Theorem C (Theorem A [Kar21])**.** *Assume Conjecture R. Suppose (K,v) and (L,w) are henselian and tamely ramified over a discrete valued field (F,v_0) with $\mathcal{O}_F$ excellent. If $RV(K) \equiv_{\exists,RV(F)} RV(L)$, then $K \equiv_{\exists,F} L$ in L_r.*

Theorem C specializes to well-known Ax-Kochen-Ershov results in residue characteristic 0 and in the mixed characteristic *unramified* setting. Moreover, these Ax-Kochen-Ershov principles are known not only for the existential theories but also for the full-first order theories. The case of *finite* tame ramification in mixed characteristic and with perfect residue fields was proven recently in Corollary 5.9 [Lee20].
At the same time, Theorem C implies conditional existential decidability results for $\mathbb{F}_p((t))$ and its finite extensions, which were already known by the work of Denef-Schoutens [DS03]. Our proof does not use Greenberg's approximation theorem, which is an essential ingredient in [DS03].
On the other hand, Theorem C applies also to the setting of *infinite* ramification, providing us with an abundance of examples of *infinitely* ramified extensions of $\mathbb{Q}_p$ and $\mathbb{F}_p((t))$ whose theory is existentially decidable. This is the content of the next section.

[1] A field extension l/k (not necessarily algebraic) is said to be separable if l is linearly disjoint from k^{1/p^∞} (see §2.6 [FJ04]).

3.1 Decidability

In Remark 7.6 [AF16], the authors write:
"*At present, we do not know of an example of a mixed characteristic henselian valued field* (K, v) *for which* k *and* (Γ, vp) *are* $\exists$*-decidable but* (K, v) *is* $\exists$*-undecidable.*"
The existence of such an example is proved in Observation 1.2.2 [Kar20]. However, if we restrict ourselves to the tamely ramified setting, we indeed get such an Ax-Kochen style statement:

Corollary (Mixed characteristic)**.** *Assume Conjecture R. Suppose* (K, v) *and* (L, w) *are henselian and tamely ramified over* $(\mathbb{Q}, v_p)$*, admitting cross-sections that extend a given cross-section of* $(\mathbb{Q}, v_p)$*. If* $k \equiv_\exists l$ *in* L_r *and* $(\Gamma, vp) \equiv_\exists (\Delta, wp)$ *in* L_{oag}*, then* $K \equiv_\exists L$ *in* L_r*.*

In particular, if (K, v) is henselian and tamely ramified over $(\mathbb{Q}, v_p)$, admitting a cross-section extending one of $(\mathbb{Q}, v_p)$, then K is existentially decidable in L_r relative to k in L_r and (Γ, vp) in L_{oag} (see Corollary 4.1.4 [Kar21]). Similarly, we obtain a positive characteristic analogue:

Corollary (Positive characteristic)**.** *Assume Conjecture R. Suppose* (K, v) *and* (L, w) *are henselian and tamely ramified over* $(\mathbb{F}_p(t), v_t)$*, admitting cross-sections that extend a given cross-section of* $(\mathbb{F}_p(t), v_t)$*. If* $k \equiv_\exists l$ *in* L_r *and* $(\Gamma, vt) \equiv_\exists (\Delta, wt)$ *in* L_{oag}*, then* $K \equiv_\exists L$ *in* L_t*.*

Among the fields that are existentially decidable, the maximal tamely ramified extensions of $\mathbb{Q}_p$ and $\mathbb{F}_p((t))$ are of arithmetic significance.

Corollary. *Assume Conjecture R. Then the fields* $\mathbb{Q}_p^{tr}$ *and* $\mathbb{F}_p((t))^{tr}$ *are existentially decidable in* L_r*.*

3.2 Tweaking Abhyankar's example

Finally, we discuss a tame variant of the following famous example, essentially due to Abhyankar [Abh56]. It is also presented by Kuhlmann in a model-theoretic context in Example 3.13 [Kuh11]:

Example. *Let* $(K, v) = (\mathbb{F}_p((t))^{1/p^\infty}, v_t)$ *and* $(L, w) = (\mathbb{F}_p((t^{1/p^\infty})), v_t)$ *be the Hahn series field with value group* $\frac{1}{p^\infty}\mathbb{Z}$ *and residue field* $\mathbb{F}_p$*. We observe that* $RV(K) \cong_{RV(\mathbb{F}_p((t)))} RV(L)$ *but* $(K, v) \not\equiv_{\exists, \mathbb{F}_p((t))} (L, w)$ *since the Artin-Schreier equation* $x^p - x - \frac{1}{t} = 0$ *has a solution in* L *but not in* K*.*

Our version of Abhyankar's example is obtained by replacing p-power roots of t with l-power roots and exhibits a totally different behaviour:

Example. *Fix any prime* $l \neq p$*. Consider the valued fields* $(K, v) = (\mathbb{F}_p((t))(t^{1/l^\infty}), v_t)$ *and* $(L, w) = (\mathbb{F}_p((t^{1/l^\infty})), v_t)$*, with the latter being the Hahn series field with value group* $\frac{1}{l^\infty}\mathbb{Z}$ *and residue field* $\mathbb{F}_p$*. We observe that* $RV(K) \cong RV(L)$ *and by Theorem C we get that* $(K, v) \equiv_{\mathbb{F}_p((t^{1/l^n})), \exists} (L, w)$*, for all* $n \in \mathbb{N}$*. It follows that* $\mathbb{F}_p((t))(t^{1/l^\infty}) \prec_\exists \mathbb{F}_p((t^{1/l^\infty}))$ *in* L_r*.*

Acknowledgements

I would like to thank E. Hrushovski and J. Koenigsmann for invaluable guidance. Also many thanks to T. Scanlon for his hospitality while I was a visitor in UC Berkeley, where part of this work was completed.

References

[Abh56] Shreeram Abhyankar. Two notes on formal power series. *Proceedings of the American Mathematical Society Vol. 7, No. 5, pp. 903-905*, 1956.

[AF16] Sylvy Anscombe and Arno Fehm. The existential theory of equicharacteristic henselian valued fields. *Algebra & Number Theory Volume 10, Number 3 (2016), 665-683.*, 2016.

[AJ19] Sylvy Anscombe and Franziska Jahnke. The model theory of Cohen rings. *arXiv https://arxiv.org/abs/1904.08297*, (2019).

[AK65] James Ax and Simon Kochen. Diophantine problems over local fields II. *Amer. J. Math. 87*, 1965.

[Bas78] Serban A. Basarab. Some model theoery for henselian valued fields. *Journal of Algebra*, 1978.

[Bél99] Luc Bélair. Types dans les corps valués munis d'applications coefficients. *Illinois J. Math. 43(2):410- 425*, 1999.

[DS03] Jan Denef and Hans Schoutens. On the decidability of the existential theory of $\mathbb{F}_p[[t]]$. *Fields Institute Communications Volume 00, 0000*, (2003).

[Ers65] Ju.L. Ershov. On elementary theories of local fields. *Algebra i Logika 4, No. 2, 5-30*, 1965.

[FJ04] Michael D. Fried and Moshe Jarden. *Field Arithmetic*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. 11 (2nd revised and enlarged ed.). Springer-Verlag, 2004.

[Gro65] Alexander Grothendieck. *Éléments de géométrie algébrique : IV. Étude locale des schémas et des morphismes de schémas, Seconde partie*. Publications mathématiques de l' I.H.É.S., tome 24, p. 5-231, 1965.

[Kar20] Konstantinos Kartas. Decidability via the tilting correspondence. *arXiv https://arxiv.org/abs/2001.04424*, (2020).

[Kar21] Konstantinos Kartas. Diophantine problems over tamely ramified fields. *arXiv https://arxiv.org/abs/2103.14646*, (2021).

[Koc74] Simon Kochen. The model theory of local fields. *ISILC Logic Conference (Proc. Internat. Summer Inst. and Logic Colloq., Kiel, 1974), pp. 384-425. Lecture Notes in Math., Vol. 499, Springer, Berlin, 1975.*, 1974.

[Koe18] Jochen Koenigsmann. Decidability in local and global fields. *Proc. Int. Cong. of Math. Rio de Janeiro, Vol. 1 (45-60)*, 2018.

[Kuh11] Franz-Viktor Kuhlmann. The Defect. *Commutative algebra - Noetherian and non-Noetherian perspectives, 277-318, Springer, New York*, 2011.

[Lee20] Junguk Lee. Hyperfields, truncated DVRs, and valued fields. *Journal of Number Theory, Volume 212, July 2020, Pages 40-71*, 2020.

[LL21] Junguk Lee and Wan Lee. On the structure of certain valued fields. *Annals of Pure and Applied Logic Volume 172, Issue 4*, 2021.

[Mac86] Angus Macintyre. Twenty years of p-adic model theory. *Logic Colloquium '84 Elsevier Science Publishers B. K (North-Holland)*, 1986.

[Sch12] Peter Scholze. Perfectoid spaces. *Publ. Math. Inst. Hautes Études Sei. 116*, November 2012.

[Zie72] Martin Ziegler. *Die elementare Theorie der henselschen Körper*. Dissertation thesis, 1972.

A Triple Uniqueness of the Maximum Entropy Approach

Jürgen Landes*

Munich Center for Mathematical Philosophy
LMU Munich Munich Germany
juergen_landes@yahoo.de

Abstract

Inductive logic is concerned with assigning probabilities to sentences given probabilistic constraints. The maximum entropy approach to inductive logic assigns probabilities to all sentences of a first order predicate logic. This assignment is built on an application of the Maximum Entropy Principle. This paper puts forward two different modified applications of this principle and shows that the original and the modified applications agree in many cases. A third promising modification is studied and rejected.

1 Introduction

Inductive logic is a formal approach to model uncertain inferences. It seeks to analyse the degree to which premisses entail putative conclusions. Given uncertain premisses $\varphi_1, \dots, \varphi_k$ with attached uncertainties $X_1, \dots, X_k$ an inductive logic provides means to attach an uncertainty Y to a conclusion ψ, where the X_i and Y are non-empty subsets of the unit interval. Using $\mid\!\approx$ to denote an inductive entailment relation this can be represented as

$$\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \mid\!\approx \psi^Y \quad ,$$

where $\mid\!\approx$ denotes an inductive entailment relation [4].

The main early proponent of inductive logic was Rudolf Carnap [2]. Nowadays, the spirit of his approach today continues in the Pure Inductive Logic approach [7, 8, 14]. In this paper, I however consider uncertain inference within the Maximum Entropy Principle, which goes back to Edwin Jaynes [5]. Roughly speaking, the Maximum Entropy Principle compels rational agents to use a probability function consistent with the evidence for drawing uncertain inferences. In case there is more than one such probability function, a rational agent ought to use one of those probability functions that has maximal entropy.

If the underlying domain is finite, then applying the Maximum Entropy Principle for inductive entailment is straight-forward and well-understood due to the work of Alena Vencovská & Jeff Paris [11, 12, 13]. Matters change dramatically for infinite domains. Naively replacing the sum by an integral in the definition of Shannon Entropy produces probability functions with infinite entropy. But then there is no way to pick a probability function with maximal entropy out of a set in which all functions have infinite entropy.

There are two different suggestions for inductive logic on an infinite first order predicate logic explicating the Maximum Entropy Principle. The entropy limit approach [1] precedes the maximum entropy approach [17, 18]. It has been conjectured, that both approaches agree in cases in which the former approach is-well defined [18, p. 191]. This conjecture has been shown to hold in a number of cases of evidence bases with relatively low quantifier-complexity [6, 9, 16].

This paper introduces modifications of the maximum entropy approach and studies their relationships. I next properly introduce this approach, the modifications and investigate their relationships.

*Many thanks to Soroush Rafiee Rad and Jon Williamson for continued collaboration on these topics.

2 The Maximum Entropy Approach and Modifications

The formal framework and notation is adapted from [9].

Given is a fixed first-order predicate language $\mathcal{L}$ with countably many constant symbols $t_1, t_2, \ldots$ and finitely many relation symbols, $U_1, \ldots, U_n$. The atomic sentences are sentences of the form $U_i t_{i_1} \ldots t_{i_k}$, where k is the arity of the relation U_i, will be denoted by $a_1, a_2, \ldots$, ordered in such a way that atomic sentences involving only constants among $t_1, \ldots, t_n$ occur before those atomic sentences that also involve t_{n+1}. The set of sentences of $\mathcal{L}$ is denoted by $S\mathcal{L}$.

The finite sublanguages $\mathcal{L}_n$ of $\mathcal{L}$ are those languages, which only contain the first n constant symbols $t_1, \ldots, t_n$ and the same relation symbols as $\mathcal{L}$. The sentences of the form $\pm a_1 \wedge \ldots \wedge \pm a_{r_n}$ are called the *n-states*. Let Ω_n be the set of n-states for each n. Denote the sentences of $\mathcal{L}_n$ by $S\mathcal{L}_n$.

Definition 1. *A* probability function *P on $\mathcal{L}$ is a function $P : S\mathcal{L} \longrightarrow \mathbb{R}_{\geq 0}$ such that:*

P1*: If τ is a tautology, i.e., $\models \tau$, then $P(\tau) = 1$.*

P2*: If θ and φ are mutually exclusive, i.e., $\models \neg(\theta \wedge \varphi)$, then $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$.*

P3*: $P\left(\exists x \theta(x)\right) = \sup_m P\left(\bigvee_{i=1}^{m} \theta(t_i)\right)$.*

A probability function on $\mathcal{L}_n$ is defined similarly (the supremum in P3 is dropped and m is equal to n). $\mathbb{P}$ denotes the set of all probability functions on $\mathcal{L}$. The set of probability functions consistent with all premisses is denoted by $\mathbb{E}$, $\mathbb{E} := \{P \in \mathbb{P} \ : \ P(\varphi_i) \in X_i \text{ for all } 1 \leq i \leq k\}$.

A probability function $P \in \mathbb{P}$ is determined by the values it gives to the quantifier-free sentences, a result known as *Gaifman's Theorem* [3]. Consequently, a probability function is determined by the values it gives to the n-states, for each n. It is thus sensible to measure entropy of P via n-states with varying n.

Definition 2 (n-entropy)**.** *The* n-entropy *of a probability function P is defined as:*

$$H_n(P) := -\sum_{\omega \in \Omega_n} P(\omega) \log P(\omega) \ .$$

The usual conventions are $0 \log 0 := 0$ and log *denoting the natural logarithm. The second convention is inconsequential for current purposes. $H_n(\cdot)$ is a strictly concave function.*

The key idea is to combine the n-entropies defined on finite sublanguages into an overall notion of comparative entropy comparing probability functions P and Q defined on the entire first order language. So far, the literature has only studied such inductive logics with respect to the first binary relation in the following definition.

Definition 3 (Comparative Notions of Entropy)**.** *That a probability function $P \in \mathbb{P}$* has greater (or equal) entropy than *a probability function $Q \in \mathbb{P}$ could be defined in the following three ways.*

1. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$.*
2. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) \geq H_n(Q)$ and there are infinitely many n such that $H_n(P) > H_n(Q)$, denoted by $P]Q$.*

3. *If and only if there is some natural number N such that for all $n \geq N$ it holds that $H_n(P) \geq H_n(Q)$, denoted by $P)Q$.*

The lower two definitions are alternative ways in which one could explicate the intuitive idea of comparative entropy.

Definition 4 (Maximum Entropy)**.** *The set of probability functions on $\mathcal{L}$ with maximal entropy relative to a notion of comparative entropy $>$ defined on $\mathbb{P}^2$ can then be defined as*

$$\operatorname{maxent}_> \mathbb{E} := \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q > P\} \ . \tag{1}$$

Definition 5 (Maximum Entropy Inductive Logics)**.** *An* inductive logic *with respect to $>$ is induced by attaching uncertainty $Y_>(\psi) \subseteq [0,1]$ to the sentences ψ of $\mathcal{L}$ via*

$$Y_>(\psi) := \{r \in [0,1] \mid \text{ there exists } P \in \operatorname{maxent}_> \mathbb{E} \text{ with } P(\psi) = r\} \ .$$

In case there are two or more different probability functions in $\operatorname{maxent}_> \mathbb{E}$, there are some sentences of ψ of $\mathcal{L}$ to which multiple different probabilities attach.

In the next section, I study (the relationships of) these binary relations and the arising inductive logics. Particular attention is paid to the case of a unique probability function for inference, $|\operatorname{maxent}_> \mathbb{E}| = 1$.

3 Maximal (Modified) Entropy

I first consider notions of refinement relating these three binary relations.

Definition 6 (Strong Refinement)**.** *$>$ is called a* strong refinement *of $\gg$, if and only if the following hold*

- *$>$ is a refinement of $\gg$, for all $P, Q \in \mathbb{P}$ it holds that $P \gg Q$ entails $P > Q$,*
- *for all $R, P, Q \in \mathbb{P}$ it holds that, if $R \gg P$ and $P > Q$, then $R \gg Q$ and $R \neq Q$.*

Definition 7 (Centric Refinement)**.** *I call a refinement $>$ of $\gg$* centric*, if and only if for all different $R, P \in \mathbb{P}$ with $R > P$ it holds that $(R + P)/2 \gg P$.*

Clearly, not all binary relations possess strong refinements; not all binary relations possess centric refinements.

Proposition 1. *$]$ is a strong and centric refinement of $\succ$. $)$ is a strong and centric refinement of $]$ and of $\succ$.*

Proof. I now display the three notions of comparative entropy line by line. The second conjunct in the first definition is superfluous as is the second conjunct in the third definition:

$$\begin{aligned} P \succ Q &:\Longleftrightarrow (H_n(P) \leq H_n(Q) \text{ not infinitely often } \& \ H_n(P) > H_n(Q) \text{ infinitely often}) \\ P]Q &:\Longleftrightarrow (H_n(P) < H_n(Q) \text{ not infinitely often } \& \ H_n(P) > H_n(Q) \text{ infinitely often}) \\ P)Q &:\Longleftrightarrow (H_n(P) < H_n(Q) \text{ not infinitely often } \& \ H_n(P) \geq H_n(Q) \text{ infinitely often}) \ . \end{aligned}$$

By thusly spelling out both comparative notions of entropy one observes that $P \succ Q$ entails $P]Q$, and that $P]Q$ entails $P)Q$. This establishes the refinement relationships.

Strong Refinements Next note that, if $R \succ Q$ or if $R]Q$, then $R \neq Q$.

$]$ is a strong refinement of $\succ$: Let $R \succ P$ and $P]Q$. Then $R \neq Q$. Furthermore, $H_n(R) \leq H_n(Q)$ is true for at most finitely many n, since from some N onwards P has always greater or equal n-entropy than Q. So, $R \succ Q$.

$)$ is a strong refinement of $]$: Let $R]P$ and $P)Q$. Then $R \neq Q$. From some N onwards P has always greater or equal n-entropy than Q. There are also infinitely many $n \in \mathbb{N}$ such that $H_n(R) > H_n(P)$. So, $R]Q$.

$)$ is a strong refinement of $\succ$: Let $R \gg P$ and $P)Q$. Then $R \neq Q$. From some N onwards P has always greater or equal n-entropy than Q. From some N' onwards R has always greater n-entropy than P. Hence, $H_n(R) \leq H_n(Q)$ can only be the case for finitely many $n \in \mathbb{N}$. So, $R \succ Q$.

Centric Refinement First, note that different probability functions disagree on some quantifier free sentence $\varphi \in L_N$ (*Gaifman's Theorem* [3]). Since $\varphi \in L_{n+N}$ for all $n \geq 1$, these probability functions also disagree on all more expressive sub-languages L_{n+N}.

$]$ is a centric refinement of $\succ$: Fix arbitrary probability functions R, P defined on $\mathcal{L}$ with $R]P$. $R \neq P$. From the concavity of the function H_n it follows that $H_n(\frac{R+P}{2}) > H_n(P)$, whenever $H_n(R) \geq H_n(P)$. By definition of $]$, there are only finitely many n for which $H_n(R) \geq H_n(P)$ fails to hold. Hence, $\frac{R+P}{2} \succ P$ by definition of $\succ$.

$)$ is a centric refinement of $\succ$: Fix arbitrary probability functions R, P defined on $\mathcal{L}$ with $R)P$. Note that R may be equal to P. From the concavity of the function H_n it follows that $H_n(\frac{R+P}{2}) > H_n(P)$, whenever $H_n(R) \geq H_n(P)$. By definition of $)$, there are only finitely many n for which $H_n(R) \geq H_n(P)$ fails to hold. Hence, $\frac{R+P}{2} \succ P$ by definition of $\succ$.

$)$ is a centric refinement of $]$: Fix arbitrary probability functions R, P defined on $\mathcal{L}$ with $R)P$. Note that R may be equal to P. Since $\frac{R+P}{2} \succ P$ (see above case) and since $]$ is a refinement of $\succ$, it holds that $\frac{R+P}{2}]P$. □

Remark 1 (Properties of Comparative Entropies)**.** *If $H_n(P) = H_n(Q)$ for all even n and $H_n(P) > H_n(Q)$ for all odd n, then $P]Q$ and $P \not\succ Q$. Hence, $]$ is a proper refinement of $\succ$.*

For $P = Q$ it holds that $P)Q$ and $Q)P$. Hence, $)$ is a proper refinement of $]$ and thus a proper refinement of $\succ$.

$]$ is transitive, irreflexive, acyclic and asymmetric. $)$ is transitive, reflexive and has non-trivial cycles, e.g, for all probability functions P, Q with zero-entropy, $H_n(P) = 0$ for all $n \in \mathbb{N}$, it holds that $P)Q$.

I now turn to entropy maximisation and the induced inductive logics.

Proposition 2. *Let $>$ be a strong refinement of $\gg$. If $\{Q\} = \operatorname{maxent}_{\gg} \mathbb{E}$, then $\{Q\} = \operatorname{maxent}_{\gg} \mathbb{E} = \operatorname{maxent}_{>} \mathbb{E}$.*

Proof. Note at first that since $>$ is a refinement of $\gg$ it holds that

$$\operatorname{maxent}_{>} \mathbb{E} \subseteq \operatorname{maxent}_{\gg} \mathbb{E} \ . \tag{2}$$

Maximal elements according to $\gg$ may not be maximal according to $>$ and all maximal elements according to $>$ are also maximal according to $\gg$.

Assume for the purpose of deriving a contradiction that $Q \notin \operatorname{maxent}_{>} \mathbb{E}$. Then, there has to exist a $P \in \mathbb{E} \setminus \{Q\}$ such that $P > Q$ but $P \gg Q$ fails to hold ($\{Q\} = \operatorname{maxent}_{\gg} \mathbb{E}$).

However, since $\{Q\} = \operatorname{maxent}_{\gg} \mathbb{E}$ and $Q \notin \operatorname{maxent}_{>} \mathbb{E}$ hold, there has to exist some $R \in \mathbb{E} \setminus \{P\}$ such that $R \gg P$, P cannot have maximal $\gg$-entropy. We hence have $R \gg P$ and $P > Q$. Since $>$ is a strong refinement of $\gg$, we obtain $R \gg Q$ and $R \neq Q$. Since $R \in \mathbb{E}$ it follows from the definition of $\operatorname{maxent}_{\gg}$ that $Q \notin \operatorname{maxent}_{\gg} \mathbb{E}$. Contradiction. So, $Q \in \operatorname{maxent}_{>} \mathbb{E}$.

Since $\{Q\} = \text{maxent}_{\gg}\,\mathbb{E} \overset{(2)}{\supseteq} \text{maxent}_{>}\,\mathbb{E} \ni Q$, it follows that $\text{maxent}_{>}\,\mathbb{E} = \{Q\}$. □

Proposition 3. *If $\mathbb{E}$ is convex, $>$ is a centric refinement of $\gg$ and $\{Q\} = \text{maxent}_{>}\,\mathbb{E}$, then $\{Q\} = \text{maxent}_{\gg}\,\mathbb{E} = \text{maxent}_{>}\,\mathbb{E}$.*

Proof. Assume for contradiction that there exists a $P \in \mathbb{E}\setminus\{Q\}$ such that P is not $\gg$-dominated by the probability functions in $\mathbb{E}$ but $>$-dominated by some $R \in \mathbb{E} \setminus \{P\}$, $R > P$. Now define $S = \frac{1}{2}(P + R)$ and note that $S \in \mathbb{E}$ (convexity) and that S, P, R are pairwise different, $|\{S, P, R\}| = 3$.

Since $>$ is a centric refinement of $\gg$, conclude that $S \gg P$, which contradicts that $P \in \text{maxent}_{\gg}\,\mathbb{E}$ and $P \neq Q$. So, only Q can be in $\text{maxent}_{\gg}\,\mathbb{E}$.

Since $Q \in \text{maxent}_{>}\,\mathbb{E}$ and $\text{maxent}_{>}\,\mathbb{E} \overset{(2)}{\subseteq} \text{maxent}_{\gg}\,\mathbb{E}$ it follows that $\{Q\} = \text{maxent}_{\gg}\,\mathbb{E}$. □

Theorem 1 (Triple Uniqueness)**.** *If $\mathbb{E}$ is convex and at least one of $\text{maxent}_{)}\,\mathbb{E}$, $\text{maxent}_{]}\,\mathbb{E}$ or $\text{maxent}_{\succ}\,\mathbb{E}$ is a singleton, then*

$$\text{maxent}\,\mathbb{E}_{)} = \text{maxent}_{]}\,\mathbb{E} = \text{maxent}_{\succ}\,\mathbb{E}\ .$$

Proof. Simply apply the above three propositions. □

Having studied refinements of $\succ$, I now briefly consider how $\succ$ could refine a binary relation. Closest to the spirit of Definition 3 would be to consider $P\}Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$. Clearly, the other three notions of comparative entropy are refinements of $\}$.

Neither of these three binary relations is a strong refinement and neither is a centric refinement. To see this, consider three pairwise different probability functions P, Q, R with i) $H_n(P) > H_n(Q)$ for all n , ii) $H_n(P)/H_n(Q) \approx 1$, iii) $H_1(Q) = H_1(R) - \delta$ for large $\delta > 0$ and iv) $H_n(Q) > H_n(R)$ for all $n \geq 1$. Then $P\}Q$ and $Q \succ R, Q]R, Q)R$. Now note that $H_1(P) < H_1(R)$ and thus $P\}R$ fails to hold. None of $\succ,],)$ is a strong refinement of $\}$. Finally, observe that $\frac{Q+R}{2}\}R$ fails to hold. None of $\succ,],)$ is a centric refinement of $\}$.

The binary relation $\}$ induces a different inductive logic than $\succ,],)$:

Example 1. *Let U be the only and unary relation symbol of $\mathcal{L}$. Suppose there is no evidence, $\mathbb{E} = \mathbb{P}$. Then every $P \in \mathbb{P}$ with $P(Ut_1) = P(\neg Ut_1) = 0.5$ has maximal 1-entropy. Hence, all such P are members of $\text{maxent}_{\}}\,\mathbb{E}$. For $\square \in \{\succ,],)\}$ it holds that $\text{maxent}_{\square}\,\mathbb{E} = P_{=}$, where $P_{=}$ denotes the equivocator function, which for all n assigns all n-states the same probability of $1/|\Omega_n|$. So, $\text{maxent}_{\square}\,\mathbb{E} \neq \text{maxent}_{\}}\,\mathbb{E}$.*

This leads to the following more general observation:

Proposition 4. *If there exists an $n \in \mathbb{N}$ such $H_n(P) = \max\{H_n(Q) : Q \in \mathbb{E}\}$, then $P \in \text{maxent}_{\}}\,\mathbb{E}$.*

This strong focus on single sublanguages $\mathcal{L}_n$ makes $\text{maxent}_{\}}$ unsuitable as an inductive logic for infinite predicate languages.

4 Conclusions

Maximum entropy inductive logic on infinite domains lacks a paradigm approach. The entropy limit approach, the maximum entropy approach as well as the here studied modified maximum entropy approaches induce a unique inductive logic in a number of natural cases. This points

towards a, perhaps surprisingly, unified picture of maximum entropy inductive logics – in spite of the number possible ways to define such inductive logics.

The Maximum Entropy Approach fails to provide probabilities for uncertain inference for certain evidence bases of quantifier complexity Σ_2 [15, § 2.2]. In these cases, for all $P \in \mathbb{E}$ there exists a $Q \in \mathbb{E}$ such that $Q \succ P$ and maxent $\mathbb{E}$ is hence empty [10]. One way to sensibly define an inductive logic could be to consider a binary relation which is refined by $\succ$. Unfortunately, the most obvious way fails to deliver a sensible inductive logic (Proposition 4). Finding a way to define such a sensible inductive logic must be left to further study.

References

[1] Owen Barnett and Jeff B. Paris. Maximum Entropy Inference with Quantified Knowledge. *Logic Journal of IGPL*, 16(1):85–98, 2008.

[2] Rudolf Carnap. The Two Concepts of Probability: The Problem of Probability. *Philosophy and Phenomenological Research*, 5(4):513–532, 1945.

[3] Haim Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics*, 2(1):1–18, 1964.

[4] Rolf Haenni, Jan-Willem Romeijn, Gregory Wheeler, and Jon Williamson. *Probabilistic Argumentation*, volume 350 of *Synthese Library*. Springer, 2011.

[5] Edwin T Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, 2003.

[6] Jürgen Landes. The Entropy-limit (Conjecture) for Σ_2-Premisses. *Studia Logica*, 109:423–442, 2021.

[7] Jürgen Landes, Jeff B. Paris, and Alena Vencovská. Some Aspects of Polyadic Inductive Logic. *Studia Logica*, 90(1):3–16, 2008.

[8] Jürgen Landes, Jeff B. Paris, and Alena Vencovská. A survey of some recent results on Spectrum Exchangeability in Polyadic Inductive Logic. *Synthese*, 181:19–47, 2011.

[9] Jürgen Landes, Soroush Rafiee Rad, and Jon Williamson. Towards the Entropy-Limit Conjecture. *Annals of Pure and Applied Logic*, 172, 2021.

[10] Landes, Jürgen and Rafiee Rad, Soroush and Williamson, Jon. Determining maximal entropy functions for objective Bayesian inductive logic. Forthcoming.

[11] Jeff B. Paris. Common Sense and Maximum Entropy. *Synthese*, 117:75–93, 1998.

[12] Jeff B. Paris. *The Uncertain Reasoner's Companion: A Mathematical Perspective*, volume 39 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, 2 edition, 2006.

[13] Jeff B. Paris and Alena Vencovská. In Defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17(1):77–103, 1997.

[14] Jeff B. Paris and Alena Vencovská. *Pure Inductive Logic*. Cambridge University Press, 2015.

[15] Soroush Rafiee Rad. Equivocation axiom on first order languages. *Studia Logica*, 105(1):121–152, 2017.

[16] Soroush Rafiee Rad. Maximum Entropy Models for Σ_1 Sentences. *Journal of Applied Logics - IfCoLoG Journal of Logics and their Applications*, 5(1):287–300, 2018.

[17] Jon Williamson. Objective Bayesianism with predicate languages. *Synthese*, 163(3):341–356, 2008.

[18] Jon Williamson. *Lectures on Inductive Logic*. Oxford University Press, Oxford, 2017.

Decidability of the theory of addition and the Frobenius map in certain rings of rational functions

Dimitra Chompitaki[1], Manos Kamarianakis[1]*, and Thanases Pheidas[1]

Department of Mathematics & Applied Mathematics, University of Crete, Heraklion, Greece

Abstract

Let p be a prime number, $\mathbb{F}_p$ a finite field with p elements, z a variable, $\mathbb{F}_p[z]$ the ring of polynomials in z with coefficients in $\mathbb{F}_p$ and $\mathbb{F}_p(z)$ the field of rational functions of z over $\mathbb{F}_p$. We consider the existential theory of addition and the Frobenius map of a ring $R \subset \mathbb{F}_p(z)$, where R is generated over $\mathbb{F}_p[z]$ by inverting finitely many irreducible polynomials of $\mathbb{F}_p[z]$. We prove that it is model-complete and hence decidable. We also prove that, if the existential theory, in the same language, of $\mathbb{F}_p(z)$ is decidable, then its first-order theory is also decidable.

1 Introduction

Our work is part of the on-going research revolving around the fact that the ring-theory and even the existential ring-theory of any field of rational functions $\mathbb{F}_p(z)$ over a finite field $\mathbb{F}_p$ with p elements is undecidable (see [Phe91] and [Vid94]). So, proving decidability of structures weaker than the ring-structure of such a field and its subrings is desirable (cf. [PZ00]).

Consider R as a structure (model) of the language $\mathcal{L} := \{+, =, x \mapsto x^p, x \mapsto zx, 0, 1\}$, where $=$ and $+$ denote regular equality and addition, $x \mapsto zx$ denotes the multiplication-by-z map and $x \mapsto x^p$ is the *Frobenius map*. In [PZ04], the authors proved that the $\mathcal{L}$-theory of $\mathbb{F}_p[z]$ is model-complete (meaning that every formula is equivalent in $\mathbb{F}_p[z]$ to an existential $\mathcal{L}$-formula), and, hence, decidable. A similar (model completeness) result has been proved for the $\mathcal{L}$-structure of the ring of power series $\mathbb{F}_p[[z]]$ of z (see [Ona18]). It is a natural question to ask whether the $\mathcal{L}$-theory of $\mathbb{F}_p(z)$ is model-complete. For the moment, this problem seems inapproachable with current means and may demand the use of novel tools. Those that we use here suffice to prove model completeness for subrings of $\mathbb{F}_p(z)$, generated over $\mathbb{F}_p[z]$ by the inverses of finitely many irreducible polynomials.

The structure of addition and the Frobenius map is interesting, not only for its own sake, but also because it is connected to various important mathematical and logical domains and problems. For example, the derivative of a function (polynomial, rational or power series) is positive-existentially definable in $\mathcal{L}$ (see [PZ04]). So, the structure of $\mathbb{F}_p(z)$ as a model of addition and differentiation is encodable in its $\mathcal{L}$-structure. It is also interesting to study this structure with p as a parameter, cf., the open *Gröthendieck-Katz conjecture* [Ber91].

In another direction, it is a long-standing (and famous) problem whether there is *resolution of singularities* of algebraic varieties in positive characteristic. In the zero characteristic case, it has been proved to always exist, in the algebraic and the analytic sense, by Hironaka [Hir64]. But in positive characteristic it is an open problem. Although it has been thought, for a long time, that such resolution always exists (no counter-example is known), all efforts to prove it have failed so far and, recently, experts have expressed doubts. From the investigations so far, it seems that the main obstacle in characteristic $p > 0$ is the existence of the Frobenius map .

*Corresponding Author, Email: kamarianakis@uoc.gr

2 Our results

2.1 The main theorems

The existential $\mathcal{L}$-theory of the ring $R \subset \mathbb{F}_p(z)$, where R is generated over $\mathbb{F}_p[z]$ by inverting finitely many irreducible polynomials of $\mathbb{F}_p[z]$, is the set of *existential sentences* of $\mathcal{L}$, i.e., the first-order sentences of the language $\mathcal{L}$ which are of the form $\exists x_1 \dots \exists x_m \in R : \phi$, where ϕ is a boolean combination of equations that may be written in the language $\mathcal{L}$.

We construct an algorithm which, given an existential formula σ of $\mathcal{L}$, finds an equivalent universal formula, thus we prove model-completeness.

Theorem 2.1. *Let R be the subring of $\mathbb{F}_p(z)$ generated over $\mathbb{F}_p[z]$ by inverting finitely many irreducibles polynomials of $\mathbb{F}_p[z]$. The $\mathcal{L}$-theory of R is model-complete.*

For the case $R = \mathbb{F}_p(z)$, we observe that our structure is a module over the non-commutative ring $\mathbb{F}_p(z)[P]$, where P is defined by $Px := x^P$, but with constant symbols that are not contained in the language of modules that is used in the existing bibliography. We make a variant of the well-known theorem of Baur and Monk [Bau76, Mon75], using work of Van den Dries and Holly [VdDH92], and obtain the following theorem.

Theorem 2.2. *Assume that the existential $\mathcal{L}$-theory of $\mathbb{F}_p(z)$ is decidable. Then the $\mathcal{L}$-theory of $\mathbb{F}_p(z)$ is decidable.*

But the decidability (or not) of the existential $\mathcal{L}$-theory remains an open problem.

2.2 The main technical theorem

We present an outline of a new method that we introduced in order to prove Theorem 2.1.

Definition 2.3. *An* additive polynomial *is a polynomial of the form*

$$f(\bar{x}) = \sum_{i=1}^{n} f_i(x_i),$$

where $\bar{x} = (x_1, \dots, x_n)$ and, for each i,

$$f_i(x_i) := b_i x_i^{p^{s(i)}} + \sum_{j=1}^{s(i)-1} c_{i,j} x_i^{p^{s(i)-j}},$$

with $b_i, c_{i,j} \in \mathbb{F}_p(z)$.

For $s \in \mathbb{N}$, let $\mathcal{V}_s$ be $\mathbb{F}_p(z)$, considered as a vector space over the field $\mathbb{F}_p(z^s)$. An additive polynomial f as above is called *normalized* if all degrees $s(i)$ are equal to some s and the set of leading coefficients $\{b_i \mid i = 1, \dots, n\}$ is linearly independent over $\mathcal{V}_s$.

Theorem 2.4. *Let f be a normalized additive polynomial of the variables $\bar{x} = (x_1, \dots, x_n)$, which has positive degree in all the variables. Let $u \in \mathbb{F}_p(z)$. Then the set $\{\bar{x} \in R^n \mid f(\bar{x}) = u\}$ is finite.*

The method of proof involves diverse tools, such as the *Hasse derivative* D_i [Has36], where i denotes the order of derivation. This "hyperderivative" is used to create a matrix operator W which generalizes the concept of the Wronskian operator in characteristic zero.

Note that Theorem 2.4 is indicative of the usefulness of our methodology; the inverse image of a rational function through a multi-variate polynomial is, in general, infinite. We also show that Theorem 2.4 is not true if one replaces R by $\mathbb{F}_p(z)$ - and this shows the limits of our method: We can construct a normalized additive polynomial f and choose a $u \in \mathbb{F}_p(z)$ such that the set $\{\bar{x} \in \mathbb{F}_p(z)^n \mid f(\bar{x}) = u\}$ is infinite.

Acknowledgments

This research work was supported from Greek and European Union resources, through the National Strategic Reference Framework (NSRF 2014-2020), under the call "Support for researchers with emphasis on young researchers (EDBM103)" and the funded project "Problems of Diophantine Nature in Logic and Number Theory" with code MIS 5048407.

References

[Bau76] Walter Baur. Elimination of quantifiers of modules. *Israel J. Math.*, 25:64–70, 1976.

[Ber91] Daniel Bertrand. Groupes algébriques et équations différentielles linéaires. *Astérisque*, 206:183–204, 1991.

[Has36] Helmut Hasse. Theorie der höheren Differentiale in einem algebraischen Funktionenkörper mit vollkommenem Konstantenkörper bei beliebiger Charakteristik. *J. Reine. Ang. Math.*, 1936(175):50–54, January 1936.

[Hir64] Heisuke Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero. I. *Ann. of Math.*, 79(1):109–203, 1964.

[Mon75] Leonard Gaines Monk. Elementary-recursive decision procedures. *PhD Dissertation, Berkeley*, 1975.

[Ona18] Gönenç Onay. $\mathbf{F}_p((\mathrm{X}))$ is decidable as a module over the ring of additive polynomials. *arXiv.org*, June 2018.

[Phe91] Thanases Pheidas. Hilbert's tenth problem for fields of rational functions over finite fields. *Invent. Math.*, 103(1):1–8, 1991.

[PZ00] Thanases Pheidas and Karim Zahidi. Undecidability of existential theories of rings and fields: a survey. *Contemp. Math.*, 270:49–106, 2000.

[PZ04] Thanases Pheidas and Karim Zahidi. Elimination theory for addition and the Frobenius map in polynomial rings. *J. Symb. Log.*, 69(4):1006–1026, 2004.

[VdDH92] Lou Van den Dries and Jan Holly. Quantifier elimination for modules with scalar variables. *Ann. Pure Appl. Logic*, 57:161–179, 1992.

[Vid94] Carlos Videla. Hilbert's tenth problem for rational function fields in characteristic 2. *Proc. Amer. Math. Soc.*, 120(1):249–253, 1994.

The V-logic Multiverse and MAXIMIZE

Matteo de Ceglie

Universität Salzburg
Salzburg, Österreich
decegliematteo@gmail.com

Abstract

In this paper, I show that $ZFC + LCs$ is restrictive compared to the V-logic multiverse, characterised as $ZFC + LCs$+multiverse axioms. This means showing that the V-logic multiverse proves the existence of an extra object that it is unavailable in $ZFC + LCs$ and that, in turn, this object realises a new isomorphism type. I argue that such an object is a class-iterable sharp, that can only be found if there are proper, uncountable, width extensions of V. Such extensions are present in the V-logic multiverse, but not in classical set theory.

1 Introduction

Classical set theory (ZFC), as instantiated by the cumulative hierarchy V, has been a very successful foundation of mathematics for over a century. Nonetheless, there are some problems with it. The first, and foremost, problem is Gödel's Incompleteness: in the context of set theory, it entails that some set theoretic statements are *independent* from ZFC. This means that we cannot prove neither that they are true, nor that they are false. The main example of such statements is the Continuum Hypothesis (CH), as proved by Cohen in the 1960s using forcing. For some time it was thought that adding new axioms to ZFC would solve this problem (Gödel's program), with large cardinals axioms (LCs) being the main candidate for addition. However, it was proved that they do not settle CH and, moreover, that they are incompatible with the Axiom of Constructibility.[1] In the end, the impossibility of solving independent questions by the simple addition of new axioms, coupled with the multiplication of incompatible models generated through forcing, led to several possible expansions of ZFC, each one giving rise to interesting mathematics. But which one should be chosen as the new foundations of mathematics? The set theoretic multiverse (as introduced by J. D. Hamkins (2012)) side-steps the question: there is no need to choose, we can integrate all these different set theories in just one *multiverse conception*. Such a solution is not appealing to the advocates of *universism*, that instead defend the idea that there is only one set theoretic universe, V, that contains all the possible sets and cannot be further expanded. Moreover, they point out that everything that can be done in the multiverse can actually already be done in the Single Universe. For example, they argue that the main tool used to generated new universes, i.e. forcing, can be interpreted as taking place entirely within the Single Universe.[2] The only possible argument against this kind of universism and in favour of multiversism entails proving that in the set theoretic multiverse we can some object that we cannot have in the Single Universe at all.

In this paper, I argue exactly that. In particular, I contend that classical set theory, $ZFC(+LCs)$, is *restrictive* compared to the V-logic multiverse (a novel set theoretic multiverse conception developed by the present author and Claudio Ternullo). This multiverse conception is based upon Friedman's Hyperuniverse[3] and Steel's set-generic multiverse[4]: like the Hyperuniverse, it uses the infinitary V-logic as background logic (this logic admits formulas of length less than the first successor of the least inaccessible cardinal, but only a finite block of quantifiers in front of them) and admits all kinds of outer models of V (produced by set-generic, class-generic, hyperclass forcing, etc.). Like

[1] The Axiom of Constructibility says that $V = L$, i.e. that all sets are constructible from simpler ones.

[2] Here and throughout the paper I will refer to models of set theory and universes interchangeably, as done in the literature.

[3] See S. Friedman (2016).

[4] See Steel (2014).

Steel's set-generic multiverse, it is recursively axiomatisable and is rooted on a ground universe that satisfies ZFC. For this proof, I compare $ZFC + LCs$ and the V-logic multiverse, characterised as $ZFC + LCs+$ the multiverse axioms, following Maddy's methodological principle MAXIMIZE (as introduced in Maddy (1997)). According to this principle, when choosing between two theories T and S we should prefer the one that can prove more isomorphism types. I claim that the V-logic multiverse, as opposed to $ZFC + LCs$, does exactly that. This is because in the V-logic multiverse theory we can prove the existence of proper, uncountable, extensions of V, that we cannot have in $ZFC + LCs$ (see Neil Barton (2019)). In turn, this extra object means we can realise more isomorphism types, since in the V-logic multiverse we can prove the existence of iterable class sharps and, more importantly, maps between them (see Antos, N. Barton, and S.-D. Friedman (nd)). Moreover, when moving from $ZFC + LCs$ to the V-logic Multiverse we are not losing anything: ZFC, all the large cardinals, inner models and V are still there. On the other hand, when moving from the V-logic multiverse to $ZFC + LCs$ we lose the actual outer models of V, iterable class sharps and iterable class sharp generated models. Thus, this latter theory is restrictive compared to the V-logic multiverse theory.

This paper is structured as follows. First, I describe the infinitary V-logic and the V-logic multiverse (section 2). After that, I show that classical set theory is restrictive compared to the V-logic multiverse (section 3). Finally, some concluding remarks sketching the road ahead end the paper (section 4).

2 The V-logic Multiverse

I now proceed to mathematically describe the V-logic multiverse. The system to be adopted allows to address universes arising from extending V in width and height.[5] More specifically, it is able to:

1. code representations of the "canonical" relationship between V and its outer models;
2. incorporate all kinds of outer-model constructs (e.g. extensions produced by various kinds of forcing);
3. formulate what one should easily acknowledge as multiverse axioms.

This multiverse conception is, philosophically, a refinement of Hamkins' broad multiverse. The key difference is that, instead of admitting all the possible universes, without any hierarchy (as done in Hamkins' multiverse), the V-logic Multiverse only admits the universes that can be defined and described in a certain, uniform way. While this is philosophical starting point aims at restricting Hamkins' philosophical conception (goal shared with other multiverses), the mathematical implementation ends up being more open.[6] In order to satisfy these requirements, I adopt the infinitary V-logic, i.e. a logic whose language $\mathcal{L}_{\kappa^+,\omega}$ is that of first-order logic, admitting formulas of length less than κ^+ (the first successor of the least inaccessible cardinal) and with a finite number (less than ω) of quantifiers in front, and supplemented with the membership relation symbol $\in$ and the following constant symbols:

- $\bar{a}$, one for each $a \in V$;
- $\bar{V}$, denoting the *ground universe* (that is, our initial V).

Proofs in V-logic are infinitary, because of the addition of the following inference rules::

Set-rule $\{\varphi(\bar{b})|b \in a\} \vdash (\forall x \in \bar{a})\varphi(x)$

[5] An *height* extension of V is produced by adding new ordinals, while a *width* extension by adding new subsets.

[6] Hamkins's multiverse, as implemented by the axioms introduced in Gitman and Joel David Hamkins (2011), is composed by only the countable computably saturated models of ZFC, while the V-logic multiverse axioms admits any kind of model.

V-rule $\{\varphi(\bar{a})|a \in V\} \vdash (\forall x)\varphi(x)$

A sentence of V-logic may also use additional symbols.[7] For example, a case of special interest is when $\bar{W}$ is introduced as a new predicate symbol (variable) ranging over "generic *outer* models of V", and one considers sentences of the form "$\bar{W} \models ZFC + \psi$" for some sentence ψ, possibly containing constants $\bar{a}$ for $a \in V$. The following fact is fundamental for my purposes:

Fact 1 (Barwise)**.** *If $\bar{V}$ is countable then a theory T of $\mathcal{L}_V$ has a $\bar{W}$-structure for a model iff T is consistent in V-logic.*

This means that, in V-logic, one can produce a sentence about *any* outer model W of V which, if consistent in V-logic, then *really* expresses a property of an outer model W of V.[8] As Barwise has shown, it turns out that structures satisfying the axioms of $\mathfrak{M}$-logic, that is, $\mathfrak{M}$-structures (V-structures, in the case of V-logic), are models of Kripke-Platek set theory (KP), a weak fragment of ZFC.[9] In turn, models of KP are called "admissible sets" (as these models are related to admissible ordinals[10] in recursion theory). The least such model, which contains $\mathfrak{M}$ as an object, is called $\mathfrak{M}^+$.[11] If we turn to consider V-logic, a V-structure is the *least admissible set* beyond V, that is the least model of KP containing V as an object, which is called V^+. In V^+, we finally have codes for proofs in V-logic, which allows one to express *syntactic* facts that are essential for axiomatising the V-logic multiverse.

I start by considering a set-theoretic sentence φ, which expresses, in V-logic, that "$\bar{W} \models \psi$". Now, consider the theory $T = ZFC +$ "$\bar{W} \models \psi$" and let Con(T) be the statement "T is consistent" (in V-logic).[12] Then, by Fact 1 above, if $Con(T)$ holds, then $\bar{W}$ really is an outer model W of V enjoying the property ψ. This outer model, identified by Con(T), is, thus, a member of the V-logic multiverse. The procedure may be generalised to *all* kinds of ψ and all kinds of outer models W,[13] which leads to the formulation of the first, and key, "multiverse axiom" of the new theory $MZFC$ – more precisely, an *axiom schema*:

Axiom 1 (Multiverse Axiom Schema)**.** *For any first-order φ with parameters from V, if the sentence of V-logic expressing "$\bar{W}$ is an outer model of V satisfying φ" is consistent in V-logic, then there is a universe W which is an outer model of V that satisfies φ.*[14]

In addition to this axiom schema, and in analogy with ZFC, $MZFC$ also features axioms describing how the *sets* and *universes* of the multiverse behave.[15] To this end, $MZFC$ contains all of ZFC, together with the following axiom:

Axiom 2 (Core Axiom)**.** *Every universe of the multiverse models ZFC.*

Thus, $MZFC$, as of now, consists of:

1. ZFC;
2. the *Multiverse Axiom Schema*;
3. additional V-logic "axioms" (the "V-rule" and the "set rule");

[7] The general features of infinitary logics, and their relationships with admissible sets (structures) are discussed in, among other works, Keisler (1974), Barwise (1975), and Dickmann (1975).

[8] This was proved in Barwise (1975).

[9] KP is the theory which results from removing the Power-Set and Infinity Axioms from ZF, and admitting restricted forms of the Separation Axiom (Δ_0-Separation) and of the Replacement Axiom (Δ_0-Replacement).

[10] An ordinal α is admissible iff the corresponding constructible universe L_α is a transitive model of KP.

[11] Barwise's original notation is $Hyp(\mathfrak{M})$, but to avoid confusion, the alternative notation $\mathfrak{M}^+$ is preferable.

[12] Technically, Con(T) is the V-logic statement: "$T \nvdash_V \varphi \wedge \neg\varphi$", where '$\nvdash_V$' is the V-logic provability relation.

[13] For instance, Con(T) above may further specify that $\bar{W}$ contains a filter $\bar{G}^C \subseteq \mathbb{P}^C$, where $\mathbb{P}^C$ is a class-poset, which would mean that $\bar{W}$ is a class-forcing extension of V.

[14] This also means that each φ consistent in V-logic has a model in the multiverse. Note that the consistency of such a φ is Π_1-expressible in a first-order way, not over V but over V^+.

[15] Of course, there are distinct variables for sets and universes in our multiverse theory.

4. the *Core Axiom*.

Although the multiverse axioms clearly describe semantic constructs, what we have at this stage is just a collection of theories. In particular, if we take an incremental approach to $MZFC$, we may informally view it as a tree made up of *branches* corresponding to alternative set-theoretic statements, and of *nodes* where alternative V-logic theories extending ZFC appear. Thus, the V-logic multiverse may be seen as *the collection of all the combinatorially conceivable consistent V-logic theories of outer models*.

Note that the V-logic multiverse maximises over outer-model constructs, as no constraint upon the nature of the outer models has been placed in the formulation of the Multiverse Axiom Schema. I argue that this represents a significant improvement over the set-generic multiverse conception, which only allows for *certain kinds* of outer models.[16]

3 $ZFC+LCs$ is restrictive compared to the V-logic multiverse

In this section I present the result that I have hinted to in the Introduction: classical set theory (ZFC) is *restrictive* compared to the V-logic Multiverse. To do so I use the methodological principle MAXIMIZE as discussed by Maddy (1997). This principle states that when comparing two theories, the one that proves more isomorphism types is preferable. The theory that proves more isomorphism types *maximizes* over the other (or, equivalently, the theory that proves less isomorphism types is *restrictive* compared to the other). Maddy (1997) applies this principle to argue against the addition of $V = L$ to ZFC, and I plan to apply the same line of reasoning to the V-logic Multiverse.

The argument consists of the following steps:

1. first of all, prove that one theory proves the existence of an extra object that cannot exists in the (claimed) restrictive one;
2. prove that this extra object realises a new isomorphism type;
3. if the two above steps are done, then we can conclude that one theory *maximizes* (in Maddy's sense) over the other (or, equivalently, that one theory is *restrictive* over the other).

I contend that this is true in the case of the V-logic multiverse and classical set theory.

First of all I need to precise the terms of this comparison. On the one hand, for the Single Universe framework I am taking classical set theory in its usual axiomatization ZFC plus the addition of large cardinals axioms, as instantiated by the cumulative hierarchy V. According to universists, this, together with the restriction of set-generic forcing to countable transitive models, is enough for set theoretic practice.[17] On the other hand, the V-logic Multiverse is characterised as $ZFC+LCs+$ the Multiverse Axioms. Note that, as usually argued by the universist, the addition of the Multiverse Axioms do not add any "real" power to $ZFC+LCs$, since everything we need is already in the latter theory, at least according to universists.

We can now proceed to the first step of my argument, i.e. showing that the V-logic multiverse can prove the existence of an extra object that it is unavailable in $ZFC + LCs$. This object is a proper, uncountable, outer model of V. Such an object cannot exists in the universist's framework of $ZFC + LCs$: indeed, the application of forcing in that usual setting is done only to countable transitive models.[18] This is because to do it we need the existence of generic filters, and for the universist there are no V-generic filters.

However, in the V-logic multiverse framework we can prove the following theorem:

Theorem 1. *Let φ be a V-logic sentence (for instance, a sentence which says "$Con(T)$" for some V-logic theory T). The following are equivalent:*

[16] In particular, models obtained through *set-forcing*.
[17] This point is argued by N. Barton (2019).
[18] See Nik (2014), Antos, N. Barton, and S.-D. Friedman (nd) and N. Barton (2019).

1. *φ is consistent in V-logic.*

2. *φ is consistent in v-logic (this is essentially the V-logic build upon a transitive countable model v, instead of the full uncountable V).*

3. *$\bar{V}$ has an outer model, $\bar{W}$, such that $\bar{W} \models \varphi$.*

4. *There exists a $\bar{W}^*$, elementarily equivalent to $\bar{W}$, such that $\bar{W}^* \models \varphi$.*[19]

This theorem implies that, in the V-logic multiverse, even if we start with a countable model of ZFC inside V, we can then end up with a proper, uncountable outer model of an *uncountable* V[20]

Consequently we have, in the V-logic Multiverse, an object that cannot be found in the universist's framework. We now need to prove that this new object realises a new isomorphism type. And this is exactly my claim.

To see this, consider the technique of #-generation.[21] As stated by Antos, Barton and Friedman, this method is very useful in encapsulating several large cardinals consequences of reflection properties.[22] It is based upon the existence of *class-iterable sharps*: these are transitive structures that are amenable (i.e. $x \cap U \in N$ for any $x \in N$), with a normal measure and iterable in the sense that all successive ultrapower iterations along class well-orders are well-founded.[23] If such an object exists, then we can have *class iterated sharp generated* models, i.e. models that arise through collecting together each level indexed by the largest cardinal of the model that result from the iteration of a class-iterable sharp.[24] Finally, we can claim that V is such class iterably sharp generated, and enjoy all advantages of this fact (the main advantage is that any satisfaction obtainable in height extensions of V adding ordinals is already reflected to an initial segment of V itself). However, in V we cannot find such a class-iterable sharp, since, if it were the case, then we would be able to prove the existence of a cardinal that is both regular *and* singular[25], but this is impossible.[26] So in the classical set theoretic framework V is not a class iterably sharp generated model, and all of the above is unattainable.

This situation is fundamentally different in the V-logic multiverse. Indeed, since in the V-logic multiverse we can have proper, uncountable, extensions of V, we can also have, in these extensions, a class-iterable sharp. And thus, in the V-logic multiverse, we can claim that V is, in fact, class iterably sharp generated. This result opens a new realm of isomorphisms types between all the various iterated ultrapowers, and models of different heights that are provided by #-generation. Thus, we can claim that $ZFC + LCs$ is restrictive compared to the V-logic multiverse, since in the latter we can find a new object that realises a new isomorphism type.

4 Concluding remarks

I have shown that the V-logic multiverse, characterised as $ZFC+LCs+$ Multiverse Axioms, and with V-logic as the background logic, proves more isomorphism types than classical set theory ($ZFC + LCs$), and thus we can say that classical set theory is *restrictive* compared to the V-logic multiverse.

[19] This theorem has been proved by the present author and Claudio Ternullo in the paper *Outer Models, V-logic and the Multiverse*, currently in preparation, and based on related results from Antos, N. Barton, and S.-D. Friedman (nd) and N. Barton (2019).

[20] The V-logic multiverse is not the only multiverse conception that claims the existence of proper outer models of V, the other being the Hyperuniverse. However, the latter assume that V is countable, thus simplifying the setting by a lot.

[21] See Antos, N. Barton, and S.-D. Friedman (nd) for a discussion of it.

[22] A reflection property is a property of a model that can be proved to be true already in an initial segment of that model.

[23] Here I am following the definition from Antos, N. Barton, and S.-D. Friedman (nd). The original definition in S. Friedman (2016) is slightly different.

[24] Again, the precise definition can be found in Antos, N. Barton, and S.-D. Friedman (nd).

[25] A regular cardinal is a cardinal which cofinality is equal to the cardinal itself, otherwise it is singular.

[26] See Antos, N. Barton, and S.-D. Friedman (nd) for the details.

The argument I presented is compelling, but it is only one step of a much wider research program. Other than the already mentioned UNIFY principle, it must be noted that my argument uses an intuitive definition of restrictiveness and isomorphism type that can both be refined. This can be done by following first and foremost the definitions present in Maddy (1997), and then the subsequent work done by Benedikt Löwe, Luca Incurvati and especially Albert Visser.[27]

In conclusion, showing that the V-logic multiverse is better than classical set theory concerning the principle MAXIMIZE is the first, necessary step to a better understanding of the set theoretic multiverse and the requirements for a good foundational framework for mathematics.

References

Antos, C., N. Barton, and S.-D. Friedman (nd). "Universism and Extensions of V". Pre-print.

Barton, N. (2019). "Forcing and the Universe of Sets: Must We Lose Insight?" In: *Journal of Philosophical Logic*. Forthcoming.

Barton, Neil (2019). "Forcing and the Universe of Sets: Must we lose insight?" In: *Journal of Philosophical Logic*, pp. 1–38.

Barwise, Jon (1975). *Admissible Sets and Structures*. Springer Verlag, Berlin.

Dickmann, M. (1975). *Large Infinitary Languages*. North Holland, Amsterdam.

Friedman, S. (2016). "Evidence for Set-Theoretic Truth and the Hyperuniverse Programme". In: *IfCoLog Journal of Logics and their Applications* 3.4, pp. 517–555.

Gitman, Victoria and Joel David Hamkins (2011). "A natural model of the multiverse axioms". In: *arXiv preprint arXiv:1104.4450*.

Hamkins, J. D. (2012). "The Set-Theoretic Multiverse". In: *Review of Symbolic Logic* 5.3, pp. 416–449.

Incurvati, Luca and Benedikt Löwe (2016). "Restrictiveness relative to notions of interpretation". In: *The Review of Symbolic Logic* 9.2, pp. 238–250.

Keisler, H. J. (1974). *Model Theory for Infinitary Logic*. North Holland, Amsterdam.

Löwe, Benedikt (2001). "A first glance at non-restrictiveness". In: *Philosophia Mathematica* 9.3, pp. 347–354.

— (2003). "A Second Glance at Non-restrictiveness†." In: *Philosophia Mathematica* 11.3.

Maddy, P. (1997). *Naturalism in Mathematics*. Oxford University Press, Oxford.

Nik, Weaver (2014). *Forcing for mathematicians*. World Scientific.

Steel, J.R. (2014). "Gödel's Program". In: *Interpreting Gödel. Critical Essays*. Ed. by J. Kennedy. Cambridge University Press, Cambridge, pp. 153–179.

Visser, Albert et al. (2006). "Categories of theories and interpretations". In: *Logic in Tehran* 26, pp. 284–341.

[27] See Löwe (2001), Löwe (2003), Incurvati and Löwe (2016), and Visser et al. (2006).

A total Solovay reducibility and totalizing of the notion of speedability

Ivan Titov, Wolfgang Merkle

March 2021

Abstract

While the set of Martin-Löf random left-c.e. reals is equal to the maximum degree of Solovay reducibility, Miyabe, Nies and Stephan [5] have shown that the left-c.e. Schnorr random reals are not closed upwards under Solovay reducibility. Recall that for two left-c.e. reals α and β, the former is Solovay reducible to the latter in case there is a partial computable function f and constant c such that for all rational numbers $q < \alpha$ we have

$$\alpha - f(q) < c(\beta - q).$$

By requiring the translation function f to be total, we introduce a total version of Solovay reducibility that implies Schnorr reducibility.Accordingly, by Downey and Griffiths [1], the set of Schnorr random left-c.e. reals is closed upwards relative to total Solovay reducibility.

Furthermore, we observe that the the notion of speedability introduced by Merkle and Titov [4] can be equivalently characterized via partial computable translation functions in a way that resembles Solovay reducibility. By requiring the translation function to be total, we obtain the concept of total speedability. Like for speedability, this notion does not dependent on the choice of the speeding constant.

1 A total version of Solovay reducibility

We first review the usual definition of Solovay reducibility in terms of a partial recursive function [3].

Definition 1.1 (Solovay reducibility, $\leq_{S,c}$)**.** *Let α and β be reals and let $c > 0$ be a rational number. Then α is* Solovay reducible *to β* with respect to a constant *c, written $\alpha \leq_{S,c} \beta$, if there is a partial computable function $\varphi\colon \mathbb{Q} \to \mathbb{Q}$ such that for all $q < \beta$ it holds that $\varphi(q) \downarrow < \alpha$ and $\alpha - \varphi(q) < c(\beta - q)$. The real α is* Solovay reducible *to β, written $\alpha \leq_S \beta$, if α is Solovay reducible to β with respect to some c.*

In case $\alpha \leq_S \beta$, we will also say that α is S-reducible to β, and similarly notation will be used for other reducibilities introduced in what follows.

Definition 1.2 (total Solovay reducibility, $\leq^{tot}_{S,c}$)**.** *A real α is* total Solovay reducible *to a real β* with respect to a constant *c, written $\alpha \leq^{tot}_{S,c} \beta$, if there is a computable function $f\colon \mathbb{Q} \to \mathbb{Q}$ such that for all $q < \beta$ it holds that $f(q) < \alpha$ and $\alpha - f(q) < c(\beta - q)$. The real α is* total Solovay reducible *to β, written $\alpha \leq^{tot}_{S} \beta$, if α is total Solovay reducible to β with respect to some c.*

The total Solovay reducibility obviously implies the normal one, thus, the Martin-Löf random left c.e. reals are closed upwards relative to the total Solovay reducibility.

2 The structural properties of the $\leq_S^{tot}$ lattice of left-c.e. reals

In this section, we argue that total Solovay reducibility is in Σ_3^0 but is not a standard reducibility in the sense of Downey and Hirschfeldt [3] because neither is addition a join operator nor is there a least degree.

Proposition 2.1. *Total Solovay reducibility is in Σ_3^0.*

Proof. Let $\alpha^0, \alpha^1, ...$ be an effective enumeration of left-c.e. reals, where we can assume that for given n on can compute a recursive index for a nondecreasing approximation $a_0^n, a_1^n, \ldots$ to α^n from below. Then we have

$$\begin{aligned}\alpha^a \leq_S^{tot} \alpha^b :\iff\ & \exists\langle e,c\rangle\forall\langle q,s\rangle\exists\langle r,t\rangle\colon (\varphi_e(q)[t]\downarrow \\ & \wedge (q < b_s \implies (a_r - \varphi_e(q) > 0 \wedge a_r - \varphi_e(q) < c(b_s - q)))).\end{aligned}$$

□

Proposition 2.2. *Let α be a left-c.e. real and let $r > 0$ be a rational number. Then it holds that $r\alpha \equiv_S^{tot} \alpha$.*

Proof. It holds that $r\alpha \leq_S^{tot} \alpha$ via the identify function and constant r, and similarly for a reduction in the reverse direction with constant $1/r$.

□

Next we review the notion of a hyperimmune set.

Definition 2.3. *Let A be an infinite set. By p_A, we denote the principal function of A, i.e., the members of A are $p_A(0) < p_A(1) < \cdots$. Let $k_A(n)$ be the least member of $A \setminus \{0, \ldots, n-1\}$.*

Recall that a set A is HYPERIMMUNE if p_A is not majorized by a computable function, i.e., for no computable function g we have $p_A(n) \leq g(n)$ for all n.

Lemma 2.4. *For any set A, the following assertions are equivalent.*

(i) p_A is not majorized by any computable function

(ii) k_A is not majorized by any computable function

Proof. In case the computable function $g(n)$ majorizes $k_A(n)$, where we can assume that h in nondecreasing, then a computable function that majorizes $p_A(n)$ is given by

$$n \mapsto \underbrace{g(g(...(g(0)...))}_{n\text{-fold application of } g}$$

Conversely, in case the computable function $g(n)$ majorizes $p_A(n)$, then the function $n \mapsto g(n+1)$ majorizes $k_A(n)$. □

Proposition 2.5. *There exists no least degree in the total Solovay degrees.*

Every least set with respect to total Solovay reducibility is also a least set with respect to Solovay reducibility. Since the sets of the latter type are exactly the computable sets, the proposition is immediate from the following lemma.

Lemma 2.6. *Let $\alpha = 0.A(0)\ldots$ and $\beta = 0.B(0)\ldots$ be reals where the set A is computable and infinite. Then α is total Solovay reducible to β if and only if the set B is not hyperimmune.*

Proof. First assume that B is not hyperimmune. For a dyadic rational q that can be written as $q = 0.\sigma$ where the last letter of σ is equal to 1, define $|q| = |\sigma|$. Then for any such q and σ where $q < \beta$, we have

$$2^{-k_B(|q|)} \leq \beta - 0.\sigma = \beta - q.$$

By Lemma 2.4, we can fix a computable function g that majorizes k_B. We obtain a computable function f witnessing $\alpha \leq_S^{tot} \beta$ by choosing $f(q) < \alpha$ such that we have

$$\alpha - f(q) < 2^{-g(|q|)}.$$

Next assume that α is total Solovay reducible to β via some function f and constant c. Then for every n and for $q_n = 0.B(0)\ldots B(n-1)$, we have $q < \beta$, thus, for some appropriate constant d one holds that

$$g(n) := \min_{|\sigma_n|=n} \{\alpha - f(0.\sigma_n) : \alpha - f(0.\sigma_n) > 0\} \leq \alpha - f(q) < c(\beta - q) \leq 2^{-k_B(n)+d}.$$

Consequently, the function $n \mapsto d + \lceil \log g(n) \rceil$ is a computable upper bound for k_B, hence B is not hyperimmune. □

Indeed, the total Solovay-lattice satisfies the following stronger property, which we state here without proof.

Proposition 2.7. *There exists a countably infinite antichain of mutually $\leq_S^{tot}$-incomparable left-c.e. reals such that each of them is incomparable with every computable real.*

Before proving that addition is not a join operator, we recall the notion of a Schnorr reducibility, namely, the uniform version of it.

Definition 2.8 (UNIFORM SCHNORR REDUCIBILITY, $\leq_{uSch,c}$)**.** *A real α is* UNIFORM SCHNORR REDUCIBLE*, or uSch-*REDUCIBLE*, to a real β* WITH RESPECT TO A CONSTANT *c, written $\alpha \leq_{uSch,c} \beta$, if there is a computable functional φ that, given a description of a computable measure machine (or, shortly, cmm) B, returns a description of another computable measure machine $\varphi(B)$, so that*

$$K_{\varphi(B)}(\alpha \upharpoonright n) \leq K_B(\beta \upharpoonright n) + c.$$

The real α is UNIFORM SCHNORR REDUCIBLE *to β, written $\alpha \leq_{uSch} \beta$, if α is uniform Schnorr reducible to β with respect to some c.*

Obviously, the uniform Schnorr reducibility implies the Schnorr reducibility, with respect to which the Schnorr random reals are closed upwards.

Proposition 2.9. *For all left-c.e.* α, β, $\alpha \leq_S^{tot} \beta$ *implies* $\alpha \leq_{uSch} \beta$

Corollary 2.10. *The Schnorr random left-c.e. reals are closed upwards relative to the total Solovay reducibility.*

Proof. Let f be a total computable function, such that $\alpha \leq_{S,c}^{tot} \beta$ via f. Given a cmm machine B computing β, we construct a cmm machine A computing α in the following uniform way:

Input:$(x \in \mathbb{Q}, w \in \{0,1\}^{\lceil log(c)+1 \rceil})$

- compute $\sigma := B(x)$ (the computation halts iff $x \in dom(B)$)
- compute τ, so that $0.\tau := (f(0.\sigma) \upharpoonright n)$

  If $0.\sigma < \beta$, then on holds

  $$\alpha - f(0.\sigma) < c(\beta - 0.\sigma)$$

  In particular, if $0.\sigma = \beta \upharpoonright n$, then $\beta - 0.\sigma < 2^{-n}$, so $\alpha - f(0.\sigma) < c2^{-n} = 2^{\lceil log(c) \rceil - n}$.

  Thus,

  $$\alpha \upharpoonright n - 0.\tau < \alpha - f(0.\sigma) + f(0.\sigma) - (f(0.\sigma) \upharpoonright n) < c2^{-n} + 2^{-n} = 2^{\lceil log(c+1) \rceil - n}$$

- return $y \in \{0,1\}^n$, so that $0.y = 0.\tau + 2^{-n} \cdot 0, w$

The constructed machine A has the following properties:

- prefix-freeness (since B is prefix-free)
- computable measure of the domain (the following relation:

  $$B(x)\downarrow \implies A((x,w))\downarrow \forall w \in \{0,1\}^{\lceil log(c+1) \rceil}$$

  implies, that $\mu(dom(A)) = \mu(dom(B))$
- $K_A(\alpha \upharpoonright n) \leq K_B(\beta \upharpoonright n) + log(c+1) + O(1)$ (since there always exists a word $w \in \{0,1\}^{\lceil log(c+1) \rceil}$ such that

  $$\alpha \upharpoonright n - 0.\tau = 2^{-n} \cdot 0.w$$

  For that w, on holds $A(x,w) = y$, such that

  $$0.y = 0.\tau + 2^{-n} \cdot 0, w = \alpha \upharpoonright n$$

  that implies $K_A(\alpha \upharpoonright n) \leq |x| + |w|$, where x may be the shortest code of $\beta \upharpoonright n$.

□

Proposition 2.11. *There is a pair of left-c.e.reals* α, β *where* $\alpha \not\leq_S^{tot} \alpha + \beta$.

Proof. Miyabe, Nies and Stephan [5, Paragraph 3] demonstrated that there exists a pair of left-c.e. reals α and β such that $\alpha \not\leq_{Sch} \alpha + \beta$. Thus we also have $\alpha \not\leq_S^{tot} \alpha + \beta$ because total Solovay reducibility implies Schnorr reducibility. □

Remark. The uniform Schnorr-reducibility is, due to the similar argumentation, also implied by the weaken version of the total Solovay reducibility, whose requirement for f differs from the original one in the additional term:

$$\alpha - f(q) < c(\beta - q) + 2^{-|q|}.$$

The motivation of this weakening is that now its lattice on the field of left-c.e. reals has a minimal degree containing all the computable reals.

3 Speedability of left-c.e. numbers.

Definition 3.1. *A function $f\colon \mathbb{N} \to \mathbb{N}$ is a* SPEED-UP FUNCTION *if it is nondecreasing and $n \leq f(n)$ holds for all n. A left-c.e. number α is* ρ-SPEEDABLE WITH RESPECT TO ITS GIVEN LEFT APPROXIMATION *$a_0, a_1, \ldots \nearrow \alpha$ for some real number $\rho \in (0,1)$ if there is a computable speed-up function f such that we have*

$$\liminf_{n\to\infty} \frac{\alpha - a_{f(n)}}{\alpha - a_n} \leq \rho, \tag{1}$$

and SPEEDABLE *if it is ρ-speedable with respect to some its left-c.e. approximation for some $\rho \in (0,1)$. Otherwise we call α nonspeedable.*

Whether a real is speedable depends neither on the left-c.e. approximation nor on the constant ρ one considers.

Theorem 3.2 (Merkle and Titov [4])**.** *Every speedable left-c.e. real number is ρ-speedable for any $\rho > 0$ with respect to any of its left approximations.*

The following theorem is immediate from the main result of Barmpalias and Lewis-Pye [2].

Theorem 3.3 (Barmpalias and Lewis-Pye [2])**.** *Martin-Löf random left-c.e. real numbers are never speedable.*

By the following proposition, the notion of speedability can be equivalently characterized as a Solovay reduction of a real number to itself via a special partial computable functions on the rational numbers. By applying the same characterization to computable functions, in what follows we obtain a variant of speedability, similar to the introduction of total Solovay reducibility.

Proposition 3.4. *Let α be a left-c.e. real and let ρ be a real number such that $0 < \rho < 1$. Then α is speedable if and only if there is a partial computable function $g\colon \mathbb{Q} \to \mathbb{Q}$ that is defined and nondecreasing on the interval $(-\infty, \alpha)$, maps this interval to itself and satisfies*

$$\liminf_{q\nearrow\alpha} \frac{\alpha - g(q)}{\alpha - q} \leq \rho. \tag{2}$$

Proof. Fix some left approximation $a_0, a_1, \ldots$ of α. First assume that α is speedable. By Theorem 3.2 there is then a computable speed-up function f that witnesses that α is ρ-speedable with respect to its left approximation $a_0, a_1, \ldots$. Let n be the partial computable function on the set of rational numbers that maps every $q < \alpha$ to the least index i such that $q \leq a_i$, and is undefined for all other q. Here we assume that rational numbers are represented in a form such that equality is a computable predicate. Then the partial function g defined by

$$g(q) = a_{f(n(q))},$$

by choice of n and f, is partial computable, is defined and nondecreasing on the interval $(-\infty, \alpha)$ and maps this interval to itself. Furthermore, the sequence $a_0, a_1, \ldots$ witnesses that (2) holds, because we have $g(a_i) = a_{f(i)}$.

Next assume that there is a function g as stated in the proposition. Then there is a not necessarily computable left approximation $q_0, q_1, \ldots$ of α such that we have

$$\liminf_{j\to\infty} \frac{\alpha - g(q_j)}{\alpha - q_j} \leq \rho.$$

Let f be the computable speed up function that maps i to the least index $n > i$ such that $g(a_{i+1}) < a_n$. Then for all q and i such that q is an element of the half-open interval $[a_i, a_{i+1})$, we have

$$\frac{\alpha - a_{f(i)}}{\alpha - a_i} \leq \frac{\alpha - g(a_{i+1})}{\alpha - q} \leq \frac{\alpha - g(q)}{\alpha - q}.$$

In particular, this chain of inequalities holds true with q replaced by any of the q_j, which by choice of the q_j implies that α is ρ-speedable via its left approximation $a_0, a_1, \ldots$ and the speed-up function f. □

From Proposition 3.4 it is immediate that the the equivalent characterization of speedability stated there does not depend on the choice of ρ in the interval $(0, 1)$. In particular, the characterization holds for some ρ in this interval if and only if it holds for all ρ in this interval.

In a same way as the totalizing of translation function for the Solovay reducibility, we can totalize the concept of speedability by requiring the function g from the latter definition to be total.

Definition 3.5. *Let ρ be a real number such that $0 < \rho < 1$. A left-c.e. real α is called* TOTAL ρ-SPEEDABLE *if there exists a nondecreasing computable function $g\colon \mathbb{Q} \mapsto \mathbb{Q}$ that maps every q in the interval $(-\infty, \alpha)$ to a value $g(q) > q$ in this interval and satisfies*

$$\liminf_{q\nearrow\alpha} \frac{\alpha - g(q)}{\alpha - q} \leq \rho. \tag{3}$$

Such a function g is called TOTAL SPEED-UP FUNCTION.

By the following proposition, the total version of speedability does again not depend on the choice of the constant. The proof is omitted due to space considerations.

Proposition 3.6. *Whether a left-c.e. real is total ρ-speedable does not depend on the choice of $\rho \in (0, 1)$.*

Barmpalias and Lewis-Pye [2] have shown that speedability implies Martin-Löf nonrandomness. We currently research the characteristics of the total speed-up function via which the total speedability will imply Schnorr nonrandomness.

References

[1] Rodney Downey and Evan Griffiths, *Schnorr randomness*, Journal of Symbolic Logic 69(2):533–554 (2004).

[2] George Barmpalias and Andrew Lewis-Pye, *Differences of halting probabilities*, Journal of Computer and System Sciences 89:349–360 (2017).

[3] Rodney Downey and Denis Hirschfeldt, *Algorithmic Randomness and Complexity*, Springer, Berlin (2010).

[4] Wolfgang Merkle and Ivan Titov, *Speedable left-c.e. numbers*, CSR 2020: Computer Science – Theory and Applications pp 303–313 (2020).

[5] Kenshi Miyabe, André Nies and Frank Stephan, *Randomness and Solovay degrees*, Journal of Logic and Analysis 10(3):1–13 (2018).

[6] Joseph Miller, *On work of Barmpalias and Lewis-Pye: A derivation on the d.c.e. reals*, Lecture Notes in Computer Science 10010:644—659 (2016).

[7] André Nies, *Computability and Randomness*, Oxford University Press (2012).

Generalized bunched implication logic

Nick Galatos*[1] and Peter Jipsen[2]

[1] University of Denver
ngalatos@du.edu
[2] Chapman University
jipsen@chapman.edu

1 Introduction

Substructural logics are non-classical logics that include Boolean, Intuitionistic, Relevance, Many-valued and Linear logic, among others. They are further connected to philosophy, mathematical linguistics, and computer science, while their algebraic semantics have been studied extensively in order algebra, in the context of algebraic logic; see [1] for more on residuated lattices and substructural logics. In this presentation we discuss (generalized) Bunched Implication logic. The initial motivation comes from pointer management, memory allocation and concurrent programming in computer science where it serves as the Hoare logic for Separation logic. The algebraic semantics (which we suitably generalize to the non-commutative case, so as to include algebras of relations) are known as GBI-algebras (and Heyting residuated lattices) and they combine two different implication connectives: the dynamic and the logical; see [8] for an account of bunched implication logic and [7] for the connections between separation logic, bunched implication logic and residuated structures.

2 Proof theory

We present a Gentzen-style sequent **GBI** calculus for GBI-logic, which has two structural connectives, corresponding to conjunction $\wedge$ and strong conjunction $\cdot$, also known as fusion.

Theorem 1. *The system* **GBI** *enjoys cut elmination.*

We present a semantical proof of cut elimination; the proof proceeds by considering distributive residuated frames, two-sorted structures that form relational semantics for GBI-algebras; see [2] and [4]. This allows us to prove cut elimination for any extension of GBI with equations over the signature $\{\vee, \wedge, \cdot, 1\}$. In particular we recover the known cut elimination for the system of (commutative) BI logic as a special case. We also discuss a contraction-controlled form of the proofs of the system, based on the notion of 3-reduced proofs.

3 Congruences

A *residuated lattice* is an algebra $(A, \wedge, \vee, \cdot, \backslash, /, 1)$, where $(A, \wedge, \vee)$ is a lattice, $(A, \cdot, 1)$ is a monoid and $x \cdot y \leq z$ iff $x \leq y/z$ iff $y \leq x\backslash z$, for all $x, y, z \in A$. If $\cdot$ is equal to $\wedge$, then $\mathbf{A}$ is called a Brouwerian algebra (these are the bottom-free subreducts of Heyting algebras) and in this case we write $x{\to}y$ for $x\backslash y$; it also follows that $y/x = x\backslash y$ so we suppress this operation. A *generalized bunched implication algebra*, or *GBI-algebra*, is an algebra $\mathbf{A} = (A \wedge, \vee, \cdot, \backslash, /, 1, \to, \top)$, where $(A, \wedge, \vee, \cdot, \backslash, /, 1)$ is a residuated lattice and $(A, \wedge, \vee, \to, \top)$ is a Brouwerian algebra.

Congruences in residuated lattices are determined by certain subsets (in a way similar to the fact that congruences in groups are determined by normal subgroups);see [1]. Given $a, x \in A$

we define $\rho'_a(x) = ax/a$ and $\lambda'_a(x) = a\backslash xa$ (which are akin to conjugates in group theory). A subset is called *normal* if it is closed under ρ'_a and λ'_a for all $a \in A$. A *(RL)-deductive filter* of a residuated lattice $\mathbf{A}$ is defined to be a normal upward closed subset of A that is closed under multiplication and meet and contains the element 1. It is known that if θ is a congruence on $\mathbf{A}$ then $\uparrow[1]_\theta$, the upset of the equivalence class of 1, is a deductive filter. Conversely, if F is a deductive filter of a residuated lattice $\mathbf{A}$, then the relation θ_F is a congruence on $\mathbf{A}$, where $a\ \theta_F\ b$ iff $a\backslash b \wedge b\backslash a \in F$. Note that if $\mathbf{A}$ is a Brouwerian or a Heyting algebra, then deductive filters are usual lattice filters.

Theorem 2. *The GBI-deductive filters are exactly the RL-deductive filters that are further closed under* $r_{a,b}(x) = (a \to b)/(xa \to b)$ *and* $s_{a,b}(x) = (a \to bx)/(a \to b)$*, for all* a, b*.*

Alternatively, congruences are characterized by their equivalence classes of $\top$. These are usual lattice filters that are closed under the terms
$t_{a,b}(x) = a/b \to (a \wedge x)/b$, $u_{a,b}(x) = a/(b \wedge x) \to a/b$, $v_{a,b}(x) = ab \to (a \wedge x)b$ and
$t'_{a,b}(x) = b/a \to b/(a \wedge x)$, $u'_{a,b}(x) = (b \wedge x)\backslash a \to b\backslash a$, $v'_{a,b}(x) = ab \to a(b \wedge x)$ for all a, b.

4 Weakening relation algebras

In their seminal on Boolean algebras with operators (BAOs), Jónsson and Tarski showed that many varieties of BAOs, including the variety of relation algebras, are closed under canonical extensions, and that a relation algebra is complete and atomic with all atoms as functional elements if and only if it is the complex algebra of a generalized Brandt groupoid. We show that results about relation algebras can also be generalized to certain involutive GBI-algebras, which we call weakening relation algebras. These can be thought as intuitionistic analogues to classical/Boolean relation algebras.

A *relation algebra* $(A, \wedge, \vee, ', \top, \bot, \cdot, ^\smile, 1)$ consists of a Boolean algebra $(A, \wedge, \vee, ', \top, \bot)$ and a monoid $(A, \cdot, 1)$ such that $xy \leq z \iff x^\smile \cdot z' \leq y'$. A *cyclic involutive GBI-algebra* (CyGBI-algebra) is an expansion $(B, \wedge, \vee, \to, \top, \cdot, 1, 0)$ of a GBI-algebra with an additional constant 0; we further define the connectives $\sim x = x\backslash 0$, $\bot = \sim \top$, $\neg x = x \to \bot$ and $x^\smile = \sim\neg x$. A relation algebra is a CyGBI-algebra where $x \to y = x' \vee y$ and $\sim x = x'^\smile$. It turns out that a CyGBI-algebra is a relation algebra iff it satisfies the identities $\neg\neg x = x$ and $(xy)^\smile = y^\smile x^\smile$. In other words, CyGBI-algebras provide a natural setting which shows all the hiden symmetries in the definition of relation algebras.

We define algebras of binary relations that are cyclic involutive GBI-algebras and generalize representable relation algebras. Let $\mathbf{P} = (P, \sqsubseteq)$ be a partially ordered set, $Q \subseteq P^2$ an equivalence relation that contains $\sqsubseteq$, and define the set of *weakening relations* on $\mathbf{P}$ by $\mathrm{Wk}(\mathbf{P}, Q) = \{\sqsubseteq \circ R \circ \sqsubseteq : R \subseteq Q\}$. Note that this set is closed under intersection $\cap$, union $\cup$ and composition $\circ$, but not under complementation $R' = Q - R$ or converse $R^\smile$.

Weakening relations are the natural analogue of binary relations when the category **Set** of sets and functions is replaced by the category **Pos** of partially ordered sets and order-preserving functions. Since sets can be considered as discrete posets (i.e. ordered by the identity relation), **Pos** contains **Set** as a full subcategory, which implies that weakening relations are a substantial generalization of binary relations. They have applications in sequent calculi, proximity lattices/spaces, order-enriched categories, cartesian bicategories, bi-intuitionistic modal logic, mathematical morphology and program semantics, e.g. via separation logic.

Let $\mathbf{P} = (P, \sqsubseteq)$ be a poset, Q an equivalence relation that contains $\sqsubseteq$, and for $R, S \in \mathrm{Wk}(\mathbf{P}, Q)$ define $\top = Q$, $\bot = \emptyset$, $1 = \sqsubseteq$, $\sim R = R^{\smile\prime}$ and $R \to S = (\sqsupseteq \circ (R \cap S') \circ \sqsupseteq)'$ where $S' = Q - S$.

Theorem 3. $\mathbf{Wk}(\mathbf{P},Q) = (\mathrm{Wk}(\mathbf{P},Q), \cap, \cup, \to, \top, \bot, \circ, 1, \sim)$ *is a CyGBI-algebra.*

Algebras of the form $\mathbf{Wk}(\mathbf{P},Q)$ are called representable weakening relation algebras, and if $Q = P \times P$, then we write $\mathbf{Wk}(\mathbf{P})$ and call this algebra the *full weakening relation algebra on* $\mathbf{P}$. If $\mathbf{P}$ is a discrete poset then $\mathbf{Wk}(\mathbf{P})$ is the full representable set relation algebra on the set P, so algebras of weakening relations play a role similar to representable relation algebras. We define the class wRRA of *representable weakening relation algebras* as all algebras that are embedded in a weakening relation algebra $\mathbf{Wk}(\mathbf{P},Q)$ for some poset $\mathbf{P}$ and equivalence relation Q that contains $\sqsubseteq$. In fact the variety RRA of representable relation algebras is finitely axiomatized over wRRA.

Theorem 4. wRRA *is a discriminator variety closed under canonical extensions. Also,* RRA *is the subvariety of* wRRA *defined by* $\neg\neg x = x$. *Finally, the class* wRRA *is not finitely axiomatizable relative to the variety of all CyGBI-algebras.*

5 Weakening relations via conuclei on GBI-algebras

A *weak conucleus* on a residuated lattice $\mathbf{A}$ is an interior operator σ on $\mathbf{A}$ such that $\sigma(x)\sigma(y) \leq \sigma(xy)$, for all $x, y \in \mathbf{A}$. Then $\sigma[\mathbf{A}] = (\sigma[A], \wedge_\sigma, \vee, \cdot, \backslash_\sigma, /_\sigma)$ is a residuated lattice-ordered semigroup, where $x \bullet_\sigma y = \sigma(x \bullet y)$, for $\bullet \in \{\wedge, \backslash, /\}$; we are interested in the cases where this algebra also has an identity element e and hence $(\sigma[A], e)$ is a residuated lattice. A *topological weak conucleus* on a GBI-algebra $\mathbf{A}$ is a conucleus on both the residuated lattice and the Brouwerian algebra reducts of $\mathbf{A}$.

Given a residuated lattice $\mathbf{A}$ and a positive idempotent element p we define the map σ_p by $\sigma_p(p) = p\backslash x/p$. Then σ_p is a topological weak conucleus (which we call the *double division conucleus by* p), and p is the identity element $\sigma_p(\mathbf{A})$; we denote the resulting residuated lattice $(\sigma_p(\mathbf{A}), p)$ by $p\backslash\mathbf{A}/p$. If $\mathbf{A}$ is involutive then so is $p\backslash\mathbf{A}/p$ and the latter is a subalgebra of $\mathbf{A}$ with respect to the operations $\wedge, \vee, \cdot, +, \sim, -$; recall that an *involutive* residuated lattice is an expansion of a residuated lattice with an extra constant 0 such that $\sim(-x) = x = -(\sim x)$, where $\sim x = x\backslash 0$ and $-x = 0/x$; we also define $x + y = \sim(-y \cdot -x)$.

Note that given a poset $\mathbf{P} = (P, \sqsubseteq)$, we have $Wk(\mathbf{P}) = \mathcal{O}(\mathbf{P} \times \mathbf{P}^\partial)$, where $\mathcal{O}$ denotes the downset operator. Recall that a map on f on a poset $\mathbf{P}$ is called *residuated* if there exists a map f^* on P such that $f(x) \sqsubseteq y$ iff $x \sqsubseteq f^*(y)$, for all $x, y \in P$. For a complete join semilattice $\mathbf{L}$, $Res(\mathbf{L})$ denotes the residuated lattice of all residuated maps on $\mathbf{L}$.

Theorem 5. $Wk(\mathbf{P}) \cong Res(\mathcal{O}(\mathbf{P}))$.

Given a poset $\mathbf{P} = (P, \sqsubseteq)$, we set $\mathbf{A} = Rel(P)$, to be the involutive GBI algebra of all binary relations on the set P. Note that $p = \sqsubseteq$ is a positive idempotent element of $\mathbf{A}$. It is easy to see that $p\backslash\mathbf{A}/p$ is exactly $Wk(\mathbf{P})$. Since $\mathbf{A}$ is an involutive GBI-algebra, so is $Wk(\mathbf{P})$.

References

[1] N. Galatos, P. Jipsen, T. Kowalski and H. Ono. Residuated lattices: an algebraic glimpse at substructural logics. *Studies in Logic and the Foundations of Mathematics*, 151. Elsevier B. V., Amsterdam, 2007. xxii+509.

[2] N. Galatos and P. Jipsen. Residuated frames with applications to decidability. *Trans. Amer. Math. Soc.* 365 (2013), no. 3, 1219–1249.

[3] N. Galatos and P. Jipsen. Relation algebras as expanded FL-algebras. *Algebra Universalis*, 69 (2013), no. 1, 1–21.

[4] N. Galatos and P. Jipsen. Distributive residuated frames and generalized bunched implication algebras. *Algebra Universalis*, 78 (2017), no. 3, 303–336.

[5] N. Galatos and P. Jipsen. The structure of generalized BI-algebras and weakening relation algebras. *Algebra Universalis*, 81 (2020), no. 3, Paper No. 35, 35 pp.

[6] N. Galatos and P. Jipsen. Weakening relation algebras and FL^2-algebras. *Relational and algebraic methods in computer science, LNCS* 12062, 117—133, 2020.

[7] P. Jipsen and T. Litak. An algebraic glimpse at bunched implications and separation logic, *Hiroakira Ono on Substructural Logics. Outstanding Contributions to Logic*, vol 23. Springer, Cham. 2022.

[8] D. Pym. The semantics and proof theory of the logic of bunched implications. *Applied Logic Series*, 26. Kluwer Academic Publishers, Dordrecht, 2002.

Quid Verificabit Ipsos Verificatores? A Model for Self-Applicable Exact Truthmaking

Simone Picenni*

University of Bristol
picenni.simone@gmail.com

Abstract

In the present paper, we will show how to construct an exact truthmaking model for a rich enough first order language containing two unary predicates S, A whose intended reading is *being (the code of) a state* and *being (the code of) an actual state* respectively, and a binary predicate $\Vdash$, which bears (codes of) states to (codes of) sentences of the language, and whose intended meaning is *making exactly true*. In other, more informal terms: we will show how to construct an exact truthmaking model for a language rich enough to talk about its own exact truthmaking semantics.

1 Introduction

Recent years have seen a rise of interest in *exact truthmakers* in philosophical logic and semantics. A *state* (of affairs, action, event, ...), s, is an exact truthmaker for a statement, ϕ, just in case *s necessitates ϕ 's truth* while being *wholly relevant to it* [9, 14]. The state of the ball being red, for instance, is an exact truthmaker for "the ball is red", while the complex state of the ball being red and round is not – the ball's shape has nothing to do with its color [14]. The concept of exact truthmaking gives rise to a very fine-grained semantics, *exact truthmaking semantics*, according to which we individuate content of sentences by means of their exact truthmakers [see 9, 13, 17].

Exact truthmaking semantics has proved useful in the reconstruction of the notion of *aboutness* [see 21, 7, 8], in the reconstruction of the semantics of *hyperintensional contexts* (see [15] for an overview, on hyperintensional contexts, and see for instance [16, 17]) – e.g.: propositional attitude reports (Alice believes that ϕ, Bob knows that ϕ, ...), in which substituting sentences that are true in all and only the same possible worlds may not preserve the truth-value of the report –, and as a semantics for large portions of natural language [see e.g. 17].

However, there are technical and philosophical problems that call for a solution. First of all, as Barwise puts it [2], a rich enough semantics should be able "to be turned on itself, and provide an account of its own information content, or rather, of the statements made by the theorist using the theory", i.e.: provide a semantics for the language we are using to do semantics. A model in which to do this is not yet present in the literature about exact truthmaking.

Secondly, we may want a hyperintensional semantics, especially one used to give an analysis of the notion of *aboutness*, to distinguish between the semantic content of a sentence ϕ and the semantic content of a sentence that ascribes truth to ϕ, "'ϕ' is true" [several arguments to support this claim are present, for instance, in 20]. For instance, it seems that "'Snow is white' is true" says something about a statement, 'Snow is white', while the sentence "Snow is white" says something about snow. We would like to have a model that would give an account of this difference, while maintaining the intensional equivalence of ϕ and "'ϕ' is true". Such a model

*ERC Starting Grant "Truth and Semantics" (TRUST 803684)

is not present in the literature on exact truthmaking, yet.
Thirdly, exact truthmaking semantics comes with philosophical problems related to paradoxes of truthmaking. Sentences like "This very sentence doesn't have truthmakers", or "The situation described by this very sentence is not actual" [see 1], generate difficulties for any truthmaking semantics. In order to diagnose how statements like these are problematic, and to provide a solution to these paradoxes of truthmaking, a good semantics for a language containing predicates whose intended reading is *being a truthmaker*, *making true*, ..., needs to be constructed. In the following sections, we will show how to build a model that could be used to address these challenges.

2 Exact Truthmaking: Technical Preliminaries

Definition 2.1. A ***state space*** is an ordered pair $(S, \sqsubseteq)$ such that

- S is a non-empty set – the set of *states*, objects that can act as truthmakers;
- $\sqsubseteq \subseteq S \times S$ is a partial order (i.e.: a reflexive, antisymmetric and transitive relation) on S, meant to represent a *parthood* relation [9].

As a notational convention: we will use the notation $\bigsqcup X$ to indicate the $\sqsubseteq$-least upper bound of the subset $X \subseteq S$ (i.e.: an $s \in S$ such that, for every $r \in X$, $r \sqsubseteq s$, and such that, for all $s' \in S$ having this property, $s \sqsubseteq s'$), and we will use the notation $s \sqcup r$ to indicate $\bigsqcup\{s, r\}$. We do not require that every subset $X \subseteq S$ has a least upper bound (in S).

Definition 2.2. An ***exact truthmaking model*** for a first order language $\mathscr{L}$ is an ordered tuple $(S, \sqsubseteq, D, A, |\cdot|)$ such that

- $(S, \sqsubseteq)$ is a state space.
- $D \neq \emptyset$, the domain, is a set (the set over which our quantifiers will range).
- $A \subseteq S$, the set of "actual" states, is a *downward closed* non-empty subset of S – that is: if $s \in A$ and $t \sqsubseteq s$, then $t \in A$.
- $|\cdot| \subseteq \mathscr{L} \times (\wp(S) \times \wp(S))$ is a function sending atomic sentences of the language $\mathscr{L}$ into ordered pairs of sets of states.

In several papers, Fine [see e.g. 7, 9] has shown how to extend a function similar to $|\cdot|$ to exact truthmaking and falsemaking relations $\Vdash$ and $\dashv\Vdash$ for a propositional language. However, there is currently no consensus on what the exact verification (amd falsification) clauses for a first-order language should be.
Above I accepted the idea that, in order to express the truth conditions of quantified sentence, we must provide a collection of objects on which the variables bound by the quantifiers of our language vary, a domain of discourse [see 12]. For reasons of simplicity, we will use a language *à la* Robinson, in which each object of the domain has a constant that denotes it [see 3]. Given these assumptions, we can consider the following semantics of exact truthmaking for a first order language to be quite natural:

Definition 2.3. Let $(S, \sqsubseteq, D, A, |\cdot|)$ be an exact truthmaking model for a first order language containing a closed term d for every object $d \in D$. The relations $\Vdash$ and $\dashv\Vdash$ of ***truthmaking and falsemaking*** are recursively defined on the set of sentences of the language $\mathscr{L}$ as follows (we assume disjunction and existential quantifiers are defined as usual in terms of other connectives/quantifiers):

1. $s \Vdash \mathsf{P}(\mathsf{t_1}, ..., \mathsf{t_n})$ iff $s \in proj_1|\mathsf{P}(\mathsf{t_1}, ..., \mathsf{t_n})|$, and $s \dashv\Vdash \mathsf{P}(\mathsf{t_1}, ..., \mathsf{t_n})$ iff $s \in proj_2|\mathsf{P}(\mathsf{t_1}, ..., \mathsf{t_n})|$
2. $s \Vdash \neg\phi$ iff $s \dashv\Vdash \phi$, and $s \dashv\Vdash \neg\phi$ iff $s \Vdash \phi$
3. $s \Vdash \phi \wedge \psi$ iff there are $s_1, s_2 \in S$ such that $s_1 \Vdash \phi$ & $s_2 \Vdash \psi$ & $s_1 \sqcup s_2$ is defined, and $s = s_1 \sqcup s_2$, and $s \dashv\Vdash \phi \wedge \psi$ iff $s \dashv\Vdash \phi$ or $s \dashv\Vdash \psi$
4. $s \Vdash \forall \mathsf{v_i} \,.\, \phi[\mathsf{v_i}]$ iff there is an $X \subseteq S$ such that
   (a) For all $d \in D$, there is an $s_1 \in X$ such that $s_1 \Vdash \phi[\mathsf{v_i} := \mathsf{d}]$, and
   (b) For all $s_1 \in X$, there is a $d \in D$ such that $s_1 \Vdash \phi[\mathsf{v_i} := \mathsf{d}]$, and
   (c) $\bigsqcup X$ is defined, and $s = \bigsqcup X$

   and $s \dashv\Vdash \forall \mathsf{v_i} \,.\, \phi[\mathsf{v_i}]$ iff there is a $d \in D$ such that $s \dashv\Vdash \phi[\mathsf{v_i} := \mathsf{d}]$

3 The Construction of the Model

Our aim is to construct a non-trivial[1] exact truthmaking model $(S, \sqsubseteq, D, A, |\cdot|)$ for a rich enough first order language containing two unary predicate symbols S and A (whose intended reading is *being a state* and *being an actual state* respectively), and a binary predicate symbol $\Vdash$ (whose intended reading is *making exactly true*), such that

$$\exists s \in A : s \text{ makes } \phi \text{ exactly true} \Leftrightarrow \exists s \in A : s \text{ makes } (\exists \mathsf{v} \,.\, \mathsf{A}(\mathsf{v}) \wedge \mathsf{v} \Vdash \ulcorner \phi \urcorner) \text{ exactly true}$$

(Where $\ulcorner \phi \urcorner$ is a quotation name for the sentence ϕ). Here's, informally speaking, an overview of our strategy. For sake of simplicity, we will consider as *states* objects similar to *Carnapian state descriptions* [see 4] – that is: non-empty sets of (codes of) literals (i.e.: atomic sentences or negations of atomic sentences) of a language. In deference to the spirit of [6, 7, 9], however, unlike what happens in [4], we will not require that states be complete (i.e.: that, for each state and each atomic statement, the state contains either the statement itself or its negation), nor consistent (ie: that an atomic sentence and its negation are not present at the same time in a state)[2]. Importantly, we want to be able to *talk about states*, and about *ways in which states and sentences of our language are related*. To this effect, we need at least a base structure in which we can effectively code expressions of our language, and in which "syntactical properties, relations, and operations can be reflected" [19]. To keep the discussion general enough, we will start the construction of our model making use of a language with denumerable signature supported by a *strongly acceptable structure* [18, 11] as syntax theory (recall that an acceptable[3] structure $\mathfrak{A} = (A, R_1, ..., R_n)$ is *strongly acceptable* [see 11] iff there is an hyperelementary coding of A into an elementary ordering of members of A, $\mathcal{N}^{\mathfrak{A}}$, isomorphic to the natural numbers, which exists because the structure is acceptable). The language of the structure will then be expanded with semantical vocabulary, and we will consider as states *the sets of literals of the expanded language that are definable in the structure*. From a technical point of view, as we will see, this restriction will make it possible to *encode our states* in the domain of our structure. From a philosophical point of view, the restriction could be viewed as a way to capture the intuition that, in formulating a rigorous formal semantics for her language, a person is bound

[1] In the sense that the set of actual states of the model is non-empty, and there is not an actual exact truthmaker for every sentence of the language.

[2] States similar to these are used, in the propositional case, in propositional HYPE [see 15], or for the construction of the canonical model of the logic of the exact entailment of [10].

[3] The notion of *acceptable structure* is standard in model theory [see 18].

by the mathematical and formal-semantic resources available to her *before* the formulation of the semantics[4].
As for the other components of the desired model, the following apply. The parthood relation will be represented by the subset relation, restricted to the set of states thus defined. We will consider an exact truthmaker of a literal to be *the code of the singleton of the literal itself*, and the relation of truthmaking will be defined for all the sentences of our language as done above. And finally, actual states will be the ones containing only true literals – we will see how to identify them.
Let's now see how we can construct the desired model.

Let $\mathfrak{M} = (M, R_1, ..., R_n)$ be a strongly acceptable structure, and let $\mathscr{L}$ be its first order language augmented with a constant m for all the m's in $\mathfrak{M}$[5]. As a notational convention, let $\#\varepsilon$ be the object $m \in M$ that encodes the expression ε according to a fixed coding function $\#$[6], and let $\ulcorner\varepsilon\urcorner$ be the closed term of the language denoting $\#\varepsilon$ [see again 19, for an overview].

As we said, we are concerned with the *sets of (codes of) literals that are describable by a "speaker" of* $\mathscr{L}$. Let's make this intuition rigorous.
In [18], Moschovakis shows how to define a satisfaction relation $Sat_{\mathfrak{A}}$, hyperelementary over a base acceptable structure $\mathfrak{A}$, such that $Sat_{\mathfrak{A}}$ bears a code of a formula $\phi(\mathsf{v_1}, ..., \mathsf{v_n})$ in the first order language of $\mathfrak{A}$ to $(a_1, ..., a_n) \in A^n$ whenever $\mathfrak{A} \models \phi(\mathsf{v_1}, ..., \mathsf{v_n})[a_1, ..., a_n]$ – i.e.: whenever the tuple $(a_1, ..., a_n)$ satisfies the formula $\phi(\mathsf{v_1}, ..., \mathsf{v_n})$ in $\mathfrak{A}$. We can make the notion of definability in $\mathfrak{M}$ precise *via* the relation $Sat_{\mathfrak{M}}$:

Definition 3.1. A set $X \subseteq M^n$ is ***definable in*** $\mathfrak{M}$ iff there is a first order formula $\phi(\mathsf{v_1}, ..., \mathsf{v_n})$ in the first order language of $\mathfrak{M}$ such that

$$(m_1, ..., m_n) \in X \Leftrightarrow Sat_{\mathfrak{M}}(\#\phi(\mathsf{v_1}, ..., \mathsf{v_n}), (m_1, ..., m_n))$$

Then we let $L_{\mathfrak{M}}$ be the set of sets definable in $\mathfrak{M}$. Let now $Form_{\mathscr{L}}$ be the set of codes of formulae of $\mathscr{L}$. Since every set in $L_{\mathfrak{M}}$ is definable by a formula in $\mathscr{L}$, we define a (hyperelementary) coding $\pi : L_{\mathfrak{M}} \to Form_{\mathscr{L}}$ of sets of $L_{\mathfrak{M}}$ into codes of formulae of our language by setting $\pi(X)$ *be the minimal code of a formula defining* X. That is:

$$\pi(X) = \mu w \in Form_{\mathscr{L}} \,.\, ((m_1, ..., m_n) \in X \Leftrightarrow Sat_{\mathfrak{M}}(w, (m_1, ..., m_n)))$$

Thus, let D_π be the set of codes in $\mathfrak{M}$ of objects in $L_{\mathfrak{M}}$, and let $|m|_\pi$ be the object in $L_{\mathfrak{M}}$ associated to $m \in M$ by π. Let the language $\mathscr{L}^+$ be the language $\mathscr{L}$ augmented with the predicates $\mathsf{S}, \mathsf{A}, \Vdash$. Finally, let $Lit_{\mathscr{L}^+} \subseteq M$ be the set of codes of literal sentences of $\mathscr{L}^+$ – i.e.: the codes of atomic sentences and negations of atomic sentences of the language $\mathscr{L}$ extended with $\mathsf{S}, \mathsf{A}, \Vdash$ [see again 18, which shows that such a set exists and is elementary over the base structure].
Then we can define the extension of the predicate *being a state*, S.

Definition 3.2. $\mathsf{S}^{\mathfrak{M}}(x) :\Leftrightarrow D_\pi(x)\ \&\ |x|_\pi \neq \emptyset\ \&\ \forall y \in |x|_\pi \,.\, Lit_{\mathscr{L}^+}(y)$

That is, informally speaking: x is declared to be a *state* in $\mathfrak{M}$ iff it is a code of an $\mathfrak{M}$-definable non-empty set of literals. Let then[7]:

[4] A similar informal story is present in [11].
[5] We will assume $\mathscr{L}$ to have denumerable signature and to contain finitely many functors.
[6] We assume it is monotonic in $\mathcal{N}^{\mathfrak{M}}$ – see p. 3.
[7] We will use Feferman's dot notation: $\dot{f}$ in $\dot{f}(\vec{x})$ represents the primitive recursive operation f on codes $\vec{x}$ [see 5].

Definition 3.3. $x \Vdash^{\mathfrak{M}} y :\Leftrightarrow \mathsf{S}^{\mathfrak{M}}(x)$ & $Sent_{\mathscr{L}^+}(y)$ &

1. $Lit_{\mathscr{L}^+}(y)$ & $x = \pi(\{y\})$, or
2. $\exists w_1 \in Sent_{\mathscr{L}^+} \,.\, y = \dot{\neg}\dot{\neg} w_1$ & $x \Vdash^{\mathfrak{M}} w_1$
3. $\exists w_1 w_2 \in Sent_{\mathscr{L}^+} \,.\, y = w_1 \dot{\wedge} w_2$ &
   $\exists x_1 x_2 \in \mathsf{S}^{\mathfrak{M}} \,.\, x_1 \Vdash^{\mathfrak{M}} w_1$ & $x_2 \Vdash^{\mathfrak{M}} w_2$ & $x = \pi(|x_1|_\pi \cup |x_2|_\pi)$, or
4. $\exists w_1 w_2 \in Sent_{\mathscr{L}^+} \,.\, y = \dot{\neg}(w_1 \dot{\wedge} w_2)$ & $x \Vdash^{\mathfrak{M}} \dot{\neg} w_1$ or $x \Vdash^{\mathfrak{M}} \dot{\neg} w_2$, or
5. $\exists v_1 \in Var_{\mathscr{L}^+} \,.\, \exists w_1 \in Form_{\mathscr{L}^+} \,.\, y = \dot{\forall} v_1 w_1$ &
   (a) $\forall m_1 \in CTerm_{\mathscr{L}^+} \,.\, \exists z \in \mathsf{S}^{\mathfrak{M}} \,.\, |z|_\pi \subseteq |x|_\pi$ & $z \Vdash^{\mathfrak{M}} w_1[v_1 := m_1]$, &
   (b) $\forall u \in |x|_\pi \,.\, \exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, \exists z \in \mathsf{S}^{\mathfrak{M}} \,.\, |z|_\pi \subseteq |x|_\pi$ &
   $z \Vdash^{\mathfrak{M}} w_1[v_1 := m_1]$ & $u \in |z|_\pi$
6. $\exists v_1 \in Var_{\mathscr{L}^+} \,.\, \exists w_1 \in Form_{\mathscr{L}^+} \,.\, y = \dot{\neg}\dot{\forall} v_1 w_1$ & $\exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, x \Vdash^{\mathfrak{M}} \dot{\neg} w_1[v_1 := m1]$

This extension mimics the one of the exact truthmaking that is present in the previous section. However, the fact that we can *quantify into what's inside a state* (provably) makes the second order quantification unnecessary.

Now: as we said, if a truthmaker is a set of literals, we would like actual truthmakers to be *sets of true literals*. As for literals of the base language, we know that there is a diagram formula $\delta_{\mathfrak{M}}(x)$ which is true of all and only the codes of true literals of the base language [see 18, 11]. We would like to know, however, which literals of the expanded language should be considered true. We will use a Kripkean strategy to establish this.

Let $Opp : Lit_{\mathscr{L}^+} \to Lit_{\mathscr{L}^+}$ be the function that, for any $x \in Lit_{\mathscr{L}^+}$, outputs the code of the negation of x if x is the code of an atomic sentence, and the atomic sentence that results from erasing the negation in front of x if x is the negation of an atomic sentence. Furthermore, let $Val : M \to M$ be the function that sends the code of a closed term to the object in M the term denotes, and let the formula $\zeta(X, x)$ be the following:

Definition 3.4. $\zeta(X, x) :\Leftrightarrow Lit_{\mathscr{L}^+}(x)$ &

1. $Lit_{\mathscr{L}}(x)$ & $\delta_{\mathfrak{M}}(x)$, or
2. $\exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, x = \dot{\mathsf{S}}(m_1)$ & $\mathsf{S}^{\mathfrak{M}}(Val(m_1))$, or
3. $\exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, x = \dot{\neg}\dot{\mathsf{S}}(m_1)$ & $\neg\mathsf{S}^{\mathfrak{M}}(Val(m_1))$, or
4. $\exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, x = \dot{\mathsf{A}}(m_1)$ & $\forall u \in |Val(m_1)|_\pi \,.\, X(u)$, or
5. $\exists m_1 \in CTerm_{\mathscr{L}^+} \,.\, x = \dot{\neg}\dot{\mathsf{A}}(m_1)$ & $\exists u \in |Val(m_1)|_\pi \,.\, X(Opp(u))$, or
6. $\exists m_1 m_2 \in CTerm_{\mathscr{L}^+} \,.\, x = m_1 \dot{\Vdash} m_2$ & $Val(m_1) \Vdash^{\mathfrak{M}} Val(m_2)$, or
7. $\exists m_1 m_2 \in CTerm_{\mathscr{L}^+} \,.\, x = m_1 \dot{\nVdash} m_2$ & $Val(m_1) \nVdash^{\mathfrak{M}} Val(m_2)$

Let now Γ_ζ be the operator induced by $\zeta(X, x)$ and defined by transfinite recursion as follows:

Definition 3.5.

$$\Gamma^{\alpha}_{\zeta}(X) = \begin{cases} X, & \text{if } \alpha = 0 \\ \{m \in M : \zeta(\Gamma^{\alpha-1}_{\zeta}(X), m)\}, & \text{if } \alpha \text{ is a successor ordinal} \\ \bigcup\limits_{\beta<\alpha} \Gamma^{\beta}_{\zeta}(X), & \text{if } \alpha \text{ is a limit ordinal} \end{cases}$$

Since $S^{\mathfrak{M}}$ and $\Vdash^{\mathfrak{M}}$ are not more complex than hyperelementary on the base structure, the standard results of [18] apply, and:

Observation 3.1. *Γ is monotone: if $X \subseteq Y$, then, for every $\alpha \in On$, $\Gamma^{\alpha}_{\zeta}(X) \subseteq \Gamma^{\alpha}_{\zeta}(Y)$*

Observation 3.2. *There is a $\beta \in On$ such that $\Gamma^{\beta}_{\zeta}(\emptyset) = \Gamma^{\beta+1}_{\zeta}(\emptyset)$*

Let β be the minimal ordinal such that $\Gamma^{\beta}_{\zeta}(\emptyset) = \Gamma^{\beta+1}_{\zeta}(\emptyset)$. Then $\Gamma^{\beta}_{\zeta}(\emptyset)$ is the set of true literals we were looking for. Then let

Definition 3.6. $\mathsf{A}^{\mathfrak{M}}(x) :\Leftrightarrow \mathsf{S}^{\mathfrak{M}}(x)\ \&\ \forall y \in |x|_{\pi}\, .\, y \in \Gamma^{\beta}_{\zeta}(\emptyset)$

That is, an object $m \in M$ is an the extension of A iff it is the code of a state and the state encoded by it contains only codes of true literals. Then the following propositions are provable:

Proposition 3.1. *Let ϕ be a formula of the language of $\mathfrak{M}$. Then ϕ is true in $\mathfrak{M}$ iff there is an $m \in \mathsf{A}^{\mathfrak{M}}$ such that $m \Vdash^{\mathfrak{M}} \#\phi$.*
In other terms: $\lambda\xi \in Sent_{\mathscr{L}}\, .\, \exists \mathsf{v}\, .\, \mathsf{A}(\mathsf{v}) \wedge \mathsf{v} \Vdash \ulcorner\xi\urcorner$ is a truth-predicate for $\mathscr{L}$ interpreted in $\mathfrak{M}$.

Proposition 3.2. *Let ϕ be a formula of the extended language. Then it is true in $(\mathfrak{M}, \mathsf{S}^{\mathfrak{M}}, \mathsf{A}^{\mathfrak{M}}, \Vdash^{\mathfrak{M}})$ that there is an $m \in \mathsf{A}^{\mathfrak{M}}$ such that $m \Vdash^{\mathfrak{M}} \#\phi$ iff there is an $m \in \mathsf{A}^{\mathfrak{M}}$ such that $m \Vdash^{\mathfrak{M}} \#(\exists \mathsf{v}\, .\, \mathsf{A}(\mathsf{v}) \wedge \mathsf{v} \Vdash \ulcorner\phi\urcorner)$.*

Proposition 3.3. *Let $S := \{x \in L_{\mathfrak{M}} : \mathsf{S}^{\mathfrak{M}}(\pi(x))\}$; $\sqsubseteq := \subseteq\restriction S$; $D := M$; $A := \{x \in L_{\mathfrak{M}} : \mathsf{A}^{\mathfrak{M}}(\pi(x))\}$; and, finally let $|\cdot| := \{(\phi, (\{\{\phi\}\}, \{\{\neg\phi\}\}))\ :\ \phi$ is an atomic sentence of $\mathscr{L}^{+}\}$. Then $(S, \sqsubseteq, D, A, |\cdot|)$ is an exact truthmaking model for $\mathscr{L}^{+}$, and the model is such that:*

- *for any sentence ϕ of $\mathscr{L}$, there's a state $s \in S \cap A$ making ϕ exactly true iff $\mathfrak{M} \models \phi$, and*
- *for any sentence ϕ of $\mathscr{L}^{+}$, there's a state $s \in S \cap A$ making ϕ exactly true iff there's a state $s \in S \cap A$ making $\exists \mathsf{v}\, .\, \mathsf{A}(\mathsf{v}) \wedge \mathsf{v} \Vdash \ulcorner\phi\urcorner$ exactly true.*

References

[1] Eduardo Barrio and Gonzalo Rodriguez-Pereyra. Truthmaker maximalism defended again. *Analysis*, 75(1):3–8, 2015.

[2] Jon Barwise. *The situation in logic*, volume 4. Center for the Study of Language (CSLI), 1989.

[3] Tim Button and Sean Walsh. *Philosophy and model theory*. Oxford University Press, 2018.

[4] Rudolf Carnap. *Meaning and Necessity: a Study in Semantics and Modal Logic*. University of Chicago Press, 1947.

[5] Solomon Feferman. Axioms for determinateness and truth. *The Review of Symbolic Logic*, 1(2):204–217, 2008.

[6] Kit Fine. Truth-maker semantics for intuitionistic logic. *Journal of philosophical logic*, 43(2-3):549–577, 2014.

[7] Kit Fine. A theory of truthmaker content i: Conjunction, disjunction and negation. *Journal of Philosophical Logic*, 46(6):625–674, 2017.

[8] Kit Fine. A theory of truthmaker content ii: Subject-matter, common content, remainder and ground. *Journal of Philosophical Logic*, 46(6):675–702, 2017.

[9] Kit Fine. Truthmaker semantics. *A Companion to the Philosophy of Language*, 2:556–577, 2017.

[10] Kit Fine and Mark Jago. Logic for exact entailment. *The Review of Symbolic Logic*, 12(3):536–556, 2019.

[11] Michael Glanzberg. A contextual-hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic*, 33(1):27–88, 2004.

[12] Michael Glanzberg. Quantification and realism. *Philosophy and Phenomenological Research*, 69(3):541–572, 2004.

[13] Mark Jago. Propositions as truthmaker conditions. *Argumenta*, 2(2), 2017.

[14] Johannes Korbmacher. Proof systems for exact entailment. *The Review of Symbolic Logic*, page 1–37, 2022.

[15] Hannes Leitgeb. Hype: A system of hyperintensional logic (with an application to semantic paradoxes). *Journal of Philosophical Logic*, 48(2):305–405, 2019.

[16] Friederike Moltmann. Cognitive products and the semantics of attitude verbs and deontic modals. *Act-based conceptions of propositional content*, pages 254–290, 2017.

[17] Friederike Moltmann. Truthmaker semantics for natural language: Attitude verbs, modals, and intensional transitive verbs. *Theoretical Linguistics*, 46(3-4):159–200, 2020.

[18] Yiannis N Moschovakis. *Elementary induction on abstract structures*. North-Holland Publishing Company, 1974.

[19] Panu Raatikainen. Gödel's Incompleteness Theorems. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition, 2022.

[20] Johannes Stern. Belief, truth, and ways of believing. In *Modes of Truth*, pages 151–181. Routledge, 2021.

[21] Stephen Yablo. *Aboutness*. Princeton University Press, 2014.

NSOP in Classes of Graphs

Ioannis Eleftheriadis[1]*and Aristomenis-Dionysios Papadopoulos[2]†

[1] University of Cambridge, Cambridge, U.K.
ie257@cam.ac.uk
[2] University of Leeds, Leeds, U.K.
mmadp@leeds.ac.uc

Abstract

In recent years Shelah's classification program has found surprising applications in the context of classes of relational structures. The work of [1], [8], [12] and others has explored the relationship between model-theoretic tameness, in particular (monadic) stability and (monadic) NIP, and algorithmic tameness, e.g. nowhere density and bounded twin-width. Motivated by this, we investigate the implications of NSOP in classes of graphs and relational structures.

1 Introduction

One core aspect of Shelah's classification program is the discovery of combinatorial configurations (*dividing lines*) which separate the "tame" theories from the "wild" ones. In recent years, there has been a lot of work in drawing connections between these tameness conditions and algorithmic or combinatorial tameness conditions, for instance in the works of [1], [8], [12].

Shelah's outlook is that instability in a theory must come from either randomness (IP) or order (SOP). It is curious that the notion of NSOP (intuitively: random phenomena are allowed as long as they don't introduce order-like behaviour) has not been examined further in the context of classes of structures. Initiating this is the focus of our paper.

We start by recalling some core definitions from Shelah's classification program, in the context of classes of (possibly finite) structures. Throughout this paper classes of structures are assumed to be closed under isomorphism.

Definition 1.1. Let $\mathcal{L}$ be a first-order language and $\mathcal{C}$ a class of $\mathcal{L}$-structures. We say that an $\mathcal{L}$-formula $\phi(\overline{x};\overline{y})$ has:

1. The *Order Property* in $\mathcal{C}$ if for all $n \in \omega$ there is some $M \in \mathcal{C}$ and sequences $(\overline{a}_i)_{i<n}$ and $(\overline{b}_i)_{i<n}$ of tuples from M such that:
$$M \vDash \phi(\overline{a}_i;\overline{b}_j) \text{ if, and only if } i < j.$$
2. The *Independence Property* in $\mathcal{C}$ if for all $n \in \omega$ there is some $M \in \mathcal{C}$ and sequences $(\overline{a}_i)_{i<n}$ and $(\overline{b}_I)_{I\subseteq[n]}$ of tuples from M such that:
$$M \vDash \phi(\overline{a}_i;\overline{b}_I) \text{ if, and only if } i \in I.$$
3. The *Strict Order Property* in $\mathcal{C}$ if for all $n \in \omega$ there is some $M \in \mathcal{C}$ and a sequence $(\overline{a}_i)_{i<n}$ of tuples from $M^{|\overline{y}|}$ such that:
$$M \vDash (\exists \overline{x})(\neg\phi(\overline{x};\overline{a}_i) \wedge \phi(\overline{x};\overline{a}_j)) \text{ if, and only if } i < j.$$

*Supported by a George and Marie Vergottis Scholarship awarded by Cambridge Trust, an Onassis Foundation Scholarship, and an EPSRC fees-only studentship

†Supported by a Leeds Doctoral College Scholarship

We say that $\mathcal{C}$ is *stable* if no formula has the order property in $\mathcal{C}$. We say that $\mathcal{C}$ is *NIP* (No Independence Property) if no formula has the independence property in $\mathcal{C}$. Similarly, we say that $\mathcal{C}$ is *NSOP* (No Strict Order Property) if no formula has the strict order property in $\mathcal{C}$.

The notions defined above were introduced by Shelah [11] in the context of first-order theories, rather than classes of structures. For instance, the definition of NSOP, in this context is presented below. Stability and NIP for complete theories are defined analogously.

Definition 1.2. Let T be a complete theory in a first-order language $\mathcal{L}$. We say that an $\mathcal{L}$-formula $\phi(\overline{x};\overline{y})$ has the *Strict Order Property over* T if there is some $M \models T$ and a sequence $(\overline{a}_i)_{i\in\omega}$ of tuples from $M^{|\overline{y}|}$ such that:

$$M \vDash (\exists \overline{x})(\neg\phi(\overline{x};\overline{a}_i) \wedge \phi(\overline{x};\overline{a}_j)) \text{ if, and only if } i < j.$$

We say that T is *NSOP* if no formula has the strict order property over T.

It follows that a class of $\mathcal{L}$-structures $\mathcal{C}$ is NSOP (resp. stable, NIP) if, and only if all completions of the common theory of $\mathcal{C}$, which we will denote by $\mathrm{Th}(\mathcal{C})$, are NSOP (resp. stable, NIP), as a consequence of compactness.

We extend our analysis to dividing lines in monadic expansions of theories/classes of structures. Given a class of $\mathcal{L}$-structures $\mathcal{C}$, by a *monadic expansion of* $\mathcal{C}$ we mean a class $\mathcal{C}'$ of $\mathcal{L}'$-structures, where $\mathcal{L}'$ is an expansion of $\mathcal{L}$ by a finite set of unary predicates, containing exactly one $\mathcal{L}'$-expansion of each structure in $\mathcal{C}$. We will call an $\mathcal{L}$-theory T *monadically NSOP* (resp. *monadically stable / monadically NIP*) if all complete extensions of T in $\mathcal{L}' = \mathcal{L} \cup \{P_1, \ldots, P_n\}$, where each P_i is a unary predicate, remain NSOP (resp. stable/NIP).

Baldwin and Shelah [2] and later Shelah [10] studied monadic stability and monadic NIP. Evidently, in their work it is shown that monadic NIP gives a completely new dividing-line, but it turns out that the same is not true of monadic NSOP.

Lemma 1.1. *A theory T is monadically NSOP if and only if T is monadically stable.*

Proof. Suppose that T is not monadically stable. Then there is an expansion of the language $\mathcal{L}' \supseteq \mathcal{L}$ by finitely many unary predicates and some $\mathcal{L}'$-structure $M \models T$ which has the order property. By a standard result, this is witnessed by a $\mathcal{L}'$-formula $\phi(x;\overline{y})$ and $(a_i)_{i\in\omega}, (\bar{b}_i)_{i\in\omega} \in M$. Add a unary predicate P for $A = \{a_i : i \in \omega\} \subseteq M$. Then, it is easy to see that

$$M \models (\exists x)(\neg(\phi(x;\bar{b}_i) \wedge P(x)) \wedge (\phi(x;\bar{b}_j) \wedge P(x))) \text{ if, and only if } i < j,$$

and so T has monadic SOP. Conversely, Stable $\implies$ NSOP still holds in the monadic case. □

We recall the following standard lemma.

Lemma 1.2. *Let $\mathcal{C}$ be a class of $\mathcal{L}$-structures. Then, the following are equivalent for an $\mathcal{L}$-structure A: (i) $A \vDash \mathrm{Th}(\mathcal{C})$; (ii) Every first-order $\mathcal{L}$-sentence true in A is true in some $M \in \mathcal{C}$; (iii) There exists $\{M_i : i \in I\} \subseteq \mathcal{C}$ and an ultrafilter $\mathcal{F}$ on I such that $A \equiv \prod_I M_i/\mathcal{F}$; (iv) There exists an ultrafilter $\mathcal{U}$ on $\mathcal{C}$ such that $A \equiv \prod_{M\in\mathcal{C}} M/\mathcal{U}$*

The above lemma therefore implies that being tame depends entirely on the ultraproducts of the class over different ultrafilters. This allows us to argue about the properties of a class by looking at very common, well-studied theories.

Example. Recall the definition of Paley graphs: Let $q = p^n$ be of prime power with $q \equiv 1 \pmod 4$. We define the *Paley Graph on q vertices*, P_q, to be the graph with vertices the elements of the finite field of q elements, $\mathbb{F}_q$, and edges xEy if and only if $x \neq y$ and $(x - y)$ is a square.

Using a deep theorem of Bollobás and Thomason [3], one can show that ultraproducts of Paley graphs over non-principal ultrafilters are elementarily equivalent to the *random graph*. Recall that the random graph is the Fraïssé limit of the class of all finite graphs. It is ultrahomogeneous, $\aleph_0$-categorical, and has quantifier elimination (QE).

Lemma 1.3. *Let $I = \{q \in \mathbb{N} : q \text{ is a prime power and } q \equiv 1 \pmod 4\}$, and $\mathcal{U}$ be a non-principal ultrafilter on I. Then $R := \prod_{q \in I} P_q/\mathcal{U}$ is a model of the theory of the random graph.*

Since each finite half graph embeds into the random graph, the infinite half graph embeds into the random graph by compactness and $\aleph_0$-categoricity, and hence the edge relation is unstable in the random graph. However, it does not have the strict order property, and in fact no formula does as it is well known that RG is NSOP. This trivially implies the following.

Corollary 1.1. *The class **Pal** of all Paley graphs is NSOP and unstable.*

In the next sections we investigate how additional assumptions on the class affect the relationship between stability and NSOP, with a focus on different closedness conditions and amalgamation.

Definition 1.3. Let $\mathcal{L}$ be a relational language and $\mathcal{C}$ be a class of $\mathcal{L}$-structures. We say that $\mathcal{C}$ is *monotone* if it is closed under weak substructures, that is, if $(M, R_i)_{i \in I} \in \mathcal{C}$ then $(M', R'_i)_{i \in I} \in \mathcal{C}$ for any $M' \subseteq M$ and $R'_i \subseteq R_i \cap (M')^{\mathrm{ar}(R_i)}$. We also say that $\mathcal{C}$ is *hereditary* if it is closed under substructures, that is, if $N \in \mathcal{C}$ and $f : M \hookrightarrow N$ is an embedding then $M \in \mathcal{C}$.

2 Monotone Classes

We briefly discuss monotone classes of graphs. The key to the relationship between stability and NSOP is the notion of *superflatness* from structural graph theory.

Definition 2.1. A class $\mathcal{C}$ of graphs is *superflat* if for all $r \in \mathbb{N}$ there is $m \in \mathbb{N}$ such that for all $G \in \mathcal{C}$ we have that the r-subdivision of the complete graph on m vertices, K^r_m, is not an induced subgraph of G.

The following is a well-known result, essentially due to Podewski and Ziegler [9].

Theorem 2.1. *Let $\mathcal{C}$ be a class of graphs. If $\mathcal{C}$ is superflat, then it is stable.*

Furthermore, Adler and Adler show in [1] that if $\mathcal{C}$ is monotone then the converse also holds. In fact, they show that monotone classes of graphs are superflat if and only if they are stable if and only if they are NIP. This collapse of NIP to stability for monotone graph classes is also true in the case of NSOP classes. The proof is essentially an adaptation of their argument.

Proposition 2.1. *Let $\mathcal{C}$ be a monotone class of graphs. If $\mathcal{C}$ is NSOP then $\mathcal{C}$ is superflat.*

Proof. We show the contrapositive. Indeed, suppose that $\mathcal{C}$ is not superflat. Then there exists $r \in \mathbb{N}$ such that for all $m \in \mathbb{N}$ there is $G \in \mathcal{C}$ for which K^r_m is an induced subgraph of G. Let H_n denote the half graph on $2n$ vertices and write H^r_n for the r-subdivision of H_n. Clearly H^r_n is a subgraph K^r_{2n-1} so by monotonicity $H^r_n \in \mathcal{C}$ for all $n \in \mathbb{N}$. Let $\phi(x, y)$ express that "there is a path of length exactly $r + 1$ from x to y". It follows that:

$$H^r_n \vDash (\exists x)(\phi(x, j) \wedge \neg\phi(x, i)) \text{ if, and only if } i < j$$

and hence $\mathcal{C}$ has the strict order property. □

In [1] the results for NIP in monotone classes of graphs are extended to monotone classes of binary structures. This can be easily done for NSOP as well. This hints to the fact that from a model-theoretic perspective, superflatness is a very strong tameness condition for monotone classes of binary structures. Naturally, one may ask if there is an analogue of superflatness for hypergraphs that has the same consequences. We briefly touch upon this in the Section 4.

Question 1. Is there a notion of sparsity for arbitrary relational structures, generalising the notion of superflatness for graphs, which in the case of monotone classes implies strong model theoretic tameness conditions?

It is not known if this collapse of NSOP to stability occurs for hereditary classes as well. In the case of NIP this is trivially false: the class of finite linear orders is hereditary, (monadically) NIP, but not stable. We further explore this question in the next section.

3 Hereditary Classes

So far we have not been able to find any examples of NSOP and unstable hereditary classes of relational structures. In fact, we conjecture that for hereditary classes $\mathcal{C}$ of relational structures, $\mathcal{C}$ is stable if and only if $\mathcal{C}$ is NSOP. We firstly restrict the scope of the question to hereditary classes with the *joint embedding property*. These are exactly the classes of finite induced substructures of countable structures.

Definition 3.1. Let M be an $\mathcal{L}$-structure. We denote by $\mathrm{Age}(T)$ the class of isomorphism types of finitely generated substructures of some $M \models T$.

This definition does not depend on the choice of M. Recall that a structure M is said to be *ultrahomogeneous* if every isomorphism between two finitely generated substructures extends to an automorphism of M. By Fraïssé, the following holds for complete theories in finite relational languages.

Fact. T is $\aleph_0$-categorical and has QE if and only if all countable $M \models T$ are ultrahomogeneous.

It is clear that passing down to substructures can be achieved by adding unary predicates. Hence if a theory is monadically tame then $\mathrm{Age}(T)$ is also tame. By proving an equivalent characterisation of monadic NIP in terms of *tuple-encoding configurations* Braunfeld and Laskowski have recently shown in [4] that the converse also holds for relational theories with QE.

Theorem 3.1 (Braunfeld, Laskowski)**.** *Let T be a complete theory in a relational language with finitely many constants and QE. Then* $\mathrm{Age}(T)$ *is NIP if and only if T is monadically NIP, and* $\mathrm{Age}(T)$ *is stable if and only if T is monadically stable.*

A natural question to ask is whether the same holds for the strict order property. Since monadic NSOP is in fact the same as monadic stability, we are led to the following equivalent question.

Question 2. *Let T be a theory in a relational language with finitely many constants and QE. Does it hold that* $\mathrm{Age}(T)$ *is stable if and only if* $\mathrm{Age}(T)$ *is NSOP?*

Motivated by this, we further restrict our investigation to ages of countable ultrahomogeneous structures. As mentioned previously, the complete theories of ultrahomogeneous relational structures have QE and are $\aleph_0$-categorical. In this light, we are specialising the context of our question to the case of $\aleph_0$-categorical structures, or equivalently, to hereditary classes with the joint embedding property and the amalgamation property. For this, we use the classification programme of ultrahomogeneous structures.

Homogeneous Graphs In the case of graphs, we have the following classification result proved by Lachlan and Woodrow in [7].

Theorem 3.2 (Lachlan, Woodrow)**.** *Every countably infinite ultrahomogeneous undirected graph is isomorphic to one of the following or their complement (i) The random graph RG; (ii) The generic graph $\mathcal{R}_n$ for the class of all countable K_n-free graphs for a given $n \geq 3$; or (iii) The disjoint union of m copies of K_n, where $m, n \leq \omega$ and at least one of m or n is ω.*

This allows us to prove the following.

Theorem 3.3. *Let G be a countable ultrahomogeneous undirected graph. Then* Age(G) *is stable if and only if* Age(G) *is NSOP.*

Proof. We use Lachlan and Woodrow's classification result to argue in a case-by-case manner. The class of all finite graphs contains the finite half graphs which witness the SOP in Age(RG). Since RG is isomorphic to its complement, this finishes the first case.

All classes defined by forbidden substructures are by definition monotone, and hence NSOP collapses to stability. This applies to $\mathcal{R}_n$. Consider the complement of $\mathcal{R}_n$, the generic K_n-full graph, which we will denote by $\mathcal{F}_n$. We show that Age$(\mathcal{F}_n)$ has the strict order property. Indeed, let G_k be with vertices $\{1, \dots, k\}$ and k disjoint copies of K_n: $A_1, A_2, \dots A_k$. We connect the singleton i with all the vertices of A_j if and only if $i \leq j$. Let $\phi(x, y_1, \dots, y_n) = \bigwedge_{i=1}^{n} E(x, y_i)$. For the sequence $(\bar{a}_i)_{i \in k}$ with $\bar{a}_i = A_i$, it holds that $G_k \models \exists x(\neg\phi(x, \bar{a}_i) \wedge \phi(x, \bar{a}_j))$ if, and only if $i < j$, and since $G_k \in$ Age(F_n) for all $k \in \omega$, Age(F_n) has SOP.

Finally, if G is a disjoint union of ω copies of K_n for a fixed $n \in \omega$, then we have that Age$(G) = \{K_{i_1} \sqcup K_{i_2} \sqcup \dots \sqcup K_{i_k} : k \in \omega, i_1, \dots i_k \in n\}$. This is trivially superflat, and hence stable. The complement G^c is the Turán graph $T(n, \omega)$, the complete ω-partite graph with parts of size n. This has the property that whenever two vertices are not connected by an edge, they must belong to the same part. It follows that G^c is P_4-free, and so Age(G^c) contains only P_4-free graphs. By results in [8] we can see that this class has bounded *rankwidth* and is therefore stable if and only if it has stable edge relation. Since the finite half-graphs are not members of Age(G^c) this class is therefore stable. Essentially the same argument applies to the other two possibilities in the third case.

□

Homogeneous Tournaments. We continue with *tournaments*, a special case of digraphs, which have the property that for any two vertices v, u there exists a single edge connecting them, that is, a graph G is a tournament if is satisfies:

$$(\forall v)(\forall u)((E(u, v) \vee E(v, u)) \wedge (E(u, v) \to \neg E(v, u))).$$

In [6], Lachlan proved the following classification result, which is analogous to Theorem 3.2.

Theorem 3.4 (Lachlan)**.** *Every countably infinite homogeneous tournaments is isomorphic to one of the following: (i) The* Universal Tournament*, which we will denote as T^∞; (ii) The* Dense Linear Order *with no endpoints, which we will denote as $(\mathbb{Q}, <)$; or (iii) The* Dense Local Order *which we will denote as $S(2)$.*

With this classification result, we deduce the following.

Theorem 3.5. *Let $\mathcal{T}$ be a countable homogeneous tournament. Then* Age$(\mathcal{T})$ *is stable if and only if* Age$(\mathcal{T})$ *is NSOP.*

Proof. Case-by-case argument akin to the proof of Theorem 3.3. □

Narrowing down Question 2, we are interested in the following:

Question 3. *Let M be an ultrahomogeneous relational structure. Is it true that* $\mathrm{Age}(M)$ *is NSOP if and only if* $\mathrm{Age}(M)$ *is stable?*

4 Future Work

The exploration of NSOP in classes of structures raises many interesting questions. We conclude by briefly discussing some possible future directions.

Firstly, much as in [1], we are interested in seeing if the collapse of NSOP (and NIP) to stability extends to monotone classes in arbitrary relational languages. The bridge to this collapse for the case of graphs was superflatness. We believe that by extending this notion to uniform hypergraphs, and interpreting classes of relational structures in classes of these, we can answer this question positively.

As discussed above, we also believe that this collapse of NSOP to stability applies to hereditary classes as well. This would extend Theorem 3.1 to account for NSOP.

Finally, a question we have not mentioned yet concerns the structural properties of classes of (possibly dense) graphs with NSOP. The intuition behind NSOP is that it allows for random configurations and it would be interesting to examine the interplay between NSOP and various known notions of pseudorandomness from combinatorial graph theory, as in [5].

References

[1] Hans Adler and Isolde Adler. Interpreting nowhere dense graph classes as a classical notion of model theory. *European Journal of Combinatorics*, 36:322–330, 02 2014.

[2] John Baldwin and Saharon Shelah. Second-order quantifiers and the complexity of theories. *Notre Dame Journal of Formal Logic*, 29, 07 1985.

[3] Béla Bollobás and Andrew Thomason. Graphs which contain all small graphs. *European Journal of Combinatorics*, 2(1):13–15, 1981.

[4] Samuel Braunfeld and Michael Laskowski. Characterizations of monadic NIP. *Transactions of the American Mathematical Society, Series B*, 8(30):948–970, Nov 2021.

[5] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, December 1989.

[6] Alistair H Lachlan. Countable homogeneous tournaments. *Transactions of the American Mathematical Society*, 284(2):431–461, 1984.

[7] Alistair H. Lachlan and Robert E. Woodrow. Countable ultrahomogeneous undirected graphs. *Transactions of the American Mathematical Society*, 262(1):51–94, 1980.

[8] Jaroslav Nesetril, Patrice Ossona de Mendez, Michal Pilipczuk, Roman Rabinovich, and Sebastian Siebertz. Rankwidth meets stability. *CoRR*, abs/2007.07857, 2020.

[9] Klaus-Peter Podewski and Martin Ziegler. Stable graphs. *Fundamenta Mathematicae*, 100:101–107, 1978.

[10] Saharon Shelah. *Monadic logic: Hanf Numbers*, pages 203–223. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986.

[11] Saharon Shelah. *Classification Theory and the Number of Non-Isomorphic Models (second ed.)*. North Holland, 1990.

[12] Pierre Simon and Szymon Toruńczyk. Ordered graphs of bounded twin-width. *arXiv preprint arXiv:2102.06881*, 2021.

Proof-relevance in Bishop-style constructive mathematics

Iosif Petrakis

Ludwig-Maximilians-Universität, Munich Germany
petrakis@math.lmu.de

Abstract

Bishop's presentation of his informal system of constructive mathematics BISH was on purpose closer to the proof-irrelevance of classical mathematics, although a form of proof-relevance was evident in the use of several notions of moduli (of convergence, of uniform continuity, of uniform differentiability etc.). Based on Bishop Set Theory (BST), a reconstruction of Bishop's theory of sets developed in [8], we associate to many formulas ϕ of BISH a set $\mathtt{Prf}(\phi)$ of "proofs" or witnesses of ϕ, providing in this way a BHK-interpretation of a large part of BISH within BST. Abstracting from several examples of totalities in BISH we define the notion of a set with a proof-relevant equality, and of a Martin-Löf set, a special case of the former, the equality of which corresponds (partially) to the equality type of a type in intensional Martin-Löf Type Theory (MLTT). Notions and facts of MLTT and its extensions (either with the axiom of function extensionality, or with Vooevodsky's axiom of univalence) are translated into BST. While BST is standardly understood through its translation to MLTT, a partial translation in the converse direction is shown to be possible.

1 On Bishop Set Theory (BST)

Bishop set theory (BST), elaborated in [8], is an informal, constructive theory of totalities and assignment routines that serves as a "completion" of Bishop's original theory of sets in [1, 2]. Its first aim is to fill in the "gaps", or highlight the fundamental notions that were suppressed by Bishop in his account of the set theory underlying BISH. Its second aim is to serve as an intermediate step between Bishop's theory of sets and an *adequate* and *faithful* formalisation of BISH in Feferman's sense [4]. To assure faithfulness, we use concepts or principles that appear, explicitly or implicitly, in BISH. The following features of BST in [8] "complete" Bishop's theory of sets.

1. Explicit use of a universe of sets.
2. Clear distinction between sets and proper classes.
3. Explicit use of dependent operations.
4. Elaboration of the theory of families of sets.

Here we apply the general theory of families of sets, in order *to reveal proof-relevance in* BISH. For all notions and results that are used without explanation or proof we refer mainly to [8, 10], and also to [6, 9, 12].

The set $\mathbb{N}$ of natural numbers is a primitive set in BST. The universe $\mathbb{V}_0$ of (predicative) sets is a special "open-ended" defined totality, which is not defined through a membership-construction, but in an open-ended way. When we say that a defined totality X is a set we "introduce" X as an element of $\mathbb{V}_0$, but we do not use a corresponding induction principle for $\mathbb{V}_0$. The universe $\mathbb{V}_0$ is not a set, but a proper class. To define a set X in $\mathbb{V}_0$ we need to provide a membership-construction for it (without appealing to $\mathbb{V}_0$) and to define an equality formula $x =_X x'$ that satisfies the properties of an equivalence relations. A formula $P(x)$ on a set X is

an *extensional property* on X, if $\forall_{x,y\in X}\big([x =_X y \;\&\; P(x)] \Rightarrow P(y)\big)$. The totality X_P generated by $P(x)$ is defined by $x \in X_P :\Leftrightarrow x \in X \;\&\; P(x)$, and the equality of X_P is inherited from $=_X$. We also write $X_P := \{x \in X \mid P(x)\}$, and we call X_P the *extensional subset* of X generated by P. E.g., the *diagonal* of a set $(X, =_X)$ is the following extensional subset of $X \times X$

$$D(X) := \{(x, y) \in X \times X \mid x =_X y\}.$$

If X, Y are totalities, a *non-dependent assignment routine* f from X to Y, in symbols $f\colon X \rightsquigarrow Y$, is a finite routine that assigns an element y of Y to each given element x of X. The membership-condition for their totality is considered to be primitive in BST, and their equality is defined pointwise, w.r.t. a given equality on Y. If X, Y are sets, a *function* from X to Y, in symbols $f\colon X \to Y$, is an assignment routine from X to Y that respects equality i.e.,

$$\forall_{x,x'\in X}\big(x =_X x' \Rightarrow f(x) =_Y f(x')\big).$$

Their set is denoted by $\mathbb{F}(X, Y)$. The canonical equality on $\mathbb{V}_0$ is defined by

$$X =_{\mathbb{V}_0} Y :\Leftrightarrow \exists_{f\in\mathbb{F}(X,Y)}\exists_{g\in\mathbb{F}(Y,X)}\big(g \circ f = \mathrm{id}_X \;\&\; f \circ g = \mathrm{id}_Y\big).$$

In this case we write $(f, g) : X =_{\mathbb{V}_0} Y$. If $X, Y \in \mathbb{V}_0$ such that $X =_{\mathbb{V}_0} Y$, we define the set

$$\mathtt{PrfEql}_0(X, Y) := \big\{(f, g) \in \mathbb{F}(X, Y) \times \mathbb{F}(Y, X) \mid (f, g) : X =_{\mathbb{V}_0} Y\big\}$$

of all objects that "witness", or "realise", or prove the equality $X =_{\mathbb{V}_0} Y$. The equality of $\mathtt{PrfEql}_0(X, Y)$ is the canonical one i.e., $(f, g) =_{\mathtt{PrfEql}_0(X,Y)} (f', g') :\Leftrightarrow f =_{\mathbb{F}(X,Y)} f' \;\&\; g =_{\mathbb{F}(Y,X)} g'$. In general, not all elements of $\mathtt{PrfEql}_0(X, Y)$ are equal. As in [14], Example 3.1.9, if $X := Y := \mathbf{2} := \{0, 1\} := \{x \in \mathbb{N} \mid x =_{\mathbb{N}} 0 \;\vee\; x =_{\mathbb{N}} 1\}$, then $(\mathrm{id}_{\mathbf{2}}, \mathrm{id}_{\mathbf{2}}) \in \mathtt{PrfEql}_0(\mathbf{2}, \mathbf{2})$, and if $\mathtt{sw}_{\mathbf{2}} : \mathbf{2} \to \mathbf{2}$ maps 0 to 1 and 1 to 0, then $(\mathtt{sw}_{\mathbf{2}}, \mathtt{sw}_{\mathbf{2}}) \in \mathtt{PrfEql}_0(\mathbf{2}, \mathbf{2})$, while $\mathtt{sw}_{\mathbf{2}} \neq \mathrm{id}_{\mathbf{2}}$. It is expected that the proof-terms in $\mathtt{PrfEql}_0(X, Y)$ are compatible with the properties of the equivalence relation $X =_{\mathbb{V}_0} Y$. This means that we can define a distinguished proof-term $\mathtt{refl}(X) \in \mathtt{PrfEql}_0(X, X)$ that proves the reflexivity of $X =_{\mathbb{V}_0} Y$, an operation $^{-1}$, such that if $(f, g) : X =_{\mathbb{V}_0} Y$, then $(f, g)^{-1} : Y =_{\mathbb{V}_0} X$, and an operation of "composition" $*$ of proof-terms, such that if $(f, g) : X =_{\mathbb{V}_0} Y$ and $(h, k) : Y =_{\mathbb{V}_0} Z$, then $(f, g) * (h, k) : X =_{\mathbb{V}_0} Z$. Let

$$\mathtt{refl}(X) := \big(\mathrm{id}_X, \mathrm{id}_X\big) \;\;\&\;\; (f, g)^{-1} := (g, f) \;\;\&\;\; (f, g) * (h, k) := (h \circ f, g \circ k).$$

It is immediate to see that these operations satisfy the *groupoid laws* and *compatibility condition*:
(i) $\mathtt{refl}(X) * (f, g) =_{\mathtt{PrfEql}_0(X,Y)} (f, g)$ and $(f, g) * \mathtt{refl}(Y) =_{\mathtt{PrfEql}_0(X,Y)} (f, g)$.
(ii) $(f, g) * (f, g)^{-1} =_{\mathtt{PrfEql}_0(X,X)} \mathtt{refl}(X)$ and $(f, g)^{-1} * (f, g) =_{\mathtt{PrfEql}_0(Y,Y)} \mathtt{refl}(Y)$.
(iii) $\big((f, g) * (h, k)\big) * (s, t) =_{\mathtt{PrfEql}_0(X,W)} (f, g) * \big((h, k) * (s, t)\big)$.
(iv) If $(f, g), (f', g') \in \mathtt{PrfEql}_0(X, Y)$, $(h, k), (h', k') \in \mathtt{PrfEql}_0(Y, Z)$, and if $(f, g) =_{\mathtt{PrfEql}_0(X,Y)} (f', g')$ and $(h, k) =_{\mathtt{PrfEql}_0(Y,Z)} (h', k')$, then $(f, g) * (h, k) =_{\mathtt{PrfEql}_0(X,Z)} (f', g') * (h', k')$.
Let I be a set and $\lambda_0\colon I \rightsquigarrow \mathbb{V}_0$ a non-dependent assignment routine from I to $\mathbb{V}_0$. A *dependent operation* Φ over λ_0, in symbols

$$\Phi\colon \bigdownarrow_{i\in I} \lambda_0(i),$$

is an assignment routine that assigns to each element i in I an element $\Phi(i) =: \Phi_i$ in the set $\lambda_0(i)$. The membership-condition for their totality $\mathbb{A}(I, \lambda_0)$ is primitive in BST, and its equality is defined pointwise w.r.t. the given equalities of each $\lambda_0(i)$.

Definition 1. *If I is a set, a family of sets indexed by I, or an I-family of sets, is a pair $\Lambda := (\lambda_0, \lambda_1)$, where $\lambda_0 \colon I \rightsquigarrow \mathbb{V}_0$, and λ_1, a modulus of function-likeness for λ_0, is given by*

$$\lambda_1 \colon \bigwedge_{(i,j)\in D(I)} \mathbb{F}\big(\lambda_0(i), \lambda_0(j)\big), \quad \lambda_1(i,j) := \lambda_{ij}, \quad (i,j) \in D(I),$$

such that the transport maps λ_{ij} of Λ satisfy the following conditions:

(a) *For every $i \in I$, we have that $\lambda_{ii} := \mathrm{id}_{\lambda_0(i)}$.*

(b) *If $i =_I j$ and $j =_I k$, the following diagram commutes*

$$\begin{array}{ccc} \lambda_0(i) & & \\ {\scriptstyle \lambda_{ij}}\big\downarrow & \searrow^{\lambda_{ik}} & \\ \lambda_0(j) & \xrightarrow[\lambda_{jk}]{} & \lambda_0(k). \end{array}$$

I is the index-set of the family Λ. If X is a set, the constant I-family of sets X is the pair $C^X := (\lambda_0^X, \lambda_1^X)$, where $\lambda_0(i) := X$, for every $i \in I$, and $\lambda_1(i,j) := \mathrm{id}_X$, for every $(i,j) \in D(I)$. The exterior union, or disjoint union, or the Sigma-set $\sum_{i\in I} \lambda_0(i)$ of Λ, and its canonical equality are defined by

$$w \in \sum_{i\in I} \lambda_0(i) :\Leftrightarrow \exists_{i\in I} \exists_{x\in\lambda_0(i)} \big(w := (i,x)\big),$$

$$(i,x) =_{\sum_{i\in I}\lambda_0(i)} (j,y) :\Leftrightarrow i =_I j \ \&\ \lambda_{ij}(x) =_{\lambda_0(j)} y.$$

The Sigma-set of the $\mathbf{2}$-family $\Lambda^{\mathbf{2}}$ of the sets X and Y, where $\lambda_0^{\mathbf{2}}(0) := X$, $\lambda_0^{\mathbf{2}}(1) := Y$, $\lambda_1^{\mathbf{2}}(0,0) := \mathrm{id}_X$ and $\lambda_1^{\mathbf{2}}(1,1) := \mathrm{id}_Y$, is the coproduct of X and Y, and we write $X + Y := \sum_{i\in\mathbf{2}} \lambda_0^{\mathbf{2}}(i)$. The totality $\prod_{i\in I} \lambda_0(i)$ of dependent functions over Λ, or the Pi-set of Λ, is defined by

$$\Theta \in \prod_{i\in I} \lambda_0(i) :\Leftrightarrow \Theta \in \mathbb{A}(I, \lambda_0) \ \&\ \forall_{(i,j)\in D(I)} \big(\Theta_j =_{\lambda_0(j)} \lambda_{ij}(\Theta_i)\big),$$

and it is equipped with the canonical equality of the set $\mathbb{A}(I, \lambda_0)$.

Proposition 2 (Membership-with-Evidence I (MwE-I))**.** *Let X, Y be sets, and let $P(x)$, where*

$$P(x) :\Leftrightarrow \exists_{p\in Y} \big(Q(x,p)\big),$$

and $Q(x,p)$ is an extensional property on $X \times Y$. Let $\mathtt{PrfMemb}_0^P : X \rightsquigarrow \mathbb{V}_0$, defined by

$$\mathtt{PrfMemb}_0^P(x) := \{p \in Y \mid Q(x,p)\},$$

for every $x \in X$, and let $\mathtt{PrfMemb}_1^P : \bigwedge_{(x,x')\in D(X)} \mathbb{F}\big(\mathtt{PrfMemb}_0^P(x), \mathtt{PrfMemb}_0^P(x')\big)$, where $\mathtt{PrfMemb}_{xx'}^P := \mathtt{PrfMemb}_1^P(x,x') : \mathtt{PrfMemb}_0^P(x) \to \mathtt{PrfMemb}_0^P(x')$ is defined by the identity map-rule $\mathtt{PrfMemb}_{xx'}^P(p) := p$, for every $p \in \mathtt{PrfMemb}_0^P(x)$ and every $(x,x') \in D(X)$.

(i) *The property $P(x)$ is extensional.*

(ii) *The pair $\mathtt{PrfMemb}^P := \big(\mathtt{PrfMemb}_0^P, \mathtt{PrfMemb}_1^P\big) \in \mathtt{Fam}(X)$.*

Extending the operations between sets to operations between families of sets, we define the following BHK-interpretation of various formulas of BISH within BST.

Definition 3 (BHK-interpretation of BISH in BST - Part I)**.** *Let membership conditions $x \in X_P$ as e.g., in Proposition 2. We define*

$$\mathtt{Prf}(x \in X_P) := \mathtt{PrfMemb}_0^P(x).$$

Let ϕ, ψ be formulas in BISH *such that $\mathtt{Prf}(\phi)$ and $\mathtt{Prf}(\psi)$ are already defined. We define*

$$\mathtt{Prf}(\phi \ \& \ \psi) := \mathtt{Prf}(\phi) \times \mathtt{Prf}(\psi),$$

$$\mathtt{Prf}(\phi \vee \psi) := \mathtt{Prf}(\phi) + \mathtt{Prf}(\psi),$$

$$\mathtt{Prf}(\phi \Rightarrow \psi) := \mathbb{F}\big(\mathtt{Prf}(\phi), \mathtt{Prf}(\psi)\big).$$

Let $\phi(x)$ be a formula on a set X, and let $\mathtt{Prf}^{\phi} := \big(\mathtt{Prf}_0^{\phi}, \mathtt{Prf}_1^{\phi}\big) \in \mathtt{Fam}(X)$, where $\mathtt{Prf}_0^{\phi} \colon X \rightsquigarrow \mathbb{V}_0$ is given by the rule $x \mapsto \mathtt{Prf}_0^{\phi}(x) := \mathtt{Prf}\big(\phi(x)\big)$, for every $x \in X$. The $\mathtt{Prf}$-sets of the formulas $\forall_{x \in X}\phi(x)$ and $\exists_{x \in X}\phi(x)$ with respect to the given family $\mathtt{Prf}^{\phi}$, where $\exists_{x \in X}\phi(x)$ is a formula that does not express a membership condition or a relation, are defined by

$$\mathtt{Prf}\Big(\forall_{x \in X}\phi(x)\Big) := \prod_{x \in X} \mathtt{Prf}_0^{\phi}(x) := \prod_{x \in X} \mathtt{Prf}\big(\phi(x)\big),$$

$$\mathtt{Prf}\Big(\exists_{x \in X}\phi(x)\Big) := \sum_{x \in X} \mathtt{Prf}_0^{\phi}(x) := \sum_{x \in X} \mathtt{Prf}\big(\phi(x)\big).$$

Due to the definition of the coproduct in Definition 1, the $\mathtt{Prf}$-sets for $\exists_{x \in X}\phi(x)$ and for $\forall_{x \in X}\phi(x)$ are generalisations of $\mathtt{Prf}$-sets for $\phi \vee \psi$ and for $\phi \ \& \ \psi$, respectively.

Example 1.1. Let the fact: if $(x_n)_{n \in \mathbb{N}^+} \in \mathbb{F}(\mathbb{N}^+, \mathbb{R})$ and $x_0 \in \mathbb{R}$, then

$$x_n \xrightarrow{n} x_0 \Rightarrow (x_n)_{n \in \mathbb{N}^+} \text{ is Cauchy.}$$

If $\chi(x_n, x_0)$ is the above implication, then $\chi(x_n, x_0)$ of the form $\phi(x_n, x_0) \Rightarrow \psi(x_n)$. Its proof (see [2], p. 29) can be seen as a rule that sends a modulus of convergence $C \colon x_n \xrightarrow{n} x_0$ of $(x_n)_{n \in \mathbb{N}^+}$ at x_0 to a modulus of Cauchyness $D \colon \mathrm{Cauchy}\big((x_n)_{n \in \mathbb{N}^+}\big)$ for $(x_n)_{n \in \mathbb{N}^+}$, where $D(k) := C(2k)$, for every $k \in \mathbb{N}^+$. This operation from $\mathtt{PrfMemb}_0^{\mathrm{Conv}_{x_0}}\big((x_n)_{n \in \mathbb{N}^+}\big)$ to $\mathtt{PrfMemb}_0^{\mathrm{Cauchy}}\big((x_n)_{n \in \mathbb{N}^+}\big)$ is a function, and

$$\mathtt{Prf}(\chi(x_n, x_0)) := \mathbb{F}\Big(\mathtt{Prf}\big(\phi(x_n, x_0)\big), \mathtt{Prf}\big(\psi(x_n)\big)\Big),$$

$$\mathtt{Prf}\big(\phi(x_n, x_0)\big) := \mathtt{PrfMemb}_0^{\mathrm{Conv}_{x_0}}\big((x_n)_{n \in \mathbb{N}^+}\big), \qquad \mathtt{Prf}\big(\psi(x_n)\big) := \mathtt{PrfMemb}_0^{\mathrm{Cauchy}}\big((x_n)_{n \in \mathbb{N}^+}\big).$$

2 Martin-Löf sets

The universe $\mathbb{V}_0$, the proper class "powerset" $\mathcal{P}(X)$ of a set X, the impredicative set $\mathtt{Fam}(I)$ of families of sets indexed by I, the set $\mathtt{Fam}(I, X)$ of families of subsets of X indexed by I, are some of the many examples of totalities studied in [8] equipped with an equality defined through an existential formula. Next we introduce Martin-Löf sets, as a generalisation of these totalities, that helps us also to transfer results and concepts from MLTT [5] or HoTT [14] into BST. So far, only the transition of results and concepts from BISH to MLTT was considered.

Definition 4. *Let Y be a set, and $(X, =_X)$ a set with an equality condition of the form*

$$x =_X x' :\Leftrightarrow \exists_{p\in Y}\big(p : x =_X x'\big),$$

where $\theta^{xx'}(p) :\Leftrightarrow p : x =_X x'$ is an extensional property on Y. Let also the non-dependent assignment routine $\mathtt{PrfEql}_0^X : X \times X \rightsquigarrow \mathbb{V}_0$, defined by

$$\mathtt{PrfEql}_0^X(x, x') := \{p \in Y \mid p : x =_X x'\}; \qquad (x, x') \in X \times X,$$

together with dependent operations

$$\mathtt{refl}^X : \bigwedge_{x\in X} \mathtt{PrfEql}_0^X(x, x), \qquad {}^{-1_X} : \bigwedge_{x,x'\in X} \mathbb{F}\Big(\mathtt{PrfEql}_0^X(x, x'), \mathtt{PrfEql}_0^X(x', x)\Big),$$

$$*_X : \bigwedge_{x,x',x''\in X} \mathbb{F}\Big(\mathtt{PrfEql}_0^X(x, x') \times \mathtt{PrfEql}_0^X(x', x''), \mathtt{PrfEql}_0^X(x, x'')\Big).$$

We call the structure $\widehat{X} := (X, =_X, \mathtt{PrfEql}_0^X, \mathtt{refl}^X, {}^{-1_X}, *_X)$ a set with a proof-relevant equality. If X is clear from the context, we may omit the subscript X from the above dependent operations. We call $\widehat{X}$ a Martin-Löf set, if the groupoid laws and the compatibility condition, described for the universe $\mathbb{V}_0$ in section 1, hold.

If $\widehat{X}$ has a proof-relevant equality, by Definition 3 we get $\mathtt{Prf}(x =_X x') := \mathtt{PrfEql}_0^X(x, x')$.

Example 2.1. The set of families $\mathtt{Fam}(I, X)$ of subsets of the set X indexed by the set I is a Martin-Löf set. Similarly, one can show that $\mathtt{Fam}(I, \mathbf{X})$, the set of families of complemented subsets of the set X indexed by the set I (see section 4.9 in [8]), and $\mathtt{Fam}(I, X, Y)$, the set of families of partial functions from the set X to the set Y indexed by the set I (see section 4.8 in [8]) are Martin-Löf sets. We get trivial examples of Martin-Löf sets by using the truncation of a set (see Definition 8(iv)).

The *maps* between Martin-Löf sets and the notion of a family of sets over a set $\widehat{I}$ with a proof-relevant equality, and the corresponding Sigma- and Pi-sets, are defined in [10].

Proposition 5. *If $\widehat{\Lambda} := (\lambda_0, \mathtt{PrfEql}_0^I, \lambda_2)$ is a function-like family of sets over the Martin-Löf set $\widehat{I}$, then a structure of a Martin-Löf set is defined on $\widehat{\sum}_{i\in I}\lambda_0(i)$.*

Lemma 6. *Let $\widehat{X}$ be a Martin-Löf set, $x_0 \in X$ and let $\mathtt{PrfEql}_0^{x_0} : X \rightsquigarrow \mathbb{V}_0$ be defined by $x \mapsto \mathtt{PrfEql}_0^X(x, x_0)$, for every $x \in X$. Moreover, let*

$$\mathtt{PrfEql}_1^{x_0} : \bigwedge_{(x,y)\in D(X)} \bigwedge_{p\in \mathtt{PrfEql}_0^X(x,y)} \mathbb{F}\big(\mathtt{PrfEql}_0^X(x, x_0), \mathtt{PrfEql}_0^X(y, x_0)\big),$$

be defined, for every $(x, y) \in D(X)$, $p \in \mathtt{PrfEql}_0^X(x, y)$ and $r \in \mathtt{PrfEql}_0^X(x, x_0)$, by

$$\mathtt{PrfEql}_1^{x_0}\big((x, y), p\big) := \mathtt{PrfEql}_{xy}^{x_0} : \mathtt{PrfEql}_0^X(x, x_0) \to \mathtt{PrfEql}_0^X(y, x_0)$$

$$r \mapsto p^{-1} * r.$$

Then $\widehat{\mathtt{PrfEql}}^{x_0} := (\mathtt{PrfEql}_0^{x_0}, \mathtt{PrfEql}_1^{x_0})$ is a function-like family of sets over $\widehat{X}$.

Theorem 7. *Let $\widehat{X}$ be a proof-relevant set, $x_0 \in X$ and let $\widehat{\mathtt{PrfEql}}^{x_0} := (\mathtt{PrfEql}_0^{x_0}, \mathtt{PrfEql}_1^{x_0})$ be the function-like family of sets over $\widehat{X}$ from Lemma 6. Let $\widehat{\sum}_{x\in X}\mathtt{PrfEql}_0^X(x,x_0)$ be equipped with its canonical structure of a Martin-Löf set, according to Proposition 5. Then for every $(x,p) \in \widehat{\sum}_{x\in X}\mathtt{PrfEql}_0^X(x,x_0)$ we have that*

$$(x,p) =_{\widehat{\sum}_{x\in X}\mathtt{PrfEql}_0^X(x,x_0)} \big(x_0, \mathtt{refl}_{x_0}\big).$$

Theorem 7 is a translation of the type-theoretic contractibility of the singleton type (see [3]) into BST. If M is the term expressing this contractibility (see also [7]), Martin-Löf's J-rule trivially implies M, and it is equivalent to M and the transport. In BISH we do not have the J-rule, but we have transport in a definitional way only. As Theorem 7 indicates, a definitional form of M is provable in BST, although there is no translation of the J-rule in BST.

3 Contractible sets and subsingletons in BST

Definition 8. *Let $(X, =_X)$ be a set.*

(i) *X is inhabited, if $\exists_{x\in X}\big(x =_X x\big)$.*

(ii) *X is a singleton, or contractible, or a (-2)-set, if $\exists_{x_0\in X}\forall_{x\in X}\big(x_0 =_X x\big)$. In this case, x_0 is called a centre of contraction for X.*

(iii) *X is a subsingleton, or a mere proposition, or a (-1)-set, if $\forall_{x,y\in X}\big(x =_X y\big)$.*

(iv) *The truncation of $(X, =_X)$ is the set $(X, {}_{||=_X||})$, where*

$$x \;{}_{||=_X||}\; y :\Leftrightarrow x =_X x \;\&\; y =_X y.$$

We use the symbol $||X||$ to denote that the set X is equipped with the truncated equality ${}_{||=_X||}$.

Clearly, a set X is contractible iff X is an inhabited subsingleton iff $X =_{\mathbb{V}_0} \mathbf{1}$. If $f\colon X \to Y$ and $y \in Y$, the fiber $\mathtt{fib}^f(y)$ of f at y is defined by

$$\mathtt{fib}^f(y) := \{x \in X \mid f(x) =_Y y\},$$

and f is contractible, if $\mathtt{fib}^f(y)$ is contractible, for every $y \in Y$. If $f\colon X \to Y$, let $\mathtt{fib}^f\colon Y \rightsquigarrow \mathbb{V}_0$, with $y \mapsto \mathtt{fib}^f(y)$, for every $y \in Y$. If f is contractible, every fiber $\mathtt{fib}^f(y)$ is contractible. A *modulus of centres of contraction* for a contractible function f is a dependent operation $\mathtt{centre}^f\colon \bigwedge_{y\in Y}\mathtt{fib}^f(y)$, with $\mathtt{centre}^f_y := \mathtt{centre}^f(y)$ is a centre of contraction for $\mathtt{fib}^f(y)$.

Proposition 9. *Let $\Lambda := (\lambda_0, \lambda_1) \in \mathtt{Fam}(I)$.*

(i) *If $\Theta\colon \bigwedge_{i\in I}\lambda_0(i)$ is a modulus of centres of contraction for λ_0 i.e., Θ_i is a centre of contraction for $\lambda_0(i)$, then $\Theta \in \prod_{i\in I}\lambda_0(i)$ is a centre of contraction for $\prod_{i\in I}\lambda_0(i)$ and $\sum_{i\in I}\lambda_0(i) =_{\mathbb{V}_0} I$.*

(ii) *If $i_0 \in I$ is a centre of contraction for I, then $\sum_{i\in I}\lambda_0(i) =_{\mathbb{V}_0} \lambda_0(i_0)$.*

Proposition 10. *Let $\Lambda := (\lambda_0, \lambda_1) \in \mathtt{Fam}(I)$, $\Theta\colon \bigwedge_{i\in I}\lambda_0(i)$ a modulus of centres of contraction for λ_0, and X, Y sets.*

(i) *If $h\colon I \rightsquigarrow \sum_{i\in I}\lambda_0(i)$ is defined by $h(i) := \big(i, \Theta_i\big)$, for every $i \in I$, then h is a function and $(\mathtt{pr}_1^\Lambda, h)\colon \sum_{i\in I}\lambda_0(i) =_{\mathbb{V}_0} I$.*

(ii) *$\mathbb{F}\big(I, \sum_{i\in I}\lambda_0(i)\big) =_{\mathbb{V}_0} \mathbb{F}(I, I)$.*

(iii) *If X is contractible and Y is a retract (see [10], Definition 6) of X, then Y is contractible.*

Theorem 11. *Let* $\Lambda := (\lambda_0, \lambda_1) \in \mathtt{Fam}(I)$*, and let* $\Theta\colon \bigwedge_{i\in I} \lambda_0(i)$ *be a modulus of centres of contraction for* λ_0*. If* $(\phi, \theta)\colon \mathbb{F}\big(I, \sum_{i\in I} \lambda_0(i)\big) =_{\mathbb{V}_0} \mathbb{F}(I, I)$*, where* ϕ *and* θ *are defined in the proof of Proposition 10*(ii)*, then* $\prod_{i\in I} \lambda_0(i)$ *is a retract of* $\mathtt{fib}^{\phi}(\mathrm{id}_I)$*.*

Theorem 11 is the translation of Theorem 4.9.4 in book-HoTT, where in the hypothesis of the latter the universe is univalent. Next corollary is the translation in BST of the fact that the univalence axiom implies the principle of weak function extensionality.

Corollary 12. *If* $\Lambda := (\lambda_0, \lambda_1) \in \mathtt{Fam}(I)$ *and* $\Theta\colon \bigwedge_{i\in I} \lambda_0(i)$ *is a modulus of centres of contraction for* λ_0*, then* Θ *is centre of contraction for* $\prod_{i\in I} \lambda_0(i)$*.*

Conclusion. We briefly described an external realisability interpretation of some part of the informal theory BISH in BST, where the corresponding realisability relation is $\mathtt{Prf}(p, \phi) :\Leftrightarrow p \in \mathtt{Prf}(\phi)$. An important consequence of revealing the witnessing information is the avoidance of choice. *The use of the axiom of choice in constructive mathematics is an indication of missing data.* A BHK-interpretation of a negated formula $\neg\phi$ is missing. Strong negation introduced in BISH in [11], facilitates its BHK-interpretation. Proof-relevance is not a priori part of BISH, but it can be revealed a posteriori. In MLTT and its univalent extensions proof-relevance is built-in, and many facts are generated or hold automatically by the presence of the J-rule, or the univalence axiom of Voevodsky. BST-notions, such Martin-Löf sets, permit the translation of interesting "parts" of MLTT and HoTT into BST in a "definitional", non-axiomatic way.

References

[1] E. Bishop: *Foundations of Constructive Analysis*, McGraw-Hill, 1967.

[2] E. Bishop and D. S. Bridges: *Constructive Analysis*, Grundlehren der math. Wissenschaften 279, Springer-Verlag, Heidelberg-Berlin-New York, 1985. Cambridge University Press, 1987.

[3] T. Coquand: A remark on singleton types, manuscript, 2014, available at http://www.cse.chalmers.se/∼coquand/singl.pdf, 2014.

[4] S. Feferman: Constructive theories of functions and classes, in Boffa et al. (Eds.) *Logic Colloquium 78*, North-Holland, 1979, 159–224.

[5] P. Martin-Löf: An intuitionistic theory of types, in [13], 127–172.

[6] I. Petrakis: Dependent Sums and Dependent Products in Bishop's Set Theory, in P. Dybjer et. al. (Eds) TYPES 2018, LIPIcs, Vol. 130, Article No. 3, 2019.

[7] I. Petrakis: A Yoneda lemma-formulation of the univalence axiom, unpublished manuscript, 2019. Available at http://www.mathematik.uni-muenchen.de/∼petrakis/content/Preprints.php

[8] I. Petrakis: *Families of Sets in Bishop Set Theory*, Habilitation Thesis, LMU, 2020. Available at https://www.mathematik.uni-muenchen.de/ petrakis/content/Theses.php.

[9] I. Petrakis: Direct spectra of Bishop spaces and their limits, Logical Methods in Computer Science, Volume 17, Issue 2, 2021, pp. 4:1-4:50.

[10] I. Petrakis: Proof-relevance in Bishop-style constructive mathematics, Mathematical Structures in Computer Science, 2022, 1–43 doi:10.1017/S0960129522000159.

[11] I. Petrakis: Strong negation in constructive mathematics, in preparation, 2022.

[12] I. Petrakis, D. Wessel: Algebras of complemented subsets, in U. Berger et.al. (Eds): Revolutions and Revelations in Computability, CiE 2022, LNCS 13359, Springer,

[13] G. Sambin, J. M. Smith (Eds.): *Twenty-five years of constructive type theory*, Oxford University Press, 1998.

[14] The Univalent Foundations Program: *Homotopy Type Theory: Univalent Foundations of Mathematics*, Institute for Advanced Study, Princeton, 2013.

Frege's Concept of Completeness

Fabian Pregel

Abstract

Existing literature suggests that Frege did not have the concept of completeness. Yet, Frege's project is usually understood as finding a formal system from which all arithmetical truths could be proven. Furthermore, Frege is credited with devising the first calculus complete for first-order logic. How are we to reconcile these three claims? I argue that Frege did not just stumble across this complete calculus, but in fact had an early conception of theory- and calculus-completeness. Heck had briefly suggested this—I pick up on Heck's work and substantially strengthen its case. The paper offers several passages of text in support. Furthermore, I maintain that the reading to the contrary is based on an outdated overall reading of Frege (logocentricity).

1 Introduction

Over the past two decades, there has been a remarkable shift in scholarly attitude towards the role of semantics and metatheory in Frege's work. The debate can be broadly divided into two phases, each dominated by one school of thought: between the mid-1960s and mid-1990s, authors such as Dreben, van Heijenoort, Goldfarb and Ricketts defended what I shall call the 'no metatheory' school of thought. In their view, semantics and metatheory have no place in Frege's work. Between the mid-1990s and mid-2000s, authors such as Stanley, Tappenden, Antonelli, May, Heck and Blanchette defended largely an opposite position, which I shall call the 'proto-metatheory' school.

While the proto-metatheory school has cleared the path towards the *possibility* and, to a more limited extent, the *existence*, of metatheory in Frege's system, one specific assertion by the 'no metatheory' school seems to have gone relatively unaddressed: that Frege had no notion of completeness.

Dreben, van Heijenoort and Goldfarb all maintained that Frege never raised the question of completeness [van Heijenoort 1967, 326; Dreben and van Heijenoort 1986, 44–45; Goldfarb 1979, 353]. Existing proto-metatheory literature is largely silent on this claim—the only author that disputes it head-on is Heck [Heck 2007, 33]. Yet, even Heck discusses this claim only briefly, and their position has not become universally accepted. Dummett, for example, again asserted that 'Frege did not have the concept of the completeness of a logical system' [Dummett 2008, 11].

Thus, in this paper, I will explore the extent to which Frege had the concept of a complete formal system and the desire to achieve completeness. Let us distinguish between the completeness of a deductive calculus ('calculus-completeness') and the completeness of an axiomatisation of arithmetic ('theory-completeness'). I will argue that Frege already had 'proto'-concepts of both calculus-completeness and theory-completeness, and that these conceptions are central to his work. By a 'proto-concept' I shall mean a concept that captures the fundamental idea but is not fully developed or formalised.

2 Initial Passages

We begin with a passage Heck quotes from Frege's 1897 *On Mr. Peano's Conceptual Notation and My Own* [Heck 2007, 33]. In this quotation, Frege speaks about his *Begriffsschrift*. Frege says:

> *In order to test whether a list of axioms is complete* [original: 'vollständig'], we have to try and derive from them all the proofs of the branch of learning to which they relate. And in doing this it is imperative that we draw conclusions only in accordance with purely logical laws. [...] If we try to list all the laws governing the inferences that occur when arguments are conducted in the usual way, we find an almost unsurveyable multitude which apparently has no precise limits. The reason for this, obviously, is that these inferences are composed of simpler ones. And hence it is easy for something to intrude which is not of a logical nature and which consequently ought to be specified as an axiom. This is where the difficulty of discerning the axioms lies: for this [to discern the axioms] the inferences have to be resolved into their simpler components. By so doing we shall arrive at just a few modes of inference, with which we must then attempt to make do at all times. And if at some point this attempt fails, then we shall have to ask whether we have hit upon a truth issuing from a non-logical source of cognition, *whether a new mode of inference has to be acknowledged*, or whether perhaps the intended step ought not to have been taken at all [emphasis and bracketed comments mine]. [Heck 2007, 33; Heck is using the

translation from Frege (1897) 1984, 235]

In the quotation, Frege is looking to ascertain whether a particular axiomatisation of a science is 'complete'. To determine whether it is, Frege suggests that we have to look at all the inferences in the science and break those inferences down into small steps. If we are unable to 'derive' an 'inference', we have to add a non-logical axiom to the axiomatisation, acknowledge an additional mode of inference, or question the validity of the inference we are attempting to formalise. The phrase 'whether a new mode of inference has to be acknowledged' suggests that Frege understood that calculus-completeness was not a given, and that the deductive calculus may need to be amended.

Unfortunately, the discussion of completeness in Heck is relatively brief, and the passage from *On Mr. Peano's Conceptual Notation and My Own* is the only textual evidence they present for their view. Thus, let me pick up from where Heck left off and see if we can strengthen their argument. To add to the plausibility of the reading that Frege had a proto-notion of calculus-completeness, it is worth considering a passage in *Begriffsschrift*:

> Since in view of the boundless multitude of laws that can be enunciated we cannot list them all, *we cannot achieve completeness except by searching out those that, by their power, contain all of them.* Now it must be admitted, certainly, that the way followed here is not the only one in which the reduction can be done. That is why not all relations between the laws of thought are elucidated by means of the present mode of presentation. There is perhaps another set of judgements from which, when those contained in the rules are added, *all laws of thought could likewise be deduced* [emphasis mine]. [Frege (1879) 1967, §13]

Here, again, Frege appears to be aware of the possibility that a deductive calculus may not be complete ('we cannot achieve completeness except by...'). Frege also seems to assert the completeness of his *Begriffsschrift* system ('all laws of thought could [...] be deduced'). He also shows awareness that there may be different deductive calculi ('another set of judgements') that are equivalent.

The prima facie case for Frege having had a proto-concept of theory-completeness seems even more straightforward. The passage from *On Mr. Peano's Conceptual Notation and My Own* already mentioned testing 'whether a list of axioms is complete'. However, this passage leaves it somewhat open what constitutes completeness of axioms. Fortunately, other passages are far more explicit. Frege's very project was to find an axiomatisation from which all of arithmetic could be proven. For example, also in *On Mr. Peano's Conceptual Notation and My Own, Frege* writes

> I became aware of the need for a *Begriffsschrift* when I was looking for the *fundamental principles or axioms upon which the whole of mathematics rests* [emphasis mine]. [Frege (1897) 1984, 235; Heck 2007, 30]

Similar views are also expressed in Frege's other works. For example, in *Formal Theories of Arithmetic,* Frege states:

> I here want to consider two views, both of which bear the name 'formal theory'. *I shall agree with the first*; the second I shall attempt to confute. *The first has it that all arithmetical propositions can be derived from definitions alone using purely logical means, and consequently that they must be derived in this way* [emphasis mine]. [Frege 1984, 114]

3 Objections

What, if anything, speaks against taking these passages at face value? That Frege was searching for a complete axiomatisation of arithmetic seems hard to dispute. However, to the suggestion that Frege was aiming for a complete deductive calculus, it may be objected that Frege was using terms like 'complete' in the earlier quotes in a different sense from how we use the term today. The objection that Frege used a different notion than our calculus-completeness can be formulated as follows: today, we have a notion of calculus-completeness that relates semantic to syntactic consequence. By (strong) completeness, we mean that, for all sets of sentences Γ and sentence ϕ: If $\Gamma \models \phi$, then $\Gamma \vdash \phi$. Spelling out the definition of $\Gamma \models \phi$ further:

$$\text{If for all models } \mathcal{M} \text{ we have } \left(\text{If } \mathcal{M} \models \gamma \text{ for all } \gamma \in \Gamma, \text{ then } \mathcal{M} \models \phi\right), \text{ then } \Gamma \vdash \phi.$$

Given that Frege appears not to have had a notion of 'for all models $\mathcal{M}$', the argument would go, surely he must have had in mind a different notion of completeness than today's [Dummett 1991, 30].

This objection must be carefully addressed. To begin with, let us distinguish between having a concept of calculus-completeness and having an idea of how to prove that a calculus is complete. What can then be conceded is that Frege had no idea of how to prove *Begriffsschrift*'s calculus-completeness. Frege's approach to establishing calculus-completeness seems to involve a sort of systematic search to assure ourselves that, in all branches of learning, given the right initial non-

logical axioms, we are able to derive every truth — and if not to amend the calculus with new modes of inference until we are. Frege suggests that we 'try and derive from them [the axioms] all the proofs of the branch of learning' and 'search out those that, by their power, contain all of them [the logical laws]'. But, as noted, whether he had a sense of how to prove calculus-completeness is separate from the question of whether Frege had a notion of completeness of a calculus. Frege in fact displays an awareness that he did not have the means to demonstrate calculus-completeness. Instead, Frege acknowledges in *Boole's logical Calculus and the Concept-script* that he solely managed to make *Begriffsschrift*'s completeness probable:

> The fundamental principle of reducing the number of primitive laws as far as possible wouldn't be fully satisfied without a demonstration that *the few left are also sufficient.* It is this consideration which determined the form of the second and third sections of my book. Here too it would be wrong to suppose that a direct comparison with Boole's work is possible. In his case there is nothing remarkable in *the attempt to manage everything with the fewest possible primitive laws.* His only object is to find a brief and practical way to solve his problems. I sought as far as possible to translate into formulae everything that could also be expressed verbally as a rule of inference, so as not to make use of the same thing in different forms. Because modes of inference must be expressed verbally, I only used a single one by giving as formulae what could otherwise have also been introduced as modes of inference. [...] it wasn't my intention to provide a sample of how to carry out such derivations in a brief and practical way: *it was to show that I can manage throughout with my basic laws. Of course the fact that I managed with them in several cases could not render this more than probable* [emphasis mine]. [Frege (1880) 1979, 37–38]

Given that Frege admits that his work renders the completeness of his calculus no more than probable, it is shared ground between those holding that Frege had no concept of calculus-completeness and those holding that he had a proto-concept that Frege did not know how to prove calculus-completeness. What divides the two sides are instead questions such as: (i) What evidence is there that Frege had no concept of 'for all models $\mathcal{M}$', rather than a proto-concept? (ii) If Frege had no proto-concept of calculus-completeness, then how are we instead to understand Frege's talk of 'completeness' in the passages quoted from *Begriffsschrift*, *On Mr. Peano's Conceptual Notation and My Own*, and *Boole's logical Calculus and the Concept-script*? (iii) If on the other hand Frege had a proto-concept of calculus-completeness, what exactly did it look like?

Thus to address the challenge that Frege had no concept of 'for all models $\mathcal{M}$', I will address these

questions in order: In section 4, I will first consider arguments that purport to establish that Frege had no proto-concept of 'for all models'. In section 5, I will then argue that alternative accounts of Frege's talk of 'completeness' that have been offered lack textual backing and cannot explain how Frege in fact developed a calculus that is complete for first-order logic. Finally, in section **??**, I will offer a positive account of Frege's proto-concept of calculus-completeness based on Blanchette's interpretation of Frege's notion of logical consequence.

4 No Proto-Concept of 'For All Models'

Members of the 'no metatheory' school shared a view of Frege's philosophy of logic. This shared view is variously called 'universality of logic' or the 'logocentric view' [van Heijenoort 1967, 326; Ricketts 1985, 3]. Goldfarb articulates this view as 'logic was the system: the results of logic were simply the logical truths, and were to be arrived at by deriving them in the system' [Goldfarb 1979, 353]. For van Heijenoort, 'another important consequence of the universality of logic is that nothing can be, or has to be, said outside of the system' [van Heijenoort 1967, 326].

Given this perception of Frege's philosophy of logic, the 'no metatheory' advocates declared that notions such as completeness are not admissible in Frege's logocentric system. Goldfarb, for example, states that 'metasystematic considerations are illegitimate rather than simply undesirable' [Goldfarb 1979, 353]. van Heijenoort maintained that 'Frege never raises any metasystematic question (consistency, independence of axioms, *completeness*)' [van Heijenoort 1967, 326, emphasis mine]. In a later paper, Dreben and van Heijenoort claimed that the question of completeness could not even arise in Frege's system [Dreben and van Heijenoort 1986, 44]. They write:

> *To raise the question of semantic completeness* the Frege-Russell-Whitehead view of logic as all embracing had to be abandoned, and Frege's notion of a formal system had to become itself an object of mathematical inquiry and be subjected to the model-theoretic analyses of the algebraists of logic [emphasis mine]. [Dreben and van Heijenoort 1986, 45]

For Goldfarb, Dreben and van Heijenoort, the absence of a concept of completeness is therefore not coincidental but systematic. Studying completeness requires a notion of metatheory and semantics, and such a concept is not possible in Frege's logocentric world view.

Yet, what exactly rules out meta-theory? Goldfarb and van Heijenoort claim that formal reas-

oning such as in the *Begriffsschrift* is conducted under a fixed interpretation [Goldfarb 1979, 352; van Heijenoort 1967, 325]. Call this view *Fixed*. Goldfarb expresses this view as follows:

> For Frege and Russell the propositions of logic contain no non-logical vocabulary: there are no schematic placeholders which may be assigned one value or another. Every logical formula has a fixed meaning; there is no question of reinterpreting any sign. [Goldfarb 1979, 352]

Under their reading, there is no useful sense of 'for all models $\mathcal{M}$' for Frege. Consequently, the argument proceeds, there is no notion of semantic consequence, and thus also no useful notion of completeness [Stanley 1996, 64]. In this light, Frege's talk of 'completeness' just refers to the need to ensure that the deductive calculus captures the truths of a fixed semantics, and Frege's talk of 'acknowledging a new mode of inference' just means the need to modify the deductive calculus if the deductive calculus does not capture all truths of this fixed semantics.

This 'fixed interpretation' view has been extensively challenged by the 'proto-metatheory' school and can now be considered outdated [Antonelli and May 2000; Heck 2007; Blanchette 2012, 156–171]. Since addressing the fixed interpretation objection is critical for the purposes of claiming that Frege had a proto-conception of calculus-completeness, I will briefly outline my position here. Frege did not hold this 'fixed interpretation' view. Instead, Frege held a similar-sounding view that does permit a 'for all models' conception or a close analogue, and consequently a proto-conception of calculus-completeness. The two views that need to be distinguished are the following:

Fully For Frege, sentences in the *Begriffsschrift* are to be 'fully interpreted' in order to conduct actual inferences.

Fixed For Frege, formal reasoning such as in the *Begriffsschrift* is conducted under a single, fixed interpretation.

We already saw Goldfarb endorse *Fixed* above. Ricketts similarly held that 'any sort of talk of different possible interpretations of an unambiguous statement is, at best, a confused way to get at what should be expressed by variables' [Ricketts 1985, 4].

While *Fully* and *Fixed* sound similar, the key difference is that under *Fully* the calculus of the *Begriffsschrift* is initially uninterpreted until some interpretation is specified. Different interpretations for the symbolism could be specified. As such, metatheoretical questions can be raised about the calculus itself. Under Fixed, by contrast, there is a single interpretation of the calculus.

The reason *Fully* is more plausible then *Fixed* is that Frege does not specify an interpretation at the beginning of *Begriffsschrift*. Rather the *Begriffsschrift* appears in schematic form, with Frege providing possible interpretations as illustrations. For example, when discussing theorem (5), in modern notation $\vdash (B \to A) \to ((C \to B) \to (C \to A))$, Frege provides an illustration in terms of the magnetisation of a piece of iron and galvanic current flowing through a wire [Frege (1879) 1967, §15 p. 35]. We find similar examples from multiple disciplines in various places in the *Begriffsschrift*. The illustrations are not just provided for propositions but also for constants and functions. For example, when deducing that $\vdash g(b) \to (\neg f(b) \to (\neg \forall x (g(x) \to f(x))))$, Frege provides the example

> Let b mean an ostrich, that is, an individual animal belonging to the species, let $g(A)$ mean "A is a bird", and let $f(A)$ mean "A can fly". Then we have the judgment "If this ostrich is a bird and cannot fly, then it can be inferred from this that some birds cannot fly".[Frege (1879) 1967, §21 p. 51]

Thus, following Frege's practice, propositions, constants and functions in the deductive calculus do not come with a fixed interpretation like Goldfarb suggested. Instead they can be assigned a specific interpretation. Therefore, Frege seems to be aware of our modern alternative, a formal language with separate interpretations. Stanley expresses this observation well:

> Now, Frege's conception of formal theory was remarkably modern. [...] Frege speaks of the project of giving signs reference. Frege simply does not speak as if his *Begriffsschrift* expressions already come with a meaning. Rather, he speaks of semantical stipulations (Festsetzungen) which assign *Begriffsschrift* expressions reference. This demonstrates that Frege simply did not think of the *Begriffsschrift* as an already interpreted language, for one does not stipulate interpretations for already interpreted signs. Rather, Frege was treating his theory as an uninterpreted set of syntactic operations on strings of symbols. [Stanley 1996, 63]

Frege specifies an example interpretation of the *Begriffsschrift* calculus in part I of *Grundgesetze* [Stanley 1996, 64]. Frege's first-order functions range over courses-of-values and truth-values [Blanchette 2012, 74; Frege (1893) 2016, GG I, §31, p. 48]. However, the specification of an intended interpretation in a specific application—the derivation of arithmetic from logic—does not mean that other interpretations could not be specified. In fact, to apply the *Begriffsschrift* calculus to areas like

physics and geology as Frege had intended, a broader interpretation would need to be specified [Frege (1879) 1967, p. 7; §4, p. 13].

A possible reason for confusion as to whether Frege held *Fixed* and *Fully* is Frege's insistence in *On the Foundations of Geometry* that an 'actual' inference is an inference between 'thought[s] recognised as true', rather than a relationship between signs:

> [...] There we read, "In this way, one sequence of formal inferences can sometimes be '*interpreted*' in different ways." What can be interpreted is perhaps a sign or a group of signs, although the univocity of the signs—which we must retain at all cost—excludes different interpretations. *But an inference does not consist of signs. We can only say that in the transition from one group of signs to a new group of signs, it may look now and then as though we are presented with an inference. An inference simply does not belong to the realm of signs; rather, it is the pronouncement of a judgment made in accordance with logical laws on the basis of previously passed judgments. Each of the premises is a determinate thought recognized as true; and in the conclusion, too, a determinate thought is recognized as true. There is here no room for different interpretations [emphasis mine].* [Frege (1906) 1971, 82; Ricketts 1985, 4]

Given an inference is between thoughts recognised as true, propositions expressing these thoughts need to be true simpliciter—they cannot be true under one interpretation and false under another. Yet, this passage does not entail that Frege held *Fixed* rather than *Fully*. Frege is just insisting that an inference is between thoughts rather than signs, and those thoughts ought to be fully determinate. In fact, as Antonelli and May highlight, part III of the 1906 *On the Foundations of Geometry* provides a sketch of how metatheoretic results such as independence of axioms may be derived in Frege's framework. We will outline and employ Frege's proposal in section **??**.

5 Alternative Understandings of Frege's Talk of 'Completeness'

Let us suppose though that we had been convinced of the argument that Frege had no proto-concept of 'for all models', and thus no proto-concept of calculus-completeness. Thus if we cannot take Frege's talk of 'completeness' at face value, how are we to understand his use of 'completeness' or similar talk in the passages quoted so far, and the additional ones to come? In this section, I will argue that no viable alternative interpretation has been put forward.

Among the authors arguing that Frege had no concept of completeness, only van Heijenoort

provides a hint of what he perceives the correct interpretation of Frege's mentions of completeness to be. According to van Heijenoort, for Frege, 'the only question of completeness that may arise is, to use an expression of Herbrand's, an experimental question. As many theorems as possible are derived in the system. Can we exhaust the intuitive modes of reasoning actually used in science?' [van Heijenoort 1967, 327]

van Heijenoort's comment seems to pick up on Frege's suggestion in *On Mr. Peano's Conceptual Notation and My Own*. The picture Frege paints here seems to be the following: pick a specific science ('branch of learning'). Consider a set of non-logical axioms of this science. Then see if, with this set of axioms plus the logical calculus, we can derive all the truths in the science that we recognise as true. If not, one of the following three has to hold: either we are missing a non-logical axiom ('hit upon a truth issuing from a non-logical source of cognition'), our calculus is incomplete ('a new mode of inference has to be acknowledged') or the truth in the science we believed is not actually a truth ('perhaps the intended step ought not to have been taken at all').

Yet how exactly is van Heijenoort's interpretation a different concept of calculus-completeness than ours? For example, if we hit upon a case in which a new mode of inference does have to be acknowledged, this is presumably because the calculus is, also in our sense, incomplete. van Heijenoort does not provide an answer to this. However, we can imagine what he might say. He could respond that, according to his reading of Frege, Frege operated under a fixed semantics. Thus semantic implication and calculus-completeness are with respect to that fixed semantics, and questions of calculus-completeness in the modern sense do not arise.

We have already seen the textual evidence against the fixed interpretation view in section 4. However, there is an additional argument against van Heijenoort's alternative reading of 'completeness': if Frege had no concept of calculus-completeness or an entirely different concept, it would be an extraordinary coincidence that Frege nonetheless formulated his nine axioms and two modes of inference so as to yield a complete first-order calculus in our sense [Liu 2017; Dummett 2008, 11]. What could, other than Frege having a proto-concept of calculus-completeness, explain this phenomenon?

Providing any account of how Frege came up with the calculus in the *Begriffsschrift* is difficult because little is known about its genesis. Any notes or drafts Frege made during the *Begriffsschrift*'s creation have either been disposed of by him or burned subsequently in the 1945 fire that destroyed

his manuscripts [Wille 2018, 154]. Thus to determine whether Frege could have come up with a calculus whose first-order part is complete without a proto-concept of completeness, we seemingly have to resort to estimating how difficult it was for Frege to 'stumble upon' such a calculus by chance, in pursuit of other goals. Note that by 'stumble upon by chance', we need to distinguish between (a) stumbling upon such a calculus with the intention of finding a complete calculus but without proof that the calculus is indeed complete and (b) stumbling upon such a calculus without such an intention. (a) would still be confirmation that Frege had a proto-concept of calculus-completeness, and thus only (b) is an alternative explanation how Frege arrived at a calculus which was indeed complete for first-order logic. This is what 'by chance' is meant to express. In the following, I want to consider two such 'stumbling-across' explanations: one historical and one aim-based.

The first, historical, stumbling-across explanation would be to maintain that (i) Frege's contemporaries had already done most of the required work for a complete first-order calculus, and (ii) Frege had a detailed understanding of his contemporaries' writings at the time of writing the *Begriffsschrift*. If assumptions (i) and (ii) were true, it could seem plausible that Frege stumbled across a complete first-order calculus. The most relevant contemporaries or immediate predecessors are the Booleans working on algebraic logic—in particular George Boole, Augustus De Morgan and Ernst Schröder. Boole had pioneered the idea of modelling logical connectives on arithmetic operations such as $+$, - and $\cdot$. Thus, to use an example from Boole, if y represents that 'Gravitation is necessarily present' and v that 'A vacuum is necessary', then 'If gravitation is necessarily present, a vacuum is necessary' is presented as $y\,(1-v) = 0$ [Boole 1854, 219–220]. Either $y = 0$ (the antecedent is false) or $v = 1$ (the consequent is true).

However, it appears that, with respect to the Booleans' work, neither (i) nor (ii) can credibly proven to be the case. With respect to (i), there are several major differences between algebraic and Fregean logic: For example, Boole had drawn a distinction between primary and secondary propositions. Primary propositions 'express a relation among things', while secondary propositions 'express a relation among propositions' [Boole 1854, 52]. An example of a primary proposition is 'the sun shines', whereas 'if the sun shines the earth is warmed' is a secondary proposition [Boole 1854, 52]. Boole then developed separate 'methods' for primary and secondary propositions [Burris 2018, 6.1, 6.2]. The *Begriffsschrift*, by contrast, avoids such a division and constitutes a single logic [Frege (1880) 1979, 14].

With respect to (ii), that Frege had a detailed understanding of his contemporaries' writings at the time of writing the *Begriffsschrift*, there also appears to be a lack of evidence. Wille and Sluga, for example, argue that there is no evidence to support the view that Frege, at the time of writing *Begriffsschrift*, had any knowledge of algebraic logic [Sluga 1987, 96; Wille 2018, 73]. All one can say for sure is that, after the *Begriffsschrift* was published and criticised for not taking into account the work of the algebraic school, Frege wrote pieces such as *Boole's logical Calculus and the Concept-script* to provide a detailed defence of the merits of the *Begriffsschrift* over algebraic logic.

Thus, without clear evidence for (i) and (ii), there seems to be little support for the alternative explanation of the *Begriffsschrift*'s calculus-completeness that Frege largely copied what his contemporaries had already done. As Wille puts it, 'Frege did not have anything comparable to resort to, which makes the resulting axioms system even more impressive, also in its systematic elegance' [Wille 2018, 154].

We now move to the second, aim-based, 'stumbling-across' explanation of the calculus-completeness of the first-order part of *Begriffsschrift*. According to this explanation, what Frege really aimed to achieve was to prove a number of propositions regarding sequences in part III of *Begriffsschrift*. The deductive system Frege provided was then merely what was required to prove the propositions of part III, and consequently the completeness of the first-order part of *Begriffsschrift* is simply a by-product of Frege's ambition to prove the propositions in part III. The question is therefore: was the calculus of *Begriffsschrift's* aim solely to prove the propositions in part III, or are the propositions in part III partially to illustrate the power of the calculus, in lieu of a completeness proof? Fortunately Frege provides the answer. In *Boole's logical Calculus and the Concept-script*, Frege tells us about the intent of part II and III of *Begriffsschrift:*

> The fundamental principle of reducing the number of primitive laws as far as possible wouldn't be fully satisfied without a demonstration that the few left are also sufficient. It is this consideration which determined the form of the second and third sections of my book. [Frege (1880) 1979, 37]

Frege makes a similar remark at the beginning of part III of the *Begriffsschrift*, where he says that 'the derivations that follow are intended to give a general idea of the way in which our ideography is handled, even if they are perhaps not sufficient to demonstrate its full utility' [Frege (1879) 1967, p. 55, §23]. That the propositions in part III are partially to illustrate the power of the calculus makes sense because, as noted earlier, Frege acknowledged that he did not have the means to prove

the completeness of his calculus. Thus part of the reason Frege uses part III of the *Begriffsschrift* to demonstrate propositions about sequences is to illustrate the ability of his calculus to prove even fairly complex propositions.

Consequently, both the historical and the aim-based 'stumbling-across' explanations of the calculus-completeness of the first-order part of *Begriffsschrift* appear to lack textual backing, and can therefore not explain how Frege in fact arrived at a complete first-order calculus.

It is often said that, for Frege, the distinction between first- and second order logic was less important than it is for us today. One may thus wonder: if Frege attributed no special relevance to first-order logic, then how can the completeness of the first-order part of *Begriffsschrift* be of explanatory value in the debate about whether Frege had a proto-conception of completeness? However, the response is straightforward: the claim is not that Frege was trying to devise a calculus that was complete for first-order logic specifically. Rather, Frege was trying to formalise a deductive system that was complete. By his own admission, Frege never obtained a proof that his deductive system was complete [Frege (1880) 1979, 37–38]. With the benefit of Gödel's incompleteness theorems, we know that the best possible result was a deductive system that was complete for first-order logic. This best-possible result is what Frege has achieved with the *Begriffsschrift,* and it is a legitimate question what explains this achievement.

It is important to keep in mind though that this is just additional circumstantial evidence for the view that Frege had a proto-concept of calculus-completeness—the primary evidence are the earlier text passages where Frege employs this proto-concept. Thus, I believe, alternative interpretations that do not take Frege's talk of completeness at face-value are not convincing.

References

Antonelli, Aldo, and Robert May. 2000. 'Frege's New Science'. *Notre Dame Journal of Formal Logic* 41 (3): 242–270.

Blanchette, Patricia. 2012. *Frege's Conception of Logic.* Oxford, New York: Oxford University Press.

Boole, George. 1854. *An Investigation of the Laws of Thought.* London: Walton and Maberly.

Burris, Stanley. 2018. 'George Boole'. In *The Stanford Encyclopedia of Philosophy,* Summer 2018, edited by Edward Nouri Zalta. Metaphysics Research Lab, Stanford University.

Dreben, Burton, and Jean van Heijenoort. 1986. 'Gödel 1929: Introductory Note to 1929, 1930 and 1930a'. In *Kurt Gödel: Collected Works,* by Kurt Gödel, edited by Solomon Feferman, I:44–59. New York: Oxford University Press.

Dummett, Michael. 1991. *Frege: Philosophy of Mathematics.* London: Bloomsbury Academic.

———. 2008. 'Gottlob Frege (1848-1925)'. In *A Companion to Analytic Philosophy,* edited by Aloysius Patrick Martinich and Ernest David Sosa, 6–20.

Frege, Gottlob. (1879) 1967. 'Begriffsschrift, a Formula Language, Modeled upon That of Arithmetic, for Pure Thought'. In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931,* edited by Jean van Heijenoort, translated by Stephen Bauer-Mengelberg, 1–82. Harvard University Press.

———. (1880) 1979. 'Boole's Logical Calculus and the Concept-script'. In *Posthumous Writings,* edited by Hans Hermes, Friedrich Kambartel and Friedrich Kaulbach, translated by Peter Long and Roger White, 9–46. Oxford: Basil Blackwell.

———. (1893) 2016. *Basic Laws of Arithmetic.* Translated by Philip A. Ebert and Marcus Rossberg. Oxford: Oxford University Press.

Frege, Gottlob. (1897) 1984. 'On Mr. Peano's Conceptual Notation and My Own'. In *Collected Papers on Mathematics, Logic, and Philosophy,* edited by Brian McGuinness, 234–248. Oxford: Basil Blackwell.

———. (1906) 1971. 'On the Foundations of Geometry'. In *On the Foundations of Geometry and Formal Theories of Arithmetic,* translated by Eike-Henner W. Kluge, 49–112. New Haven and London: Yale University Press.

———. 1984. 'Formal Theories of Arithmetic'. In *Collected Papers on Mathematics, Logic, and Philosophy,* edited by Brian McGuinness, 112–121. Oxford: Basil Blackwell.

Goldfarb, Warren David. 1979. 'Logic in the Twenties: The Nature of the Quantifier'. *The Journal of Symbolic Logic* 44 (3): 351–368.

Heck, Richard Kimberly. 2007. 'Frege and Semantics'. *Grazer Philosophische Studien* 75 (1): 27–63.

Liu, Yang. 2017. 'Frege's Begriffsschrift Is Indeed First-Order Complete'. *History and Philosophy of Logic* 38 (4): 342–344.

Ricketts, Thomas Grant. 1985. 'Frege, The Tractatus, and the Logocentric Predicament'. *Noûs* 19 (1): 3–15.

Sluga, Hans. 1987. 'Frege against the Booleans'. *Notre Dame Journal of Formal Logic* 28 (1): 80–98.

Stanley, Jason. 1996. 'Truth and Metatheory in Frege'. *Pacific Philosophical Quarterly* 77 (1): 45–70.

Van Heijenoort, Jean. 1967. 'Logic as Calculus and Logic as Language'. *Synthese* 17 (3): 324–330.

Wille, Matthias. 2018. *Gottlob Frege: Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Klassische Texte der Wissenschaft. Springer Spektrum.

Dynamic Approximation of Self-Referential Sentences (for Philosophical Logic)

Vladimir A. Stepanov[1]

CC FIC CSC RAS, Moscow, Russia
vastvast@yandex.ru

Abstract

Non-classical logic via approximation of self-referential sentences by dynamical systems are consistently presented. The new 6-valued truth values <T,va,A,V,av,F> (here A=Liar, V=TruthTeller) are presented as a function of the classical truth values $x_i \in \{0, 1\}$, which resulted in a philosophical standpoint known as Suszko's Thesis. Three-valued truth tables were created corresponding to Priest's tables of the same name (Priest, 1979). In the process of constructing 4-valued truth tables, two more new truth values (va, av) were revealed that do not coincide with the four original ones. Therefore, the closed tables turned out to be 6-valued. Prof Dunn's 4-valued truth tables are compared with our 4-valued truth tables. De Morgan's laws are confirmed by six-valued truth tables. Constructed 3-, 4- and 6-valued lattices obeying De Morgan's laws.

Many of the results are new.

Introduction

Sentences that refer to themselves are called self-referential. The most popular of these is the "Liar" sentence. It can be noted that the study of self-referencing admits two possible approaches:

- external - which describes the reaction of self-referential sentences to the system under study. These include the popular studies of Priest in 1978 (LP), see (Priest, 1978); and (Dunn, 2019);
- internal - when the emphasis is on the study of the structure of self-referential sentences, which began with Peirce in 1855 (Michael, 1975). We will devote our article to this last approach.

The constructive analysis of the Liar sentence was carried out by Charles Peirce (Michael, 1975), who, as far as we know, was the first to notice in his lectures in 1864-1865, that self-referential sentences generate an infinite sequence of substitutions into themselves. This is the first application of the principle, which in the second half of the 20th century was called "turning a vicious circle into a generating circle".

We are talking about the **S** icon, which first appeared in the article (Johnstone, 1981): $Q =_{df} \mathbf{S}_Q P$. According to the meaning, **S** indicates that the entire expression belongs to self-referencing, and introduces the entire self-referential construction to the rank of WFF. The Liar sentence: $\mathbf{S}_Q \sim TQ$. (Formulas in the author's annotation; we do not decipher them).

1 Basic definitions

We define a dynamic approximation of self-referential sentences, which for the Liar and the TruthTeller, generates three-valued Kleene logic, and allows us to obtain new 4- and 6-valued

truth tables (Stepanov, 2021). We fix the self-referencing of the sentence using a special self-referencing icon: $\mathbf{S}x$, which is placed in front of the predicate $P(x)$. We call the predicate $P(x)$ the core of a self-referential sentence. A self-referential sentence looks like this:

$$\mathbf{S}xP(x). \tag{1}$$

Expression (1) obeys the axiom of self-reference (Feferman 1984):

$$\mathbf{S}xP(x) \leftrightarrow P(\mathbf{S}xP(x)). \tag{2}$$

Peirce (Michael, 1975) applied (2) to infinite Liar sentence:

$$\mathbf{S}xP(x) \leftrightarrow P(P(P(...\mathbf{S}xP(x)...))). \tag{3}$$

Let's break it down into iterative steps:

$$\mathbf{S}xP(x) \approx \mathsf{S}xP(x) =< x, P(x), P(P(x)), \dots > . \tag{4}$$

The right-hand side of expression (4) will be considered as an approximation $\approx$ of a real self-referential sentence $\mathbf{S}xP(x)$. The sign $\mathsf{S}xP(x)$ is used to denote the result of the approximation. Expression (4) is the definition of the trajectory of a dynamical system of the form $(\{0,1\}, P(x))$ with orbits $< P^n(x), n \in Z^+ >$, where $P^n(x) = P(P^{n-1}(x))$, (Konev et al., 2006). Consider the case when the kernels of self-referential sentences $P(x)$ are composed of $Tr(x)$ using the propositional connectives equivalence and negation:

$$P(x) \in \{Tr(x), \neg Tr(x), Tr(x) \leftrightarrow Tr(x), Tr(x) \leftrightarrow \neg Tr(x)\}. \tag{5}$$

It is easy to see that expression (4) is periodic, with a maximum period of 2. This means that the second and third terms of the sequence (4) determine the entire remaining infinite sequence. Therefore, in our case, we rightfully shorten the definition of a self-referencing quantifier as follows:

$$\mathsf{S}xP(x) =< x, P(x), P(P(x)) > . \tag{6}$$

The variable x and the predicates $P(x)$ from (5) in our case take values from $\{0,1\}$.

Definition 1: For $\mathsf{S}xP(x) \rightleftharpoons \{< 1, P(1), P(P(1)) > , < 0, P(0), P(P(0) > .\}$:
$\neg\mathsf{S}xP(x) \rightleftharpoons \neg\{< 1, P(1), P(P(1)) > , < 0, P(0), P(P(0) >\}$
$\neg\mathsf{S}xP(x) \rightleftharpoons \{\neg < 1, P(1), P(P(1)) > , \neg < 0, P(0), P(P(0) >\}$ (7)
$\neg\mathsf{S}xP(x) \rightleftharpoons \{< \neg 1, P(\neg 1), P(P(\neg 1)) > , < \neg 0, P(\neg 0), P(P(\neg 0) >\}$

This is the table for the negation symbol:

$\mathsf{S}xP(x)$	$\neg\mathsf{S}xP(x)$
$\{< 1,1,1 >; < 0,1,1 >\} = T$	$F = \{< 1,0,0 >; < 0,0,0 >\}$ (False)
$\{< 1,0,1 >; < 0,1,0 >\} = A$	$A = \{< 0,1,0 >; < 1,0,1 >\}$ (Antinomy, Liar)
$\{< 1,1,1 >; < 0,0,0 >\} = V$	$V = \{< 0,0,0 >; < 1,1,1 >\}$ (Void,TruthTeller)
$\{< 1,0,0 >; < 0,0,0 >\} = F$	$T = \{< 0,1,1 >; < 1,1,1 >\}$ (True)

Definition 2:We define two-place connectives $o \in \{\wedge, \vee, \rightarrow, \leftarrow\}$ for two S-formulas $\mathsf{S}xP(x)$ and $\mathsf{S}xQ(x)$. We study such a variant of two-place connectives, when the trajectories of estimates of the formula $\mathsf{S}xP(x)$ of the one branch (x = 1 or x = 0) interact with the trajectories of the formula $\mathsf{S}xQ(x)$ of the same branch ($x = 1$ or $x = 0$, respectively):

$$\mathsf{S}xP(x) \text{ o } \mathsf{S}xQ(x) \rightleftharpoons$$
$$\{<1,P(1),P(P(1))>,<0,P(0),P(P(0)>\} o \{<1,Q(1),Q(Q(1))>,<0,Q(0),Q(Q(0))>\} =$$
$$\{<1,P(1),P(P(1))>o<1,Q(1),Q(Q(1))>,<0,P(0),P(P(0))>o<0,Q(0),Q(Q(0))>\} =$$
$$\{<1o1,P(1)oQ(1),P(P(1))oQ(Q(1))>,<0o0,P(0)oQ(0),P(P(0))oQ(Q(0))>\}.$$

Example.: F∧V=
$= \{<1,0,0>,<0,0,0>\} \wedge \{<1,1,1>,<0,0,0>\}$
$= \{<1,0,0> \wedge <1,1,1>,<0,0,0> \wedge <0,0,0>\}$
$= \{<1,0,0>,<0,0,0>\}$=F.

2 Main results

Let's compare Kleene-Priest tables with our tables on our rules for A and V:

Kleene-Priest p

∧	t	p	f
t	t	p	f
p	p	p	f
f	f	f	f

≡

Hypothesis: p = A

∧	T	A	F
T	T	A	F
A	A	A	F
F	F	F	F

≡

Hypothesis: p = V

∧	T	V	F
T	T	V	F
V	V	V	F
F	F	F	F

Lemma 1: 1. The sentences *Liar* (A) has the tabular model, coinciding with tabular model *Liar* (*p*) of Priest (Priest, 1979) and, accordingly, the same evidential theory.
2. The sentences *TruthTeller* (V) has the same configuration tabular model, coinciding with configuration tabular model *Liar* (p) of Priest (Priest, 1979).
Lemma 2: When constructing the interaction of V and A, new truth values were obtained:
A∧V=$\{<1,0,1>,<0,0,0>\}$=av=¬(va),
A∨V=$\{<1,1,1>,<0,1,0>\}$=va=¬(av).
The author has not come across any statement in the literature that the sentences A∧V and A∨V have a similar assessment of the truth of av and va. This is a new result!

Our table

∧	T	A	V	F
T	T	A	V	F
A	A	A	av	F
V	V	av	V	F
F	F	F	F	F

Our table

∨	T	A	V	F
T	T	T	T	T
A	T	A	va	A
V	T	va	V	V
F	T	A	V	F

For comparison, here are the Dunn tables:

(Dunn, 2019)

∧	T	B	N	F
T	T	B	N	F
B	B	B	F	F
N	N	F	N	F
F	F	F	F	F

(Dunn, 2019)

∨	T	B	N	F
T	T	T	T	T
B	T	B	T	B
N	T	T	N	N
F	T	B	N	F

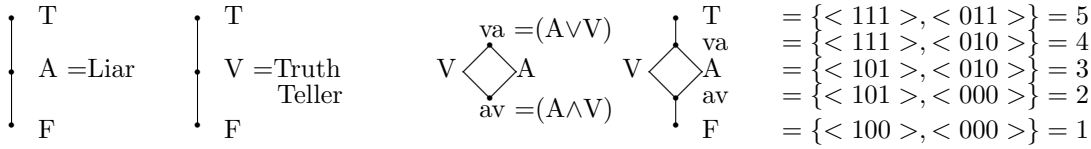
These are the complete 6-valued tables:

¬	
T	F
va	av
A	A
V	V
av	va
F	T

∧	T	va	A	V	av	F
T	T	va	A	V	av	F
va	va	va	A	V	av	F
A	A	A	A	av	av	F
V	V	V	V	av	av	F
av	av	av	av	av	av	F
F	F	F	F	F	F	F

∨	T	va	A	V	av	F
T	T	T	T	T	T	T
va	T	va	va	va	va	va
A	T	va	A	va	A	A
V	T	va	va	V	V	V
av	T	va	A	V	av	av
F	T	va	A	V	av	F

Lemma 3: The next four lattices are DeMorgan lattices, á la (Leitgeb, 1999):
{ F ≤ av ≤ A ≤ V ≤ va ≤ T } ; (1 ≤ 2 ≤ 3 ≤ 3 ≤ 4 ≤ 5) :



DM3A **DM3V** **DM4VA** **DM6**

Our truth-values are finite estimates of infinite periodic classical sequences of kernels of self-referential statements. This is consistent with Suszko's principles of transforming non-classical truth-values through a set of classical truth-values.

References

Dunn, 2019 – Dunn, J.M., "Two, three, four, infinity: The path to the four-valued logic and beyond", New Essays on Belnap-Dunn Logic, ed. by H. Omori and H. Wansing, Cham: Springer, 2019, 77–97.

Feferman, 1984 - Feferman S. Toward Useful Type-Free Theories I.
The Journal of Symbolic Logic, 1984, 49, 75-111.

Johnstone, 1981 - Johnstone A. Self-reference, the Double Life and Godel
Logique et Analyse, 1981, 24, 35-47.

Kleene, 1950 - Kleene, S. C., Introduction to Metamathematics,
D. Van Nostrand Co., Princeton, 1950.

Konev at al., 2006 – Konev, B. Kontchakov, R., Wolter, F., Zakharyaschev, M.
On Dynamic Topological and Metric Logics. Studia Logica, 84, (2006), 129-160.

Leitgeb, 1999 – Leitgeb, H., Truth and the Liar in De Morgan-Valued Models
Notre Dame Journal of Formal Logic, Vol. 40(4),1999, 496–514.

Michael, 1975 - Michael E. Peirce's Paradoxical Solution to the Liar's Paradox
Notre Dame Journal of Formal Logic, Vol.XII, No. 3, 1975. 369-374.

Priest, 1979 - Priest G. The Logic Paradox.
Journal of Philosophical Logic, 8 (1979) 219-241.

Stepanov, 2021– Stepanov V.: In defense of the self-referencing quantifier Sx. Approximation of self-referential sentences by dynamic systems.
Logico-Philosophical Studies, 19(2), 145-150 (2021), DOi: 10.52119/LPHS.2021.49.50.014.

8-valued non-deterministic semantics for modal logics

Pawel Pawlowski and Elio La Rosa

In this paper, we show that modal logics obtained by combinations of axioms **K,T,D,4,5,B** have an intuitive *non-deterministic* characterization that is at most *8-valued*.

Our starting point is the propositional modal system **H** which does not have any axioms for modal language except the Dual axiom. One can think about this system as the propositional logic in an extended language by two modal operators that are interdefinable. We construct an *8-valued* non-deterministic semantics for **H**, where the values represent modal status of the formula. So they not only convey the information whether a given formula is true but also whether it is possible or/and necessary. These is summarized by the following table:

Table 1: Meaning of values

Value	**Status of the sentence**
$\mathtt{T}_\Diamond$	$\Box\varphi, \Diamond\varphi, \varphi$ (necessary, possible and true)
$\mathtt{T}$	$\Box\varphi, \neg\Diamond\varphi, \varphi$ (necessary, not possible and true)
$\mathtt{t}_\Diamond$	$\neg\Box\varphi, \Diamond\varphi, \varphi$ (not necessary, possible and true)
$\mathtt{t}$	$\neg\Box\varphi, \neg\Diamond\varphi, \varphi$ (not necessary, possible and true)
$\mathtt{f}_\Diamond$	$\neg\Box\varphi, \Diamond\varphi, \neg\varphi$ (not necessary, possible and false)
$\mathtt{F}$	$\Box\varphi, \neg\Diamond\varphi, \neg\varphi$ (necessary, not possible and false)
$\mathtt{F}_\Diamond$	$\Box\varphi, \Diamond\varphi, \neg\varphi$ (necessary, possible and false)
$\mathtt{f}$	$\neg\Box\varphi, \neg\Diamond\varphi, \neg\varphi$ (not necessary, not possible and false)

Based on this semantics we will show how to extend it for logics with arbitrary combinations of axioms: D, T, B, 4, 5. In order to regain the rule of necessitation we apply the Kearn's m-th level valuations. By doing so, we obtain semantics for all the normal modal logics definable by the mentioned axioms. As a byproduct of this procedure we also obtain a semantics for logics where the rule of necessitation is restricted only to the theorems that does not use necessitation[1]

While some of these results are already known, we complete the overall picture and simplify it by providing *reductions* that minimize the number of values. We also emphasize on the use of more economical axiomatizations that do not incorporate axioms that become redundant in the presence of the *necessitation rule*. In this process, some logics are coupled with *non-normal* (in the sense of lacking the necessitation rule) companions.

[1] Roughly we get the rule of necessitation used in the logic **S1**.

To get a better grasp of non-deterministic semantics consider the following example: 2-valued (V, V') table for the connective $\circ$:

$\circ$	V	V'
V	V	V, V'
V'	V'	V, V'

Notice how, for the value associated with the V' column, the table does not single out a single value, but a set $\{V, V'\}$ of them, indicated in the abbreviated form by V, V' in our table. We call these tables *non-deterministic*. Non-deterministic *semantics* are based on such a generalization of (many-valued) tables. Since the interpretation of connectives can give a non-empty set of truth-values instead of a single one, the valuation function singles out one of the possible values given by the set. In other words, this means that the interpretation of a connective assigns a non-empty set of values (i.e an element of the power set of values minus the empty set) to the complex formula.[2]

This allows for introducing new interpretations for an otherwise extensional reading of a connective, making it possible to semantically characterize logics that cannot be characterized by finitely many valued (deterministic) approaches. Examples of this can be found in [Jorge and Holik, 2020], [Avron and Zamansky, 2007], [Pawlowski and Urbaniak, 2018] and [Coniglio et al., 2019]. In this paper, we deal with finitely many-valued characterizations of modal logics that cannot be captured by deterministic tables, as those studied in [Dugundji, 1940]. According to Dugundji's theorem one cannot provide a finitely-many valued deterministic semantics for modal logics between **S1** and **S5**.[3]

The way to bypass the Dugundji's theorem is to start with a non-deterministic characterization of a modal logics that does not validate the rule of necessitation and strengthen it with the method on mth-level valuations. Roughly speaking, this method reduces the set of admissible valuations by removing those, according to which the tautologies of the logic are not necessary.

So far, approaches of this kind were developed for capturing modal logics weaker than **K** [Ivlev, 1988], for **K**, **T**, **S4**, **S5**, **KD**, **KB** and **KTB** [Kearns, 1981], [Coniglio et al., 2015], [Omori and Skurt, 2016].

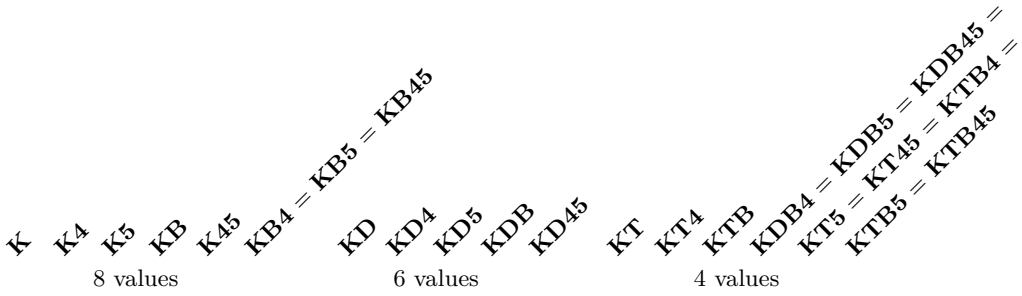
No systematic study of non-deterministic semantics has hence been conducted for such variety of modal logic that includes all the logics of the modal cube portrayed in the *Stanford Encyclopedia of Philosophy* entry on modal logic [Garson, 2021].

[2]The relation between this type of semantics and the possible world semantics or neighbourhood semantics has not yet been studied. However, it seems that non-determinsitic semantics allows one to capture modalities that are too weak to be capturable in the neighbourhood semanitcs.

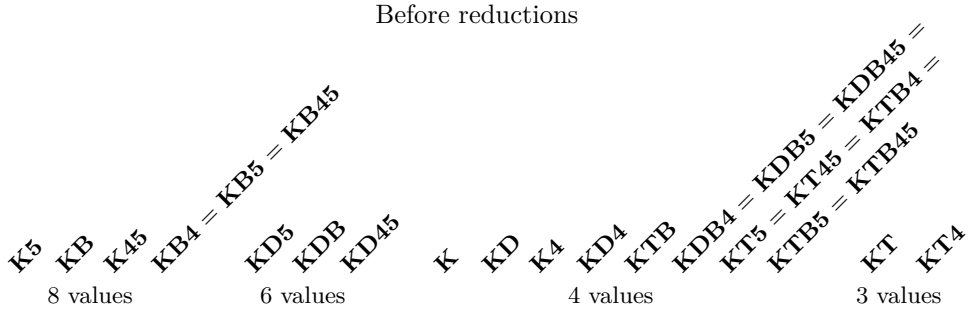
[3]The question whether one can provide a direct non-deterministic semantics for those systems has been answered negatively in [Grätz, 2021]

In this paper, we fill this gap and rework the logics found in the works cited above in order to guarantee a more economical axiomatization while preserving modularity over the presence of the necessitation rule and, in some cases, a reduction of the number of values will be given. As a consequence, axiomatizations which do not include necessitation are not closed under the rule of *substitutivity of equivalents*.[4] The following table summarizes all the results w.r.t the entry on *The Stanford Encyclopedia of Philosophy*:

Figure 1: *Table of results with respect to SEP entry*



Before reductions



After reductions

References

[Avron and Zamansky, 2007] Avron, A. and Zamansky, A. (2007). *Many-Valued Non-deterministic Semantics for First-Order Logics of Formal (In)consistency*, pages 1–24. Springer Berlin Heidelberg, Berlin, Heidelberg.

[Coniglio et al., 2015] Coniglio, M. E., Fariñas del Cerro, L., and Peron, N. M. (2015). Finite non-deterministic semantics for some modal systems. *Journal of Applied Non-Classical Logics*, 25(1):20–45.

[4] This was first noticed by [Omori and Skurt, 2016], where the authors spotted a mistake in the work of [Ivlev, 1988]. The proposed axiomatic system provided there is not sound with respect to the substitution of a subformula φ for $\neg\neg\varphi$ and vice versa in a formula ψ. This mistake carries over to [Coniglio et al., 2015], but was later acknowledged in [Coniglio et al., 2016].

[Coniglio et al., 2016] Coniglio, M. E., Fariñas del Cerro, L., and Peron, N. M. (2016). Errata and addenda to 'Finite non-deterministic semantics for some modal systems'. *Journal of Applied Non-Classical Logics*, 26(4):336–345.

[Coniglio et al., 2019] Coniglio, M. E., Fariñas del Cerro, L., and Peron, N. M. (2019). Modal logic with non-deterministic semantics: Part I—Propositional case. *Logic Journal of the IGPL*, 28(3):281–315.

[Coniglio and Peron, 2014] Coniglio, M. E. and Peron, N. M. (2014). Dugundji's theorem revisited. *Logica Universalis*, 8(3-4):407–422.

[Dugundji, 1940] Dugundji, J. (1940). Note on a property of matrices for Lewis and Langford's calculi of propositions. *Journal of Symbolic Logic*, 5(4):150–151.

[Garson, 2021] Garson, J. (2021). Modal Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2021 edition.

[Grätz, 2021] Grätz, L. (2021). Truth tables for modal logics T and S4, by using three-valued non-deterministic level semantics. Accepted for publication.

[Ivlev, 1988] Ivlev, Y. V. (1988). A semantics for modal calculi. *Bulletin of the Section of Logic*, 17(3/4):114–121.

[Jorge and Holik, 2020] Jorge, J. P. and Holik, F. (2020). Non-deterministic semantics for quantum states. *Entropy*, 22(2):156.

[Kearns, 1981] Kearns, J. T. (1981). Modal semantics without possible worlds. *The Journal of Symbolic Logic*, 46(1):77–86.

[Omori and Skurt, 2016] Omori and Skurt (2016). More modal semantics without possible worlds. *IfCoLog Journal of Logics and their Applications*, 3(5):815–845.

[Pawlowski and Urbaniak, 2018] Pawlowski, P. and Urbaniak, R. (2018). Many-valued logic of informal provability: A non-deterministic strategy. *The Review of Symbolic Logic*, 11(2):207–223.

The Foundations of Arithmetic: Peano vs Dedekind

Katerina Petsi
Department of History and Philosophy of Science
National and Kapodistrian University of Athens.
oretakh@yahoo.gr

Abstract

The axioms of Arithmetic are widely known among mathematicians. They have been named Dedekind-Peano axioms, although there was no collaboration between the two mathematicians. But there is no concrete evidence that one knew the other's work at the time he was writing his own. The aim of this work is to study the similarities and differences in the presentations of the texts of the two mathematicians and to investigate any influences between each other.

1 Introduction

In 1889 the *Arithmetices principia, nova methodo exposita* by Giuseppe Peano was published. A year earlier, in 1888, *Was sind und was sollen die Zahlen?* was published; by Richard Dedekind. These two works presented the foundation of Arithmetic. Peano's Arithmetic is the most generally accepted official basis of arithmetic, but the initial presentation of the axioms by Peano differs considerably from the form known today.

Some writers talk about Dedekind-Peano's axioms instead of Peano's axioms. How much did Dedekind's work affect Peano's work? What are the similarities and differences between the two texts? According to Kennedy, Peano did not see Dedekind's book until his own was published (Kennedy, 2002). But why does the author himself, in the introduction to his work states that the work of Dedekind was very useful (Peano, 1889)? In the introduction to the first edition of his work, Dedekind states that the preparation of his presentation was done before the publication of his work in sequels, but only from 1872 to 1878, he was able to commit to a text, which many mathematicians considered and discussed with him (Dedekind, 1888). In short, Dedekind had a public debate on the foundation of Arithmetic for at least five years, during which Peano according to Kennedy (Kennedy, 2002), studied at the University of Turin (1876-1880). It is therefore very likely that Peano had come into contact with Dedekind's attempt to establish Arithmetic at that time. After all, according to Volterra in the second half of the 19th century, there were contacts between Italian and German universities (Volterra, 1908). Below we will study the similarities and differences between the two texts.

2 Method and Objective

In the second half of the 19th century, the foundations of most mathematical theories were laid. Many mistakes were made while trying to establish the foundations and questions were asked in an attempt to correct the mistakes. Peano in the early years of his career was involved in correcting mistakes, to the point that Segre calls him a “mistake hunter” (Segre, 1994). In his introduction to the *Arithmetices principia, nova methodo exposita*, Peano expresses the view that the difficulty of foundations is due to the ambiguity of language and that it is very important to examine carefully every word we use (Peano, 1889). Thus, he sets as the goal of his work the presentation of a method that emerged after the examination of the problem as well as an application on Arithmetic. As he wrote, his book was an introduction to logical symbolism. His method is nothing more than the use of a symbolic language.

Dedekind, on the other hand, shows particular interest in the study of the axiomatic properties of numbers as well as in isolating the properties from their numerical character so that they can be incorporated into more general concepts, as he himself mentions in his letter to Keferstein (Dedekind, 1890). In that letter, Dedekind responds to Keferstein's critique of *Was sind und was sollen die Zahlen?* and especially the concept of the chain[1], defending his work. Dedekind begins with informal references to some basic principles of set theory, beginning with the definition of the important concept of chain. A few years later, in the introduction to the third edition of his book, he would express his concern, because in the meantime doubts had arisen about the credibility of the important foundations of his conception. (Dedekind, 1888).

3 The number, the axioms and the propositions

Peano did not give definitions for fundamental concepts of arithmetic, such as number, the unit, and successor. As he pointed out "from a practical point of view, the question seems to have been resolved, that is, it is not convenient in a teaching to give a definition of number, as this idea is very clear to students and each definition has only the result of confusion." (Peano, 1891). The definition of number was also the point of Russell's strong criticism of Peano's Arithmetic (Segre, 1994).

In contrast, Dedekind gives an extremely complex definition using the concept of chain:

> «71. Definition. A system N is said to be *simply infinite* when there exists a similar[2] mapping φ of N into itself such that N appears as the chain (44) of an element not contained in $\varphi(N)$. We call this element, which we shall denote in what follows by the symbol 1, the *base-element* of N, and say that the simply infinite system N is *ordered* by this mapping φ. If we retain the earlier convenient symbols for images and chains (§4) then the essence of a simply infinite system N consists in the existence of a mapping φ of N and an element 1 which satisfy the following conditions α, β, γ, δ:
>
> α. $N' \ni N$.
>
> β. $N = 1_0$.
>
> γ. The element 1 is not contained in N'.
>
> δ. The mapping φ is similar.
>
> […]

[1] 37. Definition. K is called a *chain* when $K' \ni K$. (Dedekind, 1888) (Let ϕ be a mapping of the system K. Then $K' = \phi(K)$).

[2] 26. Definition. A mapping ϕ of a system S is said to be *similar* or *distinct*, when to different elements a, b of the system S there always correspond different images $a' = \phi(a)$, $b' = \phi(b)$. (Dedekind, 1888)

73. Definition. If in the consideration of a simply infinite system N ordered by a mapping φ we entirely neglect the special character of the elements, simply retaining their distinguishability and taking into account only the relations to one another in which they are placed by the ordering mapping φ, then these elements are called *natural numbers* or *ordinal numbers* or *simply numbers*, and the base-element 1 is called the *base-number* of the number-series N.» (Dedekind, 1888)

The axioms given by Peano in his original text are nine of which four refer to the relation of equality. A few years later, in *Rivista di matematica 1* he somewhat transforms his system by giving the five known axioms that are still characterized today as Peano axioms:

«1. 1 ε[3] N

2. + ε N\[4]N

3. a, b ε N .[5] a + = b + : Ɔ[6] . a = b

4. 1 -[7] ε N +

5. s ε K . 1 ε s . s + Ɔ s : Ɔ . N Ɔ s» (Peano, 1891)

It is worth noting here that Peano makes the following comment: "The previous sentences are due to Dedekind, however, there is a slight difference in the statement of sentence 5" (Peano, 1891). It is obvious that the first four axioms of Peano are indeed identical to those which Dedekind incorporated in his definition of natural numbers. The 5th axiom, which is nothing more than mathematical induction, is given by Peano as an axiom while Dedekind proves it as a proposition:

«80. Theorem of complete induction (inference from n to n'). In order to show that a theorem holds for all numbers n in a chain m_0, it is sufficient to show,

ρ. that holds for $n = m$, and
σ. that from the validity of the theorem for a number n of the chain m_0 its validity for the following number n' always follows.

This results immediately from the more general theorems (59) or (60). The most frequently occurring case is when $m = 1$ and therefore m_0 is the complete number-series N.» (Dedekind, 1888)

Then, both authors give and prove a series of properties of natural numbers as well as basic operations. Peano does not define addition while all its properties are proven using induction. Additionally, subtraction, multiplication, and powers are defined recursively. Dedekind follows a similar path for properties and operations within natural numbers, with the main difference being the recursive definition of addition.

A comparison in the style of the two texts one can easily see that Peano's text uses the language of logic as a tool, while Dedekind instead wanted to reduce these "axioms" to deep logical principles, so that propositions could be made for him. (Kahle, 2021). Moreover, on one hand *Arithmetices principia, nova methodo exposita* is easier to read but does not fully prove all the properties of numbers considering that the reader can practice the method on his own by proving the sentences. On the other hand, Dedekind's book is much more detailed but also more

[3] «The sign ε means *is*.» (Peano, 1889)

[4] «Essendo a e b delle classi, con a\b intenderemo «segno che messo dopo un a produce un b.»» (Since a and b are classes, with a \ b we mean «a sign that put after a produces a b.») (Peano, 1891)

[5] «To show the order in which they should be taken, we use *parentheses*, as in algebra, or dots, ., :, ∴, ::, and so on. […] Then *ab.cd* means (*ab*)(*cd*)» (Peano, 1889)

[6] «The sign Ɔ means *one deduces*.» (Peano, 1889)

[7] «The sign – is read *not*.» (Peano, 1889)

difficult for the reader. Finally, Peano's presentation is much closer to what we know today about Arithmetic.

4 Conclusions

When *Was sind und was sollen die Zahlen?* was published, Dedekind was a famous mathematician. He worked on the foundations of Arithmetic for several years and for which he exchanged views with other mathematicians of his time. It is very likely that the discussion was transmitted through the various editions to Turin where the young Peano lived and studied. Peano in the early years of his career worked as an assistant professor at the University of Turin and more specifically as an assistant to Genocchi. At that time, his research interest was to look for errors in the notes of Calculus, which he taught, to correct them and to offer a lesson as comprehensible as possible to his students. In his view, the errors were due to natural language, so he studied a way to express the theories he studied in symbolic language, which was also Leibniz's vision. It is not excluded that all this time he came to contact with Dedekind's effort. Immediately after the *Arithmetices principia, nova methodo exposita*, Peano publishes the *I Principii Di Geometria Logicamente Esposti* in which he attempts to apply his method in Geometry as well. In 1897, at the 1st World Mathematical Congress in Zurich, he presented his papers, and claimed to have answered Leibniz's question:

> «After two centuries, this “dream” of the inventor of the infinitesimal calculus has become a reality…. We now have the solution to the problem proposed by Leibniz. I say “the solution” and not “a solution”, for it is unique. Mathematical logic, the new science resulting from this research, has for its object the properties of the operations and relations of logic. Its object, then, is a set of truths, not conventions.» (Kennedy, 2002)

While Peano's international exposure was mainly at the 1900 Conference in Paris where his school dominated the discussions, it then focused on constructing a technical language as well as teaching.

It is obvious that the two mathematicians started from a different starting point and with a different goal. Peano's main goal was to apply a symbolic language to a theory, within the framework of formal logic, while Dedekind was interested in the foundation of Arithmetic, something he had been studying and discussing for about 17 years. Their work presents the same theory with different methods of approach. It is possible that Peano knew Dedekind's studies and used them to apply his method of translating a mathematical theory into symbolic language. What is certain is that after the publication of *Arithmetices principia, nova methodo exposita*, Peano read *Was sind und was sollen die Zahlen?*, given that he "omitted" four of his original positions, resulting in 4 of the final 5 to being similar to those of Dedekind. As well as that he went into the process of explaining why he does not consider it necessary to get involved in trying to grasp the meaning of number, something that did not seem to bother him in its original version.

One of the main differences between the two texts is the definition of number. While Peano insists on the non-necessity of definition, Dedekind creates a new concept, that of the chain, aimed at definition in set theoretic terms. Another important difference between the two texts has to do with mathematical induction. While Peano accepts it as an axiom, Dedekind presents it as a theorem, which he proves.

Finally, I think it is remarkable that although the names of both mathematicians are inextricably linked to Number Theory, and while Peano mentions Dedekind in both the *Arithmetices principia, nova methodo exposita* and *Rivista di matematica 1*, Dedekind makes no mention of Peano, not even to the introduction of the third edition of his work in 1911, 27 years after first edition and 22 years after the publication of *Arithmetices principia, nova methodo exposita*. And last but not least, no evidence was found to prove any meeting of the two mathematicians nor any attempt to communicate each othter, in any way.

References

Dedekind, Richard. 1888. Was sind und was sollen die Zahlen? [συγγρ. βιβλίου] William Bragg Ewald. *From Kant to Hilbert, Volume 2: A Source Book in the Foundations of Mathematics. 787-833.* Oxford University UK : Clarendon Press.

—.1890. Letter to Keferstein. [συγγρ. βιβλίου] Jean van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931: 98-103*.

Ewald, William Bragg. 1996. From Kant to Hilbert, Volume 2: A Source Book in the Foundations of Mathematics. s.l. : Clarendon Press, Oxford University, Vol. 2, pp. 787-833.

Ferreirós, José. 2005. Richard Dedekind (1888) and Giuseppe Peano (1889), booklets on the Foundations of Arithmetic. *Landmark Writings in Western Mathematics 1640-1940.* s.l. : Ivor Grattan-Guinness, 47, pp. 613-626.

Grattan-Guinness, Ivor. 2011. Giuseppe Peano: a Revolutionary in Symbolic Logic? [ed.] Skof F. *Giuseppe Peano between Mathematics and Logic.* Milan : Springer.

Kahle, Reinhard. 2021. Dedekinds Sätze und Peanos Axiomata. *Philosophia Scientiæ*, Τόμ. 25/1.

Kennedy, Hubert. 2002. *Peano Life and Works of Giuseppe Peano.* San Francisco : Peremptory Publications.

Lolli, Gabriele. 2011. Peano and the Foundations of Arithmetic. [ed.] Skof F. *Giuseppe Peano between Mathematics and Logic.* Milan : Springer.

Peano, Giuseppe. 1889. The principles of arithmetic, presented by a new method. [book auth.] Jean van Heijenoort. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931.* pp. 83-97.

—.1891. Sul concetto di numero. Nota I. *Rivista di matematica 1.* pp. 87-102.

—.1891. Sul concetto di numero. Nota II. *Rivista di matematica 1.* pp. 256-267.

Segre, Michael. 1994. Peano's Axioms in their Historical Context. *Archive for History of Exact.* 48(3/4), pp. 201-342.

Volterra, Vito. 1908. Mathematics in Italy in the second half of the 19th century. *https://mathshistory.st-andrews.ac.uk/Extras/Volterra_history/.* [Ηλεκτρονικό]

Some Philosophical Remarks on the Current Definitions of Algorithms

Philippos Papayannopoulos

IHPST (UMR 8590), CNRS and Université Paris 1 Panthéon-Sorbonne, Paris, France
fpapagia@uwo.ca

Abstract

There has been a resurgent interest in formalizing the notion of 'algorithm'. In this paper, I discuss the relation between algorithms and computations, point to some tensions inherent in our informal concept of an algorithm, and discuss some trade-offs between competing desiderata for any proposed formal definition.

1 Introduction

The idea of an algorithmic procedure is almost as old as mathematics itself (see, e.g., [3]). Nevertheless, despite the long-standing prevalence of algorithmic methods in mathematics, attempts to formalize the concept of an 'algorithm' *itself* are relatively recent, and they are mostly a spin-off from the impressive conceptual advancements in understanding how to demarcate the computable functions. As is well known, this understanding came about as the result of seminal work in the 1930s, by Turing, Gödel, Church, Kleene, Rosser, Herbrand, and others.

The formalisms of Church, Rosser and Kleene (λ-calculus), Gödel and Herbrand (general recursion) and Kleene (μ-recursion) were soon proved equivalent and turned out later to capture what we now consider the correct class of number-theoretic computable functions. However, from a conceptual point of view, these particular formalisms lacked convincing power regarding their completeness, for there seemed to be no compelling reason why (e.g.) the general recursive or the λ-definable functions would include all and only those functions that can be calculated by purely mechanical means. A great reluctance to accept Church's thesis (in this form) was expressed by Gödel's famous comment to Church that such approaches were "thoroughly unsatisfactory". The situation changed radically when Turing's [19] analysis came along, which focused on the *process* of computation itself, by breaking it down into its conceptual constituents; this provided a low-level analysis of what can (and cannot) ultimately be achieved by purely mechanical and elementary steps, carried out by an (idealized) human agent. Turing's analysis was widely conceived as conclusive, and the Church-*Turing* thesis (CTT) became a universally accepted foundation for computer science, and especially computability and complexity.

The fact that Turing's analysis focused on the *process* of computation, together with the (seemingly innocuous) tacit assumption that what is meant by a "mechanical process of computing a function" (aka "effective procedure") coincides with what is meant by "execution of an algorithm" led to the widely held view that the CTT and the Turing Machine (TM) formalism explicate the notion of algorithm. As a result, this view has become part of the folklore of logic and computer science (CS).[1] However, Turing does not mention 'algorithms' at all in [19], and while Church [4] does use the term, he is not concerned with the process of computation itself (only with the extension of the concept of 'computable function'). In actual fact, then, the 1930s developments did not concern the (intentional) concept of algorithm per se but solely the demarcation of the class of computable functions (which is an extensional concern).

[1] See, e.g., [16, 3,102] and [11, 246] for two examples of this view being clearly articulated.

When did the interest in the idea of algorithm itself come along? Markov's [12, 13] seem to be among the very first works that claimed to define '*algorithms*'. But Markov's definition was a narrow one (not too different from a Turing program), unable to capture the informal notion in its generality. But, without an intentionally good definition of algorithms, it was still a conceptual possibility that one could follow a procedure that is much more permissible than Markov's —yet would still seem algorithmic— and get to compute a function that's beyond the class of partial recursive ones. To rule out such scenarios Kolmogorov and Uspenskii (K&U) set out to give the first full-fledged formal definition of algorithms, in a work so influential [10] that some of the ideas it introduced are still found today (even implicit) in almost every work in the area. But what is exactly the relation between algorithms and computable functions?

2 Algorithms and computation: A marriage made in heaven?

Computability is a semantic notion. A function is computable if it is such that its values can be identified by a process of computation; that is, by following a mechanical procedure. But the process of computation is syntactic and symbolic. In carrying out a computation, an agent (human or otherwise) deals with concrete entities (symbols on paper, physical voltages, etc.). Insofar as algorithms are understood as specifying mathematical computations, then, they specify procedures over symbols. Shapiro echoes exactly this view:

> Mechanical devices engaged in computation and humans following algorithms[..] do not encounter numbers themselves, but rather physical objects such as ink marks on paper. Since strings are the relevant abstract forms of these physical objects, *algorithms should be understood as procedures for the manipulation of strings*, not numbers. ([18, p.14]; emphasis added)

Thus, on a view that sees algorithms as specifying actual computations, algorithms are procedures for manipulating symbols and, hence, synonymous to effective procedures. They are tightly interlocked with the representations of the data they operate upon. Given some vocabulary and a representation of the input by strings of symbols from this vocabulary, an algorithm is a stepwise procedure for combinatorily manipulating these symbols and obtaining a result, which is a representation of the computed function's output. Since a schoolchild in ancient Rome would be taught a different combinatory sequence of steps for multiplying two 3-digit integers from a schoolchild in ancient Greece (owing to the different notation systems), the two children would have mastered *different* algorithms for obtaining the product of two numbers (and the same holds for multiplying two integers today in, say, decimal and binary notations).

This presupposition is clearly seen embedded in Markov's as well as in K&U's approach to defining algorithms: "Without fixing a standard way of writing numbers, to speak of the algorithm computing [the value of a function from its input] would make no sense." [10, fn.2]. What is more, if one goes further and simply *identifies* algorithms to Turing programs, then the above presupposition becomes also reflected in the dominant contemporary approach to real computability, in terms of Type-2 Turing Machines (TTE) [21]. As some key results in this area indicate, when Turing computations are extended to uncountable domains (such as $\mathbb{R}$) computability of functions acquires a strong dependence on the employed representations.[2]

The strong association of algorithms with computation has placed the concept of 'algorithm' at the heart of computer science. Since it is indeed a dominant view that *algorithms specify*

[2] For example, there is no (Type-2) Turing machine —hence no algorithm either— that computes the function: $g(x) = 3x$ ($g : \mathbb{R} \to \mathbb{R}$) on the decimal representation. Yet, the same function is (almost trivially) computable on the base-3 representation.

computations, statements about specific properties of algorithms (including existence) are staples in areas such as computability and complexity theory. Consider, e.g., assertions like "there is no algorithm that decides the validity for any first-order sentence" or "if $\mathbf{P} \neq \mathbf{NP}$, there is no algorithm that solves the Boolean satisfiability problem in polynomial time."

3 The third in the marriage: Mathematical practice

Despite the well-received view of algorithms and computations as being tightly interlocked, one quickly notices non-negligible conceptual problems. The view of algorithms as specifying computations does not quite square with prominent uses of the term in certain areas of mathematical practice. To wit, recall that algorithms have also been the subject-matter of the long-standing field of numerical analysis. Numerical algorithms concern continuous problems, and their purposes, very often, include identifying (exactly or approximately) solutions of (systems of) equations, guiding linear interpolation, etc. Typical examples include the bisection method, least-squares fitting, Gaussian elimination, Newton's method, and many others.

It does not seem natural to say that algorithms like the aforementioned specify computations (i.e., syntactic procedures) in the same sense that we saw in the previous section. They definitely do not specify exact sequences of steps to the smallest detail, in the sense that any specified sequence would have to change, had the employed notation (or representation) changed as well. Rather, numerical algorithms are developed and analyzed without any consideration of notation or representation, and they are naturally thought of as each one possessing a natural structure and identity of its own; a structure and identity that are invariable under changes in how data are represented and in what the exact order of operations is.[3] This attitude toward the fundamental idea of an 'algorithm' is an explicit motivation behind the Blum-Shub-Smale (BSS) model of real computability [2] as well as Moschovakis's foundational approach to algorithms (e.g., [14, 15]). As L. Blum puts it:

> We want a model of computation which is more natural for describing algorithms of numerical analysis, such as Newton's method [..] Translating to bit operations would wipe out the natural structure of this algorithm. [1, 1028]

Indeed, attending to the long-standing study of numerical algorithms and their purposes ([3], [6]) shows that it would be stretching a point to say that (e.g.) Newton's method is a mechanical procedure for pushing symbols around, in the sense found in works on algorithms like [13], [10], [9], and others. In stark contrast to the multiplication example from above —i.e., different algorithms for different notations— Newton's algorithm arguably remains the same (abstract) entity, regardless of what notation is used or what the exact order of operations is.

One might try to remedy this apparent discrepancy between the two understandings of algorithms just described by saying that numerical algorithms still specify computations, albeit at some "higher level" of abstraction. That is, they still report mechanical procedures, but by abstracting away from any particular details of the process (such details can *always* be filled in later). But, as usual, the devil hides in the details: consider that a good many numerical algorithms specify as *essential* steps comparisons between real numbers; think, for example, the bisection algorithm. This creates a conceptual gap between numerical algorithms and actual mechanical computations, because comparing two reals is in fact not effectively decidable. More precisely, comparisons (and identity) are not decidable by a (Type-2) TM [21]; they are

[3] This holds also for CS algorithms; think, e.g., of the MERGESORT algorithm. But, for reasons that will only briefly be touched upon here, the case for numerical algorithms is stronger, because the existence of their natural structure is orthogonal to whether they are realizable by a TM (which is not the case with CS algorithms).

only (negatively) semi-decidable. As a result, if we adopt an extended version of the CTT (an '*Uncountable-CTT*') to the effect that "the effectively computable real-valued functions are exactly the functions that are computable by a Type-2 TM" —which is a very natural assumption—, then most numerical algorithms will turn out not to be effectively computable. This indicates that the link between algorithms and effective procedures/mechanical computations has to be severed. But now a fundamental question arises: how, and to what extend, should the above considerations be taken into account by any attempt to define algorithms?

4 A tension in definitions: inclusive or specialized?

Mathematical definitions often face challenges imposed by a strong tension between generality and inclusiveness on the one hand and domain-specific fecundity on the other. Regarding algorithms, we have, on the one hand, the desire to include under some unified formal concept both algorithms over countable and uncountable domains; in particular, both effective procedures and numerical algorithms; but, ideally, the definition might also go some distance toward subsuming additional related notions such as parallel algorithms and geometric constructions, so that a uniform study of all these notions could become possible. This desideratum for inclusiveness pushes in the direction of a formal concept that is as abstract as possible. More specifically, we would like to have a formal explicatum of 'algorithms' such that: (a) the identity of any algorithm is not essentially dependent on the representations of the data it operates upon or on the finest-grained details of its evolution (so it is not essentially affected by implementation details); (b) it lends itself to a spectrum of primitive operations of variant strengths (a desideratum that is best served by a model/structure/level-theoretic view of algorithms, since in that case a step can be any primitive operation defined as such by the model/structure/level itself); (c) it retains its applicability to particular domains, so it does not contradict fundamental results of the more specific instances of the same concept (i.e., it should preserve basic theorems of computability theory or of numerical analysis). The formal approaches by Blum et al. [2], Gurevich (e.g., [7, 8]), and Moschovakis (e.g., [14, 15]) all satisfy the first two conditions, for they all offer formal concepts that are representation-invariant and level-relative (though not necessarily effective), trying purposely to capture algorithms that go beyond Turing programs, while TTE is mainly the only model with a wide scope (it applies to both countable and uncountable domains) which satisfies (c) (though it does not satisfy a and b).

On the other hand, the desire for the formal counterpart to be such that it lends itself to interesting relations with well-entrenched concepts pushes in the direction of a formal concept with a significantly narrower domain of application than in the previous case. More specifically, we would like to have a formal explicatum of algorithms such that (a′) it retains as much as possible the intentional character of the informal processes it purports to formalize; (b′) more importantly, it would feature in deep theorems and connect to other well-established concepts, such as those from complexity theory. Formal approaches like K&U machines [10], (ordinary or Type-2) TMs [21], and BSS machines [2] are successful models in these regards but, predictably, each one meets only one of the two conditions and in a particular domain. As I discuss next, K&U fares better in (a′) and applies to discrete algorithms, while BSS and (ordinary or Type-2) TMs fare better in (b′) and apply solely to either discrete or real algorithms but not both.

5 How the existing definitions meet the different needs

A main upshot of the above discussion is that it seems a Herculean task to find a formal explicatum of 'algorithms' that respects all our intuitions and expectations together. This is

not an unusual situation in logic and mathematics. In many other cases, however, a certain formal concept among the rivals (capturing *some* of what we consider the essential features of the intuitive idea) catches on and becomes the "orthodoxy", on account of being successful in providing interesting results ('continuity' being a case in point). But in the case of algorithms most formal frameworks have been fruitful already, even though some of them are genuinely incompatible (e.g., TTE vs. BSS).

The way I see the situation, then, is that 'algorithm' is a cluster concept; this means that in order to be able to give preference to any particular framework, the community first needs to have decided on which intuitions and goals to prioritize. In what follows, I will consider some of the (conflicting) *intuitions* and possible different *goals*. Undoubtedly, there are many intuitions about algorithms that most mathematicians would agree on. Here I will focus only on those that I think practitioners might rather disagree.

Symbolic vs. Abstract: Is the identity of an algorithm relative to the vocabulary of symbols it operates upon? For example, by changing from a decimal to a binary notation, would we have different algorithms of (say) multiplication or one algorithm with different implementations? At its heart, the question concerns the extent to which the precise sequence of steps bears on the identity of an algorithm. Based on common informal characterizations of algorithms in logic texts (commonly to the effect that "an algorithm is *a precise, step-by-step procedure...*") any difference in the exact sequence of steps (caused by the different notations) would give rise to a different algorithm. But based on the (also) common practice of assigning specific names (e.g., 'Euclid's algorithm') and properties (e.g., asymptotic running costs) to various algorithms, "small variations" in steps should not affect the algorithm's identity. To give an example of what is at stake: in sequentially executing a MERGESORT, does the algorithm change if we stipulate that the left-most possible merge operation is to be executed first (and the right-most one second) or if we stipulate the opposite? Intuitively, we might want to say that it is always the same algorithm, which is just implemented differently; so algorithms are abstract objects in a sense. But, then, such an abstract notion with no additional constraints may be too broad to underpin algorithmic analyses, for, it may allow of "algorithms" that trivially accomplish complicated tasks within just one step.[4] This is because, in practice, the way to exclude such unrestricted cases is by assuming that the algorithms that are suitable for underpinning complexity analyses are those that are easily couched in some formal model of computation ([5]) from the first machine class (fn.5). But this leads us back to granting conceptual priority to machine models, i.e., entities that have sensitive dependence on notational choices. The formal concepts of (Type-2) TMs, K&U machines, ASMs (Gurevich), BSS machines (Blum et al.) and recursors (Moschovakis) tackle these questions differently. But there seems to be no way of ranking our preferences for these concepts on the basis of how well they address the above questions, unless one has already decided on answers to the above questions pre-formally.

Absolute vs. Relative: Are algorithms absolute entities, whose existence is a yes-or-no matter, or relevant with respect to some structure/model/level of abstraction, whose existence is dependent on the defined primitive operations over the stipulated entities in the structure's/model's/level's universe? While a choice on this matter may have no significant bearing on algorithms over countable domains, it does make a difference in the case of uncountable domains, for the latter approach may give rise to algorithmic steps that even involve infinitary labor on a symbolic configuration within one step —e.g., a complete operation between two irrational numbers— and to functions that *would* be deemed algorithmically computable with-

[4] Consider a TRIVIALSORT(B) algorithm for sorting a list B, whose sole instruction reads: "Return sort(B)". The unique step of this algorithm is effective (since effective sorting algorithms exist) and the algorithm is very efficient (running time is $O(1)$). Clearly this is an undesirable "algorithm" for purposes of algorithmic analysis, for, if accepted, it would lead astray our analyses of the complexity of sorting tasks. The example is from [5].

out being effectively computable (an example is the floor function). A related dilemma has to do with the notion of an algorithmic step, and whether any such step is required to be "local" in the sense of some pre-fixed suitable metric or just in relation to the stipulated primitive operations within the given structure/model/level in which the algorithm lives. TMs (ordinary and Type-2) and K&U machines can be seen as formalizing an "absolute" view of algorithms and steps, while BSS, ASMs and recursors can be seen as capturing a "relative" view.

Turning now to trade-offs between *goals*, it seems difficult to single out a formal explicatum on hopes that it would be responsive to all the linguistic and technical practices in mathematics and computer science. An important issue is complexity theory. Can we single out a formal explicatum that would underlie a unified complexity theory for both computer science and numerical analysis? To answer, consider that in computational and mathematical practice we grant (discrete and numerical) algorithms intrinsic asymptotic running time costs. As Dean [5] notes for the discrete case (but also holds for the numerical one), such asymptotic costs must be preserved by any particular machine model that aspires to formalize these algorithms. Ordinary TMs (or equivalent models) satisfy this condition for the discrete case. Therefore, such models support a rich theory of classical complexity and a network of powerful theorems. But the TM model is too narrow to express algorithms in their generality, and it violates the first two inclusiveness desiderata from above (a and b).[5] And when it comes to computations over uncountable domains, although the TTE-framework offers also a relevant complexity theory for real computation [21], it is however too "low-level" to be naturally used by practitioners [17]. Recall after all that Type-2 TMs cannot compute comparisons between reals, which are staples in numerical algorithms. Finally, the BSS formalism, which accepts highly-idealized TMs that operate on exact real numbers as unanalyzed entities in an algebra (so it permits comparisons in a single step) does provide a rich complexity theory for numerical analysis (so it satisfies a, b, and c′). But a BSS machine is far too powerful to be a first machine class, so it cannot be a formal concept that relates to the concepts of classical complexity theory for discrete problems.

The upshot is that formal concepts that turn out to be successful in theorem-generation (in particular, those that support a rich complexity theory in some particular domain) achieve this goal at the expense of generality, violating either (a) or (b) or even (c). On the other hand, Gurevich's and Moschovakis's frameworks fare much better at the generality desideratum. But, as Dean [5, p.54] notes, their achieved generality comes at the cost of severing the foundational link between the practice of informal algorithmic analysis (concerning discrete algorithms) and the complexity costs of first class machine models.

6 Conclusions

I have proposed that there is no unambiguous and uniform way in which the concept of algorithm functions in mathematical and computational practice. Consequently, there is no unique informal concept that could serve as the yardstick by which we evaluate the success of the formal concepts purporting to explicate it. The inherent tensions in our long-time use of algorithms can be alleviated by deliberately sharpening the informal concept *in advance*. And yet there is a number of different ways of trading off inclusiveness against strength of results, which makes it possible that in the end we will have more than one formal notion of 'algorithm' established in the theoretical and practical discourse.

[5] In fact, the situation is worse, because there are additional restrictions for those TMs that found complexity classes. Such machines form an (equivalence) class, called the *first machine class*, which imposes restrictions on the computational power of its members. First machine class models must be powerful enough to handle representation of numbers in binary, but no so powerful as to allow parallel computations with arbitrary branching (see [20]). As it becomes apparent, this pushes even stronger in the direction of a specialized formal concept.

Acknowledgments

This work has been partially supported by the ANR project *The Geometry of Algorithms – GoA* (ANR-20-CE27-0004). I am thankful to Alberto Naibo for many stimulating discussions on algorithms and to the anonymous reviewers of PLS13 for their comments and suggestions.

References

[1] Lenore Blum. Computing over the reals: Where Turing meets Newton. *Notices of the AMS*, 51(9):1024–1034, 2004.

[2] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. *Complexity and Real Computation*. Springer Science, 1997.

[3] Jean-Luc Chabert, editor. *A History of Algorithms: From the Pebble to the Microchip*. Springer-Verlag, 1999.

[4] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363, 1936.

[5] Walter Dean. Algorithms and the mathematical foundations of computer science. In L. Horsten and P. Welch, editors, *Gödel's Disjunction: The scope and Limits of Mathematical Knowledge*, pages 19–66. Oxford University Press, 2016.

[6] Herman H. Goldstine. *A History of Numerical Analysis from the 16th Through the 19th Century*. Studies in the History of Mathematics and Physical Sciences. Springer-Verlag, 1977.

[7] Yuri Gurevich. Sequential abstract-state machines capture sequential algorithms. *ACM Transactions on Computational Logic (TOCL)*, 1(1):77–111, 2000.

[8] Yuri Gurevich. What is an algorithm? In M. Bieliková, G. Friedrich, G. Gottlob, S. Katzenbeisser, and G. Turán, editors, *SOFSEM: Theory and Practice of Computer Science 7147*, pages 31–42. Springer, Berlin, 2012.

[9] Hans Hermes. *Enumerability, Decidability, Computability: An Introduction to the Theory of Recursive Functions*. Springer-Verlag, 2nd. edition, 1969.

[10] Andrey N. Kolmogorov and Vladimir A. Uspenskii. On the definition of an algorithm. *American Mathematical Society Translations*, 29:217–245, 1963. Translated from the Russian by Elliott Mendelson. Original publication 1958.

[11] Harry R. Lewis and Christos H. Papadimitriou. *Elements of the Theory of Computation*. Prentice-Hall., 2nd edition, 1998.

[12] Andrey A. Markov. The theory of algorithms. *American Mathematical Society Translations*, (Series 2)(15):1–14, 1960. Original publication 1951.

[13] Andrey A. Markov. *Theory of algorithms*. Israel Program for Scientific Translations, 1962. Translated from the Russian ed. by Jacques J. Schorr-Kon and PST Staff. Original publication 1954.

[14] Yiannis N. Moschovakis. On founding the theory of algorithms. In H. G. Dales and Gianluigi Oliveri, editors, *Truth in Mathematics*, pages 71–104. Clarendon Press, Oxford, 1998.

[15] Yiannis N. Moschovakis. What is an algorithm? In Björn Engquist and Wilfried Schmid, editors, *Mathematics Unlimited — 2001 and Beyond*, pages 929–936. Springer, 2001.

[16] Piergiorgio Odifreddi. *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers*. Elsevier, 2nd. edition, 1999.

[17] Philippos Papayannopoulos. Unrealistic models for realistic computations: How idealisations help represent mathematical structures and found scientific computing. *Synthese*, 199:249–283, 2021.

[18] Stewart Shapiro. Acceptable notation. *Notre Dame Journal of Formal Logic*, 23(1):14 – 20, 1982.

[19] Alan M Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42(1):230–265, 1936.

[20] Peter van Emde Boas. Machine models and simulations. In *Handbook of Theoretical Computer Science (Vol. A): Algorithms and Complexity*, page 1–66. MIT Press, 1990.

[21] Klaus Weihrauch. *Computable Analysis: An Introduction*. Springer Science, 2000.

Non-monotonic rule-based logical programming for modelling legal reasoning

the example of Answer Set Programming

Evangelos Iatrou

Institute of Logic, Language, and Computation, University of Amsterdam,
Amsterdam, The Netherlands
evangelos.iat@gmail.com

Abstract

In this paper, I will present the motivation for using non-monotonic rule-based logical programming to model legal reasoning and its main limitations caused by the notion of *truth* in legal practice. I use Answer Set Programming (ASP) as a toy example of such programming methods. In §1, I introduce the origins of the problem of modelling legal reasoning using logic and the basics of ASP. In §2.1, I argue why legal reasoning can be construed as rule-based reasoning focusing on the *subsumptive-deductive* character of the former. In §2.2, I demonstrate how non-monotonicity can model *defeasible reasoning* and the *presumption of innocence* - both inherent core characteristics of legal reasoning. Moreover, in §2.3, I contend that rule-based logical modelling is one of the few programming methods that render modelling of legal reasoning useful for legal practice, in contrast with more popular and efficient black-box AI methods. Moving to §3, I focus on the most persistent and critical limitations to the proposed modelling method induced by the conception of truth in legal practice; that of *interpretation* and the notorious in normative logic(s) (?) *Jørgensen's dilemma*. Finally, in §4, I conclude on the future direction of the proposed modelling method.

Keywords: rule-based logical programming, non-monotonicity, legal reasoning, subsumption, defeasible reasoning, interpretative concepts, Jørgensen's dilemma, explainable models - XAI

1 Introduction

The idea of using *logic* to systematize legal reasoning precedes the Fregean revolution of formalising logic and the burst of logical systems that spawned from it. Probably the most characteristic such effort was Christopher Columbus Langdell's[1] "*Selection of Cases on the Law of Contracts*"[2] in which Langdell construes *Law* as a logical system of principles and doctrines like any "*proper*" science. Langdell's ideas were immediately met with criticism with some even characterising them as "*logical theology*" [17]. Oliver Wendell Holmes[3] - one of the most prominent Langdell critics - contended that "*[t]he life of the law has not been logic; it has been experience.*". Particularly, Holmes argued that since Law is shaped by moral, political, and historical factors, it can not be "*dealt with as if it contained only the axioms and corollaries of a book of mathematics*". There is something *more* to it. The gist of his criticism is perfectly summed up in Susan Haack's article about that debate: logic is for legal reasoning "*something, but not All*" [17].

It has been more than a century since Langdell's ideas, and the debate regarding the logical formalisation of legal reasoning is still going on. Two pivotal points in that debate have been: (a) the realisation that a formal logic of norms[4] - if it is even possible - would differ from the classical formal logic of propositions. Jørgensen's 1937 paper "*Imperatives and logic*" ([19]) is usually the reference point of the dichotomy between a logic of norms and a logic of propositions.; (b) the emergence of symbolic

[1] C. C. Langdell (1826-1906) was an American jurist, academic and the first Dean of Harvard Law School [20, 17].

[2] According to [17], the first edition was published in 1871. A more contemporary publication of the last edition is [21].

[3] O. W. Holmes (1841 – 1935) was a U.S. Supreme Court Justice [17] and one of the most influential American legal scholars of the 20th century (the third most-cited one) [30]

[4] Laws can be construed as norms. What differentiate them from other norms (e.g., ethical ones) is that they are *binding* [32].

logical programming that was aspiring to simulate human expert reasoning. In [25] published in 1988 - one year after the first-ever *International Conference on AI and Law* (*ICAIL-1*) - one can already find a substantive review of the pros and cons of logical programming of legal reasoning. The reader can have a look at [29] published in 1986 for an elaborate effort of formalising the *British Nationality Act* using the logical programming language of *Prolog*. In this *position paper*, I weights in on contemporary aspects of that debate taking a clear stance in favour of the use of *non-monotonic rule-based logical programming* as a modelling method of legal reasoning. Despite that, in §3, I concede that Haack is still right; rule-based modelling characterises *some* parts of legal reasoning, *"but not All"*.

1.1 Answer Set Programming (ASP) as a rule-based logical programming method

To make my case, I will use as an example a particular rule-based logical programming method that has actually been used to model legal - and normative in general - reasoning (*see* e.g., [6, 12, 23]), that of *Answer Set Programming (ASP)*. In the classical "vanilla" ASP, a programme Π is a set of rules of the form `head :- body.`, where `head` is an atom `a` and the body consists of combinations of literals L_i, where each L_i can either be an atom `a` or its *default* negation `not a` [13]. The default negation of `a` can be construed as the case in which we have *not* proven *yet* that `a` is the case, while the classical negation (symbolized as `-a`, where `-a` is also an atom in contrast with `not a`) can be construed as the case in which we have actually *proven* that `a` is not the case [2, 12]. An atom `a` (or its classical negation `-a`) is said to be *proven* whenever it appears to the head of rule whose body is satisfied. When we have the edge case where $body(rule) = \emptyset$ (i.e., `a:-.` or its shorthand `a.`) `a` is always proven. Hence, we call rule `a.` and the atom `a` itself a *fact*.

For every programme Π, its output is the *stable model* of Π, where a stable model can be construed as a regular logical model of Π which includes *only* those atoms which have been proven. Or alternatively, a stable model of Π can be construed as the $\subseteq -$*minimal* model of all logical models compatible with Π [13]. Π's stable (or $\subseteq -$minimal) models are called *answer sets*. Since facts `a.` and `-a.` are always proven, the facts of a programme Π necessarily belong to any answer set of Π. For instance, assume the programme $\Pi_1 := \{$`a:-b.,b`$\}$. Since `b` is a fact it has to belong to any model of Π_1. At the same time, since the body of `a:-b.` is satisfied, then its head (i.e., `a`) has also to belong to any model of Π_1. Since `a` and `b` are the only atoms proven, the answer set of Π_1 is $AS_1 = \{$`a`$, $`b`$\}$.

Finally, apart from atoms, ASP syntax has predicates of arity n (e.g., `loves/2` is a predicate with arity 2) and variables which in contrast with atoms start with capital letter (e.g., `Person` is a variable and `person` is an atom). A predicate whose arguments are atoms is also considered an *atom* (e.g., `loves(achilles,patroclus)` is an atom). Those details of ASP syntax are enough to make my case in the arguments that follow.

2 Modelling legal reasoning: why ASP

2.1 Legal reasoning as rule-based reasoning

The construal of legal reasoning as rule-based reasoning is quite common in the literature [5]. One way to perform such a construal is to view a legal inference as an inference whose premisses consist of both *particular facts* and *general rules* [22]. Let us borrow an example of such a rule from [16, p.679]: "*If x lives in Italy for more than 183 consecutive days over a 12-month period, then x is obliged to pay taxes in Italy on their worldwide income.*". What gives the aforementioned rule a character of *generality* is that it is applicable to any x that belongs to the concept $\mathcal{C}_1$:="*living in Italy for more than 183 consecutive days over a 12-month period*".[5] In the *particular* case where $x = Alice$, we have the following *inference*:

[5] According to MacCormick in [22], $\mathcal{C}_1$ should have been considered a predicate and not a concept. However, I would like to stay as ontologically neutral as possible, and hence, I construe $\mathcal{C}_1$ as a concept assuming that a concept is a more general

Example 1.

(n_1) If x lives in Italy for more than 183 consecutive days over a 12-month period, then x is obliged to pay taxes in Italy on their worldwide income. (general rule)

(p_1) Alice lives in Italy for more than 183 consecutive days over a 12-month period. (particular fact)

(n_2) Alice is obliged to pay taxes in Italy on their worldwide income. (conclusion)

As MacCormick argues [22], Example 1 is a *subsumptive-deductive inference*: the particular fact in (2) is subsumed by the general concept $\mathcal{C}_1$ of (1), and hence, we end up with the normative inference (3).

Let us see now how we can model Example 1 by using *rule-based* programming. A rule-based model is a model that consists of *rules* $\phi(x) \Rightarrow \psi(x)$ and *facts* $\chi(a)$, where ϕ, ψ, χ are propositional functions,[6] x is a variable and a is a term without free variables. Whenever $\phi(a)$ is true, then for every rule i of the form $\phi(x) \Rightarrow \psi_i(x)$ we have that $\psi_i(a)$ is true [16]. When they appear in the code of a programme Π, *rules model norms and facts model propositions* [16]. Specifically, $\phi(x) \Rightarrow \psi(x)$ is interpreted as "*Whenever $\phi(x)$ is the case, then $\psi(x)$* ***should be*** *the case.*" and the fact $\phi(a)$ is interpreted as "$\phi(a)$ ***is*** *the case.*". Usually, the *output* has *only* facts like $\psi(a)$. In the output (*not* in the programme), those facts can be interpreted as *norms*: "*According to the programme Π, the fact $\psi(a)$* ***must be*** *the case.*". Let's see now a modelling of Example 1's inference using rule-based programming:

Example 1 (Continuing Example 1 from p.3)**.** Assume that $\psi(x)$:="*x pays taxes in Italy on their worldwide income.*", $\phi(x)$:="*x lives in Italy for more than 183 consecutive days over a 12-month period.*". Then, a rule-based model of the subsumptive-deductive inference in p.3. would be the following: $\mathcal{M}_{taxes} := \{\phi(x) \Rightarrow \psi(x), \phi(Alice)\}$. Its output is $output_{\mathcal{M}_{taxes}} = \{\psi(Alice)\}$. In the lingo of ASP, this model can be written as $\Pi_{taxes} := \{$`taxes2Italy(X):-italy183(X).`, `italy183(alice).`$\}$ where "`taxes2Italy`" is used in the place of "ψ" and "`italy183`" is used in the place of "ϕ". Its answer set is $\mathcal{AS}_{taxes} = \{$`taxes2italy(alice)`$\}$.

2.2 The non-monotonicity of legal reasoning

A distinctive characteristic of ASP is that it is *non-monotonic*. *Non-monotonic logic* can be summed up in the following way: assume that you have a set of premisses P and from them, you can infer a set of conclusions C. In non-monotonic logic, if you add more premisses $p_1, p_2, ...p_n$ to P, there is a possibility that C will change. This is not possible in logics like classical propositional logic or classical first-order logic which are both monotonic [12, 34].

Non-monotonicity is necessary for incorporating in our model the *presumption of innocence* which is central in the legal tradition of liberal democracies [35]. Specifically, in every trial, the proposition p :="*The defendant is innocent.*" is considered a true conclusion of the facts of the case $\mathcal{F}$ *by default*. However, the Court may update its facts - e.g., the Court may accept DNA evidence f against the defendant - and based on that evidence, the truth value of p may change. The change of p's truth value (i.e., claiming that $\mathcal{F} \models p$, but $\mathcal{F} \cup \{f\} \models \neg p$) can only happen if we allow for *non-monotonic* inferences.

Let us have a look on how ASP can incorporate that rationale. Assume two ASP programmes Π_{innoc} and Π_{guilty} respectively:

```
innocent:- not evidence.        innocent:- not evidence.
guilty:- evidence.              guilty:- evidence.
                                evidence.
```

$\Pi_{guilty} = \Pi_{innoc} \cup \{$`evidence.`$\}$, i.e., Π_{guilty} includes all the premisses of Π_{innoc}. However, its answer set is different than that of Π_{innoc}: $AS_{guilty} = \{$`guilty`$\} \neq AS_{innoc} = \{$`innocent`$\}$. Hence, the non-monotonicity. What allows the non-monotonicity in this example is the default negation of `evidence`.

notion than that of a predicate. Moreover, in §3.1, I introduce the problem of *interpretation* which is a problem about legal *concepts*. Hence, I have to introduce the notion of "*concept*" in my ontology and its role in legal reasoning before describing a problem about it. The literature I use in my construal of legal concepts and the problem of interpretation is: [9, 28, 33].

[6] Their arities can vary.

Specifically, in Π_{innoc}, `evidence` is not a fact, and hence, it is considered negated by default; if there are no *proven* evidence of the contrary, we assume that the defendant is innocent. However, once we have proven evidence of his guilt (i.e., once we include the fact `evidence.`), `evidence` is no longer negated by default and consequently, the defendant is guilty.

Apart from the presumption of innocence, what makes non-monotonicity vital for the modelling of legal reasoning is that it allows us to model the *defeasibility* of the involved rules. Specifically, there are always situations in which although prima facie the facts of a case fall under the general concepts involved in a given rule, that rule is not applicable after all (it is "*defeated*") [16]. Ganascia [12] gives such an example using Kant's landmark example about categorical imperatives:[7]

Natural Language	**Answer Set Programming**
John is hiding.	`proposition(john_hidding).`
If O.J. finds John, he will murder him.	`person(oj). consequences(know(oj, john_hidding), murder).`
Lying about John's hideout is immoral.	`immoral(lie(P, PP)) :-`
	`person(P), proposition(PP), not non_deserve(P, PP).`

output: `immoral(lie(oj, john_hidding))`

In the "*Natural Language*" column, one can see the Kantians' argument: lying to O.J. is immoral, even if by saying the truth John would be murdered. Indeed, the answer set of the ASP model of the natural language sentences contains the atom `immoral(lie(oj, john_hidding))`, whose semantical interpretation is that it is immoral to lie to O.J. about John's whereabouts. Kant critics though would argue that lying should be acceptable if saying the truth leads to undesired consequences, like John's murder. This objection can be incorporated into the ASP programme of the "*Answer Set Programming*" column by adding the following code:

```
immoral(A) :- consequences(A,murder), non_deserve(M, N).
non_deserve(P, PP) :- person(P), proposition(PP), consequences(know(P, PP),murder).
```

new output: `immoral(know(oj,john_hidding))`

The additional code makes the atom `non_deserve(oj, john_hidding)` true, while previously it was false by default. While it was false by default, the body of the rule `immoral(lie(P, PP)):- person(P), proposition(PP), not non_deserve(P, PP).` was satisfied (*see* the red-lettered not in the initial ASP programme). But now that is not the case. Hence, the atom `immoral(lie(oj,john_hidding))` is no longer proven and hence, it is not contained in the new answer set. At the same time, due to the new rule `immoral(A):- consequences(A,murder), non_deserve(M,N).`, the atom `immoral(know(oj,john_hidding))` becomes a fact and therefore, it is included in the new answer set. I.e., now it is not immoral to lie to O.J. On the contrary, what is immoral is O.J. learning about John's hideout. This case of deafisibility can be construed either as a case of *conflicting rules* (e.g., the rule of "*not killing*" conflicting with the rule of "*not lying*") leading to the defeasibility of the rule whose harm is the "*lesser*" of the two (*lex inferioris*). Or as a case of *exclusionary rules*, i.e., certain cases of lying are *excluded* from the rule of "*not lying*" and the foregoing situation is such a case [16].

2.3 The necessity of explainability

In both chapters §2.1 & §2.2, my justification for the use of non-monotonic rule-based logical programming as a modelling method of legal reasoning is based on its resemblance to the experts' reasoning process (e.g., defeasible reasoning, subsumption). A reasonable objection to that approach is to question why that resemblance is a valid justification in the first place. Why not to use other AI methods that do

[7] The fact that Ganascia's example is about ethical rules does not make it irrelevant to legal rules. Legal rules can be construed as special cases of ethical rules [33].

not resemble the experts' reasoning - like the supervised machine-learning method of SVM (see e.g., [4]) - which are more widely used, have better accuracy, and are less computationally complex [1, 15].

I do not doubt that there are applications in which such alternatives would actually be a better modelling option. Having said that, in a large number of possible applications their practical importance is essentially *zero*.[8] *Prima facie*, that sounds like a very bold statement. But is it not. I will provide two arguments for my claim:

(2.3.1) In legal practice, the truthfulness of an inference like $p :=$"*The defendant is innocent.*" is *grounded* on the *authority* of a particular group of legal experts which is *authorised* by law. E.g., in a criminal trial, it is the *judges* that have the authority to decide the truthfulness of p and not e.g. the lawyers of the defence and the prosecution. In other practices though, like the practices of empirical sciences or medical practice, the truthfulness of a proposition is independent of the beliefs of any group of domain-experts; a patient has cancer independently of whether the doctors agree or not. In other words, in legal practice, truth is *established* by a group of authorised experts *in virtue of* their *authority*, while in other practices truth is pre-exists in an expert-independent reality (*ordo essendi*[9]) [3].

This distinction has the following consequence: in most practices, an algorithm is practically useful as long as its output coincides with the *ordo essendi* independently of whether it resembles the experts' rationale since that rationale does not have any influence on the *ordo essendi*. We may even end up with algorithms whose performance is higher than that of the experts since the experts' rationale may not be the *optimal* way to discover the *ordo essendi*. This can *not* the case in legal practice. The authorised experts' reasoning *is* the optimal way to find the truth since, without that reasoning, there is *no* such thing as truth. Consequently, when an algorithm *substitutes* experts that have the authority to establish a legal truth, then that algorithm "*has*" the obligation to provide a justification for the truthfulness of its outputs similar to the justification that the authorised experts would have provided had the algorithm not substituted them.

(2.3.2) Any automated process used in legal practice needs to be able to be challenged legally. If your online speech activity is censored by upload-filters,[10] you should be able to protest to that censorship, and the platform employing the upload filters should be able to defend their output. Hence, there needs to be a *justification* of why given this specific input (your online speech activity) the algorithm's output (the censorship) should be legally acceptable. That justification should have a *form* that makes it *amenable* to evaluation by legal experts - and black-box programming methods like SVM do not satisfy that requirement [1, 15]. How could legal experts argue about complex probabilistic optimization functions? Are those functions even meaningful at all? Note that this *explainability* requirement is not simply a unilateral opinion among legal practitioners, but it has also started being incorporated into official legislation to protect the public's interest, as one can see for instance in paragraph 2 of Article 13 of the *Genereal Data Protection Regulation (GDPR)* which dictates that the "*data-subject*" whose data is processed by "*automated decision-making*" should be provided with "*...**meaningful** information about the **logic** involved*" [14, 27].

3 Is truth compatible with rule-based logical modelling?

[8] For an overview on the applications of AI in legal practice have a look at [24] for the private sector and [10] for the public sector. The applications for which argument (2.3.1) holds are mostly those of the public sector. The strong hesitance of the the public sector to adopt AI toolkits is essentially grounded on that argument.

[9] Term borrowed from [8].

[10] An upload filter is an algorithm that *identifies* and *regulates unlawful* online user-generated data like child pornography, terrorist propaganda, and hate speech [27].

3.1 The problem of *interpretation*

In §2.1, we have seen how subsumption works in the rule-based modelling of legal reasoning: if the *particular facts* of a case fall under the *general concepts* of the rules of a given normative system, we infer the conclusion that follows from the application of those rules to those facts. In this inferential schema, we need to distinguish between the inference that follows from the *application* of the rules to the facts that they subsume and the *decision* that those facts are indeed subsumable by those rules. The latter is a necessary *pre*-requisite of the former. The *decision-making process* of the authorised legal experts deciding whether the facts of a case are particular instantiations of the general concepts involved in the rules they want to apply is called *interpretation* [22]. Therefore, for the conclusion of a legal inference like that of Example 1 to be *true*, the respective interpretation has to be true. I.e., the truth of a subsumptive-deductive inference *supervenes* on the truth of the interpretation. Consequently, if a model can not decide about the truth of the latter but takes it as a given input, then that model does not provide a *self-contained* representation of legal reasoning.

Example 1's proposed rule-based model Π_{taxes} is such a non-self-contained model. One could propose to *expand* it to a rule-based model Π'_{taxes} that also incorporates interpretation. To do so, we need to introduce a new ASP programme Π_{intrp} to model interpretation and then define Π'_{taxes} as $\Pi'_{taxes} := \Pi_{taxes} \cup \Pi_{intrp}$. As a rule-based model, Π_{intrp} will include *rules* and *facts*. The rules will be rules about whether an atom `a` (e.g., `alice`) belongs to the concept $\mathcal{C}_1$. That means that we need to include a rule of the form $\phi_1(x) \wedge \phi_2(x) \wedge ... \wedge \phi_n(x) \Rightarrow \phi(x)$ (for the definitions of $\mathcal{C}_1$ and $\phi(x)$ *see* §2.1). Regarding the facts of the model, they would be the facts about the atom `a` that allegedly satisfy the foregoing rules. In case that indeed `a` belongs to the concept $\mathcal{C}_1$, the answer set of Π_{intrp} will include the atom `italy183(a)`.

In order to decide on the rules that determine whether an atom `a` belongs indeed to a concept $\mathcal{C}$, it should be the case that there are specific fixed criteria of whether `a` belongs to $\mathcal{C}$. If that is the case, $\mathcal{C}$ is what is called a *criterial* concept [33]. *Prima facie*, $\mathcal{C}_1$ is such a concept. Essentially, the criteria of whether an entity belongs to it are in its name: to belong to the concept of *living in Italy for more than 183 consecutive days over a 12-month period* one has to live in Italy for more than 183 consecutive days over a 12-month period.[11] However, in principle, legal concepts are *not* criterial, but *interpretative* ones: the criteria under which an entity `a` falls under a legal concept $\mathcal{C}$ are decided each time by a group of the authorised experts and they are inevitably influenced by their background beliefs - mainly political and ethical ones [9, 33]. A prime such example is the decision of whether fetuses belong to the concept of *human*. Therefore, those criteria are anything but fixed. What should also be noted, is that the influence of the background beliefs in the interpretation of a concept is *not* a defect of legal reasoning that legal experts want to eliminate, but quite the opposite: plurality of background beliefs and the openness to disagreements is the only way to secure a unilaterally accepted conception of justice, and for some, an objective one [9, 28, 33]. Concluding, by accepting the thesis that interpretation *should* be influenced by the background beliefs of the authorised experts, we concede that there can not be a set of fixed rules Π_{interp}. Haack is still right: rule-based logical programming models *some* aspects of legal reasoning (the subsumptive-deductive part), *but not all* of it - at least not the part of interpretation.

Let us see this limitation in an actual example in the ASP literature. Morris [23] used ASP to model a rule of conduct among legal professionals practicing in Singapore. *Inter alia*, Morris' ASP programme has to model rules that refer to businesses that "*...[detract] from, [are] incompatible with, [derogate] from the dignity of the legal profession...*". Therefore, Morris uses rules that include the predicates `detracts_from_dignity_of_legal_profession/1`, `incompatible_dignity_of_ legal_profession/1`, `derogates_from_dignity_of_legal_profession/1`. A business belongs to one of the three aforementioned unary predicates based on how the authorised experts interpret the word "*dignity*", but also the words "*detract*", "*incompatible*", and "*derogate*". In a colloquial discourse, the latter three words are often used interchangeably; if an interlocutor had said that a business' activity *detracts* from the dignity of the legal profession instead of saying that it *derogates* that dignity, we would have considered that they

[11] One can still come up with borderline cases for which that criterion is contestable. For instance, what if someone was planning to leave Italy before the 123rd day but could not do so due to an accident?

meant exactly the same thing. But in legal practice, no words in a regulatory document are redundant; interchangeability is not possible. This is a prime example of the fine-grained job legal experts have to do when they interpret and of the inherent inability of ASP to perform such a *semantical* analysis by appealing to fixed sets of rules.

3.2 Jørgensen's dilemma: is a logic of norms at all possible?

In §2.1, I recommended the use of *rules* to model *norms*. By doing so, I exposed rule-based logical programming to the same criticism that any candidate logic of *norms* face, like the notorious *Jørgensen's dilemma* [3, 16, 18, 31, 32]. Let us have a closer look at it. A logic of norms is basically a logic of imperatives, where an imperative does not have to be construed necessarily as a grammatical mood, but as the content of an imperative *speech act* - like a command [18]. E.g., it can be both s_1 ="*Don't do that!*" and s_2 ="*You ought not to do that!*", where s_2 is not in an imperative mood, but it is nonetheless an imperative speech act whose content is the same as that of s_1. Now if we construe a rule as an imperative, then the premisses of a legal inference consist of both imperative (general rules) and indicative (facts) speech acts [3, 22]. Hence, for a legal inference to be valid, it has to be the case that whenever those imperative and indicative speech acts are true, then the conclusion is always true. However, what does it mean for an imperative speech act - e.g., s_1 - to be true? Admittedly, imperatives do not take a truth value! At the same time though, there seems to be a clear distinction between normative inferences which are reasonable (e.g., Example 1) and normative inferences which are not (e.g., Example 2):

Example 2. This is an example drawn from Danish philosopher Jørgen Jørgensen's 1937 "*Imperatives and logic*" paper ([19]) in which he introduced the dilemma. The coinage "*Jørgensen's dilemma*" was coined in 1944 by Danish philosopher and jurist Alf Ross in a paper of the same name ([26]) in which he tried to provide his own answer to the dilemma [32].

1. Love your neighbor as yourself!	(imperative)
2. You love yourself.	(indicative)
3. Do not love your neighbor!	(unreasonable conclusion)

The contrast between the inability of imperatives to take truth values and the intuition that there is a clear distinction between reasonable and unreasonable normative inferences leads to the notorious Jørgensen's dilemma consisting of the following two horns:

i. Logic examines whether from *true* premisses we can infer a *true* conclusion. Since some premisses of a normative inference can not be assigned a truth value (the premisses that are imperative speech acts), normative inference are *not* amenable to logical treatment. In other words, there can *not* be a logic of norms.

ii. However, (i) seems counterintuitive to many real-life cases of normative inferences like Example 1 which seems a fairly *reasonable* inference and like Example 2 which seems a fairly *unreasonable* inference. In other words, from simple real-life examples, there seems to be a clear-cut way to *logically* distinguish between reasonable and unreasonable normative inferences. If that is the case though, we would have to concede that logic is not limited to examining whether from *true* premisses we can infer a *true* conclusion; it has to be expanded to something more than *truth-preservation*.

Both of those choices seem costly. If we choose horn (i) and reject horn (ii), then we satisfy the intuition that logic is about truth-preserving inferences but we reject the intuition that there exists a logical distinction between reasonable and unreasonable normative inferences. If we choose horn (ii) and reject horn (i), then we satisfy the intuition that there exists a logical distinction between reasonable and unreasonable normative inferences, but we reject the intuition that logic is about truth-preserving inferences forcing us to introduce new metaphysical counterparts of truth (e.g., "*valid*" inferences next

to "*true*" inferences) as well as delineate the relations among all those counterparts (e.g., which *valid* inferences are also *true* inferences).

My personal preference so far is to choose the latter; it is an intuitive decision stemming from the intuition that there are more metaphysical counterparts to "*truth*" that characterise inferences. Fox [11] introduces a *logical calculus* of such counterparts. For instance, they recommend that instead of saying that a norm n is *true*, we should say that a norm n is *satisfied* by a *subject* σ (n Satisfied$_\sigma$). They label those notions under the umbrella term "*judgements*" J ($J ::= P$ True $\mid P$ False $\mid I$ Satisfied$_\sigma \mid I$ unSatisfied$_\sigma$,[12] where P is a proposition and I is a norm) and they propose a set of inference rules regarding those judgements. E.g.:

$$\frac{I_1 \text{ Satisfied}_\sigma \qquad I_2 \text{ Satisfied}_\sigma}{(I_1 \wedge I_2) \text{ Satisfied}_\sigma} \ , \ \frac{P \text{ True} \qquad I \text{ Satisfied}_\sigma}{(P \rightarrow I) \text{ Satisfied}_\sigma} \ .$$

For the reader that wants to look for more alternative solutions to the dilemma, I recommend the following literature: [16, 18, 32].

4 Conclusion

The arguments in favour of the use of non-monotonic rule-based logical programming to model legal inference are in no way exhausted in this paper. The same goes for the description of its limitations. Having said that, I hope that I have convinced the sceptical reader about the current realistic potential of the proposed modelling method. And if I have not, due to the inevitable necessity for models of legal inference that resemble the experts' reasoning delineated in §2.3, I hope I have made a strong case as to why we should continue pursuing ways to counter the problems that hinter self-sufficiency of that method - like the problems of interpretation and Jørgensen's dilemma. For the former, it is my strong belief that we should look for more reasoning methods - apart from deduction (normative subsumptive-deductive inference in particular) - to be able to model the decision-making process of when a particular fact is subsumable by a general rule. For instance, we could use analogical and/or counterfactual reasoning to compare past cases of applications of the same rule with the current case (appealing to *legal precedent*) [7, 5]. As for Jørgensen's dilemma, I have already stated in §3.2 that my intuition is to generate a logical calculus for metaphysical counterparts of *truth* in the meta-language of normative inferences following Fox's [11] paradigm. However, considering my current lack of expertise, this may end up being the wrong way; we will only know if we try.

References

[1] Bibal Adrien, Michael Lognoul, Alexandre de Streel, and Benoît Frénay. Legal requirements on explainability in machine learning. *Artificial Intelligence and Law*, 29:149–169, 2021.

[2] Felicidad Aguado, Pedro Cabalar, Jorge Fandinno, David Pearce, Gilberto Perez, and Concepcion Vidal. Revisiting explicit negation in answer set programming. *Theory and Practice of Logic Programming*, 19(5-6):908–924, 2019.

[3] Carlos E. Alchourrón. Limits of logic and legal reasoning. In Carlos Bernal and Carla Huerta, editors, *Essays in legal philosophy*. Oxford University Press, [1992] 2015.

[4] Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoţiuc-Pietro, and Vasileios Lampos. Predicting judicial decisions of the European Court of Human Rights: a natural language processing perspective. *PeerJ Computer Science*, 2:e93, 2016.

[5] Larry Alexander and Emily Sherwin. *Demystifying legal reasoning*. Cambridge University Press, 2008.

[6] Theo Aravanis, Konstantinos Demiris, and Pavlos Peppas. Legal reasoning in answer set programming. *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 302–306, 2018.

[7] Giorgio Bongiovanni, Gerald Postema, Antonino Rotolo, Giovanni Sartor, Chiara Valentini, and Douglas Walton, editors. *Handbook of legal reasoning and argumentation*. Springer, Dordrecht, 2018.

[12] This is not the complete version of Fox's Backus-Naur form of judgements J.

[8] Willem R. de Jong and Arianna Betti. The classical model of science: a millennia-old model of scientific rationality. *Synthese*, 174(2):185–203, 2010.

[9] Ronald Dworkin. *Justice for hedgehogs*. Belknap Press of Harvard University Press, 2011.

[10] European Commission for the Efficiency of Justice. *European ethical Charter on the use of artificial intelligence in judicial systems and their environment*. printed by the Council of Europe, 2019.

[11] Chris Fox. Imperatives: a judgemental analysis. *Studia Logica: An International Journal for Symbolic Logic*, 100(4):879–905, 2012.

[12] Jean-Gabriel Ganascia. Modelling ethical rules of lying with answer set programming. *Ethics and Information Technology*, 9:39–47, 2007.

[13] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. Answer set solving in practice. 2012.

[14] Biyce Goodman and Seth Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, 38, 2017.

[15] Łukasz Górski and Shashishekar Ramakrishna. *Explainable Artificial Intelligence, Lawyer's Perspective*, page 60–68. Association for Computing Machinery, New York, NY, USA, 2021.

[16] Guido Governatori, Antonino Rotolo, and Giovanni Sartor. Logic and the law: philosophical foundations, deontics, and defeasible reasoning. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leon van der Torre, editors, *Handbook of deontic logic and normative systems*, volume 2. College Publications, 2021.

[17] Susan Haack. On logic in the law: "something, but not all". *Ratio Juris*, 20(1):1–31, 2007.

[18] Risto Hilpinen and Paul McNamara. Deontic logic: A historical survey and introduction. In Dov Gabbay, John Horty, Xavier Parent, Ron van der Meyden, and Leon van der Torre, editors, *Handbook of deontic logic and normative systems*, volume 1. College Publications, 2021.

[19] Jørgen Jørgensen. Imperatives and logic. *Erkenntnis*, 7:288–296, 1937.

[20] Bruce A. Kimball. The proliferation of case method teaching in american law schools: Mr. langdell's emblematic "abomination," 1890-1915. *History of Education Quarterly*, 46(2):191–247, 2006.

[21] Christopher Columbus Langdell. *A Selection of Cases on the Law of Contracts*. The Legal Classics Library, The Legal Classics Library edition, 1983.

[22] Neil MacCormick. Legal deduction, legal predicates and expert systems. *International Journal for the Semiotics of Law*, 5:181–202, 1992.

[23] Jason Morris. *Constraint Answer Set Programming as a Tool to Improve Legislative Drafting: A Rules as Code Experiment*, page 262–263. Association for Computing Machinery, 2021.

[24] Dana Remus and Frank S. Levy. Can robots be lawyers? computers, lawyers, and the practice of law. *Georgetown journal of legal ethics*, 30:501+, 2016.

[25] Edwina Rissland. Artificial intelligence and legal reasoning: A discussion of the field and gardner's book. *AI Magazine*, 9(3):45, 1988.

[26] Alf Ross. Imperatives and logic. *Philosophy of Science*, 11(1):30–46, 1944.

[27] Giovanni Sartor and Andrea Loreggia. *The impact of algorithms for online content filtering or moderation - upload filters*. Study requested by the JURI Committee. European Parliament Think Tank, 2020.

[28] François Schroeter, Laura Schroeter, and Kevin Toh. A new interpretivist metasemantics for fundamental legal disagreements. *Legal Theory*, 26(1):62–99, 2020.

[29] M. J. Sergot, F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. T. Cory. The british nationality act as a logic program. *Communications of the Association for Computing Machinery (ACM)*, 29(5):370–386, 1986.

[30] Fred R. Shapiro. The most-cited legal scholars. *Journal of Legal Studies*, 29(1):409—-426, 2000.

[31] Juliele Sievers and Sébastien Magnier. Reasoning with Form & Content. In *Past and Present Interactions in Legal Reasoning and Logic*. Springer Verlag, 2015.

[32] Juliele Maria Sievers. *Jørgensen's Dilemma in the Interface Between Legal Positivism and the Natural Law Tradition*, pages 397–411. Springer International Publishing, Cham, 2022.

[33] Nicos Stavropoulos. *Objectivity in law*. Oxford University Press, 1996.

[34] Christian Strasser and G. Aldo Antonelli. Non-monotonic Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2019 edition, 2019.

[35] Victor Tadros. Rethinking the presumption of innocence. *Criminal Law and Philosophy*, 1:193–213, 2007.

Argumentation: Reasoning Universalis

Antonis Kakas
Department of Computer Science, University of Cyprus

Abstract. We examine the question of whether argumentation can form the basis for any form of reasoning, informal or formal logical reasoning. We propose that argumentation provides the wider framework encompassing uniformly all reasoning with strict or formal logical reasoning being a special boundary case. We also attempt to link this unifying role of argumentation with Aristotle's original investigation into reasoning and the formation of logical systems.

1 Introduction

Logic is traditionally separated into two forms: **Formal Logic** at the foundations of Mathematics and Science and **Informal Logic** as the study of human reasoning at large. These two forms of logic are generally considered to be very different. Yet they are both concerned with understanding the nature of human thought and, in fact, they share the same roots in Aristotle's work[1]. In this work we are interested in the question of whether formal and informal logic can be placed under a single framework and, if so, to understand their distinguishing features. In other words, we are interested in finding a universal form of reasoning that would be able to capture both informal and formal reasoning. In answering this question we will also attempt to link our proposal to the origins of the study of reasoning in Aristotle and how Aristotle's study can help in forming a unified view of reasoning. In a sense, the distinction of the two forms of logic seems to have evolved with the development of these over the last few centuries drawing them more and more apart.

In order to be concrete we will consider that Formal Logic is represented by Classical Logic (or simply Propositional Logic). For the case of Informal Logic it is more difficult to select a representative example. It is important though to realize that in the study of Informal Logic, within the humanities and particularly in Philosophy, scholars have been equating informal reasoning with **argumentation**. The entry on Informal Logic in the Stanford Encyclopedia of Philosophy (`https://plato.stanford.edu/entries/logic-informal/`) states:

[1] All statements in this paper relating to Aristotle are to be understood as hypotheses posed by the author in the context of his extremely limited knowledge of Aristotle's work. They are therefore subject to disproval by any Aristotelian scholar. They are made in an attempt to understand how Aristotle, as the first logician and his general study of systematizing human reasoning, relates to current attempts in AI to formalize and automate human reasoning.

> Though contributions to informal logic include studies of specific kinds or aspects of reasoning, the overriding goal is a general account of argument which can be the basis of systems of informal logic that provide ways to evaluate arguments. Such systems may be applied to arguments as they occur in contexts of reflection, inquiry, social and political debate, the news media, blogs and editorials, the internet, advertising, corporate and institutional communication, social media, and interpersonal exchange. In the pursuit of its goals, informal logic addresses topics which include, to take only a few examples, the nature and definition of argument, criteria for argument evaluation, argumentation schemes, fallacies, notions of validity, the rhetorical and dialectical aspects of arguing, argument diagramming ("mapping"), cognitive biases, the history of argument analysis, artificial intelligence (AI), and the varying norms and rules that govern argumentative practices in different kinds of contexts.

One field which studies Informal Logic, in the sense of human reasoning at large, is that of AI, where the aim to formalize and automate common sense reasoning was set as an early foundational problem. This resulted in the search for and development of a plethora of new logics for AI, called **non-monotonic logics**, starting with the logic of Circumstantiation for formalizing the Situation Calculus, a system for common sense reasoning about the effects of actions and the change they bring about [9]. These new logics aimed to capture the non-monotonicity feature of human inference recognizing that it should be possible to abandon, in contrast to the monotonic inference of formal classical logic, earlier inferences in the face of new relevant information. Non-monotonicity was needed to render the inference flexible, in the same manner as human do when drawing inferences, to missing or ambiguous information and tolerant to (apparently) contradictory information.

Nevertheless, these new non-monotonic logics were developed based on the same formal and strict underpinnings of Classical Logic making it difficult to deliver on their promise of "AI systems with common sense" and "human-like natural intelligence". Then in the 1990s, it was shown (see e.g. [1]) that using argumentation it was possible to reformulate (and in some cases extend) most, if not all, known logical frameworks of non-monotonic reasoning in AI. This AI approach to argumentation, sometimes referred to as **Computational Argumentation**, was motivated and to some extent grounded on earlier foundational work [19, 13, 14] on argumentation in Philosophy and Cognitive Science. The result of reconciling non-monotonic logics through argumentation resulted in a strong focus on Computational Argumentation as a way of capturing human reasoning in AI along the same frame of interest as that of Informal Logic. For example, argumentation can provide a principled approach to knowledge representation and reasoning about actions and change [12] and applied to problem of narrative comprehension akin to the way humans perform this task [2].

Similarly, following recent work in the Psychology of Reasoning that strongly supports the link of argumentation to human reasoning (e.g. [11, 10]) we can synthesize the framework of computational argumentation with cognitive principles

to form a framework, called **Cognitive Argumentation**, aiming to model human reasoning in its various forms. This framework has been shown to capture well the human empirical data from several different experiments that are traditionally used in Cognitive Science to evaluate cognitive models of human reasoning. These empirical evaluation domains include "Syllogistic Reasoning" with experiments on how humans reason on the original Aristotelian syllogisms, the "Selection Task" where humans are tested on the way they reason about conditionals and the "Suppression Task" where the non-monotonic nature of human reasoning is observed [16–18]. Cognitive argumentation accounts for the data in a cognitively adequate way that also reflects well the variation of human reasoning across the population.

We will therefore accept that human or informal reasoning is a matter of argumentation and ask whether argumentation can also encompass formal logic. Hence we will be interested in whether argumentation can be given some formal structure and how this might also cover formal classical deductive reasoning. We will argue that this is possible so that both informal but also formal logic can be captured uniformly within the same formal structure of argumentation. Argumentation is the wider framework encompassing all reasoning with strict or formal logical reasoning being a special boundary case.

2 Formal Argumentation

Argumentation is a process of considering the alternative positions that we can take on some matter with the aim to justify or refute a standpoint on the matter. It can take place socially, i.e. within a group of entities, in a debate where entities argue for different standpoints, or within a single entity where the entity contemplates or reasons internally about the various standpoints on the matter, in order to decide on its own stance on the matter.

Argumentation has the general form of a **dialectical** process of (i) starting with some argument(s) directly supporting some desired standpoint or conclusion, then (ii) considering various counter-arguments against the initial argument(s) and (iii) defending against these counter-arguments, typically with the help of other arguments as allies of the initial arguments. The process repeats by considering further counter-arguments against these new allied defending arguments, until we have formed a **coalition of arguments** that stands "well" as a **case** for the standpoint or conclusion of interest.

We therefore have an "argumentation arena" where arguments attack and defend against each other in order to support their claims. This arena of argumentation can be captured by a formal **argumentation framework** which in an abstract form can be simply given as a tuple, $\langle \mathcal{A}rgs, \mathcal{A}TT \rangle$, where $\mathcal{A}rgs$ is a set of arguments, $\mathcal{A}TT$ is an attack (typically non-symmetric) relation between arguments. Note in this minimal formalization of argumentation frameworks, $\mathcal{A}TT$, serves both the purpose of identifying counter-arguments but also defense-arguments, as arguments attacking back (under $\mathcal{A}TT$) a counter-argument.

Given an argumentation framework we need to formalize, through some normative condition, the notion that a subset of arguments stands well as a case of arguments. In fact, the dialectical process of argumentation indicates how to give a suitable semantics to formal argumentation. The dialectic argumentation semantics is defined via a relation $ACC(\Delta, \Delta_0)$ between any two sets of arguments Δ, Δ_0. This relation specifies the **acceptability** of the set of arguments Δ under the context where the set Δ_0 of arguments is considered as given and so a-priori acceptable. Informally, the relative acceptability of $ACC(\Delta, \Delta_0)$ is defined recursively to hold when the argument set Δ can render all its attacking (or counter-arguments) non-acceptable in the context of accepting Δ_0 together with Δ. This acceptability relation is formally defined as the least-fixed point of a naturally associated operator satisfying the following (see [5] for the technical details):

$ACC(\Delta, \Delta_0)$ holds, iff $\Delta \subseteq \Delta_0$, or, for any A such that $(A, \Delta) \in \mathcal{ATT}$ (A attacks Δ), $A \not\subseteq \Delta_0 \cup \Delta$, and there exists D such that $(D, A) \in \mathcal{ATT}$ (D defends against A) and $ACC(D, \Delta_0 \cup \Delta \cup A)$ holds.

Then the **acceptable or case** subsets of arguments are defined as those that are acceptable in the context of the empty set, i.e. the subsets Δ for which $ACC(\Delta, \emptyset)$ holds. Such acceptable subsets of arguments can be computed following the fixed point definition of acceptability. This is illustrated by figure 1.

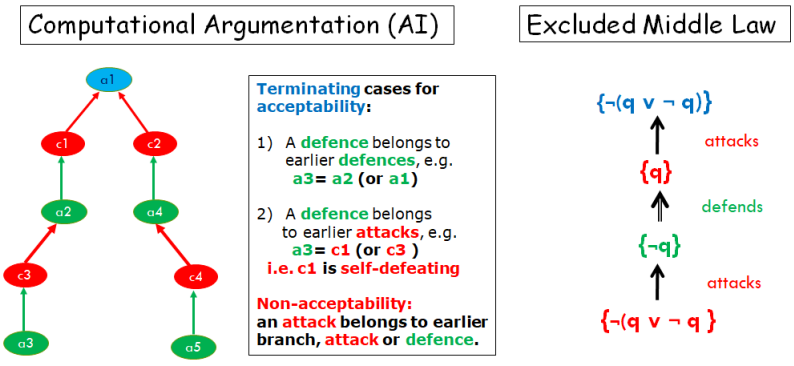


Fig. 1. Dialectic Acceptability/Non-Acceptability of Arguments

The left hand part of this figure, under the heading of Computational Argumentation in AI, shows this dialectic acceptability semantics in terms of labelled trees. Red nodes indicate attacking counter-arguments whereas green nodes indicate defending arguments. The termination conditions for the acceptability (respectively the non-acceptability) of the root argument are shown in terms of a defense (respectively an attack) node belonging to the branch above it. These

complement the base termination conditions of the non-existence of an attacking (respectively defending) argument.

3 Formal Logic as a case of Argumentation

This semantics of argumentation was used to reformulate the formal logical reasoning of classical Propositional Logic in terms of argumentation. Argumentation Logic [8] is defined as a realization of the above abstract argumentation framework and its semantics. The arguments are made up of sets of propositional formulae and the attack and defense relations are defined through the incompatibility between formulae and their negation. It is then possible to show that Argumentation Logic is logically equivalent to classical deductive reasoning whenever the given theory of premises that we are reasoning from is classically consistent [7]. The correspondence shows that classical truth models correspond to cases of acceptable subset of formulae, with cases though existing even when the given theory is inconsistent.

Non-surprisingly, as in most works that aim to bring formal logic closer to human reasoning, e.g. the early example of Intuitionistic Logic, the central element for this result of reformulating formal logical reasoning in terms of argumentation lies in the way that Reductio ad Absurdum is captured within the framework of argumentation. This is done by identifying structurally **self-defeating** (or fallacious) arguments and relating these to indirect logical proofs, i.e. proofs requiring Reductio ad Absurdum, within Propositional Logic.

Informally, a self-defeating argument, S, is one that "turns on itself" by rendering one of its attacking arguments acceptable in its own context of S. This means that the self-defeating argument renders the arguments that it needs for its defence, against some attacking counter-argument, non-acceptable. More formally, we can define a self-defeating argument S as one for which there exists a counter-argument A such that $\neg ACC(A, \emptyset)$ and $ACC(A, S)$ hold. So, although the attack A is in general (i.e. when we do not take any argument to be as given) non-acceptable under S this attack is rendered acceptable. Hence S brings about its own defeat and non-acceptability. The simplest example of a self-defeating argument is one that attacks itself, since in its own context its self-attack is acceptable.

For a more elaborate example of a self-defeating argument let us consider an example from the argumentation-based reformulation of formal logic, related to how we can derive the excluded middle law in Argumentation Logic. This is illustrated in the right part of Figure 1 where we see that the negation of the law, i.e. $\neg(q \vee \neg q)$, is shown to be non-acceptable. This is because this is attacked by the formula q, as from q we can directly derive $q \vee \neg q$. This attack by q can only be defended by taking on the opposing position of $\neg q$. But this defense is attacked by the root formula of $\neg(q \vee \neg q)$ since, as in the above attack, we can directly derive $q \vee \neg q$ from $\neg q$. Computationally, we see that an attack belongs to the branch above indicating the non-acceptability of the

root argument/formula. Hence we see that the argument $\neg(q \vee \neg q)$ renders its required defense non-acceptable and thus indirectly also itself non-acceptable.

Posing a hypothesis as a premise in a Reductio ad Absurdum proof corresponds to considering a context in which the hypothesis as an argument is accepted. Then the hypothesis leading to an inconsistency corresponds to the dialectic argumentation process leading to the non-acceptability of a (necessary) defending argument in the context of the posited argument. This correspondence is exact when the propositional theory of given premises is classically consistent in which case the non-acceptability of a formula argument also means the acceptability of the complement of the formulae, in the same way that Reductio ad Absurdum is used to derive the complement of the posited hypothesis. For the general case where the given theory under which we are reasoning is inconsistent then this latter step does not hold and we can have that both a formulae and its complement are non-acceptable. This signifies that we cannot have a position on such formulae. Nevertheless, this does not mean that the whole reasoning of Argumentation Logic trivializes but only that for some isolated formulae we are completely agnostic.

In summary, the classical formal reasoning is captured as a special case of argumentation were a logical conclusion emerges as the result of contemplating arguments for and against the conclusion. Argumentation Logic is constructed by adopting a set of direct proof rules as basic argument schemes together with the recognition of self-defeating arguments to cover the indirect proofs through Reductio ad Absurdum. Both the basic argument schemes and the notion of self-defeating arguments are structures that are content independent as it is expected from a framework of formal logical reasoning. This is in contrast with informal reasoning which, although it is captured under the same framework of argumentation, the various constructs of argument schemes, attacks and defenses depend on the content of arguments and the dynamically changing environment in which the reasoning takes place. Importantly, the paraconsistent[2] form of argumentative reasoning can be understood as a smooth extension of strict classical logical reasoning, in cases where indeed the given premise information is contradictory [6].

4 Aristotle: The origins of Systems of Reasoning

We will now briefly look into Aristotle's work on dialectic argument from a contemporary argumentation perspective. Specifically, we will examine the resemblance between the basic acceptability semantics that we have argued above unifies informal and formal reasoning, with the method of Aristotle for dialectic argumentation found in the books of *Topica*.

In these books Aristotle considers the wider context of what today we associate with informal reasoning and laid argumentation as the foundational element

[2] It is evident that Argumentation Logic is related to Paraconsistent Logics [15] which similarly consider how we can define forms of reasoning that do not trivialize under inconsistent premises.

of reasoning. His study of dialectic argument is extensive and quite thorough in an attempt to provide a pragmatically effective method of applying argumentation to support a position or a claim. He categorizes the different possible positions in terms of four types of "predicables" and goes into great length to give, for each different type of predicable, elaborate prescriptions (topoi) or strategies of how to go about supporting, attacking and defending each particular type of position.

From a contemporary point of view these topoi can be linked to the notion of argument schemes [21, 20] that associate premises to a position or to the contrary of a position, together with the pragmatics or heuristics to follow when carrying out the process of argumentation, as for example in the pragma-dialectal approach to argumentation in [3]. Interestingly, irrespective of the particular details of each topos the purpose of dialectic argumentation when applying the topoi is to arrive at a **refutation**. Aristotle states that the purpose of *Topica* ($100a^{18-22}$) is:

> To discover a method by which we shall be able to reason from generally accepted opinions about any problem set before us and shall ourselves, when sustaining an argument, avoid saying anything **self-contradictory** (copied from Rigotti and Greco, 2019:8 [4]).

At the very general level the strategy of dialectic argumentation in Aristotle is to bring the opposite view into a situation which is unacceptable because it is self-contradictory. Aristotle describes how this strategy can be executed through a process between a *Questioner* and an *Answerer*. This process can be understood as a semi-formal computational structure consisting of three stages:

(a) **Opening:** The Questioner presents a statement to which the Answerer can reply either yes or no. The overall aim of the Questioner is to force the Answerer to accept that his answer is self-contradictory and thus not reasonable. (b) **Interrogation:** The Questioner introduces questions to the Answerer to establish beliefs that the Answerer holds. The aim of the Questioner in this stage is to gather such beliefs from the Answerer that would allow him to build a strong argument against the Answerer's claim. (c) **Conclusion:** Once the Questioner has all the information he needs he reveals to the Answerer the counter argument, which he builds through a syllogism based on premises that the Answerer has accepted. The fact that this is build through a syllogism means that this is quite a strong argument and cannot be dismissed. Hence the Answerer has no option than to accept that his initial position is in contradiction with his other beliefs, i.e. his case is self-contradictory.

In this adversarial process, the goal for the Answerer is to prevent the Questioner from succeeding by reasonably rejecting the premises that would lead him in self-contradiction. The difficulty for the Answerer lies in realizing the counter-argument that the Questioner has in mind to build so that he can be careful on the beliefs he accepts during the second interrogation stage.

We can then observe a resemblance between this method of Aristotle for dialectic argumentation and the notion of acceptability and non-acceptability of

arguments that we have presented above as the unifying foundation of contemporary informal and formal reasoning. The central task in Aristotle to arrive at a self-contradiction is analogous to the identification of self-defeating arguments under the formal notion of acceptability of arguments. Just like the dialectic method of Aristotle concludes with the exposition of a contradiction in the beliefs held by the Answerer, in the same way the computational trees of acceptability (see figure 1 and termination conditions for non-acceptability) closes with an attacking argument playing also the role of a needed defending argument in the same dialectic branch of the tree, thus rendering the defending argument as self-defeating and non acceptable. Let us illustrate this correspondence through an example, shown in figure 2.

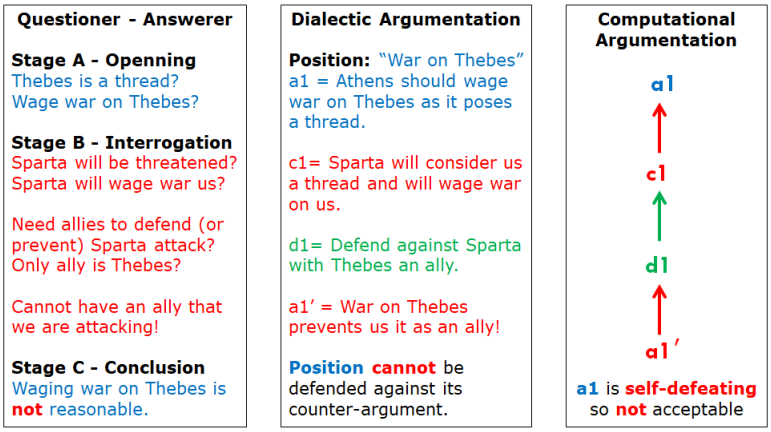


Fig. 2. Example of Aristotle's Dialectic Argument

In the leftmost box of the figure we see the questions asked by the Questioner. We assume that the Answerer has answered yes to all these questions. The Questioner can then re-construct an explicit dialectic argumentation process (seen in the middle box of the figure) where the attacking counter-argument of $c1$ is revealed together with the fact that the proposed defense $d1$ against this, i.e. to use "Thebes as an ally", is in conflict with the original position of the answerer of "waging war on Thebes" and therefore could not for a coalition with the initial argument of $a1$. The rightmost part of the figures shows the abstract computational structure of this argumentation process and how it ends up with the non-acceptability of the initial argument supporting the original position of the answerer[3].

[3] Strictly speaking the attacking argument $a1'$ is not the same as $a1$ but has the same effect of terminating the branch at an attack level. The only way to defend against $a1'$ is either by an argument against its premise of waging war on Thebes or an argument against Thebes being an ally. In either case this new defense will be attacked either by $a1$ or by $a3$ resulting in the non-acceptability of the branch.

5 Conclusions: Reasoning in AI

One of the main tasks of today's AI is to understand, formalize and effectively compute human reasoning. If we accept, as we are proposing in this paper, the universality of argumentation for reasoning, indeed that *Reasoning is Argumentation*, then we are led to *re-enact Aristotle's study of argumentation* in the *Topica*. Just like Aristotle studied how to conduct argumentation in an effective way and proposed different topoi as guidelines for achieving this we can carry out an analogous study for the effective realization of computational argumentation in AI. To do so we need to consider, as Aristotle did, the dynamic and uncertain nature of the environment in which argumentation takes place where the computational process of argumentation should adapt to new information and in many cases be guided to actively seek new relevant information. There is of course one major difference: Aristotle's argumentative reasoning was to be carried out by the "machine of the human brain" whereas in AI the machine is a poor artifact of the human brain. Nevertheless, we can draw on the study of argumentation over the centuries in philosophy, the psychology of reasoning and other disciplines to help us in this task of an effective process of reasoning through argumentation. In any case, the study from a modern perspective of Aristotle's extensive work on the good practice of argumentation, as for example in the recent work of [4], could provide us with valuable insights for the development of AI.

References

1. A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.
2. Irene-Anna Diakidoy, Antonis C. Kakas, Loizos Michael, and Rob Miller. Story comprehension through argumentation. In Simon Parsons, Nir Oren, Chris Reed, and Federico Cerutti, editors, *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 31–42. IOS Press, 2014.
3. Frans H. van Eemeren and Rob Grootendorst. *A Systematic Theory of Argumentation: The pragma-dialectical approach*. Cambridge University Press, 2003.
4. Sara Greco and Eddo Rigotti. *Inference in Argumentation: A Topics-Based Approach to Argument Schemes*. Springer Verlag, 2019.
5. A. Kakas and P. Mancarella. On the Semantics of Abstract Argumentation. *Journal of Logic and Computation*, 23(5):991–1015, 2013.
6. Antonis C. Kakas. Informalizing formal logic. *Informal Logic*, 39(2):169–204, 2019.
7. Antonis C. Kakas, Paolo Mancarella, and Francesca Toni. On argumentation logic and propositional logic. *Studia Logica*, 106(2):237–279, 2018.
8. Antonis C. Kakas, Francesca Toni, and Paolo Mancarella. Argumentation logic. In Simon Parsons, Nir Oren, Chris Reed, and Federico Cerutti, editors, *Computational Models of Argument - COMMA 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 345–356. IOS Press, 2014.

9. John McCarthy. Programs with common sense. In *Proc. of Teddington Conf. on Mechanization of Thought Processes*, pages 75–91, London, 1959.
10. Hugo Mercier. The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, 20(9):689 – 700, 2016.
11. Hugo Mercier and Dan Sperber. Why do humans reason? arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2):57–74, 2011.
12. Loizos Michael and Antonis C. Kakas. Knowledge qualification through argumentation. In Esra Erdem, Fangzhen Lin, and Torsten Schaub, editors, *Logic Programming and Nonmonotonic Reasoning, 10th International Conference, LPNMR 2009, Potsdam, Germany, September 14-18, 2009. Proceedings*, volume 5753 of *Lecture Notes in Computer Science*, pages 209–222. Springer, 2009.
13. Ch. Perelman and L. Olbrechts-Tyteca. *The new rhetoric. A treatise on argumentation*. Notre Dame/ London: University of Notre Dame Press, 1969.
14. J.L. Pollock. Defeasible reasoning. *Cognitive Science*, 11(4):481–518, 1987.
15. Koji Tanaka Priest, Graham and Zach Weber. Paraconsistent logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy (Spring 2022 Edition)*, 2022.
16. Emmanuelle-Anna Dietz Saldanha and Antonis C. Kakas. Cognitive argumentation and the selection task. *Künstliche Intell.*, 33(3):229–242, 2019.
17. Emmanuelle-Anna Dietz Saldanha and Antonis C. Kakas. Cognitive argumentation and the suppression task. *CoRR*, abs/2002.10149, 2020.
18. Emmanuelle-Anna Dietz Saldanha and Antonis C. Kakas. Cognitive argumentation and the selection task. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*. https://escholarship.org/uc/item/4n0867c7, 2021.
19. S.E. Toumlin. *The Uses of Argument*. Cambridge University Press, 1958.
20. Douglas Walton, Christopher Reed, and Fabrizio Macagno. *Argumentation Schemes*. Cambridge University Press, 2008.
21. Douglas N. Walton. *Argumentation Schemes for Presumptive Reasoning*. LEA, 1996.

The largest intrinsic disquotational definition of truth

Edoardo Rivello

University of Torino, Torino, Italy
rivello.edoardo@gmail.com

Contents

1 Introduction

Let $\mathcal{L}$ be a first-order language rich enough to express the basic laws of syntax, and let $\mathcal{L}_{\mathbf{T}}$ be $\mathcal{L}$ augmented by a fresh unary predicate $\mathbf{T}$ (for "true'). A *disquotation sentence* is a sentence of the form

$$\mathbf{T}\ulcorner\phi\urcorner \leftrightarrow \phi,$$

where ϕ is any sentence of $\mathcal{L}_{\mathbf{T}}$ and $\ulcorner\phi\urcorner$ is a structurally descriptive name of ϕ. A *disquotation set* is a set of disquotation sentences. A *disquotational theory of truth* is a theory of truth presented as the first-order consequences of a theory of syntax together with a distinguished disquotation set.

The disquotation sentences represent a central component of our naive concept of truth, however, due to the Liar paradox, the set of *all* disquotation sentences of $\mathcal{L}_{\mathbf{T}}$ is inconsistent with the basic laws of syntax.

A natural reaction to this fact is aiming to the goal of "retaining as many disquotation sentences as it is possible". We will refer to this goal in the following as to the *maximality principle*. McGee [5] first tried to formalise the maximality principle and proved a theorem which undermines it under several respects.

For the purposes of this paper, we will work in Halbach's setting [3], assuming that $\mathcal{L}$ is the first-order language of arithmetic in which the basic laws of syntax are represented *via* a fixed Gödel coding. Let PA denote Peano Arithmetic. McGee's Theorem, reworded in Halbach's setting and notation, is the following statement (cfr. [3, Theorem 19.2, p. 269]):

Thm 1.1 (McGee's Theorem 1)**.** *Let Δ be a set of sentences of $\mathcal{L}_{\mathbf{T}}$ consistent with PA. Then there is a disquotation set Γ such that*

1. *for all $\delta \in \Delta$, $\Gamma \cup \mathsf{PA} \vdash \delta$,*
2. *$\Gamma \cup \mathsf{PA}$ is consistent,*
3. *any disquotation set that properly includes Γ is inconsistent with PA,*
4. *$\Gamma \cup \mathsf{PA}$ is complete.*

From Theorem 1.1 (or from its proof) several bad consequences follow for the programme of maximising the set of all consistent (with PA) disquotation sentences:

1. (*Non-axiomatisability*) No maximal consistent disquotation set is axiomatisable.

2. (*Arbitrariness*) There are many mutually incompatible maximal consistent disquotation sets and no evident criterion for choosing among them.

3. (*Unwanted consequences*) Each maximal consistent disquotation set has unwanted consequences making it unplausible as a theory of truth.

4. (*Poorness*) Restricting ourselves to only those disquotation sentences which belong to *all* maximal consistent disquotation sets leaves us with a very poor theory of truth.

The first consequence, non-axiomatisability, obviously blocks the hope for an *axiomatic* disquotational theory of truth based on the maximality principle. Even worse, the other consequences of McGee's theorem represent bad news for *semantic* theories of truth too, for instance, for the aim of a (metatheoretic) definition of truth required to entail a maximal disquotation set. McGee's strategy to overcome the arbitrariness issue is to look at those disquotation sets which belong to *all* maximal consistent disquotation sets. This move obviously avoids arbitrariness and arguably helps to exclude some, if not all, unwanted consequences of maximal consistent disquotation sets. However, McGee shows that applying this strategy we are left with a disquotational theory only including disquotation sentences which are already provable from PA alone, like the Truth-teller sentences. By contrast, we put as a (very minimal) requirement for a theory of truth that it includes at least all *Tarski biconditionals*, namely, all those disquotation sentences $\mathbf{T}^{\ulcorner}\phi^{\urcorner} \leftrightarrow \phi$ in which ϕ is an arithmetical sentence.

McGee's theorem shows that if we want to work out a decent disquotational theory of truth from the maximality principle we need to restrict in some way the disquotation sets we allow to be maximised. We are mostly interested in investigating what happens if we apply McGee's strategy by interpreting "possible", in the statement of the maximality principle ("retaining as many disquotation sentences as it is possible"), by some metatheoretic property strengthening mere consistency.

A first step in this line of research was made by McGee himself who, in the same article, considered to replace "consistent with PA" with "consistent by ω-logic". Later, Cieśliński [2] proposed to replace "consistent with PA" with "conservative over PA". McGee and Cieśliński results show that non-axiomatisability, arbitrariness and unwanted consequences still affect these proposals. Moreover, it easily follows that applying McGee's strategy to these sets of maximal disquotation sets still leaves us with very poor theories of truth.

2 Definitions and disquotation: A general limitative result

McGee's "consistent by ω-logic" and Cieśliński's "conservative over PA" can be understood as first steps towards a disquotation set being an "implicit definition" of truth. For, every implicit definition is model-theoretically conservative over PA and in turn this property implies both consistency by ω-logic and conservativity over PA. Unfortunately, we already know that the Liar paradox together with Beth's theorem rule out the possibility of an implicit definition of truth: For, by Beth's theorem, an implicit definition of truth would be equivalent (over PA) to an explicit definition of $\mathbf{T}$ in terms of the arithmetical language; by our assumption, this definition should imply all Tarski biconditionals, so contradicting Tarski's indefinability of truth theorem.

A natural move, then, is to turn our attention to some metatheoretic property which lies "in between" being model-theoretically conservative and being an implicit definition. One idea is to move from total implicit definitions to only *partial* implicit definitions. The model-theoretic condition characterising a set of sentences Σ as being an implicit definition of $\mathbf{T}$ over PA is that for every model $\mathcal{M}$ of PA in the language of arithmetic there exists exactly one expansion $\mathcal{M}+Z$ which models Σ. In particular, this means that Σ "fixes the extension of $\mathbf{T}$" in the sense that, given a model $\mathcal{M}$ of PA the extension of $\mathbf{T}$ is forced by Σ to be represented by the set Z of individuals of $\mathcal{M}$. We can ask less than this. We can ask that Σ only fixes the extension of $\mathbf{T}$ inside its "range of significance".

The idea that a predicate comes with a "range of significance" is quite natural: For instance, when we speak about a number being "odd", we implicitly assume that the predicate "odd" is not to be applied to everything, not even to all "numbers", but only to the "natural numbers". In other words, the natural numbers constitute the *range of significance* of the predicate "odd". When we understand what the range of significance of a predicate is, we can content ourselves with a *conditional* definition of the predicate, rather than seek for a fully explicit definition. For instance, if we want to define "odd" we can say that "a natural number is odd iff it is not divisible by two". This is a conditional definition, in the sense that we give a definition which applies only to those individuals which fall under the range of significance of the predicate. We simply do not care about the meaning of the expression "x is odd" when x is not a natural number.

If ϕ is an arithmetical formula defining the range of significance of a unary predicate $\mathbf{P}$, a conditional definition of $\mathbf{P}$ (over PA) has the form

$$\forall x\,(\phi(x) \to (\mathbf{P}x \leftrightarrow \psi)), \tag{1}$$

where ψ is another arithmetical formula where only the variable x occurs free.

By Beth's theorem, a set of sentences Σ is equivalent (over PA) to a conditional definition like (1) iff Σ satisfies the following model-theoretic property:

$$\forall \mathcal{M}\,(\mathcal{M} \models \mathsf{PA} \Rightarrow \exists Z \subseteq \phi(\mathcal{M})\,\forall Y \subseteq M\,(\mathcal{M}+Y \models \Sigma \Leftrightarrow Y \cap \phi(\mathcal{M}) = Z)). \tag{2}$$

In words: For every model $\mathcal{M}$ of PA in the language of arithmetic there exists exactly one subset Z of the domain of $\mathcal{M}$ such that an expansion $\mathcal{M}+Y$ models Σ iff Y agrees with Z on the interpretation in $\mathcal{M}$ of the range of significance ϕ.

When the predicate $\mathbf{T}$ has to mean "true", the pre-theoretic intuition is that its range of significance is given by the set of all sentences of the language $\mathcal{L}_{\mathbf{T}}$ (by contrast, we do not care about the application of "true" to non-sentences, namely, to natural numbers which do not code sentences of $\mathcal{L}_{\mathbf{T}}$). However, we have already seen that this assumption leads to a contradiction with Tarski's indefinability theorem. The Liar Paradox suggests to us that the range of significance of $\mathbf{T}$ should be a proper subset of $\mathcal{L}_{\mathbf{T}}$: At the very least, a subset which excludes any Liar sentence. Since we have troubles with defining what the range of significance of $\mathbf{T}$ is to be, we can replace the formula ϕ in (2) by a variable X and try to maximise it. More precisely, we say that Σ is a *partial implicit definition* (in the following we will often omit "implicit" in order to be shorter) of $\mathbf{T}$ if and only if the following condition holds:

$$\forall \mathcal{M}\,(\mathcal{M} \models \mathsf{PA} \Rightarrow \exists X \subseteq M\,\exists Z \subseteq X\,\forall Y \subseteq M\,(\mathcal{M}+Y \models \Sigma \Leftrightarrow Y \cap X = Z)). \tag{3}$$

Clearly, the notion of a partial implicit definition is intermediate between the notion of a model-theoretic conservative theory and the notion of an implicit definition, as required: On the one hand, if Σ is a partial implicit definition, then for every model $\mathcal{M}$ simply take $Y = Z$

to obtain a model of Σ which witnesses conservativity; On the other hand, if Σ is an implicit definition, then it is also a partial implicit definition having the entire domain of $\mathcal{M}$ as its range of significance. Hence we could look for maximal disquotational partial implicit definitions of truth.

However, the above notion of partial implicit definition does not couple very well with disquotation. For one thing, according to the way we have defined this notion, the set of all Tarski biconditionals, is *not* a partial definition as it would be expected. Following Bays [1] we think that this fact is not to be ascribed to a weakness of the Tarski biconditionals, it is rather a by-product of our notion of partial definition. What we actually expect to be fixed by the Tarski biconditionals is the extension of $\mathbf{T}$ when applied to genuine arithmetical sentences, not to individuals that some (non-standard) model of PA thinks to be arithmetical sentences. In other words, the intended range of significance of $\mathbf{T}$ is the set of all sentences of $\mathcal{L}_{\mathbf{T}}$ *qua* syntactic objects: It is not given by a formula representing this set in Peano Arithmetic whose interpretation can vary from one model of PA to another.

Bays' remark about the Tarski biconditionals, extended to all disquotation sentences, leads us to modify the notion of partial definition as follows: We say that Σ is a *rigid* partial implicit definition of $\mathbf{T}$ if and only if the following condition holds:

$$\exists X \subseteq \omega \, \forall \mathcal{M} \, (\mathcal{M} \models \mathsf{PA} \Rightarrow \exists Z \subseteq M \, \forall Y \subseteq M \, (\mathcal{M} + Y \models \Sigma \Leftrightarrow Y \cap X = Z)). \tag{4}$$

It can be proved that maximising the range of significance of rigid partial definitions is the same as maximising the theories and that maximal disquotational rigid partial definitions of $\mathbf{T}$ do exist. Unfortunately, the same objections raised against the properties of being consistent with PA, consistent by ω-logic, and conservative over PA, still apply to the notion of being a rigid partial definition: There are too many incompatible disquotational rigid partial definitions, no apparent criterion for choosing among them, and the set of disquotation sentences belonging to all of them does not even include the set of all Tarski biconditionals.

This result is not surprising at all. Both McGee's and Cieśliński's limitative results rely on the fact, discovered by McGee, that "every set of sentences can be given the form of a disquotation set" or, more precisely, that for every set of sentences Σ there exists a disquotation set Γ which is equivalent to Σ modulo Peano Arithmetic. From this it follows that, as long as we are concerned with a metatheoretic property which is preserved by equivalence modulo PA, there is no gain in coupling this property with that of having the form of a disquotation set. For this reason, once one has realised that properties of sets of sentences such as consistency, conservativity or being a partial definition are too general for characterising a predicate, one adds nothing by further requiring that the sentences involved should be disquotation sentences.

3 The largest intrinsic rigid disquotational partial definition of truth

The moral we get from the previous section is that to be a disquotation set *per se* does not characterise truth. However, when we look at the set of all Tarski biconditionals we recognise that it *does* characterise truth for the language of arithmetic. Why it is so? The reason is that the set of all Tarski biconditionals not only partially fixes the extension of truth (in Bays' sense), but it does so *for the same set of sentences for which the disquotation sentences are assumed to be true*. In other words, the set of sentences for which we assume the disquotation sentences and the range of significance of $\mathbf{T}$ coincide. This is the characteristic feature of Tarski's theory of truth which is missed by the applications of the maximality principle we have seen above.

Let us modify once again our notion of "disquotational partial definition" in order to get profit from Tarski's lesson. For any set of sentences X, let $\mathrm{TB}(X)$ denote the set of all disquotation sentences built up from X, namely, $\mathrm{TB}(X) = \{\mathbf{T}\ulcorner\phi\urcorner \leftrightarrow \phi \mid \phi \in X\}$. We say that a disquotation set $TB(X)$ is a disquotational rigid partial definition *of truth* if and only if the following condition holds:

$$\forall\mathcal{M}\,(\mathcal{M} \models \mathsf{PA} \Rightarrow \exists Z \subseteq M\, \forall Y \subseteq M\, (\mathcal{M} + Y \models \mathrm{TB}(X) \Leftrightarrow Y \cap X = Z)). \tag{5}$$

Condition (4), coupled with the assumption that Σ is a disquotation set, can be used to characterise Σ as being a disquotational (rigid) partial definition "of $\mathbf{T}$", but nothing implies that "$\mathbf{T}$" has to mean "true". The predicate $\mathbf{T}$ could be used to mean, for instance, "prime number greater than two", and in this case we would obtain that the largest rigid partial definition of $\mathbf{T}$ does exist and that its range of significance is the set of all odd numbers: However, we could axiomatise this partial definition by a set of disquotation sentences as well as with any other set of sentences (in this case we could even use an explicit conditional partial definition for the same purpose). Only adding the requirement that the disquotation sentences are built from the sentences of the range of significance of $\mathbf{T}$ we obtain a condition which makes use of disquotation in a relevant way and that makes our partial definition "of $\mathbf{T}$" a partial definition "of truth" for its range of significance.

More precisely, observe that Condition (5) is obtained from Condition (4) by performing two moves. The first one is that of assuming that Σ is a disquotation set, namely, a set of sentences of the form $\mathrm{TB}(X)$ for some set X of sentences of $\mathcal{L}_{\mathbf{T}}$. This move (trivially) makes Σ "materially adequate", as a theory of truth, for the set of sentences X. The second move is that of assuming that Σ is a rigid partial definition of $\mathbf{T}$ which determines exactly X as the range of significance of $\mathbf{T}$. With Condition (5) in our hands we can try once again to apply the maximality principle: This time the general limitative result described in Section 2 no longer applies, because Condition (5), contrary to Condition (4), is a metatheoretic property of $\mathrm{TB}(X)$ which is *not* preserved by equivalence modulo PA. It is in this sense that we said above that Condition (5) "makes use of disquotation in a relevant way".

Using Condition (5) we still obtain that there are many incompatible maximal disquotational rigid partial definitions of truth and that, apparently, we do not have a criterion for choosing among them. Yet, McGee's strategy of taking the set of all disquotation sentences which belong to all maximal disquotation sets still applies and we get a disquotational rigid partial definition of truth (the largest intrinsic one, in the ordered-theoretic sense of "intrinsic"), call it $\mathrm{TB}(\Theta)$, which is not "poor" in an obvious way: Indeed we can show that $\mathrm{TB}(\Theta)$ properly includes the set of all Tarski biconditionals.

Finally, to support the claim that $\mathrm{TB}(\Theta)$ is an interesting (read: not poor and unlikely to have unwanted consequences) disquotational theory of truth, we can prove a further result. Being a disquotational rigid partial definition of truth, $\mathrm{TB}(\Theta)$ satisfies Condition 5, hence, for every model $\mathcal{M}$ of PA, $\mathrm{TB}(\Theta)$ uniquely determines the subset $Z_{\mathcal{M}}$ of those sentences of Θ which are true in every expansion of $\mathcal{M}$ to a model of $\mathrm{TB}(\Theta)$. We can prove that, in particular, taking $\mathcal{M}$ to be the standard model of arithmetic, the range of significance Θ and the valuation $Z_{\mathcal{M}}$ coincide with the domain and the extension (respectively) of the largest intrinsic fixed point of Kripke's monotone operator using Van Fraassen's supervaluation [4, p. 711] .

References

[1] Timothy Bays. Beth's theorem and deflationism. *Mind*, 118(472):1061–1073, 2009.

[2] Cezary Cieśliński. Deflationism, conservativeness and maximality. *Journal of Philosophical Logic*, 36(6):695–705, 2007.

[3] Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, 2011.

[4] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[5] Vann McGee. Maximal consistent sets of instances of Tarski's schema T. *Journal of Philosophical Logic*, 21:235–241, 1992.

Epistemic logic for verifying runtime verification communication protocols*

Antonis Achilleos[1], Elli Anastasiadi[1], Adrian Francalanza[2], and Jasmine Xuereb[1,2]

[1] ICE-TCS, Department of Computer Science, Reykjavik University, Iceland
[2] Department of Computer Science, University of Malta, Malta
antonios@ru.is, elli19@ru.is, adrian.francalanza@um.edu.mt, jasmine21@ru.is

Abstract

We build on previous work on monitors for hyperproperties, where we propose a logic for specifying properties over sets of traces that can be monitored with circuit-like non-communicating monitors for violations at runtime. In this paper, we propose an epistemic multi-agent logic framework for proving the correctness of distributed and communicating, runtime verification protocols over hyperproperties. Our protocols use monitors that can communicate and accumulate information. To verify the correctness of such a protocol, we can describe communication with epistemic statements that can be then used to derive a proof in an epistemic logic. We then present an example epistemic proof of correctness for a given communication protocol over a specific property that requires communicating monitors, and therefore is not included in the original fragment. This is a step towards a general epistemic framework for the verification of distributed monitoring systems.

1 Introduction

The field of runtime verification provides methods for checking whether a system satisfies an intended specification at runtime. This runtime analysis is done through a computing device called a *monitor* that observes the current run of a system in the form of a trace and attempts to infer the satisfaction or violation of the specification by the system or its run [1,4,5,7,12,16]. Recent work focuses on monitoring for *hyperproperties*, which are properties of sets of traces, introducing novel monitoring setups that process multiple traces [2,6,11]. A centrepiece in this line of work has been the specification logic Hyper-LTL [8]. Intuitively, Hyper-LTL uses trace variables and allows for quantifying these variables over a set of traces that can represent a set of system runs, or a collection of local executions of different system components. Hyper-LTL can use these trace variables to refer to the satisfaction of propositional variables at specific traces, and thus express relationships between local events.

We use the specification logic Hyper-μHML instead of Hyper-LTL, and we build on previous work on monitorability and monitor synthesis for μHML, which is a reformulation of the μ-calculus, and Hyper-μHML is its extension to hyperproperties [1, 3, 13]. The logic μHML allows for straightforward translations from well-known temporal logics such as LTL, and, at the same time, has an intuitive synthesis for monitors [1, 13]. The current paper extends the work from [3], where the authors give a monitor synthesis from a fragment of Hyper-μHML with good correctness and complexity guarantees. However, just like Hyper-LTL and unlike

*The work reported in this paper is supported by the projects 'Open Problems in the Equational Logic of Processes' (OPEL) (grant 196050-051), 'Mode(l)s of Verification and Monitorability' (MoVeMent) (grant no 217987) of the Icelandic Research Fund, and 'Runtime and Equational Verification of Concurrent Programs' (ReVoCoP) (grant 222021), of the Reykjavik University Research Fund.

the fragment from [3], Hyper-μHML can define dependencies over different traces, which can introduce additional latency when monitoring at runtime. The monitoring framework in [3] kept the processing-at-runtime cost minimal by restricting the type of properties it verified to a fragment of Hyper-μHML that effectively does not allow multiple traces to be referred to in the scope of the same quantifier. Therefore, local monitors do not need to communicate in order to detect violations.

In this work, we consider an extension of the circuit-like monitors from [3] that allows monitors to communicate. We observe that there can be more than one correct way to monitor for a given property, and a monitoring system can be engineered with specific goals, such as to minimise the communication overhead or to preserve certain privacy or robustness properties. Therefore, one needs to consider alternatives to a uniform monitor synthesis, which need to be proven correct. We propose a framework for using epistemic logic to prove the correctness of the communication strategy of distributed monitoring protocols. Then, we give an example of describing communications between monitors with epistemic statements and using these to prove that a monitoring protocol can detect all violations of a specification. Our goal is to extend this framework in future work, so that one can prove the correctness of monitoring systems for more general properties and for more notions of correctness.

2 Preliminaries

2.1 The Specification Logic

We present Hyper-μHML, the logic that we use to specify hyperproperties. Hyper-μHML extends the linear-time interpretation of μHML [14, 15, 17] by allowing quantification over traces. We assume two disjoint, countably infinite sets: a set Π of trace variables and a set V of recursion variables; and a finite set Act of events or actions. We define $\text{Act}_\Pi = \{a_\pi \mid a \in \text{Act} \text{ and } \pi \in \Pi\}$. A set $A \subseteq \text{Act}_\Pi$ is called *consistent* if for all $a_{\pi_1}, b_{\pi_2} \in A$, $a = b$ or $\pi_1 \neq \pi_2$. Events in a consistent set can occur simultaneously on different traces.

Definition 1. *Formulae $\varphi \in$ Hyper-μHML are constructed by the following grammar:*

$$\begin{aligned}\varphi &::= \exists\pi\varphi \mid \forall\pi\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \psi \\ \psi &::= \texttt{tt} \mid \texttt{ff} \mid [A]\psi \mid \langle A\rangle\psi \mid \psi \wedge \psi \mid \psi \vee \psi \mid \max x.\psi \mid \min x.\psi \mid x,\end{aligned}$$

where $\pi \in \Pi$, $x \in V$, and $A \subseteq \text{Act}_\Pi$ is consistent. When $A = \{a_\pi\}$, we may simply write $[a_\pi]\psi$ or $\langle a_\pi\rangle\psi$ instead of $[A]\psi$ or $\langle A\rangle\psi$.

Semantics. The semantics of Hyper-μHML is given over a finite set of infinite traces T over a finite set of actions Act and it is a natural extension of the linear-time semantics of μHML. We require an environment ρ that maps recursion variables to sets of traces, and an assignment $\tau : \Pi \to T$ of trace variables to traces in T. Let $T^0_\tau = \{a_\pi \mid \tau(\pi) = at \text{ for some } t \in \text{Act}^\omega\}$, $T^X_\tau = \{t \mid \tau(\pi) = at \text{ for some } t \in \text{Act}^\omega\}$, and let $\tau^+ : \Pi \to T$ be defined such that $\tau^+(\pi) = t$, where $\tau(\pi) = at$. We only give the case for the universal modality here:

$$T, \tau, \rho \models [A]\psi \text{ iff } A \subseteq T^0_\tau \text{ implies } T^X_\tau, \tau^+, \rho \models \psi.$$

We use the standard notation $T \models \varphi$ to denote that the set of traces T satisfies φ (and similarly for $T \not\models \varphi$). The work in [3] demonstrates how to monitor for the fragment Hyper^1-sHML of this logic, which does not allows nested trace quantification, diamonds, disjunctions, or least-fixpoints, using circuit-like monitors.

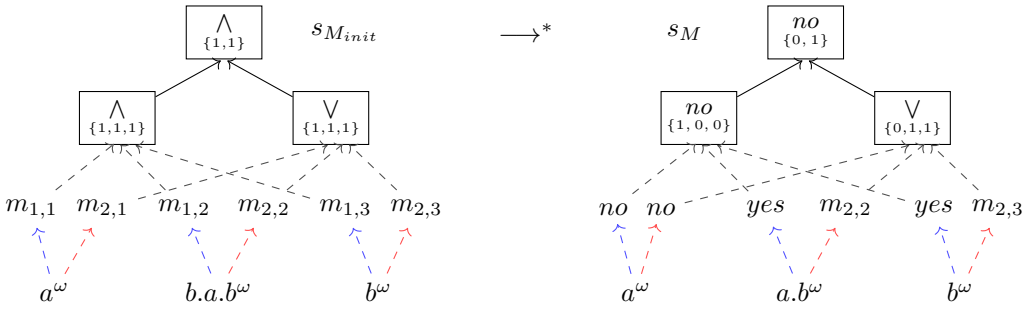


Figure 1: The circuit monitor for the formula in Example 1 over $T = \{a^\omega, b.a.b^\omega, b^\omega\}$.

Example 1. *The Hyper*^{1}*-*sHML *formula* $\forall_\pi[a_\pi]\mathtt{ff} \wedge \exists_\pi[b_\pi](\max x.([a_\pi]\mathtt{ff} \wedge [b_\pi]x))$*, over the set of actions* $\{a, b\}$*, states for a set of traces* T*, that no trace starts with* a*, and* $b^\omega \in T$*.*

2.2 The Circuit Monitors Model

In this section, we give the intuition behind the monitor design in [3]. Circuit monitors are composed of a hierarchy of gates, connected in a circuit-like structure and instrumented over a finite set of traces T. Each trace $t \in T$ is assigned a fixed set of regular monitors that correspond to the local properties to be verified and are at the bottom layer of the structure. Monitors assigned to the same trace run in parallel [1] and observe identical events, whereas those assigned to another trace also run in parallel but completely isolated from other traces. When monitors reach a verdict, *yes*, *no* or *end*, they communicate it to the smaller gates connecting them. These then evaluate to some verdict themselves and propagate their evaluation upward through logic gates until the root of the circuit reaches a verdict as well.

Definition 2. *The language* CMON$_k$ *of* k*-ary monitors, for* $k > 0$*, is given through the following grammar:*

$$M \in \textsc{Cmon}_k ::= \bigvee[m]_k \mid \bigwedge[m]_k \mid M \vee M \mid M \wedge M$$
$$m ::= \quad yes \mid no \mid end \mid a.m,\ a \in \textsc{Act} \mid m + n \mid rec\ x.m \mid x$$

CMON *is the collection of infinite sequences* $(M_i)_{i\in\mathbb{N}}$ *of terms that are generated by substituting* $k = i, \forall i \in \mathbb{N}$*, in a term* M *in* CMON$_k$*.*

The notation $[m]_k$ corresponds to the parallel dispatch of k identical regular monitors m, where $k = |T|$, with $T = \{t_1, \ldots, t_k\}$. The circuit monitor $\bigwedge[m]_k$ evaluates to a *yes* verdict if all sub-monitors evaluate to *yes* verdicts, and a *no* verdict if at least one sub-monitor evaluates to a *no* verdict. Otherwise, if all sub-monitors evaluate to some verdict but none of the previous criteria is met, it evaluates to *end*. The evaluation of $\bigvee[m]_k$ is symmetric, whereas the evaluation of the $\vee$ and $\wedge$ gates over them follows similar rules.

Figure 1 from [3] illustrates the circuit monitor and its evaluation for the formula in Example 1. The notion $m_{i,j}$ signifies that monitor m_i is instrumented with trace j, where monitors m_1 and m_2 respectively monitor for the local properties $\forall_\pi[a_\pi]\mathtt{ff}$ and $\exists_\pi[b_\pi](max\ x.([a_\pi]\mathtt{ff} \wedge [b_\pi]x))$.

Given a formula $\varphi \in$ Hyper^1-sHML and a set of traces T, we can synthesise a circuit monitor M through the recursive function $Syn_T(-)$: Hyper^1-sHML $\rightarrow$CMON defined in [3].

Proposition 1 (from [3]). *For a formula $\varphi \in$ Hyper^1-sHML and a set of traces T, we have that $Syn_T(\varphi)$ is a violation complete monitor for φ over T, in that $Syn_T(\varphi)$ outputs a verdict no if and only if $T \not\models \varphi$.*

2.3 Epistemic Logic

Definition 3 (Multi-agent modal logic). *For a set of agents $\mathcal{A}$, a formula ϕ in the multi-agent modal logic is defined as:*

$$\phi ::= \top \quad | \quad \bot \quad | \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K_i\phi \quad | \quad C_G\phi$$

where p is an atomic formula, $i \in \mathcal{A}$, and $G \subseteq \mathcal{A}$. Implication and disjunction can be defined from the other operators as usual.

We use the standard multi-agent S5 semantics for epistemic formulas with common knowledge, as seen, for example, in [10]. Later on, we also use a natural deduction proof system for multi-agent $S5$. One would need a more intricate logic to fully analyse monitoring frameworks, but as the following section demonstrates, sometimes the above epistemic logic suffices to prove the correctness of a protocol.

3 Epistemic Analysis of Communication Protocols

The fragment Hyper^1-sHML that was introduced in [3] is quite restricted. This allows for a uniform, correct monitor synthesis that does not require the monitors to communicate. In this section, we consider monitoring systems with a communication protocol, which allows us to monitor for more involved properties. In contrast to [3], instead of giving a monitor synthesis for a larger fragment of Hyper-μHML, we focus on proving the correctness of the communications part of a monitoring framework that might have not been produced by an automated synthesis.

3.1 Two Protocols for Two Quantifiers

We consider the example of the following Hyper-μHML formula, which uses two nested quantifiers:

$$\varphi = \forall\pi\forall\pi' max\ x([p_\pi, \overline{p}_{\pi'}]\mathtt{ff} \wedge [p_\pi, p_{\pi'}]x \wedge [\overline{p}_\pi, \overline{p}_{\pi'}]x)$$

Formula φ states that all traces π and π' must agree on all events p. Said otherwise, if p is observed in some trace, then all traces must have p at that time as well. The setup of circuit monitors from Section 2 cannot handle properties similar to this one. More specifically, all local monitors would only be able to observe the value of their own traces regarding p and produce the verdict *end* when asked for the transition $[p_\pi, \overline{p}_{\pi'}]$. However, the latter cannot happen as the transition specified contains at least one step on a trace that the monitors are not instrumented on.

The following monitoring setup would succeed in detecting the violations of φ. If p occurs in some trace, then there are two possible scenarios: *(i)* either all other traces have p as well or *(ii)* at least one trace does not have p. The key point is that in the case of a violation of the property φ, there must be some traces that do not agree on p. Thus, should we design a naive communication protocol where every monitor instantaneously communicates with all other monitors after each event it observes, to inform them about whether or not it observed p, at least a few of the monitors would indeed observe this violation. Naturally, such a protocol would produce a great number of messages while it runs.

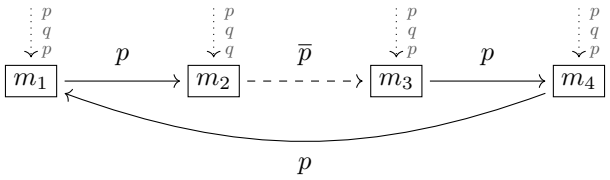


Figure 2: Communication Protocol for formula φ

In order to reduce the significant communication overhead of the above approach, we can use fewer messages and compensate for the lack of information via epistemic reasoning. For instance, consider a protocol where, after observing an event, each monitor communicates with exactly one agent (the one to its right) and receives exactly one message from another agent (the one on its left). Both messages are identical in nature, whereby they inform the receiver whether the sender observed p or not. We refine this further by allowing monitors to communicate only when p is observed since the absence of a message can convey the negation of this statement as shown by the dashed line in Figure 2.

A logician could easily recognise that the above protocol detects a violation of the property described above through the following basic epistemic reasoning. Assume that there are two traces π, π' that do not both have p. If all agents are assigned some order in which they will perform the described communication protocol, there will be two consecutive agents whose values of p do not match, and both of them will be able to deduce that the property is violated. For instance, monitors m_2 and m_3 in Figure 2 detect a violation of φ since the former received p but it observed q, whereas the latter observed p but didn't receive anything, from which it can infer that m_2 didn't observe p.

Remark 1. *In the worst case, two monitors will be able to infer the no verdict, while all the others produce the end verdict. However, this is sufficient for the gate on the higher level to produce the no verdict as well, giving us violation completeness for the specific property φ.*

In what follows, we demonstrate how to prove the correctness of the protocol that we discussed, using epistemic logic.

3.2 Proving Correctness

A first attempt at presenting a correctness proof for the communication protocol discussed above is modelling each monitor as an agent $r \in \mathcal{A}$. The protocol is modelled thought sentences in epistemic logic that are obeyed in each round, where a round is the time during which an event is observed by all agents.

As is described, a monitor (agent r) can observe the occurrence of the event p on trace r (denoted p_r), in which case it has to inform the monitor assigned to the same property on the trace "on its right" about this occurrence. We denote the "next" agent as $r+1$ *mod* k, where k is the total number of traces. Thus, in a round i we first prove that with the protocol we mentioned it is always the case that a monitor $r+1$ knows whether p_1 or $\overline{p_r}$. The natural deduction proof for this can be seen in Table 1, given in the appendix due to lack of space.

that a violation of the property φ occurs the behaviour of the monitors will be in accordance with the guarantees provided by the epistemic natural deduction proof given in Table 2. Note that since a violation has occurred it means that there exists traces (and thus monitors r_0, r_1 such that p_{r_0} and $\overline{p_{r_1}}$.

Having established this inference, we use it to prove that when a violation of the property occurs in round i, (i.e., $\exists r_0 \exists r_1 p_{r_0} \wedge \overline{p_{r_1}}$) then there is some agent j that detects the violation. The proof of this inference can be seen in Table 2 in the appendix.

We remark that one can use a similar approach to prove the correctness of monitoring setups for more interesting properties. For example, propositional variables in the epistemic syntax can be used to encode the violations of arbitrary monitorable formulas; the (eventual) detection of such a violation encoded by p by the monitor on trace i can be written as $K_i p$, and its monitorability as $p \to K_i p$. Then, we can proceed as above.

4 Conclusion and Future Work

In this work we present an initial attempt to incorporate multi-agent epistemic reasoning to the analysis of distributed runtime verification protocols. The key aspect of our approach is to first design a protocol for sharing information and then prove formally that it provides correctness guarantees.

Besides the verification of distributed monitoring setups, one of our aims is to eventually produce a sound *synthesis* algorithm for communication protocols such as the one given in [3] (Proposition 1). However, there are several obstacles that remain to be incorporated into this reasoning framework before reaching this goal. The first shortcoming is the static way in which epistemic logic has been incorporated, which constrains the proofs to be done in a round-by-round fashion as they are currently. We aim to model the exact content of a communication into an *epistemic action* that occurs and has an outcome of the models of a formula. Our approach here would be to incorporate Dynamic epistemic logic [9] so that the temporal aspect of a proof is not introduced externally.

Moreover we have not yet formally extended the monitoring setup to include monitors that can synchronise and produce these communications, and we have not assigned any sort of formal semantics to such a syntactical modification. Thus, to fully perform the upgrade we need also to adapt the implementation to mach the capabilities of the theory.

Finally in order to automate the synthesis of a communicating monitor setup, after having performed the above steps, we want be able to extract from a proof for a certain epistemic theorem into a communication protocol. For example a theorem we would like to test for Hyper-μHML formulae could be formulated as $\neg\varphi \to \exists i K_i \neg\varphi$. In such a scenario one would want to synthesise a valid monitoring setup from potential tableau proof of this statement. Of course, the semantics of the epistemic operator in the above formula would need to be defined as part a more complex language that will be able to refer to past and future situations.

Finally, our contribution, even though we only present one specific protocol-property pair, can also provide privacy guarantees. Specifically, instead of our protocol, all of the trace observation could be done in a central fashion, or in a distributed one but where every event is fully broadcast to all agents. However, it is easy to see that both such scenario allow for also security breaches, as all information is gathered in one place which can be compromised. Our alternative allows not only for less communication- and thus smaller overhead at runtime- but also minimises the amount of information exchange, which ensures that for example a compromised node will only gather partial information about the system. Thus our effort is a step towards enabling the formal verification of concurrent systems though faster and more secure distributed monitoring mechanisms.

References

[1] Luca Aceto, Antonis Achilleos, Adrian Francalanza, Anna Ingólfsdóttir, and Karoliina Lehtinen. Adventures in monitorability: From branching to linear time and back again. *Proceedings of the ACM on Programming Languages*, 3(POPL):52:1–52:29, 2019.

[2] Shreya Agrawal and Borzoo Bonakdarpour. Runtime Verification of k-Safety Hyperproperties in HyperLTL. In *2016 IEEE 29th Computer Security Foundations Symposium (CSF)*, pages 239–252, 2016.

[3] Elli Anastasiadi, Luca Aceto, Antonis Achilleos, and Adrian Francalanza. Monitoring Hyperproperties with Circuits. In *42nd International Conference on Formal Techniques for Distributed Objects, Components, and Systems (FORTE 2022)*, volume short paper, to appear, 2022.

[4] Ezio Bartocci, Yliès Falcone, Adrian Francalanza, and Giles Reger. Introduction to runtime verification. volume 10457 of *Lecture Notes in Computer Science*, pages 1–33. Springer, 2018.

[5] Laura Bocchi, Kohei Honda, Emilio Tuosto, and Nobuko Yoshida. A theory of design-by-contract for distributed multiparty interactions. In Paul Gastin and François Laroussinie, editors, *CONCUR 2010 - Concurrency Theory*, pages 162–176, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[6] Borzoo Bonakdarpour and Bernd Finkbeiner. The complexity of monitoring hyperproperties. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, pages 162–174, 2018.

[7] Ian Cassar, Adrian Francalanza, Claudio Antares Mezzina, and Emilio Tuosto. Reliability and fault-tolerance by choreographic design. In *PrePost@iFM*, 2017.

[8] Michael R. Clarkson and Fred B. Schneider. Hyperproperties. In *2008 21st IEEE Computer Security Foundations Symposium*, pages 51–65, 2008.

[9] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer Publishing Company, Incorporated, 1st edition, 2007.

[10] Ronald Fagin, Joseph Y Halpern, Yoram Moses, and Moshe Vardi. *Reasoning about knowledge*. MIT press, 2004.

[11] Bernd Finkbeiner, Christopher Hahn, Marvin Stenger, and Leander Tentrup. Monitoring hyperproperties. *Formal Methods in System Design*, 54, 11 2019.

[12] Adrian Francalanza, Luca Aceto, Antonis Achilleos, Duncan Paul Attard, Ian Cassar, Dario Della Monica, and Anna Ingolfsdottir. A foundation for runtime monitoring. volume 10548, pages 8–29, 09 2017.

[13] Adrian Francalanza, Luca Aceto, and Anna Ingolfsdottir. Monitorability for the Hennessy–Milner logic with recursion. *Formal Methods in System Design*, 51, 08 2017.

[14] Dexter C. Kozen. Results on the propositional μ-calculus. *Theoretical Computer Science*, 27:333–354, 1983.

[15] Kim G. Larsen. Proof Systems for Satisfiability in Hennessy-Milner Logic with recursion. *Theoretical Computer Science*, 72(2):265 – 288, 1990.

[16] Claudio Antares Mezzina and Jorge A. Pérez. Causally consistent reversible choreographies: A monitors-as-memories approach. In *Proceedings of the 19th International Symposium on Principles and Practice of Declarative Programming*, PPDP '17, page 127–138, New York, NY, USA, 2017. Association for Computing Machinery.

[17] Moshe Y. Vardi. A Temporal Fixpoint Calculus. In *POPL*, pages 250–259, New York, NY, USA, 1988. ACM.

Appendix 1 - Natural Deduction Proofs

1	$\forall r : C_{\mathcal{A}}(p_r \leftrightarrow K_r p_r)$	
2	$\forall r : C_{\mathcal{A}}(K_r p_r \rightarrow K_{r+1} p_r)$	
3	$\overline{p_c}$	
4	$K_{c+1} p_c \rightarrow p_c$	$S5$ axiom.
5	$\overline{p_c} \rightarrow \neg K_{c+1} p_c$	prop Thm (4)
6	$\neg K_{c+1} p_c$	$\rightarrow_e$, (3,5)
7	$K_{c+1} \neg K_{c+1} p_c$	$S5$ axiom (6)
8	$\boxed{K_{c+1}}$	
9	$p_c \rightarrow K_c p_c$	C (1)
10	$K_c p_c \rightarrow k_{c+1} p_c$	C (2)
11	$\neg K_{c+1} p_c$	$K_e e$ (7)
12	p_c	assumption.
13	$K_c p_c$	$\rightarrow_e$, (9,12)
14	$K_{c+1} p_c$	$\rightarrow_e$, (10,13)
15	$\bot$	$\neg_e$ (11,14)
16	$\overline{p_c}$	$\neg_i$ (12,15)
17	$K_{c+1}(\overline{p_c})$	$K_c i$ (8,16)

Table 1: Deduction of the non-occurrence of p_c from agent $c+1$

In Tables 1 and 2, lines 1 and 2 are using the quantification not as part of the syntax, but over the number of agents to indicate the existence of k many *real* premises corresponding to the relative line - one for each agent. There we model the communications taking place as part of the protocol though epistemic premises. Line 1 describes that all agents are operating on a distributed monitoring scenario where all agents can know whether p or $\neg p$ on each round, and are aware that this is the protocol applied ot all of them. Line 2 encapsulates the communication of the observance of p in a similar fashion. Assuming on round i we have a violation, we have that $\exists \pi$, and $\exists \pi'$ where due to the argument given above, $\pi' = \pi + 1 \; mod \; k$ such that $p_\pi \wedge \overline{p}_{\pi'}$. Thus we show that there exists some trace c, where the relative monitor (agent) will observe the appropriate events that enable it to deduce $[K_c(p_c \wedge \overline{p_c})]$.

1	$\forall r : C_{\mathcal{A}}(p_r \leftrightarrow K_r p_r)$	
2	$\forall r : C_{\mathcal{A}}(K_r p_r \rightarrow K_{r+1} p_r)$	
3	p_c	
4	$\overline{p_d}$	
5	$K_c p_c$	$\rightarrow_e$ (1,3)
6	$K_{c+1} p_c$	$\rightarrow_e$ (2,4)
7	$K_{c+1} K_{c+1} p_c$	$S5$ axiom (6)
8	$\boxed{c+1}$	
9	$(p_{r_0+1}) \vee \overline{p_{r_0+1}}$	taut.
10	$\overline{p_{c+1}}$	assumption
11	$K_{c+1}(\overline{p_{c+1}})$	$\rightarrow_e$ (1)
12	$K_{c+1} p_c$	$S5$ axiom (6)
13	p_c	$S5$ axiom (12)
14	$p_c \wedge \overline{p_{c+1}}$	$\wedge_i$ (10,13)
15	$K_{c+1}(p_c \wedge \overline{p_{c+1}})$	$S5$ axiom (14)
16	$[K_{c+1}(p_c \wedge \overline{p_{c+1}})] \vee p_{c+1}$	$\vee_i$ (15)
17	p_{c+1}	assumption
18	$[K_{c+1}(p_c \wedge \overline{p_{c+1}})] \vee p_{c+1}$	$\vee_i$ (17)
19	$[K_{c+1}(p_c \wedge \overline{p_{c+1}})] \vee p_{c+1}$	$\vee_e$ (16,18)
20	$\boxed{c+2}$	
21	$\ldots$	(8-19)
22	$[\bigwedge_{r \in \mathcal{A}} p_r] \vee [\bigvee_{r \in \mathcal{A}} [K_r(p_{r-1} \wedge \overline{p_r})]]$	
23	$[\bigwedge_{r \in \mathcal{A}} p_r]$	assumption
24	p_d	$\wedge_e$ (23)
25	$\bot$	$\neg_e$ (4,24)
26	$\neg([\bigwedge_{r \in \mathcal{A}} p_r])$	$\neg_i$ (23,25)
27	$\bigvee_{r \in \mathcal{A}} [K_r(p_{r-1} \wedge \overline{p_r})]$	prop. Thm (22,26)

Table 2: Epistemic guarantees of correctness of round i